In this paper, we study an analytical approach to selecting expansion locations for retailers selling add-on products whose demand is derived from the demand for a separate base product. Demand for the add-on product is realized only as a supplement to the demand for the base product. In our context, either of the two products could be subject to spatial autocorrelation where demand at a given location is impacted by demand at other locations. Using data from an industrial partner selling add-on products, we build predictive models for understanding the derived demand of the add-on product and establish an optimization framework for automating expansion decisions to maximize expected sales. Interestingly, spatial autocorrelation and the complexity of the predictive model impact the complexity and the structure of the prescriptive optimization model. Our results indicate that the formulated models are highly effective in predicting add-on-product sales, and that using the optimization framework built on the predictive model can result in substantial increases in expected sales over baseline policies.

Key words: add-on products; derived demand; empirical demand estimation; retail expansion optimization; predictive and prescriptive analytics

History: Received: April 2018; Accepted: January 2019 by Subodha Kumar, after 2 revisions.

1. Introduction

Retailers are constantly seeking to increase sales, and, since sales in a fixed location are inherently limited (Kumar and Leone 1988), expansion to new locations provides a significant opportunity to achieve this goal (Pancras et al. 2012). The location choice for new outlets is critical to retailers, especially for those whose services have to be offered through fixed geographic locations (Brown et al. 1981, Ketelaar et al. 2017). Access to retail outlets is a critical factor in determining patronage (Eze et al. 2015, Ghosh and Craig 1986). A good location strategy also gives the retailer strategic advantages over its competitors (Jain and Mahajan 1979, Obeng et al. 2016). The location decision usually represents a major long-run fixed investment (Craig 1984, Oner 2018). When considering adding more service locations, the retailer trades off between the expected revenue and the additional cost of providing the service (Ghosh and Craig 1986, Pancras et al. 2012). If the expected profit will not cover the cost, expansion is not pursued. Moreover, the effect on the performance of a whole retail chain of adding new outlets to an existing retail network is complex (Pancras et al. 2012). Picking the outlet that generates the largest expected sales on its own will not guarantee the largest marginal revenue for the whole retail chain. The manager should consider both a single, separate location and the whole retail network.

This paper focuses on retail network expansion in the context of add-on products. Add-on products are discretionary benefits that provide utility only if consumed with the corresponding base good (Bertini et al. 2008, Guiltinan 1987). Firms sell add-on products or features that enhance the value of the base product (Erat and Bhaskaran 2012). Add-on products or services are well-known as a special case of mixed bundling (Guiltinan 1987). The customer may purchase the base product (e.g., a car wash) alone or may select additional amenities (e.g., vacuuming) that are sold only with the base product (Guiltinan 1987). Extant papers discuss when a firm benefits from bundling (Guiltinan 1987) and how to price the bundles (Cui et al. 2018b, Ellison 2005, Geng and Shulman 2015, Geng et al. 2018, Shulman and Geng 2013). To our best knowledge, no research has studied where to sell the add-on products, that is, the add-on products’ location strategy.

When an add-on product retail manager makes expansion decisions, in addition to referring to the local demographic and socioeconomic indices,
operational values could also be provided by current operating locations and sales of the base product. To be specific, one can take advantage of the fact that the add-on products can only be sold at the same location as the base product, and the add-on product can only be purchased together with the base product.

In our research setting, the add-on product and the base product do not have to be owned by the same company. We look into this question making use of the known base-product location and sales information in addition to the demographic and socioeconomic indices. We also consider the spatial interactions of the add-on-product sales and the base-product sales at adjacent locations. We utilize a spatial-weight matrix to handle the spatial interaction if the test result of spatial autocorrelation is significant. We show empirically that incorporating the spatially weighted base-product sales would improve prediction accuracy only in the presence of statistically significant spatial autocorrelation.

Add-on products are pervasive in industry, with different defining characteristics. Examples include:

- Kiosks selling specialized good or services at malls, for example, Cellairis;
- Gas stations selling fuel additives at the pump, that is, the additives are added together with the gasoline through the same pipeline at the consumers’ choice, for example, Additech;
- Airports offering gyms to travelers during layovers, for example, Roam fitness;
- Restaurant menus listing toppings, condiments, and other ingredients that patrons can add to complement a standard order (Bertini et al. 2008);
- Fitness centers pricing amenities such as locker rentals, towels, and selected group activities separately from basic membership (Bertini et al. 2008, Guiltinan 1987);
- The seller of computer hardware offering maintenance contracts to its consumers (Guiltinan 1987);
- Hotels offering weekend packages that combine lodging and some meals at special rates (Guiltinan 1987);
- Personal care services offering body care products at an extra charge;
- Car washes offering tire polish service at an extra charge;
- Car dealers urging prospective buyers to consider accessory packages and extended warranties when they purchase a new vehicle (Bertini et al. 2008); and
- Airlines providing meal service, alcoholic beverages, and in-flight entertainment on domestic routes for an additional fee (Bertini et al. 2008).

The focus of this paper is on settings where:

1. The add-on product shares brick-and-mortar selling points with the base product.
2. The add-on product is only sold together with the base product, that is, the consumer can purchase the base product alone or both the base product and the add-on product. Consumers do not buy the focal add-on product alone at the selling points or anywhere else.
3. There is a significant capital cost in expanding the offering of the add-on product to a new location; and
4. The focal add-on product and its base product are not necessarily manufactured/owned by the same company.

In particular, the first three examples given above (kiosks selling services and products at malls, at-the-pump fuel additive sellers, and airports offering workout services) fall specifically in this context.

The problem we study is how to maximize the expected sales of add-on products when expanding a firm’s retail network within another, fixed retail network, where a company can only expand to a limited number of locations among those where the base product is currently offered. We present a decision-making framework for determining the optimal location for expansion. This requires accurate prediction of sales given a set of base-product retail locations and an optimization framework for automating the selection of an optimal collection of expansion sites.

For predicting sales, we apply several machine learning algorithms and evaluate their performance on a year of point-on-sales data at existing sites where add-on products are sold. We use basic area-specific demographic information (median household income and population) to supplement base-product and add-on-product sales data to create more accurate models. We further supplement the data set with a spatial-weight matrix, whose entries provide a measure of geographic closeness between base-product retail locations that decays as the distance grows. Including this matrix as part of an input to the learning models is shown to increase predictive accuracy only when spatial autocorrelation exists in base-product sales, as exhibited through an experimental evaluation on two different regions, one in the presence of and the other in the absence of spatial autocorrelation, as indicated by the Moran’s I Test (Bivand et al. 2008, Moran 1950). The models we develop in this paper have predictive power better than other models suggested in the literature for similar contexts; for example, the (out-of-sample) mean absolute percent error of our predictive models is around 20%, which is less...
than that reported by Glaeser et al. (2019), for predicting sales at online websites.

Equipped with an accurate predictive model for sales, we then turn to formulating and solving an optimization model for automating expansion decision making. The optimization model can take various forms, depending on the predictive model employed. We describe in detail the form that the optimization model takes for the various predictive models tested and suggest a simple general-purpose heuristic that can be used for any such model.

In comparison with a standard baseline policy for expansion, in which the add-on product company expands based on higher base-product sales, the predictive-and-optimization modeling framework developed can result in over 5% additional sales, depending on the base-product sales distribution, the number of expansion sites that the add-on product company can afford, and the presence of spatial autocorrelation, simply by choosing better expansion locations. The solutions obtained are shown to be robust, in that they result in favorable expansion decisions, even when evaluated on different predictive models than the one used in the optimization model.

The contributions of this research and the managerial insights it reveals are therefore as follows:

- The first study of add-on product location decision making;
- The design of accurate predictive models for the demand of add-on products using limited data;
- An understanding of when spatial-weight matrices can improve predictive accuracy based on spatial autocorrelation;
- An optimization framework for automating expansion decisions based on predictive models; and
- An application to a real-world data set where the predictive-and-optimization modeling framework provides substantially better solutions than baseline expansion policies.

The remainder of the paper is organized as follows. We first review related literature in section 2. In section 3, we provide details of the problem setting and context. In section 4, we describe the data set provided by our industrial partner, and the various data elements that we gathered from publicly available sources. Section 5 describes the predictive models we employ and analyzes their accuracy on the data. Section 6 presents the optimization models, both in general and for specific predictive models, suggests a general-purpose heuristic for solving the optimization models, and reports on the effectiveness of the solution obtained. We conclude in section 7.

2. Literature Review

This work is closely related to a variety of research streams. We discuss each, highlighting the papers found in the literature that are most relevant to the problem we study.

2.1. Add-on Products

Add-on products and services are common in retail and are often a major source of revenue for many firms (Ellison 2005). In the following, we will focus on add-on products, but our discussion extends to both add-on products and add-on services.

Add-on products are a form of mixed bundling. There are two types of bundling: pure bundling and mixed bundling. Pure bundling refers to settings where consumers must buy the bundle or none of the items in the bundle, whereas mixed bundling allows consumers to buy products individually or together. Add-ons are a special form of mixed bundling because add-on products are those that are offered together with base products but cannot be purchased alone. In most of the cases addressed in the literature, both the add-ons and the base product are from the same company. Hence, they are sometimes related to product-line bundling. However, in our study, we do not assume that the add-on product and the base product are necessarily from the same company. Hence, we further assume the prices of both products are not coerced.

Add-on products have been extensively discussed in the economic and marketing literature. Some research studies situations under which firms can benefit from add-ons and how firms set the price. For example, Guiltinan (1987) presents a normative framework for selecting appropriate types of services for different mixed-bundling discount forms. Ellison (2005) focuses on analyzing add-ons with unadvertised/unobservable prices. Shulman and Geng (2013) uses an analytical model to examine the consequences of add-on pricing when firms are both horizontally and vertically differentiated and there is a segment of boundedly rational consumers who are unaware of the add-on fees at the time of initial purchase. Geng and Shulman (2015) find that cost saving from add-on pricing may in fact result in profit loss for firms when consumers are heterogeneous in price sensitivity. Shugan et al. (2016) study the add-on pricing of a monopolist’s product lines. Lin (2017) examines firms’ product policies when they sell an add-on in addition to a base product under vertical differentiation. Geng et al. (2018) examine the interaction between an upstream firm’s add-on strategy and a downstream online platform’s distribution contract choice and examine how add-on pricing interacts with distribution contracts.
Other than studying add-on products from an economic perspective, some literature addresses add-on products from the consumer behavior perspective. For example, Bertini et al. (2008) argue that consumers draw inferences from the mere availability of add-ons, which in turn leads to significant changes in the perceived utility of the base good itself. Erat and Bhaskaran (2012) posit that consumers find a greater need for the value of add-ons when the “unrecovered” value associated with the base products is higher.

To the best knowledge of the authors, this paper provides the first study focusing on retail sales estimation and retail expansion for add-on products. We exploit the operational value of the location information of the add-on and base products in the add-on product’s retail expansion decision-making process. We are also distinguishable from the literature, as we do not presuppose that the base product and the add-on product belong to the same firm. One minor difference is that there is no price discount for purchasing the bundle in our research setting. Since the add-on products can reduce the rate at which the base product is consumed, this may account for the extra price incentive.

2.2. Retail Network Expansion and Facility Location Problem

The problem we address in this paper is to select a set of locations for expansion to maximize expected revenue. Location decision making in the retail industry is critical. There exists considerable literature dealing with retail location models (Craig 1984, Drezner et al. 2015). For example, Ghosh and Craig (1986) design a network of service centers by choosing the network size, the location of outlets, and operating characteristics simultaneously. In this paper, we solve an expansion problem. To be specific, we add new nodes onto the currently operating retail network. In addition, expansion decisions for the add-on product are restricted to the locations that sell the base product.

The problem we solve is closely related to the facility location problem, which is concerned with the optimal placement of facilities under various objectives. The literature on facility location problems dates back at least one hundred years (Weber et al. 1909), with perhaps an even longer history dating back to studies of Pierre de Fermat, Evangelista Torricelli (a student of Galileo), and Battista Cavalieri (Drezner and Hamacher 2002). There are many books dedicated to the subject (e.g., Church and Murray 2009, Daskin 2011, Drezner and Hamacher 2002, Handler and Mirchandani 1979, Love et al. 1988, Nickel and Puerto 2005, Rosenwein 1994, White and Francis 1974, Wolf 2011), as well as countless surveys and articles. Some authors have attempted to develop a taxonomy for the facility location problem (Daskin 2008, Hamacher et al. 1996) where models are broken down based on continuous/discrete/network demand and location decision, the number of facilities, the type of facilities, capacity restrictions, and objective functions.

The stream of literature most closely related to the problem studied in this paper is the competitive facility location (CFL) model where facilities set up by decision makers will compete for market share and profitability. Hotelling (1929) was the first to study location models with competition, and he considers the case of two ice cream shops picking locations for their respective storefronts where customers choose a shop based on their relative distances. There have been a number of papers dedicated to classifying CFL models (Eiselt et al. 1993, Karakitsiou 2015, Kress and Pesch 2012). The taxonomy is typically based on the spatial representation (continuous, discrete, network), the distance measure (Manhattan distance, $\ell_2$ norm, etc.), the nature of the competition (static, dynamic, simultaneous, sequential), the number of new locations (single vs. multiple), the nature of the customers (elastic demand vs. inelastic, customers come from a single location or multiple locations, customers are served by a single location or by any number of locations based on a probability distribution, etc.), and the objective type (customers prefer facilities nearby or far away), among other defining characteristics.

There are an increasing number of papers that consider the cannibalization effect among outlets of the same retailer in the marketing (Pancras et al. 2012), economics (Nishida 2014), operations management (Glaeser et al. 2019), and operations research (Bozkaya et al. 2010) literature. For example, Pancras et al. (2012) examines the impact of opening/closing an outlet on overall chain performance. Bozkaya et al. (2010) study CFL with location routing where a firm incurs costs due to vehicles having to service each location from a central warehouse. Other authors have investigated spatial interaction models assuming a concave demand function (Aboolian et al. 2007a,b, 2009). More recently, there have been papers investigating models using a leader–follower framework, where one firm decides where to locate facilities and the other firm follows with a decision (Drezner et al. 2015).

Among extant research, the derived nature of the add-on-product demand is yet to be examined. We show that managers for the add-on product can benefit from the location information of the base product. Due to the add-on nature of our focal product, we operate under the assumption that the cost of continuing operates is lower than that of ending the contract with the companies that operate the base product.

Competitive facility location models considering spatial interaction typically employ a demand
function for each customer. These models typically assume that the probability that a customer will choose a particular facility is based on the customer’s proximity to a facility, while in this study we do not start with a theoretical model assuming a representative consumer and his/her demand function. Instead, we utilize advanced machine learning methods to predict the expected sales of candidate outlets. Our way of incorporating the spatial interaction is that we add the base-product sales at nearby locations into the prediction model when spatial autocorrelation presents in the data. The formulation and the complexity of the optimization model depends on the predictive models we use.

2.3. Machine Learning for Predicting Sales and Optimizing Decision Making

Machine learning methods have been applied in marketing (Ma et al. 2016) and operations management (Ferreira et al. 2015) research, among others. Sales forecasting, or the prediction of future sales based on past historical data (Sun et al. 2008), is an important topic in operations management (Cui et al. 2018a) and plays a prominent role in business strategy (Kuo and Xue 1998). Effective sales forecasting can help the decision maker calculate the production and material costs and determine prices. For example, Ferreira et al. (2015) test various regression models, such as least squares regression, principal components regression, partial least squares regression, multiplicative (power) regression, semilogarithmic regression, and regression trees for predicting demand of future first exposure styles of an online retailer, and they find that regression trees provide the best prediction. They eventually utilize results from the prediction to optimize prices and see significant expected improvement in profit.

Investigations addressing sales forecasting problems classically employ statistical methods, such as a regression or autoregressive and moving average (ARMA) (Kuo and Xue 1998). In recent years, there have been studies focusing on predicting demand for products using advanced machine learning algorithms. For example, Kuo and Xue (1998) and Kuo (2001) propose a fuzzy neural network. Liang and Huang (2006) propose a demand forecast system taking inventory into consideration. Ferreira et al. (2015) utilize only explainable models for predicting demand of first exposure styles of an online retailer. Glaeser et al. (2019) propose a combined method to predict demand at potential locations, which utilizes both machine learning methods and econometric techniques.

In addition to exploring the methods, some studies are interested in showing the value of certain features. For example, Ma et al. (2016) identify influential features in demand forecasting by a multistage LASSO regression. The results have implications for promotional management. Cui et al. (2018a) implement multiple machine learning methods to forecast daily sales for an online apparel retailer and show that incorporating social media information improves prediction accuracy. In this paper, we find that incorporating spatially weighted base-product sales improves the prediction of the add-on-product sales when there exists spatial autocorrelation in the base-product sales.

Our work differentiates and builds upon this literature by studying the location strategy of add-on products, predicting their sales using machine learning algorithms incorporating the spatially weighted base-product sales when the Moran’s I test result is significant, and then using optimization to determine the best operational decisions. We also introduce the idea of performing a robustness check, where one takes the solutions obtained by a predictive-and-prescriptive model and evaluate their performance assuming other predictive models. This ensures that the solutions obtained are robust and can be trusted by the organization employing them for decision making.

3. Problem Context

The goal of this paper is to develop a hybrid predictive-and-optimization modeling framework for automating the retail expansion location decision-making process for companies selling add-on products where base products are sold in fixed geographic locations. We, therefore, seek to develop predictive models specifically designed for derived-demand products, and identify expansion locations based on those predictive models. We first develop a relationship between the profitability of retail outlets specifically for the add-on product and the characteristics of their locations, and then we optimize the profit of the entire network.

We learn the relationship between performance and the location characteristics from studying the past performance of existing locations. When retail managers make location expansion decisions, the attractiveness level of a location depends on the demand, buying power, and the competition level. In addition, the demand for our focal product is derived from the demand for the base product. Hence, in the predictive models of the focal product’s sales, we use the base-product sales, the weighted base-product sales in the neighborhood, the local median household income, and the local population as our predictive variables.

Before proceeding with the models, we fix notation. Let \( N = \{1, \ldots, n\} \) be the set of sites selling the base product. \( N \) is partitioned into two sets, \( S \) and \( Z \), representing the sites currently selling the add-on product...
(active sites) and the possible sites for expansion (candidate sites), respectively. Denote by \( n := |N|, s := |S|, \) and \( z := |Z| \) the cardinality of the sets. Let \( g_i \) be the total sales of the base product at site \( i \). Furthermore, let \( K \) be the number of expansion sites that the firm chooses. We assume that expanding operations to any site requires a fixed cost which does not vary from site to site, and so, given a fixed budget for expansion, the firm can choose any set of sites within the allowable budget to expand to, although this assumption can be relaxed.

In our setting, an add-on product is offered for purchase each time a customer purchases the base product. For any site \( i \in N \), let \( g_i \) be the number of times the base product is purchased at site \( i \) (i.e., demand for the base product at size \( i \)). In order to build accurate models for predicting purchases, we use demographic information common to the field studied. In particular, we incorporate median household income \( h_i \) and zip-code population \( p_i \). For notational convenience, we drop the index on parameters to represent the entire vector of values of that parameter (e.g., \( g \in \mathbb{R}^n \) is the \( n \)-dimensional vector of base-product demands for all sites).

The first step in the framework is to develop a robust and accurate predictive model for the sales of the add-on product, \( a \). More formally, for any \( N \subseteq N \), let \( f(N) \) be the amount of the add-on-product sales given the firm’s decision to offer the add-on product at locations \( N \). We seek a model \( \tilde{f} \) that approximates \( f \) using known data, which includes base-product sales at each location, median household income, and population in the area of each location.

Our models allow for the incorporation of location effects. In particular, the data we tested suggests that even for the same add-on product, location effects can exist in particular regions or not. This can be formally tested through checking for spatial autocorrelation through the Moran’s I test.

After determining the best predictive model, we formulate a general optimization model for finding the best expansion sites. Namely, we solve:

\[
\max \{ \tilde{f}(\tilde{N}) : \text{subject to } \tilde{N} \supset S, |\tilde{N}| = |S| + K \} \quad (\text{EXP})
\]

The complexity of this optimization problem depends on the underlying structure of \( \tilde{f} \). We discuss the various forms this model takes based on the predictive model employed in section 6. We also devise simple greedy heuristics for solving (EXP) given any functional form, which, despite its simplicity, works well and finds optimal solutions for all cases tested. For the problems in which we can find optimal solutions, the predictive models are of a form that when formulating (EXP), the problem reduces to a simple linear or quadratic binary optimization problem.

4. Data

In order to evaluate the efficiency of the framework we develop, we use data provided by an add-on product retailer. The retailer operates in over 1500 locations in the United States and partners with seven base-product companies operating in locations that currently do, and can in the future, offer the add-on. The focal product is an add-on product in the context of car maintenance that is sold at locations frequented by customers of the base product and offered on site at the time of purchase of the base product. When the consumer stops at the partnered stores for the base product, she faces the opportunity to purchase the add-on product as well. To be specific, the consumer can choose to buy the base product alone or to buy the add-on product as well. The consumer cannot buy the add-on product alone at the partnered outlets.

The data utilized is at-the-site transactional data that indicates each time a customer decides to purchase the add-on product when purchasing the base product. This option is available each time the customer purchases.

The over 1500 partnered outlets are not evenly distributed across the United States. Since the location effect decays with distance, we focus on existing clusters, defined as geographic areas where the density of the partnered outlets is relatively high. The number of outlets in a cluster varies from a dozen to several hundreds. We focus on the two largest clusters, as the remaining clusters have fewer than 30 outlets and are hence too small to derive any statistically reliable conclusion. The two clusters are separate. The distance between their closest borders is over 600 miles, so they have no location effects on one another. An outlet is selected if it has at least one neighbor within a fixed distance. This prespecified distance for Region 1 is 20 miles. The distance for Region 2 is 46 miles. These parameters are selected to provide a realistic definition of a cluster given the distribution of the locations.

Region 1 covers 11,878.59 square miles and has 89 existing operation sites and 228 candidate sites. The average distance between locations in this area is 17.2 miles. The locations are closely gathered as one cluster. Region 2 covers 132,549.2 square miles and has 146 existing operation sites and 162 candidate sites. The average distance between locations in this area is 54.8 miles. There are three smaller clusters in this area. Figure 1 depicts these regions. In both figures, the circles with crosses represent current locations, and circles with dots are candidate locations. There are 146 candidate locations in Region 1 and 162 candidate locations in Region 2. In both regions, the candidate sites are not evenly located. Some of them are
Figure 1  Spatial Distribution of Base-Product Retail Sites in Region 1 and Region 2 [Color figure can be viewed at wileyonlinelibrary.com]

more closely clustered, while some are far from any center. Our expansion strategy works well for both regions as demonstrated in section 6.

For each site \( i \in S \), we have the aggregated point-of-sale transaction records of the outlets during 2015. We aggregate the sales at an annual level. This is because the cost of establishing a selling point for the add-on product is high, and hence it is not economically wise to reverse the partnership once established. Therefore, we treat the profitability of a given collection of retail locations as a constant once the locations are determined. The relative values of annual sales should not vary much. Also, in mature economies, the effect of expansion decisions are expected to last for a reasonable period of time, in terms of their effectiveness for generating revenue for the firm (Bozkaya et al. 2010) Although the focal product is a new form of products with similar functionality, the economy of the base product is a business that lasts over a hundred years. In this vein, we are only interested in predicting the expected add-on-product annual sales at each location. By doing so, we also eliminate any temporal variances or seasonal trends.

In summary, for each outlet, we have the annual transaction records for both the base product and the add-on product. In addition, we have the geographic location (the latitude and the longitude in degrees), and the demographics, that is, the median income per household and the total population in a corresponding zip code.

The candidate locations \( Z \) we consider are those locations run by the base-product retailers where the add-on product is not currently sold. By the same mechanism used to gather data about current add-on sites, we have, for each location \( i \in Z \), \( h_i \) and \( p_i \).

The data from our partner company only lists base-product sales for the sites at which the add-on product is currently sold. This provides us an opportunity to test different base-product demand-sales profiles, in order to understand when the models we propose are particularly effective. We discuss this in section 6.

Summary statistics of the data is provided in Tables 1 and 2. For Region 1 and Region 2, for \( S \) and \( Z \), the data report the mean and standard deviation of the base-product sales \( g \) (not available for \( Z \) in both regions), the add-on-product sales \( a \) (not available for \( Z \) in both regions), the pair-wise distance between locations \( D \) (in miles), the household income \( h \), and the total population \( p \).

5. Predictive Models

This section provides details of the predictive models applied, which will be used to optimize expansion decisions.

5.1. Spatial Competition and Spatial Autocorrelation

Georeferenced observations generally are not independent of one another (Getis 2008) (e.g., product sales at locations in vicinity of one another). Strong evidence of the influence of spatial autocorrelation appears in a variety of settings, for example, in the consumption of gasoline (Dos Santos and Faria 2012).
sites. In order to test whether spatial autocorrelation exists in base-product sales and add-on-product sales at the current locations, we apply the Moran’s I test. This test is specifically designed to test whether or not spatial autocorrelation exists. The test statistics in this test, $I_g$, are calculated as a ratio of the product of the variable of interest and its spatial lag, with the cross-product of the variable of interest, and adjusted for the spatial weights used:

$$I_g = \frac{s \sum_{i \in S} \sum_{j \in S} w_{ij} (g_i - \bar{g})(g_j - \bar{g})}{\sum_{i \in S} (g_i - \bar{g})^2},$$

where $\bar{g}$ is the average of base-product sales over sites in $S$, in the example of the test on base-product sales. The test assumes, in the absence of spatial autocorrelation, that this statistic follows a normal distribution. In other words, if the test result is not significant, we cannot reject the null hypothesis that the observations are random. On the other hand, if the test statistic is positive and significant, it means that similar values are close to each other in the targeted space. In this case, we hypothesize that adding the information conveyed by the sales at nearby locations into the predictive models will increase predictive power because the sales are expected to have similar values to the ones at nearby locations. So, we incorporate a spatial-weight matrix once we detect significant and positive spatial autocorrelation in the product sales. We discuss the results of applying the test to the sites in both regions in section 5.3. The feature we include to encode the scaled base-product sales is an inverse distance matrix $W$ times $g$, the vector of base-product sales. In particular, for any two sites $i, j \in N$, define $w_{ij} := \frac{1}{d_{ij}}$, where $d_{ij}$ is the geographic distance (taken as the Euclidean distance) between the two sites. In this way, $Wg$ is a vector that, in coordinate $i$, contains the value $\sum_{j \in N \neq i} \frac{1}{d_{ij}} g_j$.

Whether or not there is an impact on sales based on proximity can impact the best choice of a predictive model. We find that the best predictive model for a given region depends on whether or not spatial autocorrelation exists in the base-product sales. As we will show, using or omitting a feature representing the total base-product sales scaled by the distance from the focal location has a positive impact on predictive models in the presence of spatial autocorrelation, and a negative impact in its absence.

### 5.2. Demand Prediction Models

In order to create the most accurate predictive models for estimating add-on-product sales, we apply linear regression models as a parsimonious baseline, and support vector regression (SVR) models as a more advanced and nonparametric method for predicting expected add-on-product sales. Our selection of

Hence, it is natural to think of competition from nearby stores.

In our research setting, the add-on product itself does not have direct competitors in that our focal add-on product is the only add-on product that offers a certain type of additional utility to the base product and is available at given outlets. The business-stealing effect could come from close-by outlets also selling this add-on product. In this paper, we do not model the spatial competition explicitly. The way we deal with the spatial competition depends on the existence of spatial autocorrelation, which we will illustrate later in this section.

Consumers are also able to obtain products with similar functions at other distributors. Those products serve similar purposes as our focal add-on product but are not offered in an add-on form at the outlets in our data set. In the paper, we do not consider them as competitors because they are not offered as an add-on product and they are not offered at locations in which we are interested.

The base product might suffer from spatial competition as well. We have considered the competition of the base product among current locations selling the add-on product via a spatial-weight matrix, if the spatial autocorrelation test result is significant. However, we do not consider competition from stores which are not selling the focal add-on product. We have their location information, as they are our candidate sites for expansion, but not their base-product sales. Future research could look into this issue.

Hence, before building the predictive models, we first look into the spatial interaction of both base-product sales and add-on-product sales at the current sites. In order to test whether spatial autocorrelation

| Table 1 | Summary Statistics of Industry Partner Data in Region 1 |
|---------|-------------------------------------------------------|
|         | $S$          | $Z$          |
|         | Mean | SD | Mean | SD |
| $g$     | 196,700.20  | 82,871.41 | —    | —   |
| $a$     | 3780.58     | 1149.45   | —    | —   |
| $D$     | 34.67        | 18.21     | 31.37 | 18.30 |
| $h$     | 67,240.60    | 18,033.64 | 62,505.84 | 11,894.45 |
| $p$     | 50,641.40    | 17,718.66 | 45,898.06 | 12,815.12 |

| Table 2 | Summary Statistics of Industry Partner Data in Region 2 |
|---------|-------------------------------------------------------|
|         | $S$          | $Z$          |
|         | Mean | SD | Mean | SD |
| $g$     | 195,246.30  | 93,927.96  | —    | —   |
| $a$     | 3087.76     | 1548.83    | —    | —   |
| $D$     | 163.81      | 88.08      | 150.61 | 98.03 |
| $h$     | 62,434.31   | 22,703.08  | 69,004.53 | 18,194.45 |
| $p$     | 47,208.74   | 17,986.66  | 53,736.55 | 12,815.12 |

© 2019 Production and Operations Management Society
predictive models was done for the following reasons. Linear regressions is a classical tool that has been used ubiquitously throughout the literature and provides a good baseline. SVRs are used because they provide another, more advanced mechanism for predicting sales, which allows input data to be transformed via kernel functions in order to accommodate nonlinearity. Our choice of a prediction model among other available options (e.g., random forests, neural networks) is because SVRs provide closed-form solutions (which are critical for optimal expansion decision making and which disqualifies decision trees and neural networks), and because in other papers estimating demand using predictive models, SVRs provide among the best predictive accuracy (Cui et al. 2018a, Glaeser et al. 2019). Random forests are shown to provide outstanding predictions (Cui et al. 2018a, Glaeser et al. 2019). However, random forests do not work well in our spatial setting where a limited set of features are used for prediction. Random forests rely on resampling both features and data in order to build prediction models. Features we employ cannot be expected to offer significantly more information when being resampled in building the random forests because we have four features at most. This also excludes any stepwise regression, principal component analysis, and partial least square regression. Second, the construction of $W$ relies on the entire training data. When data is resampled with replacement, it is unreasonable to have two sites at the same location. Consequently, random forests would be nearly identical to a single decision tree if we resample without replacement. In addition, we did not try polynomial regression or incorporate interaction terms because of the potential existence of the spatial-weight matrix in the model. Therefore, we employ SVR models and linear regression models for prediction. In our preliminary tests, the linear-kernel and radial-kernel SVR models outperform those with a sigmoid kernel or a polynomial kernel, and therefore we only apply the linear-kernel and radial-kernel SVR models for the remainder of this paper.

Our goal is to predict the add-on-product sales given any set of sites. In our derived demand scenario, the demand for the add-on product is derived from the demand for the base product. Hence, a portion of the predicted add-on-product sales comes from sales of the base product. In addition, the probability that a customer patronizes a facility is proportional to the attractiveness level and to a distance decay function (Berman et al. 2009), and the attractiveness of a location is related to its demographics. We include the base-product sales, the local median income per household, and the total population as the independent variables, along with $W_g$.

Price might be a strong indicator for sales. However, in our case, the price of the add-on product is fixed across the year we study. The price of the base product does not vary much across locations, especially after we take the average of the entire year. The mean of the base-product price in Region 1 is 2.21, while its standard deviation is 0.076. The mean of the base-product price in Region 2 is 2.09, while its standard deviation is 0.071. Hence, we do not include the price in our prediction models.

In summary, the predictive models we apply are as follows: (They are the ones we test in regions with spatial autocorrelated base-product sales. In regions without the spatial autocorrelated base-product sales, we remove $W_g$ from the model specifications below.)

Linear Regression: $a = \beta_1g + \beta_2W_g + \beta_3h + \beta_4p + \beta_0 + \epsilon$

Linear-Kernel SVR: $a = \omega \cdot \text{ker}_{LK}(g, W_g, h, p) + b_0 + \epsilon$

Radial-Kernel SVR: $a = \omega \cdot \text{ker}_{RK}(g, W_g, h, p) + b_0 + \epsilon$

where, $a$ is the add-on-product sales; $g$ is the base-product sales; $W_g$ is the spatially weighted average of the base-product sales; $h$ is the local median household income; $p$ is the local population; $\omega$ is the estimated weights in the SVR models; $\text{ker}_{LK}$ is the linear kernel function, that is, $(x, x')$; $\text{ker}_{RK}$ is the radial kernel function, that is, $\exp(-\gamma \|x - x'\|^2); b_0$ is the estimated constant term from the SVR models; and $\epsilon$ is the error.

5.3. Analysis

We begin with the results from the Moran’s I tests, obtained using moran.test (for the sales of the two products) and lm.moran.test (for the residuals of a linear regression model of the add-on-product sales on the base-product sales) (in package spdep in R, version R.3.4.2) (Bivand et al. 2008). Table 3 reports the Moran’s I test statistics and the $p$-values for the base-product sales, the add-on-product sales, and the residuals, in both regions of interest. Neither tests of the two product sales result in statistically significant spatial autocorrelation for Region 1, while the opposite is true for Region 2. We, therefore, conclude the following:

- Neither the base-product sales nor the add-on-product sales in Region 1 are spatially autocorrelated.
- Both the base-product sales and the add-on-product sales in Region 2 show spatial autocorrelation. In addition, after controlling for base-product sales, the residuals of the add-on-product sales show insignificant spatial autocorrelation, allowing us to conclude that the spatial autocorrelation in the focal product sales is inherited from the spatial autocorrelation of the base-product sales.
Generally speaking, if the Moran’s I test statistic is statistically significant and positive, then the spatial distribution of high/low values is more spatially clustered than would be expected if underlying spatial processes were random (ArcGIS 2017). Based on the test results in Table 3, this is the case for Region 2. We can observe three clusters on Figure 1. On the other hand, for Region 1, we cannot reject the possibility that the spatial distribution of feature values is the result of random spatial processes.

We conduct grid search over 50-time 10-fold cross validation to tune the hyperparameters in the two SVR models. For the linear-kernel SVR model, we need to tune the cost C and 𝜖. For the radial-kernel SVR model, we need to tune C, γ, and 𝜖. We compute the root mean squared error (RMSE) as a measurement of how the combination of the hyperparameter values performs. We pick the setting of the hyperparameter values with the lowest RMSE. We start the search with C in \{2^0, 2^1, \ldots, 2^{16}\}, 𝜖 ∈ \{0, 0.1, \ldots, 1\}, and γ ∈ \{10^{-2}, 10^{-4}, \ldots, 10^{-3}\}. If the optimal combination contains boundaries, then we adjust the search range to make sure the previous optimal hyperparameter value is within the search range. We iterate this procedure until the optimal hyperparameters found are inside the search ranges. We report the value of the hyperparameters we pick in Table 4.

We now turn to evaluating the impact of including Wg as an input variable in the models tested. As described below, we find that in the region exhibiting spatial autocorrelation, it is beneficial to have the spatial-weighted average base-product sales as one of the features.

Tables 5 and 6 report the root mean squared error (RMSE) and the mean absolute percent error (MAPE) based on a 50-time 10-fold cross validation for Region 1 and Region 2, respectively. For Region 1, which exhibits no spatial autocorrelation, models using three features are significantly better than those with four features. Radial-kernel SVR models slightly outperform the other two types of models. The out-of-sample MAPE of this radial-kernel SVR model is 22.9%, which is lower than those in Glaeser et al. (2019) and Ferreira et al. (2015). For Region 2, which has spatial autocorrelation, models using four features are slightly better than those with three features. Similar to Region 1, radial-kernel SVR models outperform the other two. Also, the out-of-sample MAPE of this model is 16.8%, which is lower than in Glaeser et al. (2019) and Ferreira et al. (2015). We therefore conclude that radial-kernel SVR models are best for both regions and that including Wg for Region 2 increases predictive accuracy.

Including Wg for all predictive models that exhibit spatial autocorrelation increases predictive accuracy in Region 2 but not in Region 1. It is interesting to see if this is also true for other predictive models, as a robustness check for the general hypothesis that adding the spatial-weight matrix increases predictive accuracy in the presence of spatial autocorrelation. To test this, we relax the restriction that the predictive model needs to have a closed form for the optimization and train a decision tree and a neural network. For features are significantly better than those with four features.

---

**Table 3 The Moran’s I Test Statistics and the p-Values (in parentheses)**

| Region | Add-on-product sales | Base-product sales | Residuals |
|--------|----------------------|-------------------|-----------|
| Region 1 | 0.00088 (0.2568) | 0.00459 (0.1986) | −0.1888 (0.6506) |
| Region 2 | 0.13851 (0.02739)** | 0.11442 (0.05601)* | −0.03278 (0.6291) |

**Table 4 Best Hyperparameters**

| Region | Model | C  | 𝜖  | γ  |
|--------|-------|----|----|----|
| Region 1 | Linear-kernel SVR (4 features) | 0.0625 | 0.3 | — |
| Region 1 | Radial-kernel SVR (4 features) | 2^{12} | 0.3 | 10^{-3} |
| Region 1 | Linear-kernel SVR (3 features) | 2^{14} | 0.5 | — |
| Region 1 | Radial-kernel SVR (3 features) | 2^{10} | 0.5 | 10^{-3} |
| Region 2 | Linear-kernel SVR (4 features) | 2^{8} | 0.1 | — |
| Region 2 | Radial-kernel SVR (4 features) | 2^{16} | 0.3 | 10^{-4} |
| Region 2 | Linear-kernel SVR (3 features) | 2^{6} | 0.1 | — |
| Region 2 | Radial-kernel SVR (3 features) | 2^{10} | 0.2 | 10^{-3} |

**Table 5 Predictive Models Performance in Region 1**

| Region 1 | Linear-kernel SVR | Radial-kernel SVR | Linear regression |
|---------|-------------------|-------------------|-------------------|
| Mean RMSE (4 features) | 751.40 | 748.99 | 835.75 |
| SD RMSE (4 features) | 14.7 | 16.2 | 18.8 |
| Mean MAPE (4 features) | 24.51% | 24.19% | 23.35% |
| SD MAPE (4 features) | 0.80% | 0.91% | 1.08% |
| Mean RMSE (3 features) | 719.99 | 718.78 | 733.80 |
| SD RMSE (3 features) | 9.3 | 9.6 | 10.4 |
| Mean MAPE (3 features) | 23.33% | 22.90% | 23.67% |
| SD MAPE (3 features) | 0.29% | 0.34% | 0.40% |

**Table 6 Predictive Models Performance in Region 2**

| Region 2 | Linear-kernel SVR | Radial-kernel SVR | Linear regression |
|---------|-------------------|-------------------|-------------------|
| Mean RMSE (4 features) | 771.38 | 756.47 | 788.43 |
| SD RMSE (4 features) | 4.8 | 8.4 | 8.5 |
| Mean MAPE (4 features) | 17.06% | 18.84% | 18.02% |
| SD MAPE (4 features) | 0.31% | 0.27% | 0.40% |
| Mean RMSE (3 features) | 772.75 | 762.03 | 791.83 |
| SD RMSE (3 features) | 2.82 | 6.40 | 6.60 |
| Mean MAPE (3 features) | 16.37% | 16.26% | 17.92% |
| SD MAPE (3 features) | 0.16% | 0.22% | 0.25% |
this robustness check, we run a 10-fold cross validation 50 times for both models using default settings in R. The means and standard deviations of the RMSEs are reported in Table 7. For the decision tree, the result is consistent with our hypothesis because adding \( W_g \) improves accuracy for Region 2 but not for Region 1. For the neural network, the result is consistent for Region 1, which means that the model without \( W_g \) for Region 1 has higher accuracy. This is, however, also true for Region 2, which is against the hypothesis. Note that, in each comparison of means, the difference in means is small relative to the standard deviation. Since we conduct 50-repetition 10-fold cross validation to obtain the estimation of the test error in each case, we expect the true test error is very close to the value we report here. Although the differences are not statistically significant, the trend is apparent, since, in general, adding \( W_g \) in Region 1 degrades the predictive power, and the opposite is true in Region 2.

In the optimization models, we use for determining expansion decisions, we will fix the best predictive models for each region and find expansion sites that will maximize expected sales. The important insight from the analysis conducted in this section is that in the presence of spatial autocorrelation, including the weighted matrix is critical for good predictive performance.

6. Expansion Optimization

Equipped with a high-performing predictive model, we now turn to determining the optimal set of expansion sites to maximize expected sales for add-on products given the set of candidate locations. The form of the optimization model depends entirely on the choice of predictive models. Simpler predictive models result in tractable optimization problems, and more complex models result in highly nonlinear optimization models that require heuristics.

As described in section 3, the optimization problem, in its most general form, is (EXP). This section first explores how this optimization model is specified for the various predictive models employed and then presents a general-purpose greedy heuristic for finding expansion decisions. We then describe how the solutions we obtain perform in practice on the data made available from our industry partners.

6.1. Models with No Spatial Effects

Consider a predictive model \( f \) taking the form LM-NS

\[
\hat{f}(\mathbf{N}) = \sum_{i \in \mathbf{N}} l_i \quad \text{(LM - NS)}
\]

This results from, for example, a linear regression model or a linear-kernel SVR model. In the former,

\[
l_i = \beta_1 g_i + \beta_3 h_i + \beta_4 p_i + \beta_0,
\]

and for a linear-kernel SVR, we have

\[
l_i = \sum_{k=1}^{m} C_k (V_{k,1} g_i + V_{k,3} h_i + V_{k,4} p_i) + b_0, \quad \text{where } C_k \text{ and } V_{k,j} \ (j = 1, 2, 4 \text{ in our example}) \text{ are the standard parameters in SVR models, and } m \text{ is the number of support vectors of a specific SVR model. In either event, (EXP) reduces to}
\]

\[
\max \sum_{i \in \mathbf{N}} l_i x_i \\
\text{s.t.} \sum_{i \in \mathbf{Z}} x_i = K \\
x_i = 1, \quad \forall i \in \mathbf{S} \\
x_i \in \{0, 1\}, \quad \forall i \in \mathbf{Z}.
\]

Notice that this expansion-optimization problem does not involve any association with existing sites at which the add-on product is sold. Furthermore, this can be solved by simply sorting the predicted add-on-product sales at the candidate sites in a non-increasing order of \( l_i \) and choosing the first \( K \) sites to expand to.

6.2. Linear Models with Spatial Effects

Suppose we incorporate the weight matrix \( W \) in a linear regression or linear-kernel SVR. Given a collection of sites \( \mathbf{N} \) offering the add-on product, the model will predict LM-WS

\[
\hat{f}(\mathbf{N}) = \sum_{i \in \mathbf{N}} l_i x_i + \sum_{i \in \mathbf{N}} \sum_{j \notin \mathbf{N} \cap i} e_{i,j} x_i x_j \quad \text{(LM - WS)}
\]

as the total sales of the add-on product. For a linear regression model, we have

\[
l_i = \beta_1 g_i + \beta_3 h_i + \beta_4 p_i + \beta_0, e_{i,j} = \beta_2 \sum_{j \neq i} \frac{1}{d_{ij}} g_j,
\]

and for a linear-kernel SVR, we have

\[
l_i = \sum_{k=1}^{m} C_k (V_{k,1} g_i + V_{k,3} h_i + V_{k,4} p_i) + b_0, e_{i,j} = \sum_{k=1}^{m} C_k \left( V_{k,2} \left( \sum_{j \neq i} \frac{1}{d_{ij}} g_j \right) \right).
\]
The expansion-optimization model can then be written as

$$\max \sum_{i \in \mathbb{N}} l_i x_i + \sum_{i \in \mathbb{N}} \sum_{j \in \mathbb{N}, j \neq i} e_{ij} x_i x_j$$

s.t. \( \sum_{i \in \mathbb{Z}} x_i = K \)

\( x_i = 1, \quad \forall i \in \mathbb{S} \)

\( x_i \in \{0, 1\}, \quad \forall i \in \mathbb{Z}. \)

The ensuing model is, therefore, a cardinality-constrained binary quadratic optimization problem (Bertsimas and Shiода 2009), which has a large literature containing dedicated methods. For this paper and the data set we use, the optimization model was easily solved by a commercial integer programming solver (GUROBI 7.5.1).

### 6.3. More General Models

Nonlinear and nonparametric predictive models often result in superior predictive power, as in the case for Region 2. The resulting optimization models are more complicated. For example, consider a radial-kernel SVR model with a spatial-weight matrix as a part of a feature. Given a collection of sites \( N \), the function form is more complicated than the linear or quadratic form above. We, therefore, devise a heuristic to solve the ensuing optimization model.

Given any function \( f(\tilde{N}) \) and a number of expansion sites \( K \), one can devise a host of heuristics to plan expansion. We suggest the following simple heuristic that can apply to any analytical function.

Start with no candidate sites, and let \( \tilde{N} = S \). Among all sites in \( Z \), choose the site \( \Phi \) that, if added to \( \tilde{N} \), the increase from \( f(\tilde{N}) \) to \( f(\tilde{N} = \{\tilde{N}, \Phi\}) \) is the greatest. Then, remove \( \Phi \) from \( Z \), and add it to \( \tilde{N} \). Continue until \( |N| - |S| = K \).

This heuristic does not guarantee an optimal selection of sites. It will guarantee an optimal selection when employing models with no spatial-weight matrix. Also, in our experimental results, this heuristic was able to match the optimal expansions for all linear models with a spatial-weight matrix. There is no guarantee that this will happen for other data sets, but, given its performance, we use this simple heuristic for the more complicated nonlinear radial-kernel SVR model as well.

### 6.4. Analysis of Results

In order to evaluate the performance of the expansion-optimization framework, we compare with a standard practice that an add-on retailer might employ, which determines expansion by choosing those sites for which the base-product sales are highest. We refer to this expansion as the baseline algorithm, or BM. For any predictive model \( P \), we let EO-P be the expansion-optimization algorithm using predictive model \( P \).

In order to simulate base-product sales in candidate locations, for each \( i \in \mathbb{Z} \), we draw, independently, \( x_i \) from a normal distribution with mean \( \mu \) and standard deviation \( \sigma \). \( \mu \) is taken as the mean of base-product sales in \( S \), and we test for \( \sigma = 4^{\nu} \), for \( \nu \in \{2, 4, 6\} \), in order to evaluate differences in performance for different base-product sales distributions. If \( g_i \) is drawn to be below the minimum base-product sales in the region, we set \( g_i \) equal to that minimum. We simulate this demand 10 times per \( \sigma \) for the following experiments. We test for \( K \in \{1, 2, \ldots, 20\} \). We additionally report the result of \( K = 40 \) to show the trend when expanding to more locations.

We first compare, for the best predictive model in each region, \( P^* \), what the average predicted total sales of add-on product will be, using both BM and EO – \( P^* \), for each \( K \) tested. \( P^* \) represents radial-kernel SVR models for both regions, with \( Wg \) included as an independent variable in Region 2.

Table 8 reports results comparing the expansion decisions. Each row corresponds to varying \( K \). The columns report the average increase in additional sales that results from employing EO – \( P^* \) over BM. In particular, let \( \zeta_i^\phi \) be the total sales of add-on product estimated by \( P^* \) without any expansion. For any \( K \), let \( \zeta_i^K \) be the total add-on-product sales achieved based on EO – \( P^* \) and \( \zeta_i^K \) be the total add-on-product sales achieved based on BM. The table reports \( \frac{\zeta_i^K - \zeta_i^\phi}{\zeta_i^K} \times 100\% \). The higher this value, the more gains that are realized by using the framework we propose.

### 6.5. Managerial Insights

Our results provide clear evidence that making expansion decisions for add-on products should not be done without more readily understanding how sales will be affected by the expansion decisions made. A standard expansion approach that only considers base-product sales can lead to far worse expansion decisions than the more complex, but effective, predictive-and-optimization framework developed in this paper.

As shown in Table 8, when \( K \) is small, we realize larger gains (percent increases) when there is no spatial autocorrelation (Region 1). However, this switches when \( K \) increases. This indicates that our suggested approach is particularly useful when the retailer is considering a large expansion strategy in the presence of spatial autocorrelation. Also, in general, gains decrease with \( K \) when there is no spatial autocorrelation (Region 1), and increase with \( K \) where there is spatial autocorrelation (Region 2).

Table 8 exhibits the gains the add-on-product seller can realize by employing our strategy instead of the
baseline policy. When comparing the results obtained with BM, the add-on-product seller can realize additional sales of over 5%, simply by better expansion decision making. This depends, however, on the distribution of the demand for base products (i.e., $\nu$). For example, if $K = 10$ in Region 1, and the variance is low ($\nu = 2$), we expect to achieve a 4.00% increase in sales simply by choosing a better set of locations. As the variance increases, the difference between BM and EO – $P^*$ becomes smaller (the percent increase is 1.75% when $\nu = 6$). For the same $K$, the percent increases get smaller when $\nu$ is larger in Region 2 as well. Our proposed framework provides much better expansion solutions when base-product sales at candidate sites do not vary much. In such a case, relying on base-product sales does not provide enough information. Hence, the information from the neighborhood become more important. Even when variation does exist among candidate sites, using a simple expansion model relying only on base-product sales will not achieve optimal results. In order to capture complex relationships between demographic information and the interplay between the demand from neighboring base-product retail sites, using spatial-weight matrices and local demographic information can improve predictive power, thereby allowing optimization models to automate the selection of desirable expansion sites.

The increased sales of the retail chain could come from two sources. One is from new customers whose utility from buying the base and add-on product bundle increases to greater than zero due to the reduced travel cost or the increased preference toward the product bundle. The other source is more frequent purchasing from existing customers. To the existing customers who buy the product bundle more frequently, the costs of going to the store and buying the base and add-on product bundle reduces or are the same as those when the new sites are not added. It is not certain that starting selling the add-on product would add sales for a single store, but it would add sales to the entire chain if the sites are chosen prudently. Comparing the sales at existing sites before and after expansion, some of their sales increase. This indicates that the increased sales are larger than the sales cannibalized by new outlets. Some outlets see decreased sales, indicating that the new sales are not large enough to cover the cannibalized.

As far as the net increase in sales, if $K = 20$ and $\nu = 2$, the difference in additional sales of the add-on product in Region 1 is 3429, and in Region 2 it is 3678. For a tight-margin business, this may have significant impacts on its bottom line, particularly because this increase in sales does not require anything more than selecting better expansion sites.

It is also interesting to view the geographic dispersion and selection of sites, comparing the expansion decisions made using BM with EO – $P^*$ for both regions. This is depicted in Figure 2 (for Region 1) and Figure 3 (for Region 2) for one instance of $K = 20$ generated with $\nu = 6$. Each figure displays a point for every site at its geographic location. The current sites

| Table 8 Average Percent Increase in Additional Sales By Using EO – $P^*$ over BM in Both Regions for Various Base-Product Demand Distributions |
|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| $K$  | $\nu = 2$ | $\nu = 4$ | $\nu = 6$ | $\nu = 2$ | $\nu = 4$ | $\nu = 6$ |
| 1   | 5.48% | 4.45% | 2.65% | 3.18% | 3.45% | 2.11% |
| 2   | 4.58% | 4.39% | 2.65% | 4.43% | 4.21% | 2.26% |
| 3   | 4.64% | 4.22% | 2.47% | 5.50% | 4.54% | 2.43% |
| 4   | 4.52% | 4.03% | 2.25% | 5.63% | 5.64% | 2.67% |
| 5   | 4.59% | 4.07% | 2.17% | 5.51% | 5.54% | 2.54% |
| 6   | 4.42% | 3.95% | 2.07% | 5.19% | 5.57% | 2.64% |
| 7   | 4.30% | 3.78% | 1.93% | 4.88% | 5.98% | 2.58% |
| 8   | 4.21% | 3.66% | 1.84% | 4.69% | 5.67% | 2.43% |
| 9   | 4.18% | 3.65% | 1.77% | 4.55% | 5.46% | 2.62% |
| 10  | 4.00% | 3.57% | 1.75% | 4.44% | 5.32% | 2.78% |
| 11  | 3.89% | 3.51% | 1.70% | 4.38% | 5.10% | 3.03% |
| 12  | 3.85% | 3.49% | 1.63% | 4.50% | 5.08% | 2.96% |
| 13  | 3.79% | 3.44% | 1.59% | 4.68% | 4.96% | 2.84% |
| 14  | 3.72% | 3.43% | 1.51% | 4.53% | 4.79% | 2.86% |
| 15  | 3.61% | 3.34% | 1.52% | 4.39% | 4.60% | 2.96% |
| 16  | 3.54% | 3.29% | 1.51% | 4.28% | 4.45% | 2.95% |
| 17  | 3.49% | 3.21% | 1.48% | 4.27% | 4.35% | 3.02% |
| 18  | 3.44% | 3.17% | 1.44% | 4.22% | 4.43% | 2.91% |
| 19  | 3.41% | 3.13% | 1.43% | 4.33% | 4.54% | 2.95% |
| 20  | 3.38% | 3.10% | 1.39% | 4.62% | 4.64% | 2.94% |
| ... | ... | ... | ... | ... | ... | ...
| 40  | 2.41% | 2.16% | 0.99% | 3.58% | 3.80% | 2.39% |
are squares and colored orange, and the candidate sites are circles, colored blue if chosen for expansion and gray otherwise. Furthermore, the points corresponding to the candidate sites are gradient sized to depict relative base-product sales (higher sales, larger circle). In both figures, the left plot depicts the solution obtained using BM, and the right plot depicts the solution obtained using EO – P∗.

These figures depict more clearly how the expansion decisions differ. In particular, we see that the expansion decisions made through EO – P∗ take advantage of local clusters, and, although selection
leans towards those sites with large base-product sales, it is often better to consider location impact over only base-product sales.

It is clear that using the optimization models for a fixed predictive model will always result in better results than using the baseline algorithm when evaluated on the predictive model used within the optimization framework. In order to ensure that the results are robust, we suggest the following procedure. Take the solutions obtained by EO – P’ and those obtained by BM, and compare them using different predictive models. Since different predictive models provide estimation using varied structure, it is not clear whether a solution that is superior for one predictive model will also be superior for other predictive models, built using alternative structural assumptions. If a solution is better across multiple predictive models, this makes the solution even more desirable.

Consider Tables 9 and 10. For both regions and each \( v \), we use the solution from EO – P’ and evaluate the number of sales predicted by other predictive models, denoted as \( P \), to see if the expansion decision remains favorable over BM when also evaluated by \( P \). This will enable managers to have more confidence in the solutions.

Let LR be the linear regression model, LK be the linear-kernel SVR model, and RK be the radial-kernel SVR model, where we include \( W_g \) as a predictor in the models only for Region 2. Each entry in the tables corresponds to the percent increase in additional expected sales realized from using the solution obtained by EO – P’ over the solution obtained by BM. Interestingly, the solutions found by EO – P’ are almost always better than BM (i.e., most of the numbers in Tables 9 and 10 are positive), independent of the predictive model on which it is evaluated. The exception is when we plug EO – P’ into LK in Region 2, but this diminishes as we increase the number of expansion sites.

One potential reason for the relatively poorly predicted quality of EO – P’ (the solution obtained by optimizing over the radial-kernel predictive model) when evaluated on LK in Region 2 is because of relatively poor predictive performance of LK in Region 2. Since the prediction is of poor quality using LK, plugging either the solution obtained by BM or by EO – P’ into LK does not provide strong estimates for expected add-on-product sales. We also observe that the expected increase in sales from evaluating the solution obtained by EO – P’ over BM in LR is more significant in Region 2 than it is in Region 1. This can perhaps be attributed to spatial autocorrelation—expanding only based on base-product sales in Region 2 is significantly worse than expanding based on a model containing spatial information even under a simplified distribution.

The demand distribution tested hints that for small variance in base-product sales at existing sites, it becomes more important to utilize the predictive-and-optimization framework presented in this paper for planning retail expansion for add-on products.

### Table 9 Robustness Check for Solutions Obtained by EO – P’ for Region 1

| \( K \) | \( v = 2 \) | \( v = 4 \) | \( v = 6 \) |
|---|---|---|---|
| | LR | LK | RK | LR | LK | RK | LR | LK | RK |
| 1 | 3.34% | 4.92% | 5.48% | 2.61% | 3.94% | 4.45% | 0.61% | 1.98% | 2.65% |
| 2 | 2.89% | 4.27% | 4.58% | 2.69% | 4.04% | 4.39% | 0.90% | 2.27% | 2.65% |
| 3 | 2.97% | 4.39% | 4.64% | 2.65% | 4.03% | 4.22% | 1.01% | 2.21% | 2.47% |
| 4 | 2.92% | 4.33% | 4.52% | 2.53% | 3.85% | 4.03% | 0.90% | 2.08% | 2.25% |
| 5 | 2.98% | 4.43% | 4.59% | 2.56% | 3.90% | 4.07% | 0.85% | 2.04% | 2.17% |
| 6 | 2.96% | 4.43% | 4.42% | 2.57% | 3.94% | 3.95% | 0.81% | 1.96% | 2.07% |
| 7 | 2.90% | 4.35% | 4.30% | 2.46% | 3.78% | 3.78% | 0.73% | 1.82% | 1.93% |
| 8 | 2.86% | 4.29% | 4.21% | 2.41% | 3.71% | 3.66% | 0.72% | 1.81% | 1.84% |
| 9 | 2.88% | 4.33% | 4.18% | 2.43% | 3.75% | 3.65% | 0.70% | 1.76% | 1.77% |
| 10 | 2.79% | 4.22% | 4.00% | 2.40% | 3.70% | 3.57% | 0.68% | 1.73% | 1.75% |
| 11 | 2.71% | 4.08% | 3.89% | 2.37% | 3.65% | 3.51% | 0.66% | 1.66% | 1.70% |
| 12 | 2.67% | 4.02% | 3.85% | 2.36% | 3.64% | 3.49% | 0.64% | 1.61% | 1.63% |
| 13 | 2.62% | 3.94% | 3.79% | 2.33% | 3.58% | 3.44% | 0.63% | 1.55% | 1.59% |
| 14 | 2.56% | 3.84% | 3.72% | 2.31% | 3.57% | 3.43% | 0.58% | 1.48% | 1.51% |
| 15 | 2.49% | 3.75% | 3.61% | 2.25% | 3.46% | 3.34% | 0.61% | 1.49% | 1.52% |
| 16 | 2.44% | 3.67% | 3.54% | 2.21% | 3.40% | 3.29% | 0.61% | 1.47% | 1.51% |
| 17 | 2.40% | 3.60% | 3.49% | 2.16% | 3.32% | 3.21% | 0.60% | 1.44% | 1.48% |
| 18 | 2.37% | 3.55% | 3.44% | 2.13% | 3.26% | 3.17% | 0.59% | 1.43% | 1.44% |
| 19 | 2.34% | 3.50% | 3.41% | 2.09% | 3.19% | 3.13% | 0.60% | 1.42% | 1.43% |
| 20 | 2.31% | 3.45% | 3.38% | 2.06% | 3.16% | 3.10% | 0.57% | 1.37% | 1.39% |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| 40 | 1.64% | 2.42% | 2.41% | 1.44% | 2.18% | 2.16% | 0.43% | 0.99% | 0.99% |
The demand distribution is however not realistic—a *shapiro* test rejects the null hypothesis that base-product sales follow a normal distribution in either region.

In order to test whether the results are consistent with a more realistic demand distribution, we run the following procedure to generate base-product sales at candidate locations. Given $s$ observed base-product sales in a region, we sort the $s$ values in ascending order and obtain an ordered list $\{s_{(1)}; \ldots; s_{(s)}\}$. For each candidate location, we randomly draw an integer $i$ from $[1, s - 1]$. Then, we randomly draw a real
number \( q \) between 0 and 1. Lastly, we take the number \( g_{(i)} + \chi \times (g_{(i+1)} - g_{(i)}) \) as the base-product sales at the candidate location. This procedure produces values following a distribution nearly identical to that of the observed distribution of base-product sales. In order to test a wide-range of demand scenarios and eliminate outliers, we can truncate the smallest and largest \( q \% \) of the \( s \) observed values so that the distribution is more concentrated around the mean (i.e., generate \( i \) from \( [s \times \frac{q}{100}, \ldots, s \cdot \frac{100-q}{100}] \)). For any such \( q \), we refer to this generation scheme as grw(\( q \)). Although this procedure overfits to the existing data, it provides an alternative and perhaps more realistic distribution for testing our algorithms.

We use this method to generate base-product sales at candidate locations, implement our framework and evaluate our framework’s performance. Figure 4 depicts the increase in sales from utilizing EO – P over BM for generation scheme grw(\( q \)) for \( q = 5, 10, 15, \ldots, 30 \) in Region 1 (left) and Region 2 (right) using the best predictive model for each region, for \( K = 1, \ldots, 20 \). The figure readily exhibits that if the base-product sales in candidate locations are more closely distributed, the gains realized by the framework in this paper increases and can be significant. As the spread of base-product sales around the mean increases, the gains diminish but are still positive. Tables 11 and 12 report the robustness check for grw(10), grw(20), and grw(30). The trend mirrors that for the previous generation scheme—for more centralized demand around the mean, the power of the framework is more substantial and should be utilized when planning add-on-product expansion.

### 7. Conclusion and Future Work

In this paper, we provide a novel framework for jointly applying predictive modeling and optimization to expansion decisions for add-on products. Through collaboration with an industrial partner, we use point-of-sales data to generate predictive models and formulate optimization models over those predictive models to automate expansion decisions. We find that the spatially weighted base-product sales are valuable in prediction when there is statistically significant spatial autocorrelation for three out of four predictive models we test. Our results indicate that making expansion decisions using the framework described in this paper can result in substantial increases in expected sales over a baseline algorithm that only takes into account base-product sales at the targeted expanding location. Specifically, if the expansion scale is small, that is, \( K \) is small, the expected gain is larger for regions without spatial autocorrelation. If the expansion scale is large, the expected gain would be larger for regions with spatial autocorrelation. In general, the gains decrease with the expansion scale, \( K \), for regions without spatial autocorrelation, and increase for regions with spatial autocorrelation. Our framework offers greater improvements over the baseline model if the variance of the base-product sales are smaller. As shown on Figures 2 and 3, our

### Table 12: Robustness Check Using a Generator Based on Real-World Data in Region 2

| \( K \) | \( q = 10 \) | \( q = 20 \) | \( q = 30 \) |
|-------|--------|--------|--------|
|       | LR     | LK     | RK     | LR     | LK     | RK     | LR     | LK     | RK     |
| 1     | 0.23%  | 1.36%  | 1.11%  | 1.78%  | 1.15%  | 1.65%  | 1.70%  | 1.17%  | 1.38%  |
| 2     | 0.73%  | 0.61%  | 1.48%  | 2.53%  | -0.48% | 1.21%  | 2.01%  | -0.03% | 1.28%  |
| 3     | 1.29%  | -0.19% | 1.46%  | 2.17%  | -0.29% | 1.14%  | 2.53%  | 0.01%  | 1.69%  |
| 4     | 1.30%  | -0.34% | 1.19%  | 2.27%  | 0.22%  | 1.45%  | 2.74%  | -0.20% | 1.53%  |
| 5     | 1.14%  | -0.32% | 0.95%  | 2.21%  | 0.10%  | 1.30%  | 2.47%  | -0.39% | 1.42%  |
| 6     | 1.36%  | -0.40% | 0.99%  | 2.87%  | 0.13%  | 1.68%  | 2.60%  | -0.22% | 1.41%  |
| 7     | 1.32%  | -0.50% | 0.82%  | 2.91%  | 0.00%  | 1.50%  | 2.37%  | -0.29% | 1.40%  |
| 8     | 1.28%  | -0.47% | 0.75%  | 2.19%  | -0.12% | 1.28%  | 2.31%  | -0.05% | 1.50%  |
| 9     | 1.95%  | -0.50% | 0.83%  | 2.02%  | 0.11%  | 1.27%  | 2.42%  | 0.14%  | 1.71%  |
| 10    | 1.98%  | -0.39% | 1.11%  | 2.12%  | 0.08%  | 1.37%  | 2.69%  | 0.07%  | 1.70%  |
| 11    | 1.31%  | -0.45% | 0.89%  | 1.60%  | 0.04%  | 1.14%  | 2.63%  | 0.06%  | 1.62%  |
| 12    | 1.39%  | -0.38% | 0.89%  | 1.68%  | -0.18% | 1.09%  | 2.49%  | -0.05% | 1.70%  |
| 13    | 1.53%  | -0.37% | 1.01%  | 1.35%  | -0.14% | 0.94%  | 2.50%  | -0.02% | 1.57%  |
| 14    | 1.23%  | -0.44% | 0.91%  | 1.17%  | -0.05% | 0.96%  | 2.69%  | -0.12% | 1.53%  |
| 15    | 1.21%  | -0.53% | 0.78%  | 1.16%  | -0.09% | 0.89%  | 2.51%  | -0.13% | 1.46%  |
| 16    | 1.49%  | -0.55% | 0.81%  | 1.27%  | -0.05% | 1.00%  | 2.44%  | -0.13% | 1.43%  |
| 17    | 1.35%  | -0.54% | 0.66%  | 1.45%  | -0.19% | 0.91%  | 2.18%  | -0.07% | 1.36%  |
| 18    | 1.41%  | -0.53% | 0.66%  | 1.60%  | -0.24% | 0.90%  | 2.08%  | -0.12% | 1.38%  |
| 19    | 1.38%  | -0.44% | 0.55%  | 1.59%  | -0.29% | 0.87%  | 2.09%  | -0.17% | 1.30%  |
| 20    | 1.49%  | -0.37% | 0.56%  | 1.63%  | -0.37% | 0.87%  | 1.81%  | -0.28% | 1.17%  |
| 21    | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
| 40    | 0.72%  | -0.29% | 0.39%  | 1.12%  | -0.38% | 0.55%  | 1.54%  | -0.34% | 0.71%  |
framework will not always pick the sites with the largest base-product sales, but instead considers the location effect so as to optimize expected sales on the entire retail chain. We also evaluate our optimal solutions using other predictive models. In most (with one exception) cases, our solutions outperform the baseline model solutions. Lastly, we test the effectiveness of the proposed framework under close to realistic base-product demand distribution. As shown in Tables 11 and 12, the above findings still hold.

This work has myriad possible directions for expansion. First, if given more complete and complex data, one could build more elaborate predictive models that provide a more fine-grained picture of what drives sales for add-on products. For example, a more detailed customer-level data on purchase behavior at different locations can be incorporated to develop improved predictive models. This can also drive the prescriptive model; for instance, expanding to a new site frequented by existing customers may increase convenience for them but may limit market expansion to access new customers. Considering such nuances in customer behavior can lead to better location choice models. Second, these results should be replicated to more settings where expansion is limited by geographic constraints. Third, the dedicated solution methodology for solving optimization models resulting from nonlinear predictive models should be investigated, so that expansion decisions are made optimally as opposed to by heuristics. Although the heuristic performs well for the company that we partnered with, there is no proof of optimality, and it is possible, and probable, that better collections of expansion sites exist. Fourth, there are many other interesting situations where add-on products are offered that do not fit precisely within our framework. Developing models for expansion in these settings is an interesting research direction. For example, if adding and removing locations in the add-on product’s retail network is feasible due to lower fixed costs for expansion, temporal effects may play a larger role in determining optimal expansion decisions. Finally, it will be interesting to investigate what other problem settings could adapt spatially weighted matrices as predictors to increase predictive power, given the increase in predictive accuracy realized in this setting.

References
Aboolian, R., O. Berman, D. Krass. 2007a. Competitive facility location and design problem. Eur. J. Oper. Res. 182(1): 40–62.
Aboolian, R., O. Berman, D. Krass. 2007b. Competitive facility location model with concave demand. Eur. J. Oper. Res. 181 (2): 598–619.
Aboolian, R., O. Berman, D. Krass. 2009. Efficient solution approaches for a discrete multi-facility competitive interaction model. Ann. Oper. Res. 167(1): 297–306.
ArcGIS. 2017. How spatial autocorrelation (Global Moran’s I) works. Available at http://pro.arcgis.com/en/pro-app/tool-reference/spatial-statistics/h-how-spatial-autocorrelation-moran-s-i-spatial-st.htm (accessed date March 29, 2018).
Berman, O., T. Dreznier, Z. Dreznier, D. Krass. 2009. Modeling competitive facility location problems: New approaches and
results. M. Oskoorouchi, eds. TutORials in Operations Research. INFORMS, San Diego, CA, 156–181.

Bertini, M., E. Ofek, D. Ariely. 2008. The impact of add-on features on consumer product evaluations. J. Consum. Res. 36(1): 17–28.

Bertsimas, D., R. Shioda. 2009. Algorithm for cardinality-constrained quadratic optimization. Comput. Optim. Appl. 43(1): 1–22.

Bivand, R. S., E. J. Pebesma, V. Gómez-Rubio. 2008. Applied Spatial Data Analysis with R, volume 747248717. Springer, New York, NY.

Bozkaya, B., S. Yanik, S. Balciyso. 2010. A gis-based optimization framework for competitive multi-facility location-routing problem. Netw. Spat. Econ. 10(3): 297–320.

Brown, L. A., M. A. Brown, C. S. Craig. 1981. Innovation diffusion and entrepreneurial activity in a spatial context: Conceptual models and related case studies. Res. Market: Res. Ann. 4: 69–115.

Church, R. L., A. T. Murray. 2009. Business Site Selection, Location Analysis, and GIS. John Wiley & Sons, Hoboken, NJ, 259–280.

Craig, C. S. 1984. Models of the retail location process: A review. J. Retail. 60(1): 5–36.

Cui, R., S. Gallino, A. Moreno, D. J. Zhang. 2018a. The operational value of social media information. Prod. Oper. Manag. 27(10): 1749–1769.

Cui, Y., I. Duenyas, O. Sahin. 2018b. Unbundling of ancillary service: How does price discrimination of main service matter? Manuf. Serv. Oper. Manag. 20(3): 455–466.

Daskin, M. S. 2008. What you should know about location modeling. Nav. Res. Log. 55(4): 283–294, ISSN 1520-6750. https://doi.org/10.1002/nav.20284

Daskin, M. S. 2011. Network and Discrete Location: Models, Algorithms, and Applications. John Wiley & Sons, New York, NY.

Dos Santos, G. F., W. R. Faria. 2012. Spatial Panel Data Models and Fuel Demand in Brazil. Available at http://www.usp.br/nereus/wp-content/uploads/TD_Nereus_10_2012.pdf (accessed date March 29, 2019).

Drezner, Z., H. W. Hamacher. 2002. Facility Location: Applications and Theory. Springer, New York.

Drezner, T., Z. Drezner, P. Kalcynski. 2015. A leader–follower model for discrete competitive facility location. Comput. Oper. Res. 64: 51–59.

Eiselt, H. A., G. Laporte, J. F. Thisse. 1993. Competitive location models: A framework and bibliography. Transp. Sci. 27(1): 44–54.

Ellison, G. 2005. A model of add-on pricing. Q. J. Econ. 120(2): 585–637.

Erat, S., S. R. Bhaskaran. 2012. Consumer mental accounts and implications to selling base products and add-ons. Market. Sci. 31(5): 801–818.

Eze, F. J., B. E. Odigbo, J. A. Ufot. 2015. The correlation between business location and consumers patronage: Implications for business policy decisions. Br. J. Econ. Manage. Trade 8(4): 294–304.

Ferreira, K. J., B. H. A. Lee, D. Simchi-Levi. 2015. Analytics for an online retailer: Demand forecasting and price optimization. Manuf. Serv. Oper. Manag. 18(1): 69–88.

Geng, X., J. D. Shulman. 2015. How costs and heterogeneous consumer price sensitivity interact with add-on pricing. Prod. Oper. Manag. 24(12): 1870–1882.

Geng, X., Y. Tan, L. Wei. 2018. How add-on pricing interacts with distribution contracts. Prod. Oper. Manag. 27(4): 605–623.

Getis, A. 2008. A history of the concept of spatial autocorrelation: A geographer’s perspective. Geogr. Anal. 40(3): 297–309.

Ghosh, A., C. S. Craig. 1986. An approach to determining optimal locations for new services. J. Mark. Res. 23: 354–362.
Shulman, J. D., X. Geng. 2013. Add-on pricing by asymmetric firms. *Management Sci.* 59(4): 899–917.

Sun, Z. L., T. M. Choi, K. F. Au, Y. Yu. 2008. Sales forecasting using extreme learning machine with applications in fashion retailing. *Decis. Support Syst.* 46(1): 411–419.

Weber, A. Über den Standort der Industrien, Tübingen. 1909. *Theory of the Location of Industries* University of Chicago Press, Chicago, IL.

White, J. A., R. L. Francis. 1974. *Facility Layout and Location: An Analytical Approach.* Prentice-Hall, Englewood Cliffs, NJ.

Wolf, G. W. 2011. Facility location: Concepts, models, algorithms and case studies. series: Contributions to management science. *Int. J. Geogr. Inf. Sci.* 25(2): 331–333. http://dblp.uni-trier.de/db/journals/gis/gis25.html#Wolf11