Titre: Title: Improving demand forecasting for customers with missing downstream data in intermittent demand supply chains with supervised multivariate clustering

Auteurs: Authors: Corey Ducharme, Bruno Agard, & Martin Trépanier

Date: 2024

Type: Article de revue / Article

Référence: Citation: Ducharme, C., Agard, B., & Trépanier, M. (2024). Improving demand forecasting for customers with missing downstream data in intermittent demand supply chains with supervised multivariate clustering. Journal of Forecasting, 43(5), 1661-1681. https://doi.org/10.1002/for.3095

Document publié chez l'éditeur officiel
Document issued by the official publisher

Titre de la revue: Journal Title: Journal of Forecasting (vol. 43, no. 5)

Maison d'édition: Publisher: Wiley

URL officiel: Official URL: https://doi.org/10.1002/for.3095

Received: 15 August 2022 | Accepted: 29 January 2024
DOI: 10.1002/for.3095

RESEARCH ARTICLE

WILEY

Improving demand forecasting for customers with missing downstream data in intermittent demand supply chains with supervised multivariate clustering

Corey Ducharme[1,2] | **Bruno Agard**[1,2] | **Martin Trépanier**[1,2]

[1]Laboratoire en Intelligence des Données, Département de Mathématiques et Génie Industriel, École Polytechnique de Montréal, Montreal, Quebec, Canada

[2]Interuniversity Research Centre on Enterprise Networks, Logistics and Transportation (CIRRELT), Montreal, Quebec, Canada

Correspondence
Corey Ducharme, Laboratoire en Intelligence des Données, Département de Mathématiques et Génie Industriel, École Polytechnique de Montréal, Montreal, QC H3C 3A7, Canada.
Email: corey.ducharme@polymtl.ca

Funding information
Natural Sciences and Engineering Research Council of Canada, Grant/Award Number: RGPIN-2019-04723

[Correction added on 25 June 2024, after first online publication: The copyright has been changed.]

Abstract

In a collaborative supply chain arrangement like vendor-managed inventory, information on product demand at the point of sale is expected to be shared among members of the supply chain. However, in practice, obtaining such information can be costly, and some members may be unwilling or unable to provide the necessary access to the data. As such, large collaborative supply chains with multiple members may operate under a mixed-information scenario where point-of-sale demand information is not known for all customers. Other sources of demand information exist and are becoming more available along supply chains using Industry 4.0 technologies and can serve as a substitute, but the data may be noisy, distorted, and partially missing. Under mixed information, leveraging existing customers' point-of-sale demand to improve the intermittent demand forecast of customers with missing information has yet to be explored. We propose a supervised demand forecasting method that uses multivariate time series clustering to map multiple sources of demand data. Members with missing downstream demand data have their resulting demand forecast improved by averaging over customers with similar delivery patterns for their final demand forecast. Our results show up to a 10% accuracy improvement over traditional intermittent demand forecasting methods with missing information.

KEYWORDS
demand forecasting, Industry 4.0, intermittent demand, multivariate time series clustering, supervised learning, supply chain forecasting

1 | INTRODUCTION

Information sharing is critical for an effective collaborative supply chain (Angulo et al., 2004). Vendor-managed inventory (VMI) is a collaborative supply chain arrangement whereby a supplier assumes the main responsibility of managing a customer's point-of-sale (POS) inventory to ensure a desired service level (Vigtil, 2007). In VMI, suppliers require accurate POS demand information to ensure adequate supply and replenishments.

Although the benefits of collaborative arrangements are well known for both suppliers and customers (Cao & Zhang, 2011; Jung et al., 2005; Zhou et al., 2017), some challenges do prevent their practical implementation. Supply chain partners may be unable or unwilling to share data due to technological limitations, lack of trust, confidentiality agreements, or antitrust laws (Colicchia et al., 2019; Hernández et al., 2014; Holweg et al., 2005; Kembro & Näslund, 2014). The cost of information systems that can collect and share data may also be prohibitive (Jung et al., 2005; Waller et al., 1999).

When a customer's POS information is missing, suppliers may simply decide to operate under a standard non-collaborative supply chain arrangement (Wang et al., 2014), or suppliers may turn to other strategies to forecast the customer's demand using whatever data are available (Ali et al., 2017). This is made even more pressing under VMI, where stocks-out may not be allowed under the terms of the arrangement.

Demand data can be captured at multiple locations along a supply chain (Holweg et al., 2005), and delivery records may be available for suppliers. These historical delivery records, however, can contain noise from *bullwhip*-type effects or logistic decisions, or have intermittent behavior, and thus may not reflect the actual demand behavior of customers (Murray et al., 2018a; Syntetos et al., 2016). Downstream, suppliers may sometimes have access to customers' demand data at the direct point of sale. Unfortunately, even these POS demand data can be intermittent due to their very granular resolution or due to the inherent behavior of the customer (Bartezzaghi et al., 1999). Thus, regardless of the source of information, intermittent demand time series can appear in supply chains and remain a persistent and pervasive challenge (Nikolopoulos, 2021).

For the scenario where a customer's POS information is missing, combined forecasting and smoothing approaches have been proposed to remove noise from a supplier's delivery demand time series to derive missing POS demand data to improve demand forecasts (Murray et al., 2018a; Nikolopoulos et al., 2011). However, when evaluating the performance of these methods, the authors did not evaluate their models on downstream POS demand information due to limitations in available data, resorting instead to either using the upstream delivery demand for the out-of-sample error or graphical evaluation.

In practice, suppliers may have multiple customers operating under similar VMI arrangements (Angulo et al., 2004). Technological advancements in both telemetry and information technology also allow suppliers to obtain multiple observations of the demand from their customers at different points along the supply chain (Januschowski et al., 2013; Li, 2007). Innovation in these types of supply chains has been driven recently by the research in the next industrial production model dubbed "Industry 4.0" (Witkowski, 2017). However, challenges, such as that of missing data, remain to be tackled to allow the full potential of exploiting varied sets of data to improve supply chain management (Ben-Daya et al., 2019; Lu, 2017).

Having access to multiple sets of demand data along the supply chain, we believe it is possible to leverage existing similar behaviors and patterns to learn a map between sets of demand data, which would improve forecasts when one set of data is missing. This problem takes inspiration from similar problems faced in missing data interpolation, which have recently used improved machine learning techniques to great success (Che et al., 2018; Luo et al., 2018). Similar techniques have also been used in direct intermittent demand forecasting (De Oliveira et al., 2020; Jiang et al., 2020; Lolli et al., 2017). We extend these methods to the partial information challenge in a collaborative supply chain. Furthermore, this problem is similar in nature to cold start demand forecasting for new products that have used machine learning methods as a proposed solution (van Steenbergen & Mes, 2020). We note that the term missing data in our context is different from the more generally understood one. Data that are *missing* refer to an entire set of observations produced at an area along a supply chain that is completely absent.

This work tackles the challenge of missing demand data in a supply chain by partnering with an industrial supplier operating a VMI arrangement with multiple customers and access to two sources of intermittent data for their customers' demand: a customer's POS demand data collected by telemetry and a customer's delivery demand data aggregated from the supplier's delivery records. This paper presents a methodology for forecasting intermittent demand data for customers with missing POS demand data by taking advantage of possible similarities in delivery patterns with other customers that are not missing any POS demand data. Key to our method is the treatment of the POS and delivery demand time series as a single multivariate time series, which represents the combined observations of both logistic variables for a given period. These multivariate data are clustered to produce a mapping function between demand time series. Customers with missing demand can be assigned to the nearest cluster whose members produce a time series prototype of the missing demand. A forecast of this prototype becomes the improved demand forecast for the customer.

In the following, the paper is subdivided in five sections. Section 2 contains the literature review. Section 3 proposes a multivariate clustering demand forecasting method. Section 4 describes the case study and data used to evaluate the performance of the proposed method. Section 5 is a discussion of the results, and Section 6 gives the conclusion, limits, and future extends of the study.

2 | LITERATURE REVIEW

The type of intermittent demand modeling depends on the use case, but the main goals are often demand forecasting or inventory management. Forecasting methods for intermittent demand can be viewed as a continuum from single model selection, which attempts to determine the optimal model to fit to the data, to model combination, which combines multiple individual models (Kourentzes et al., 2019). The individual intermittent demand forecasting models have mostly revolved around parametric and nonparametric models. Parametric models assume a certain distribution on the intermittent demand with Croston's method and its modifications (Croston, 1972; Shale et al., 2006; Syntetos & Boylan, 2005) being the most popular. Nonparametric models make no assumptions on the underlying distribution. Popular nonparametric models include the aggregation–disaggregation approach (Nikolopoulos et al., 2011; Spithourakis et al., 2012), bootstrapping (Hasni et al., 2019; Snyder et al., 2012; Willemain et al., 2004), and machine learning methods both supervised and unsupervised approaches (Kourentzes, 2013; Lolli et al., 2017; Murray et al., 2018b).

Croston's method is an extension of simple exponential smoothing. It applies exponential smoothing separately to the non-zero demand observations and the time intervals between consecutive periods of non-zero demand. The ratio of these two exponential forecasts is the final demand forecast. This means that the prediction is only updated for non-zero demand, while during periods of zero demand, the forecast remains unchanged. Corrections to this method were brought with the Syntetos–Boylan approximation (SBA) (Syntetos & Boylan, 2005), which remains the most popular and empirically tested version of Croston's methods (Syntetos et al., 2016, 2015). The SBA corrects for bias in the final demand forecast. Because Croston's method performs two separate exponential smoothings, two coefficients are required. In their original formulation of the SBA, Syntetos and Boylan (2005) recommended a value of 0.05 for both coefficients. This choice of coefficients was improved in Kourentzes (2014) where he proposed an automatic coefficient selection algorithm based on minimizing the in-sample error. However, this procedure does carry the risk of overfitting when there are few observations in the demand time series. Drawbacks to Croston's method are the lack of a proper underlying stochastic model (Shenstone & Hyndman, 2005) and the fact that the transformation can diminish or mask the demand behavior (Murray et al., 2018a).

Temporal aggregation for intermittent demand forecasting was first developed in the *aggregate–disaggregate intermittent demand approach* (ADIDA) of Nikolopoulos et al. (2011). This method begins with aggregating time series from higher to lower time frequency (e.g., daily to weekly), then forecasting the aggregated time series using an intermittent demand forecasting model, and, finally, disaggregating using weights the forecasted results down to the frequency of the original time series. Aggregating an intermittent time series solves, in part, the issue of zero demand observations, because the aggregated series will have fewer zero demand observations to affect the demand forecasting method. However, too much aggregation can cause loss of information, thereby distorting the true demand behavior (Spithourakis et al., 2012).

For the specific case of forecasting demand under missing POS demand data, Murray et al. (2018a) combined the Croston and ADIDA frameworks to propose their *Aggregate, Smooth, Aggregate, Convert to Time-series* (ASACT) method. Murray et al. (2018a) showed that previous intermittent demand forecasting and their new method could also be used to infer missing product consumption data at a customer's point of sale from a supplier's delivery records. This, in turn, improves the demand forecast. Essentially, they proposed the often implicitly assumed link that smoothed intermittent demand data offer a reasonable approximation of the real unobserved POS demand by removing noise so as to retain the characteristic behavior of the time series (Murray et al., 2018a; Nikolopoulos et al., 2011). However, due to a lack of POS data, their study did not evaluate its results on the underlying unobserved demand data, instead using simulated data and out-of-sample delivery demand observations.

Supervised machine learning approaches have been studied for intermittent demand forecasting such as neural networks (De Oliveira et al., 2020; Kourentzes, 2013; Lolli et al., 2017) and structured vector machines (Jiang et al., 2020). The advantage of these methods is that they do not rely on any underlying assumptions of the demand behavior. However, the observed improvement at the cost of added complexity shows mixed evidence of improvement over simpler methods (Mukhopadhyay et al., 2012). Furthermore, such black-box models do not provide a view of the underlying behavior of the demand, providing only a resulting time series forecast.

Unsupervised machine learning approaches have been applied to intermittent demand. Clustering of intermittent demand time series by Kalchschmidt et al. (2006) and Murray et al. (2017) was used as a method to identify characteristic behavior in the demand time series. They were able to observe distinct seasonality, trends, and cycles attributes in clusters. This challenges previous simpler methods of fitting a set parametric model to all time series. These methods were further extended in Murray et al. (2018b) to produce a demand forecast using dynamic time warping (DTW) and hierarchical agglomerative clustering. However, parameter tuning and optimization of the forecasting method, as well as the ability of unsupervised clustering to tackle missing demand data, have not been investigated.

Model combination is a forecasting method that combines the results of multiple forecasts using a weighting function (Kourentzes et al., 2019). Petropoulos and Kourentzes (2015) studied combinations of the previously presented Croston and ADIDA models. Their results showed improvements in both accuracy and robustness over a single model with both the average and median weighting function. Combining previous work on temporal aggregation and forecast combination, Kourentzes et al. (2014) proposed a forecasting method named the *Multiple Aggregation Prediction Algorithm* (MAPA). MAPA performs well on both intermittent and non-intermittent time series. MAPA combines estimates of multiple exponential smoothing forecasts done on different aggregation levels of the original time series. The aggregation level begins at 1 and increases up to a maximum level based on the seasonality of the time series. The state space exponential smoothing method (Hyndman et al., 2008) produces the forecasts for the aggregated time series.

3 | MULTIVARIATE CLUSTERING DEMAND FORECASTING

The goal of our method is to produce forecasts of a customer's demand with only partial demand data for the customer. The proposed method is a combination of supervised clustering and forecasting procedures. The clustering learns a mapping function between sets of demand data. The forecasting step assigns customers with missing demand data to an existing cluster, produces a prototype time series of the missing demand data from the other members of the cluster, and then uses a forecasting model to produce a final forecast. A flowchart of the method is depicted in Figure 1.

We begin by preprocessing (1) two sets of input demand data: a supplier's historical deliveries and the POS stock usage data. Next, the method relies on multivariate clustering (2) to create groups of customers with similar behavior for all demand data. This is our map. Forecasting (3) the demand data for a customer with missing POS data involves assigning that customer to an existing cluster using the univariate distance of its delivery demand data to the cluster's centroid. From the assigned cluster's members, we produce a time series prototype of the POS demand, which is then used to forecast that customer's future demand. To optimize the method, we calculate the error (4) and iterate over our parameter space. The final forecast and error (5) are done to compare our method with other models found in the literature.

The first core method is preprocessing (1) the two sets of input demand data. Both are time series. Preprocessing begins with categorizing each demand time series. The series are then smoothed and merged to produce a multivariate time series where each series is a variable. Finally, based on the results of the intermittent demand categorization (IDC), we use stratified samples to produce the training, validation, and testing datasets composed of multivariate time series. Next, we initiate the parameter space. This parameter space includes all the hyperparameters for the subsequent clustering and forecasting steps. Our method iterates over all possible combinations of parameters to determine the optimal model.

The second core method is multivariate clustering (2) using DTW for the distance calculation and the hierarchical clustering algorithm. Clustering is done on the training data for a chosen set of parameters. This produces our map.

The third core method is forecasting (3). Forecasting is done on members of the validation dataset under the assumption that only their delivery demand data are known. First, each customer is assigned to an existing cluster based on the shortest distance to the cluster's delivery demand time series centroid. Second, we produce a time series prototype by taking the average of the POS demand time series of all members in the assigned cluster. Lastly, we use an intermittent demand time series forecasting model on the time series prototype to forecast the customer's future demand.

The fourth core method is the error measurement (4). For the forecasts produced on the validation dataset, we perform error calculations and store the results for the model's parameters. We then select the new parameters and continue the iteration.

The fifth core method is the final step (5). Once the parameter space is empty, we determine the best model based on the error measurements that have been measured over the parameter space. With the best model for both the clustering step and the forecasting step, we perform a final forecast for all members in the test set. From these forecasts, we obtain our final error, which we can then compare to other demand forecasting models and strategies.

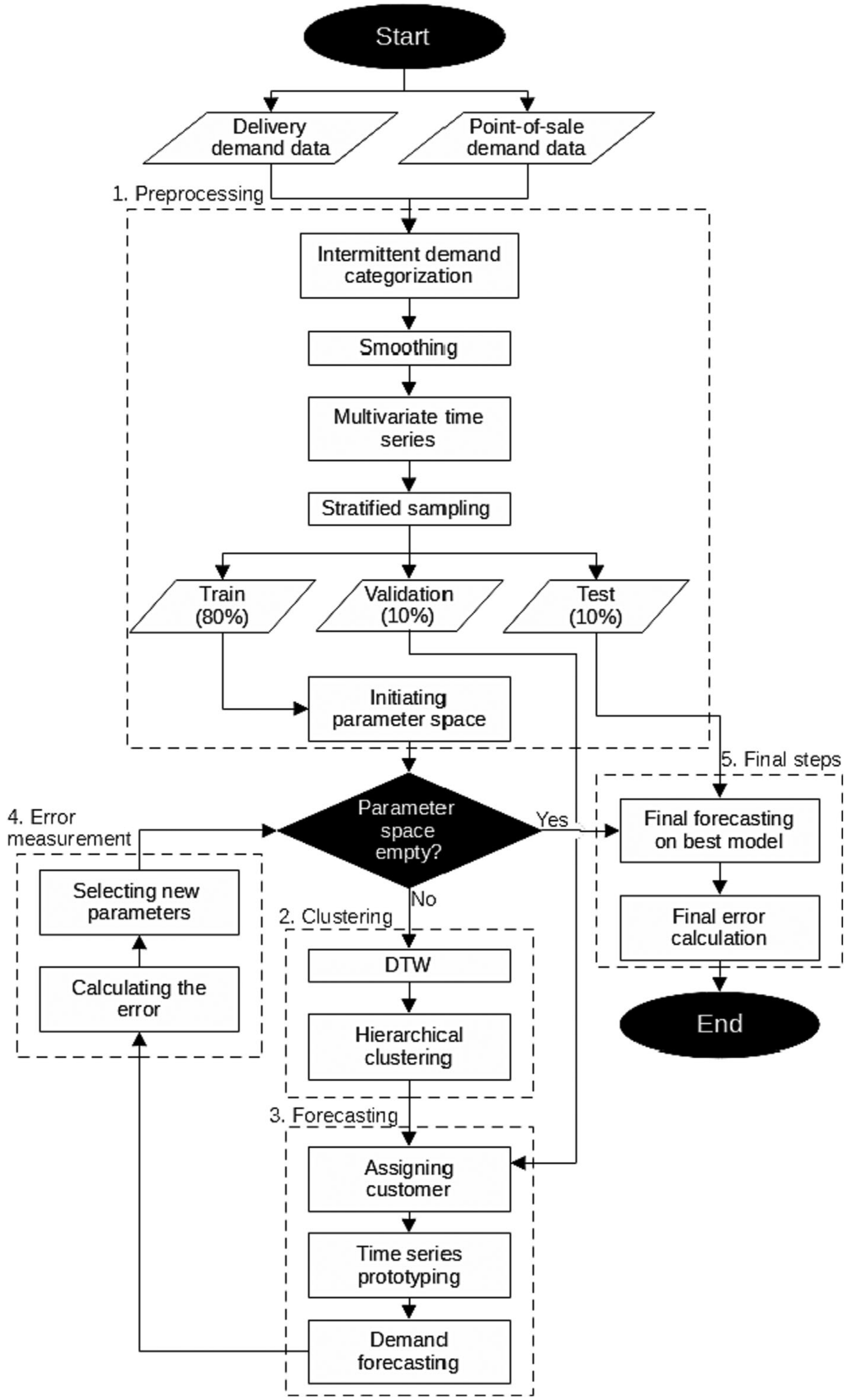

FIGURE 1 Multivariate clustering demand forecasting method. DTW, dynamic time warping.

Further details on the five core methods identified in Figure 1 are presented in the following sections.

3.1 | Preprocessing

The input data for the proposed method are demand observations for customers that can be taken any time along the supply chain. In our application of the method, demand data are observed at two locations for customers: upstream in the supplier's delivery records and downstream the customer's point of sale. Demand is observed periodically in both locations, which produces time series. Because they are taken at different locations and with different devices, they can have varying degrees of noise, or behave differently, even though fundamentally they represent the same thing: a customer's demand. Following the principles of demand information propagation presented in Holweg et al. (2005), we can assume that demand information taken closer to a customer's site will be less intermittent compared to demand information obtained further from the customer such as at the supplier's site. The differences in noise, behavior, and intermittence between the two sets of demand data provide an additional challenge when preprocessing the data.

3.1.1 | IDC

To allow for a systematic way of preprocessing the data, we used IDC as a feature engineering step to identify the intermittent behavior in the raw time series. IDC is a classification strategy for demand patterns based on statistical measurements of variability of the amount and the interval between non-zero observations. IDC was originally proposed as a systematic approach to model selection between Croston's method, SBA, and exponential smoothing (Syntetos et al., 2005). IDC was further developed by Petropoulos and Kourentzes (2015), and their Syntetos–Boylan–Croston–Kostenko–Hyndman–simple exponential smoothing (SBC-KH-SES) categorization is the one that is used in our method.

Although IDC was proposed as a method to distinguish between a small selection of parametric forecasting models, underlying IDC is the idea of determining whether a time series is intermittent or not. In our method, we propose using that aspect of the method as feature engineering to augment each demand time series by labeling them as intermittent or not. With the final goal of our forecasting method to improve demand forecasting, IDC can improve our proposed forecasting method by indicating whether smoothing is necessary as clustering of non-smoothed intermittent time series produces very poor clusters (Murray et al., 2017). When clustering intermittent demand time series, no systematic way of determining if an intermittent demand time series requires smoothing has yet to be proposed. This intermittent feature will be used when producing balanced training, test, and validation data samples.

3.1.2 | Smoothing

The second step in our preprocessing method is smoothing. Smoothing data before clustering improves the quality of the segments produced (Serban & Wasserman, 2005). However, the smoothing procedure needs to be adapted to the type and nature of the data.

In our method, we have two sets of demand data. They differ in nature due to being observed at different locations in our supply chain, as well as in purpose, because we wish to perform forecasts for customers with missing POS demand data. For these customers, our method uses their delivery demand data as input and produces POS demand data, which is then forecasted. With that in mind, we chose to only smooth the delivery data and not the POS data. Smoothing the delivery data improves the clustering, which produces a better map. Not smoothing the POS data allows the mapping function to return POS data and not *smoothed* POS data.

When clustering intermittent demand time series, Murray et al. (2017) showed that by using Croston's method to smooth the time series, they were able to produce clusters with similar demand behavior among the members and distinct behavior between clusters. As such, based on the results of the IDC, time series that are classified as intermittent are smoothed.

The smoothing procedure for the intermittent demand data is the automatic optimal parameter selection algorithm for the SBA developed by Kourentzes (2014). The SBA is presented in Equations (1)–(3).

$$\hat{y}_t = \left(1 - \frac{\alpha_x}{2}\right)\hat{Z}_t / \hat{X}_t \tag{1}$$

$\hat{Z}_t$ and $\hat{X}_t$ are the simple exponential smoothing forecasts of the non-zero demand amounts and the intervals between non-zero demand periods, respectively.

$$\hat{Z}_t = \alpha_z z_t + (1-\alpha)\hat{z}_{t-1} \tag{2}$$

$$\hat{X}_t = \alpha_x x_t + (1-\alpha)\hat{x}_{t-1} \tag{3}$$

The smoothing parameter α_x and α_z are determined optimally on an in-sample cost function described in Kourentzes (2014) and available in the *tsintermittent* R package (Kourentzes & Petropoulos, 2014). The SBA in-sample prediction is our smoothed time series. The in-sample cost function used is the mean absolute rate (MAR), which is defined as

$$MAR_n = \sum_{i=1}^{n} |r_i| \tag{4}$$

where n is the number of in-sample observations. r is the cumulative mean of the time series and is defined as

$$r_i = \hat{y}_i - i^{-1}\sum_{j=1}^{i} y_j \tag{5}$$

where y_t is the original time series and $\hat{y}_t$ is its forecast.

3.1.3 | Multivariate time series

The third step in our preprocessing method is combining the smoothed delivery data and the raw POS data into a multivariate time series. This simply requires that both time series be matched up along the time axis and excess observations be removed.

$$y_t = \left(y_t^1, y_t^2\right)^T, \ -\infty < t < \infty \tag{6}$$

Here, we use the notation, y_t^v, to indicate the time series observation for variable v at time t.

Because our method involves clustering logistic time series, all the multivariate time series are trimmed down to the same time interval. Although some clustering distance measurements are capable of working on series of different lengths, the underlying assumption is that they represent different observations of the same event, for example, a speech pattern but of varying length (Wang et al., 2019). However, this is different for logistic time series where different lengths indicate different events being observed.

3.1.4 | Stratified sampling

The fourth step of our preprocessing method is to produce the training, validation, and testing datasets. Members of the training dataset will be clustered to produce our mapping function. Errors calculated on the validation sample are used to optimize the clustering and demand forecasting parameters. The testing dataset produces our final model error to compare with different demand forecasting strategies.

Following the results of the IDC, we want to maintain an equal level of intermittence in all three datasets for both demand time series. A standard method for producing equally partitioned samples from a population composed of different groups is stratified random sampling (Särndal et al., 2003). The intermittence feature for both demand time series is the two population groups for stratified sampling.

3.1.5 | Initiating parameter space

Supervised methods have the advantage that the method's hyperparameters can be optimized on the available data. In our proposed method, two steps have hyperparameters: clustering (Section 3.2) and forecasting (Section 3.3). In the clustering step, the number of clusters is the ubiquitous hyperparameter (Milligan & Cooper, 1985). In the forecasting step, the entire demand forecasting model (Section 3.3.3) can be viewed as a hyperparameter. As such, all of the demand forecasting models that will be evaluated are placed in the parameter set. This provides two parameter sets: one for each step. We take the Cartesian product of both sets to obtain the final parameter space state; for example, $\{(k=1, F=ETS), (k=2, F=ETS), \ldots\}$.

3.2 | Clustering

Clustering creates groups of similar objects; each group is called a cluster. Ideally, members of the same cluster are as similar to each other as possible, while being as dissimilar as possible to members in other clusters. Time series clustering provides additional challenges: The data are highly dimensional (length of the time series), the dimensionality is not consistent across members (different lengths), and the series may contain multiple values (multivariate time series) (Liao, 2005).

Clustering methods involve two core steps: the distance matrix calculation and the distance between clusters calculation. Taken together, these two steps constitute a clustering algorithm. Time series clustering algorithms are classified based on what the grouping is done on: feature-based is on extracted static features, model-based is on fitted model parameters, and shape-based is on alignment distance (Aghabozorgi et al., 2015). In logistic supply chains, the goal of clustering is to create clusters of customers with similar demand behavior (Basallo-Triana et al., 2017; Murray et al., 2017). This leads to using shape-based clustering algorithms that can produce clusters where members have similar logistic profiles. Both the distance calculation and the clustering method will be chosen with those goals in mind.

The different clustering procedures that were used in this paper and explored during our research made significant use of the R package *dtwclust* (Sarda-Espinosa, 2018).

3.2.1 | DTW

The distance, also known as similarity or dissimilarity, calculation is a measurement of the distance between members. Distance calculations between all objects are represented as a distance matrix. Clustering algorithms operate on the distance matrix and produce clusters.

In the case of shape-based time series clustering, the most used distance is DTW (Aghabozorgi et al., 2015; Liao et al., 2006). The calculation of DTW distance uses a dynamic programming algorithm that finds the optimal warping path between the two series under constraints

(Sakoe & Chiba, 1978). The major criticism of the DTW algorithm is its high computational cost, in both time and memory utilization, which can limit its use (Zhang et al., 2006).

The DTW algorithm used in this method was developed by Giorgino (2009). The DTW distance supports multivariate time series. The operating parameters for the DTW algorithm are similar to previous applications of DTW on logistic time series—that is, no window size, the *symmetric2* step pattern, and the L_1 norm (Murray et al., 2017, 2018b). For these parameters, the equations for the DTW algorithm are as follows.

First is the equation for the local cost matrix (*lcm*) for the L_1 norm shown in Equation (7).

$$lcm_1(i,j) = \sum_{v} \left| x_i^v - y_j^v \right| \tag{7}$$

Here, x and y are two multivariate time series of length n, with v indicating the variable.

Second, the DTW algorithm iteratively travels through the *lcm* without any window size constraint, determining the path of least cost using the *symmetric2* step pattern and aggregating the cost—our final distance. We can define $\emptyset = \{(1,1),\ldots,(n,n)\}$ as the set containing all the points that fall on the optimum path. We can then write the final DTW distance as

$$DTW_1(x,y) = \sum \frac{m_\emptyset lcm_1(k)}{M_\emptyset}, \forall k \in \emptyset \tag{8}$$

Here, $DTW_1(x,y)$ is the L_1 norm DTW distance between two multivariate time series x and y, $m_\emptyset$ is a per-step weighting coefficient, and $M_\emptyset$ is the corresponding normalization coefficient (Giorgino, 2009).

3.2.2 | Hierarchical clustering

Hierarchical clustering is one of the most common methods for time series clustering (Liao, 2005). Hierarchical clustering is a simple clustering method that creates a hierarchy of clusters from a distance matrix. In agglomerative hierarchical clustering, as the hierarchy increases, clusters are created by agglomerating clusters from the lowest level where all members are clusters containing only themselves to the highest level where there is a single cluster containing all members. This creates an ordered sequence of groupings (Hastie et al., 2009). The main disadvantage of hierarchical clustering is that it is not capable of dealing effectively with large datasets due to its quadratic computational complexity (Wang et al., 2006).

When deciding which clusters to agglomerate, a distance between clusters is required. This is known as the linkage criterion. The chosen linkage criterion for the hierarchical algorithm is the unweighted average linkage clustering from Sokal and Michener (1958) shown in Equation (9). The proximity between two clusters is the average of the distances between all members of both clusters. Clusters produced using this linkage criterion can be viewed as a close-knit collective without any imposed shape or outline. This method has been recommended for time series clustering when using the DTW distance (Łuczak, 2016).

$$\frac{1}{|A||B|} \sum_{a \in A} \sum_{b \in B} d(a,b) \tag{9}$$

3.3 | Forecasting

Following the application of the clustering step, we obtain clustered sets of multivariate demand data. These clusters can be viewed as mapping functions. The goal of the forecasting step is to apply this produced map to a new customer for which the POS data are missing. We assume that members of the validation and test datasets are missing these data. Their real point of sale will be used when calculating the error.

3.3.1 | Assigning customer

Forecasting the demand for a customer with missing POS data begins with assigning it to an existing cluster. This requires identifying the nearest cluster to the customer. The complexity of this step is that POS data are unknown for these customers whereas the cluster contains only multivariate time series. Calculating the distance between a new member and existing clusters is thus done by taking the distance from the new member's time series to each cluster's centroid while excluding the missing variable. The centroid is a singular value that is meant to be representative of the entire cluster often at the center of the cluster. The chosen centroid function for clusters is the medoid. A medoid is an existing member in a cluster who has the smallest distance between itself and all other members in the cluster (Kaufman & Rousseeuw, 2009). Because the clustered data are multivariate, the centroid will also be multivariate. This POS variable is simply removed from the multivariate centroid when calculating the distance. The distance between a new customer's delivery data and the centroid uses the same DTW calculation as presented in Section 3.2.1.

This assignation is shown in Equation (10). We wish to find which cluster c from our set of cluster C has the centroid with the smallest DTW distance between our customer's delivery variable time series and the centroid's. The customer is then assigned exclusively to that cluster.

$$\min_{c} DTW\left(y^{del}, cent_{c}^{del}\right), \forall c \in C \tag{10}$$

3.3.2 | Time series prototyping

Once a customer is assigned to an existing cluster, we determine a time series prototype for its missing POS data based on the behavior of the other members in the cluster. A good time series prototype provides information on seasonality, trends, and cycles. Many time series prototypes exist, but the most common, and the one selected for our method, is the arithmetic mean, in part due to its robustness and ease of calculation (Aghabozorgi et al., 2015). To determine the time series prototype, the arithmetic mean is taken for the POS variable of all members in the assigned cluster.

This is shown in Equation (11), where the time series prototype for a customer y assigned to cluster c can be written as

$$y^{pos} = \frac{1}{|c|}\sum_{i=1}^{|c|} x_i^{pos} \tag{11}$$

with $|c|$ being the size of the cluster.

3.3.3 | Demand forecasting

The last step in the forecasting method is producing a demand forecast on the assigned time series prototype. The two previous steps produce what can be viewed as "substitute" or "inferred" POS demand data for a customer without it. The final forecasted demand for a new customer is a demand forecast produced on this substitute POS data. Because we are in the presence of intermittent demand time series, we chose to forecast the demand using standard robust intermittent demand forecasting models like Croston's method, the ADIDA framework, and the ASACT method. We also use the standard ETS model as a baseline comparison.

For ETS (Hyndman et al., 2008), we will use a state space ETS (ZZZ) model with an automatic selection of a possible specific model (additive or multiplicative) for the error, trend, and seasonality. With SBA, a smoothing value of 0.05 will be selected (SBA 0.05) (Syntetos & Boylan, 2005), as well as optimized in-sample parameters (SBA opt) (Kourentzes, 2014). For the ADIDA (Nikolopoulos et al., 2011) and ASACT (Murray et al., 2018a) models, they require choosing an aggregation level, a forecasting function, and a disaggregating function. We selected two different lead times (week: 7, month: 30) and a review period of 1 day: ADIDA(7 + 1, SBA opt, EQW) and ADIDA(30 + 1, SBA opt, EQW). Murray et al. (2018a) suggested daily aggregation on either calendar weeks or months without a review period: ASACT(week, ETS, EQW) and ASACT(month, ETS, EQW). SBA forecasts perform best under the ADIDA framework (Nikolopoulos et al., 2011; Petropoulos & Kourentzes, 2015). Under the ASACT method, the aggregated data are already smoothed, and therefore, a more standard time series forecasting method like ETS is appropriate. Finally, both methods recommend using the equal weight disaggregation function (EQW) to obtain the final forecast. Finally, MAPA (Kourentzes et al., 2014) will be used, considering the maximum aggregation permitted with actual seasonality in the data.

3.4 | Error measurement

The clustering and forecasting steps produce a demand forecast for customers with missing POS demand data. Our error is thus the difference between the *forecasted* demand time series and the *true* POS demand time series. Error measurements on the validation dataset allow us to optimize our method, and error measurements on the test dataset provide the final model error.

3.4.1 | Calculating the error

To measure the time series forecast error, we chose the scaled mean error (sME) for the bias, the scaled mean absolute error (sMAE) for the accuracy, and the scaled mean squared error (sMSE) for the variance. All three errors are scaled using the mean value of all in-sample observations to allow for averaging the errors of multiple series. These error measurements are standard metrics for time series (Ducharme et al., 2021; Petropoulos et al., 2016).

The mean absolute arctangent percent error (MAAPE) is a modification of the standard mean absolute percent average for use with intermittent time series. It evaluates the accuracy of forecasts (Kim & Kim, 2016). By taking the arctangent, it avoids potential issues of infinite values occurring when the actual values are zero.

Scaled periods in stock (sPIS) tracks how the cumulative amount a forecasted item has fictitiously spent in or

TABLE 1 Forecast error measurement equations.

Error	Equation	
sME	$\frac{1}{H}\sum_{h=1}^{H}\frac{y_{N+h}-\hat{y}_h}{1/N\sum_{t=1}^{N}y_t}$	(12)
sMAE	$\frac{1}{H}\sum_{h=1}^{H}\frac{\lvert y_{N+h}-\hat{y}_h\rvert}{1/N\sum_{t=1}^{N}y_t}$	(13)
sMSE	$\frac{1}{H}\sum_{h=1}^{H}\left(\frac{y_{N+h}-\hat{y}_h}{1/N\sum_{t=1}^{N}y_t}\right)^2$	(14)
MAAPE	$\frac{1}{H}\sum_{h=1}^{H}\arctan\left\lvert\frac{y_{N+h}-\hat{y}_h}{y_{N+h}}\right\rvert$	(15)
sPIS	$\frac{\sum_{h=1}^{H}\sum_{j=1}^{h}\left(\hat{y}_j-y_{N+j}\right)}{1/N\sum_{t=1}^{N}y_t}$	(16)

Abbreviations: MAAPE, mean absolute arctangent percent error; sMAE, scaled mean absolute error; sME, scaled mean error; sMSE, scaled mean squared error; sPIS, scaled periods in stock.

out of stock. As it is a cumulative error, it measures the bias of the forecast (Wallström & Segerstedt, 2010). A small sPIS highlights a good forecast. For comparison between forecasting methods, all final error measurements are computed using Table 1.

In Table 1, H is the forecast horizon, N is the number of in-sample observations, y_{N+h} is the real value of the hth out-of-sample period, and $\hat{y}_h$ is the h-steps-ahead forecast.

3.4.2 | Selecting new parameters

The error measurements are taken for the results of a model using a set of parameters. We note that our chosen parameter sets are not exhaustive but were chosen because they are known to have the most impact on the final results. For the final part of the optimization, we must choose one error. The most common choice would be the sMAE, which is a measurement of the accuracy of the forecast. The other measurements still provide added information for analyzing the results of a model. Industry-specific errors based on the use case can, and should, be substituted if available.

3.5 | Final steps

The final steps produce the final evaluation of our proposed multivariate clustering demand forecasting. The final steps are done on the test data, which have not been used to either train or validate the method. This allows for a comparison between our proposed model and other intermittent demand forecasting strategies, so long as the same test dataset is used.

3.5.1 | Final forecasting

The final forecast is the demand forecast for customers with missing POS demand using the parameters that minimized the error on the validation dataset. The forecast is produced using the map trained on the training data and the forecasting validated on the validation data.

3.5.2 | Final error calculation

We calculate the final error on the forecasts of the test dataset. This error can be compared to errors produced by different forecasting strategies or models. The final error uses the same error calculations presented in Section 3.4.1.

4 | EMPIRICAL EVALUATION

4.1 | Experimental setup

The goal of the experimental setup is to compare our proposed multivariate clustering for demand forecasting to other demand forecasting strategies. Our method uses two sources of a customer's demand data: a supplier's delivery data and POS product usage data. We propose three different strategies for performing demand forecasts based on the available sources of information: delivery data only, POS data only, and both delivery and POS data. Under the single source of data strategies, demand forecasts are performed classically with past observations used to feed an intermittent demand model. Under the mixed delivery/POS data scenario, the two sets of known demand data will be used to produce the mapping function, with the goal of improving the demand forecast of customers that have missing POS demand. This is compared to the scenario in which only the delivery data would be available and to the ideal scenario in which the real POS data are known. In terms of information, our three scenarios can be viewed as optimal (POS data only), baseline (delivery data only), and leveraged (POS + delivery data).

4.1.1 | Data

The data are provided by a large supplier of raw liquid materials. The supplier operates across the contiguous

USA with products used by thousands of customers across a wide variety of industries such as manufacturing, medical services, and the food industry. The supplier operates with some of its clients under a VMI arrangement. The supplier is solely responsible for the level of point-of-use inventory kept on the client's site. The supplier ensures that an uninterrupted flow of stock is available to the customer. To gather the data required to operate the VMI, the supplier installs dedicated reservoirs to store the product with telemetry sensors on the customer's site. The specifications for these reservoirs (volume, number, etc.) are decided by the supplier based on the needs of each customer. The sensors periodically measure the quantity of stock in the tank. However, this telemetric system is expensive, so it is not implemented for all customers. The supplier's delivery history to all customers is always available.

The anonymous data that have been made available to us are the telemetric POS demand (POS) and the supplier's historical deliveries (delivery) time series for 923 customers from July 1, 2015, to January 31, 2017. The descriptive statistics for these two time series are shown in Tables 2 and 3, respectively, for the three constituent parts of Croston's method: *demand*, *inter-demand interval*, and *demand per period*.

Following the preprocessing steps described in our method (Section 3.1), we begin with the SBC-KH-SES IDC scheme by Petropoulos and Kourentzes (2015) of the delivery demand time series in our dataset shown in Table 4. We also present the classification for the POS demand time series. Under SBC-KH-SES categorization, a series classified as SBA or Croston is intermittent and those categorized as simple exponential smoothing (SES) are not.

Somewhat unexpectedly, some of the POS data are intermittent. This occurs even with very few intervals of zero demand observations as seen in the third column of Table 3. Some series have enough variability in their demand amounts (lumpiness) to be classified as intermittent. Even as we move our downstream, observed demand data may still be intermittent.

Because all delivery data are classified as intermittent, we proceed to smooth them using the Croston smoothing method. The POS data are not smoothed even though some were classified as intermittent because we do not

TABLE 4 SBC-KH-SES intermittent demand categorization of the delivery and point-of-sale demand time series.

Category	Delivery	POS
Croston	7	74
SBA	916	156
SES	0	693

Abbreviations: POS, point-of-sale; SBA, Syntetos–Boylan approximation.

TABLE 2 Descriptive statistics of the delivery demand time series.

	Demand (units)		Inter-demand interval (days)		Demand per period (units/day)	
	Mean	SD	Mean	SD	Mean	SD
Minimum	43.79	24.33	10.11	18.95	3.27	3.57
First quartile	76.81	33.87	16.68	21.67	8.96	12.53
Median	93.82	39.82	21.48	26.10	13.09	19.11
Third quartile	109.33	44.60	26.78	30.42	19.99	29.62
Maximum	158.85	91.72	41.67	39.81	68.66	91.94

TABLE 3 Descriptive statistics of the point-of-sale demand time series.

	Demand (units)		Inter-demand interval (days)		Demand per period (units/day)	
	Mean	SD	Mean	SD	Mean	SD
Minimum	2.73	6.74	1	0	2.73	6.74
First quartile	7.86	14.31	1	0	7.86	14.31
Median	11.24	18.29	1	0	11.23	18.28
Third quartile	15.46	24.97	1.00	0.03	15.45	24.96
Maximum	72.75	118.83	1.53	2.38	72.38	118.41

wish to distort the output of the map, as previously explained in Section 3.1.2.

When creating the multivariate demand time series, all series are trimmed down to a single year,January 1, 2016, to December 31, 2016. We found through our work that trimming intermittent time series near their beginning was necessary when smoothing with Croston as it behaves inconsistently near the beginning. If there are zero observations near the beginning of time series, Croston is unable to smooth them, which leads to different starting dates among different multivariate time series.

Based on the IDC results in Table 4, we performed a stratified sampling to obtain our training, validation, and test datasets. We wish to maintain the same proportion of intermittent and non-intermittent time series in all three datasets. This results in 739 customers in the training dataset, 92 customers in the validation dataset, and 92 customers in the test dataset.

4.1.2 | Forecasting strategies

A forecasting strategy asks two core questions: What data are available, and what is the forecasting model? In our proposed model's strategy, we leverage existing POS data to improve forecasts for customers with only delivery data. Based on available data, we can compare three different supply chain demand data strategies: (1) forecasts on delivery data, (2) forecasts on POS data, and (3) forecasts based on both delivery and POS data (the proposed method). In terms of forecasting models, the same models are used for all data strategies.

With the final goal being to obtain the best demand forecasts of a customer, the error measurements are calculated on the real POS demand data for all three strategies. Furthermore, the error measurements presented in Table 1 are calculated for three horizons: 7, 14, and 28 days. These three horizons represent short-, medium-, and long-term daily forecasts. Errors are calculated on the last month (January 1, 2017, to January 28, 2017) of the real POS data.

Under single demand data scenarios 1 and 2, the forecasting models are fed the whole time series from July 1, 2015, to December 31, 2016, without trimming or preprocessing. Each model handles starting zero observations differently. Furthermore, the forecasting models are not machine learning methods and do not require a training or validation dataset. As such, results are simply computed on and presented for the test dataset.

Under mixed demand data scenario 3, our proposed multivariate clustering demand forecasting method follows the steps described in Section 3. However, to avoid overfitting, we selected an upper limit of 370 for the number of clusters, which is nearly half the number of members in the training dataset. Furthermore, to significantly reduce the number of calculations, we only computed the number of clusters at an interval of 5—that is, $k = [1, 5, 10, ..., 370]$. Of all available error measurements, we chose to optimize for the sMAE. Because the sMAE is forecasted for three horizons, the average is taken to obtain the final error. Results are presented for the members of the test dataset forecasted using the model, which minimized the error on the validation dataset.

4.2 | Results

Table 5 presents the test results for the POS data strategy. In terms of accuracy (sMAE and MAAPE), the MAPA model performs the best. The MAPA model has a maximum aggregation level of 7 because the frequency of the daily time series is 7. It is followed closely by the ETS model. With regard to bias (sME and sPIS), the ADIDA and ASACT monthly models performed well for all horizons. The more accurate MAPA and ETS show a higher bias for long-term (28) forecasts. In terms of variance (sMSE), the simpler SBA 0.05 scored the highest.

Table 6 presents the test results for the delivery data strategy. Regarding accuracy, the ADIDA(30 + 1, SBA opt, EQW) model performs the best. However, accuracy for small horizons is comparable across all models. They distinguish themselves at long horizons, for which all ADIDA and ASACT models perform well and the MAPA and ETS models perform quite poorly. In terms of both bias and variance, the SBA 0.05 model performs best.

Before presenting the detailed errors for our multivariate clustering demand forecasting models, we must determine the optimal parameters from our parameter space. This is presented in Figure 2, where the average sMAE is shown in relation to the number of clusters for each forecasting model. The optimal number of clusters is indicated with a larger sized marker for each forecasting model. Across all forecasting models and all cluster numbers, the SBA opt model with 310 clusters has the lowest error.

We can also observe that errors behave similarly for all forecasting models as the number of clusters increases. This indicates independence between the mapping function and the forecasting model. Furthermore, the high number of clusters used for optimization would seem to be justified because a quasi-optimal plateau is reached at $k = 100$ to $k = 175$, before finding another plateau at $k = 260$. Values appear to be optimal, as the error continues to increase from $k = 320$ to $k = 370$.

TABLE 5 Forecast errors under point-of-sale data strategy.

Error	*h*	SBA opt	SBA 0.05	MAPA	ETS	ADIDA (7 + 1, SBA opt, EQW)	ADIDA (30 + 1, SBA opt, EQW)	ASACT (week, ETS, EQW)	ASACT (month, ETS, EQW)
sME	7	**0.0028**	−0.0098	0.0377	0.0691	−0.0429	−0.0615	−0.0120	−0.0291
	14	0.0340	0.0214	0.0616	0.0905	−0.0117	−0.0303	0.0215	**0.0021**
	28	0.0696	0.0570	0.0826	0.1067	0.0239	**0.0053**	0.0613	0.0377
sMAE	7	0.4236	0.4107	**0.4036**	0.4123	0.4318	0.4519	0.4291	0.4446
	14	0.4386	0.4306	**0.4300**	0.4435	0.4421	0.4610	0.4411	0.4503
	28	0.4535	**0.4474**	0.4551	0.4747	0.4548	0.4698	0.4550	0.4611
sMSE	7	0.4573	**0.4420**	0.4855	0.4960	0.4564	0.4683	0.4625	0.4978
	14	0.5229	0.5267	0.5877	0.6158	0.5208	0.5349	**0.5205**	0.5417
	28	0.6202	**0.6083**	0.6638	0.6983	0.6127	0.6178	0.6190	0.6267
MAAPE	7	0.4618	0.4470	0.4155	**0.4155**	0.4852	0.5062	0.4675	0.4752
	14	0.4575	0.4454	**0.4179**	0.4197	0.47878	0.4977	0.4616	0.4659
	28	0.4471	0.4368	**0.4186**	0.4288	0.4649	0.4820	0.4498	0.4541
sPIS	7	1.765	2.118	**0.354**	−0.545	3.044	3.564	2.180	2.658
	14	**0.5468**	1.873	−3.583	−6.777	5.345	7.295	1.971	3.898
	28	−16.81	−11.69	−27.44	−38.47	**1.741**	9.278	−12.43	−3.855

Note: The bolded number represents the best performing model for the error and time horizon.
Abbreviations: ADIDA, aggregate–disaggregate intermittent demand approach; ASACT, Aggregate, Smooth, Aggregate, Convert to Time-series; MAAPE, mean absolute arctangent percent error; MAPA, Multiple Aggregation Prediction Algorithm; SBA, Syntetos–Boylan approximation; sMAE, scaled mean absolute error; sME, scaled mean error; sMSE, scaled mean squared error; sPIS, scaled periods in stock.

TABLE 6 Forecast errors under delivery data strategy.

Error	*h*	SBA opt	SBA 0.05	MAPA	ETS	ADIDA (7 + 1, SBA opt, EQW)	ADIDA (30 + 1, SBA opt, EQW)	ASACT (week, ETS, EQW)	ASACT (month, ETS, EQW)
sME	7	0.017	**0.010**	0.030	0.038	0.021	0.045	0.016	0.015
	14	0.050	**0.043**	0.062	0.070	0.054	0.078	0.048	0.048
	28	0.093	**0.086**	0.105	0.113	0.097	0.121	0.091	0.091
sMAE	7	0.512	0.521	0.512	0.512	0.513	**0.508**	0.511	0.516
	14	0.524	0.527	0.537	0.539	0.524	**0.521**	0.523	0.527
	28	0.542	0.545	0.570	0.573	0.542	**0.541**	0.542	0.543
sMSE	7	0.659	0.661	0.644	**0.639**	0.662	0.661	0.656	0.673
	14	0.734	**0.725**	0.748	0.747	0.736	0.736	0.731	0.740
	28	0.933	**0.923**	0.999	1.008	0.933	0.929	0.930	0.936
MAAPE	7	0.492	0.496	**0.478**	0.484	0.493	0.483	0.491	0.494
	14	0.485	0.485	0.480	0.487	0.485	**0.477**	0.484	0.486
	28	0.471	0.472	0.472	0.478	0.471	**0.465**	0.471	0.471
sPIS	7	1.54	1.73	1.56	1.33	1.41	**0.75**	1.57	1.58
	14	−0.81	**−0.09**	−1.37	−2.23	−1.26	−3.75	−0.67	−0.64
	28	−24.6	**−21.8**	−28.1	−31.3	−26.4	−36.0	−23.9	−23.9

Note: The bolded number represents the best performing model for the error and time horizon.
Abbreviations: ADIDA, aggregate–disaggregate intermittent demand approach; ASACT, Aggregate, Smooth, Aggregate, Convert to Time-series; MAAPE, mean absolute arctangent percent error; MAPA, Multiple Aggregation Prediction Algorithm; SBA, Syntetos–Boylan approximation; sMAE, scaled mean absolute error; sME, scaled mean error; sMSE, scaled mean squared error; sPIS, scaled periods in stock.

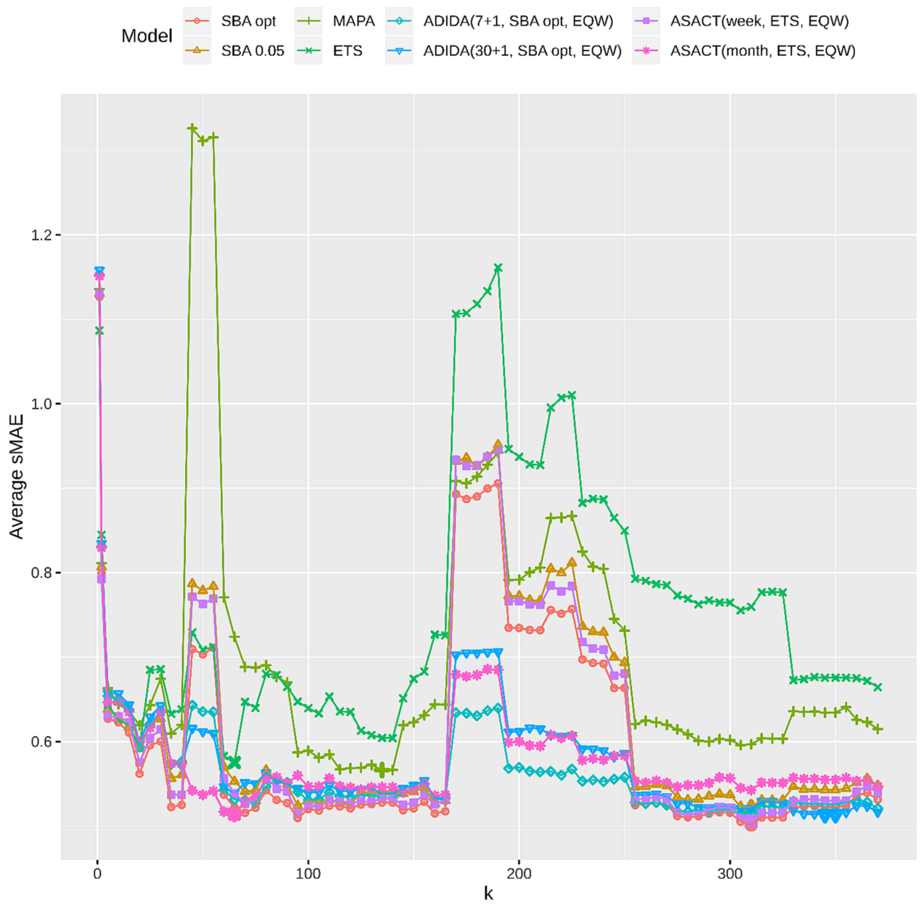

FIGURE 2 Error as a function of the number of clusters for multivariate clustering demand forecasting models. sMAE, scaled mean absolute error.

Table 7 presents the test results for the multivariate clustering model under a mixed delivery/POS data strategy. For all forecasting models, the optimal number of clusters is shown next to the forecasting model's name.

In terms of accuracy, the SBA opt model performs the best. It is followed closely by the ASACT weekly model. This is good news, because the SBA opt and ASACT weekly models were also the best and second-best performing models, respectively, on the validation dataset. This would indicate that the validation and test datasets' members were similarly representative. Regarding bias, the SBA opt model is the best performing model, especially in terms of long-horizon bias. For the variance, the ASACT monthly model scored the best, followed by the SBA opt model.

In terms of relative improvement in accuracy (sMAE), our proposed multivariate clustering with SBA opt forecasting is 7.6%, 8.5%, and 10%. This is compared to the improvement of 20.5%, 17.4%, and 15.8% for the POS data strategy, as forecasted with the MAPA model. Both are relative to the baseline delivery data strategy forecasted with the ADIDA(30 + 1, SBA opt, EQW) model. The gains of the multivariate clustering demand forecasting method are therefore equivalent to 36%, 49%, and 63% of the gains of the POS data strategy.

5 | DISCUSSION

The proposed multivariate method improves forecasting accuracy over intermittent demand models using exclusively upstream delivery demand. These gains are driven by the mapping function determined using supervised clustering. The mapping process extends previous clustering methods of intermittent demand time series done by Murray et al. (2018b). The low bias can be explained in part by a time series prototype produced by averaging multiple similar demands. This further reinforces intermittent demand forecasting methods involving model pooling and combination in line with current results (Kourentzes et al., 2019). The proposed method also improves over previous forecasting strategies for missing point of sale (Murray et al., 2018a) by learning behavior on a subset of customers for which we have all the information. We also used existing IDC schemes as a method to systematically preprocess mixed intermittent demand data to improve our demand forecasting method.

Although not necessarily the goal of the paper, the results permit a discussion of the strengths of intermittent demand models under different levels of intermittent demand data. For downstream POS demand data, which was partially intermittent, the MAPA model performed

TABLE 7 Forecast error for multivariate clustering demand forecasts under mixed demand information strategy.

	h	SBA opt, *k* = 310[a]	SBA 0.05, *k* = 310	MAPA, *k* = 135	ETS, *k* = 65	ADIDA (7 + 1, SBA opt, EQW), *k* = 310	ADIDA (30 + 1, SBA opt, EQW), *k* = 345	ASACT (week, ETS, EQW), *k* = 310	ASACT (month, ETS, EQW), *k* = 65
sME	7	**−0.068**	−0.085	−0.075	−0.206	−0.166	−0.174	−0.097	−0.304
	14	**−0.037**	−0.054	−0.062	−0.183	−0.135	−0.143	−0.061	−0.272
	28	**−0.007**	−0.024	−0.046	−0.188	−0.104	−0.113	−0.019	−0.260
sMAE	7	**0.469**	0.490	0.570	0.548	0.499	0.506	0.477	0.549
	14	**0.477**	0.498	0.579	0.571	0.510	0.518	0.485	0.561
	28	**0.486**	0.510	0.598	0.582	0.512	0.519	0.494	0.544
sMSE	7	0.521	0.570	0.853	0.607	0.537	0.539	0.525	**0.492**
	14	**0.575**	0.622	0.989	0.703	0.595	0.603	0.580	0.589
	28	0.718	0.772	1.172	0.675	0.732	0.731	0.725	**0.557**
MAAPE	7	**0.565**	0.576	0.634	0.756	0.606	0.613	0.577	0.773
	14	**0.557**	0.568	0.627	0.755	0.596	0.605	0.568	0.760
	28	**0.541**	0.553	0.615	0.755	0.576	0.586	0.550	0.736
sPIS	7	3.53	4.01	**3.20**	6.94	6.27	6.51	4.35	9.75
	14	**7.7**	9.6	8.4	22.1	18.0	18.8	10.5	32.0
	28	**12.3**	19.3	22.8	77.6	52.0	55.2	20.2	111

Note: The bolded number represents the best performing model for the error and time horizon.
Abbreviations: ADIDA, aggregate–disaggregate intermittent demand approach; ASACT, Aggregate, Smooth, Aggregate, Convert to Time-series; MAAPE, mean absolute arctangent percent error; MAPA, Multiple Aggregation Prediction Algorithm; SBA, Syntetos–Boylan approximation; sMAE, scaled mean absolute error; sME, scaled mean error; sMSE, scaled mean squared error; sPIS, scaled periods in stock.
[a]Best performing model on the validation dataset.

the best. MAPA seems quite well adapted to forecast both types of demand series, which is consistent with its proposed goals (Kourentzes et al., 2014). On completely intermittent upstream demand data, all demand forecasting models performed comparatively well for short- and medium-term forecasts in terms of accuracy. They only began to distinguish themselves at long-term forecast accuracy where the SBA, ADIDA, and ASACT methods performed significantly better than the ETS and MAPA models. Of note is the base SBA 0.05 model, which continues to perform well especially in terms of error bias and variance, further reinforcing its continued use (Syntetos et al., 2016, 2015).

Separating the initial dataset into three sets avoided the problem of arbitrarily selecting hyperparameters, a well-known difficulty in clustering-based methods (Milligan & Cooper, 1985). This type of model construction is more complicated and does have a history of being worse and more cumbersome than simpler parametric methods (Mukhopadhyay et al., 2012). They also require a significant amount of input data. However, when enough data is available, we believe that supervised learning approaches should be preferred, especially when the number of parameters is high, or are known to be difficult to determine.

In the case study when producing the training, validation, and testing datasets, we randomly selected the members through stratified sampling. In practice, we believe that customers truly missing POS data may behave differently than those with both sets of data that were made available to us for our experiment. Thus, there may be a selection bias, which was not possible to model in our experiment.

In this paper, we considered a clustering approach that extended previous work on the topic (Kalchschmidt et al., 2006; Murray et al., 2017, 2018b) to produce a mapping function. However, more typical methods of supervised learning have found use in the demand forecasting literature, such as nearest neighbor (NN) (Kück & Freitag, 2021; Nikolopoulos et al., 2016). Our proposed method was also modified to use NN as the mapping function and tested following all the same steps presented in Section 3 with the clustering (Section 3.2) being replaced with NN. Some small caveats are noted. First, NN does not support multivariate time series. Finding the NNs of a customer is done exclusively with the delivery data. Second, the NN model can be optimized directly on the training datasets. As such, the training and validation datasets were combined when optimizing the model's parameters. This could be advantageous when the number of observations is small. Results for the NN model are shown in Figure A1 and Table A1. The NN model also provided improved results compared to models using exclusively downstream delivery demand. Accuracy as measured by the sMAE is less than the proposed multivariate clustering model, but bias (sPIS and sME) is improved. Improvement in bias may be due to the consistent number of neighbors used to produce the time series prototype.

6 | CONCLUSION

We have proposed a novel forecasting model for intermittent demand under mixed demand information conditions that is based on multivariate clustering to produce a mapping function for supply chain partners with missing demand data, which is then used to forecast their future demand. A key innovation is the extension of previous univariate unsupervised clustering methods in intermittent demand in a supervised multivariate forecasting model. This leverages the recent progress in clustering intermittent demand to identify a group of customers with similar behaviors to improve forecasts. The forecasts generated are done on averaged time series, which helps lower bias and improve overall accuracy. Our method is also supervised in a way that allows optimal parameters to be automatically determined and uses previous work on IDC to systematically preprocess the data.

To evaluate the proposed method, we compared the performance of different demand information strategies in supply chains: There is full information sharing that provides delivery and POS data, no information sharing that means only delivery data are available, and then there is a realistic mixed delivery/POS data. It is in this latter more realistic context that the proposed method works. It allows suppliers to exploit existing Industry 4.0 technologies in a practical scenario and to then leverage existing downstream POS demand data to help improve demand forecasts for partners without any downstream data, instead of exclusively using upstream delivery demand data for those forecasts.

Throughout all error measurements, we found that the multivariate clustering demand model combined with the SBA opt forecast performed best in terms of forecast accuracy. It provided a 7.6%, 8.5%, and 10% accuracy improvement in sMAE for 7-, 14-, and 28-day forecast horizons compared to traditional models with missing POS demand data. Although an optimal model was determined using the proposed method, it should not be understood as a prescription. Without access to more empirical evidence, we can only propose and promote optimizing the forecasting model to the available data. However, we note that the intermittent demand models (SBA and ADIDA) performed better than the standard ETS model, which would reinforce the use of intermittent demand forecasting models when applying our proposed model in a mixed delivery/POS scenario.

An NN method was also evaluated after a small modification of the proposed clustering methodology and was found to be less accurate than the proposed method, but it does possess some advantages in terms of lowered forecast bias and being better suited to smaller datasets.

We further found that demand forecasts using POS downstream demand data are ideal, as they have the best available data, and forecasts using upstream delivery demand data are significantly worse. This indicates that the data available in supply chains are more important than the choice of model when tackling the challenge of intermittent demand in supply chains.

Regarding the improvement in accuracy for the available data strategies, our best multivariate model obtained greater improvements for longer forecast horizons. This contrasts with the downstream POS data strategy, where the improvements are greater at small horizons—that is, information sharing benefits more short-term forecasts. This would indicate to a practitioner that the choice of strategy for demand forecasting data is based on the forecasting goals (short-term vs. long-term). Collecting downstream data is not cost-free if data collection requires specialized systems, whereas our proposed method's cost is limited to the computation cost of developing and running the algorithm.

Further extensions to the proposed method could allow it to work with time series of varying lengths, which are more common in practice, especially when new partners join the supply chain. Our proposed method can still be used with a judicious choice of the distance method, but specialized supervised learning mapping functions could also be considered.

Finally, mixed-data approaches could also offer an advantage to suppliers when data collection in collaborative supply chains is costly. It could possibly allow a supplier operating under a collaborative arrangement to strategically decide from which partners to collect downstream information, while still improving forecasts for partners without such an information arrangement. Further research is required on the granular value of information collection for supply chain management.

ACKNOWLEDGMENTS

The authors would like to acknowledge our industrial partner and the Natural Sciences and Engineering Research Council of Canada (NSERC) for funding under Grant RGPIN-2019-04723. Our industrial partner also

provided the initial data that inspired the case study. The authors would like to thank the reviewers for their excellent comments, which substantially improved the quality of the manuscript.

DATA AVAILABILITY STATEMENT

Due to legal and commercial issues related to this research, supporting data are not available.

ORCID

Bruno Agard https://orcid.org/0000-0002-3653-7147
Martin Trépanier https://orcid.org/0000-0001-8408-8035

REFERENCES

Aghabozorgi, S., Shirkhorshidi, A. S., & Wah, T. Y. (2015). Time-series clustering—A decade review. *Information Systems*, *53*, 16–38. https://doi.org/10.1016/j.is.2015.04.007

Ali, M. M., Babai, M. Z., Boylan, J. E., & Syntetos, A. A. (2017). Supply chain forecasting when information is not shared. *European Journal of Operational Research*, *260*(3), 984–994. https://doi.org/10.1016/j.ejor.2016.11.046

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. https://doi.org/10.1002/j.2158-1592.2004.tb00171.x

Bartezzaghi, E., Verganti, R., & Zotteri, G. (1999). A simulation framework for forecasting uncertain lumpy demand. *International Journal of Production Economics*, *59*(1–3), 499–510. https://doi.org/10.1016/S0925-5273(98)00012-7

Basallo-Triana, M. J., Rodríguez-Sarasty, J. A., & Benitez-Restrepo, H. D. (2017). Analogue-based demand forecasting of short life-cycle products: A regression approach and a comprehensive assessment. *International Journal of Production Research*, *55*(8), 2336–2350. https://doi.org/10.1080/00207543.2016.1241443

Ben-Daya, M., Hassini, E., & Bahroun, Z. (2019). Internet of things and supply chain management: A literature review. *International Journal of Production Research*, *57*(15–16), 4719–4742. https://doi.org/10.1080/00207543.2017.1402140

Cao, M., & Zhang, Q. (2011). Supply chain collaboration: Impact on collaborative advantage and firm performance. *Journal of Operations Management*, *29*(3), 163–180. https://doi.org/10.1016/j.jom.2010.12.008

Che, Z., Purushotham, S., Cho, K., Sontag, D., & Liu, Y. (2018). Recurrent neural networks for multivariate time series with missing values. *Scientific Reports*, *8*(1), 6085. https://doi.org/10.1038/s41598-018-24271-9

Colicchia, C., Creazza, A., Noè, C., & Strozzi, F. (2019). Information sharing in supply chains: A review of risks and opportunities using the systematic literature network analysis (SLNA). *Supply Chain Management*, *24*(1), 5–21. https://doi.org/10.1108/SCM-01-2018-0003

Croston, J. D. (1972). Forecasting and stock control for intermittent demands. *Operational Research Quarterly (1970–1977)*, *23*(3), 289–303. https://doi.org/10.2307/3007885

De Oliveira, A. C. A., Jorgea, J. M., Santos, A. C. D., & Filho, G. P. R. (2020). Neural network with specialized knowledge for forecasting intermittent demand. In *27th ISTE International Conference on Transdisciplinary Engineering, TE 2020, July 1, 2020–July 10, 2020* (Vol. 12, pp. 524–533). Virtual, Online). IOS Press BV.

Ducharme, C., Agard, B., & Trépanier, M. (2021). Forecasting a customer's Next Time Under Safety Stock. *International Journal of Production Economics*, *234*, 108044. https://doi.org/10.1016/j.ijpe.2021.108044

Giorgino, T. (2009). Computing and visualizing dynamic time warping alignments in R: The dtw package. *Journal of Statistical Software*, *31*(7), 1–24. https://doi.org/10.18637/jss.v031.i07

Hasni, M., Babai, M. Z., Aguir, M. S., & Jemai, Z. (2019). An investigation on bootstrapping forecasting methods for intermittent demands. *International Journal of Production Economics*, *209*, 20–29. https://doi.org/10.1016/j.ijpe.2018.03.001

Hastie, T., Tibshirani, R., & Friedman, J. (2009). Unsupervised learning. In *The elements of statistical learning* (pp. 485–585). Springer. https://doi.org/10.1007/978-0-387-84858-7_14

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. https://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. https://doi.org/10.1016/j.emj.2005.02.008

Hyndman, R., Koehler, A. B., Ord, J. K., & Snyder, R. D. (2008). *Forecasting with exponential smoothing: The state space approach*. Springer Science & Business Media. https://doi.org/10.1007/978-3-540-71918-2

Januschowski, T., Kolassa, S., Lorenz, M., & Schwarz, C. (2013). Forecasting with in-memory technology. *Foresight: The International Journal of Applied Forecasting*, *31*, 14–20.

Jiang, P., Huang, Y., & Liu, X. (2020). Intermittent demand forecasting for spare parts in the heavy-duty vehicle industry: A support vector machine model. *International Journal of Production Research*, *59*, 7423–7440. https://doi.org/10.1080/00207543.2020.1842936

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. https://doi.org/10.1007/978-3-540-30585-9_61

Kalchschmidt, M., Verganti, R., & Zotteri, G. (2006). Forecasting demand from heterogeneous customers. *International Journal of Operations & Production Management*, *26*(6), 619–638. https://doi.org/10.1108/01443570610666975

Kaufman, L., & Rousseeuw, P. J. (2009). *Finding groups in data: An introduction to cluster analysis* (Vol. 344). John Wiley & Sons.

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. https://doi.org/10.1108/IJPDLM-09-2012-0287

Kim, S., & Kim, H. (2016). A new metric of absolute percentage error for intermittent demand forecasts. *International Journal of Forecasting*, *32*(3), 669–679. https://doi.org/10.1016/j.ijforecast.2015.12.003

Kourentzes, N. (2013). Intermittent demand forecasts with neural networks. *International Journal of Production Economics*, *143*(1), 198–206. https://doi.org/10.1016/j.ijpe.2013.01.009

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

Kourentzes, N., Barrow, D., & Petropoulos, F. (2019). Another look at forecast selection and combination: Evidence from forecast pooling. *International Journal of Production Economics*, *209*, 226–235. https://doi.org/10.1016/j.ijpe.2018.05.019

Kourentzes, N. & Petropoulos, F. (2014). tsintermittent: Intermittent time series forecasting.

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. https://doi.org/10.1016/j.ijforecast.2013.09.006

Kück, M., & Freitag, M. (2021). Forecasting of customer demands for production planning by local *k*-nearest neighbor models. *International Journal of Production Economics*, *231*, 107837. https://doi.org/10.1016/j.ijpe.2020.107837

Li, L. (2007). *Supply chain management: Concepts, techniques and practices: Enhancing value through collaboration*. World Scientific Publishing Company. https://doi.org/10.1142/6273

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

Liao, T. W., Ting, C.-F., & Chang, P.-C. (2006). An adaptive genetic clustering method for exploratory mining of feature vector and time series data. *International Journal of Production Research*, *44*(14), 2731–2748. https://doi.org/10.1080/00207540600600130

Lolli, F., Gamberini, R., Regattieri, A., Balugani, E., Gatos, T., & Gucci, S. (2017). Single-hidden layer neural networks for forecasting intermittent demand. *International Journal of Production Economics*, *183*, 116–128. https://doi.org/10.1016/j.ijpe.2016.10.021

Lu, Y. (2017). Industry 4.0: A survey on technologies, applications and open research issues. *Journal of Industrial Information Integration*, *6*, 1–10. https://doi.org/10.1016/j.jii.2017.04.005

Łuczak, M. (2016). Hierarchical clustering of time series data with parametric derivative dynamic time warping. *Expert Systems with Applications*, *62*, 116–130. https://doi.org/10.1016/j.eswa.2016.06.012

Luo, Y., Cai, X., Zhang, Y., Xu, J., & Yuan, X. (2018). Multivariate time series imputation with generative adversarial networks. In *NIPS'18: Proceedings of the 32nd International Conference on Neural Information Processing Systems* (pp. 1603–1614).

Milligan, G. W., & Cooper, M. C. (1985). An examination of procedures for determining the number of clusters in a data set. *Psychometrika*, *50*(2), 159–179. https://doi.org/10.1007/BF02294245

Mukhopadhyay, S., Solis, A. O., & Gutierrez, R. S. (2012). The accuracy of non-traditional versus traditional methods of forecasting lumpy demand. *Journal of Forecasting*, *31*(8), 721–735. https://doi.org/10.1002/for.1242

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

Murray, P. W., Agard, B., & Barajas, M. A. (2018a). ASACT—Data preparation for forecasting: A method to substitute transaction data for unavailable product consumption data. *International Journal of Production Economics*, *203*, 264–275. https://doi.org/10.1016/j.ijpe.2018.07.010

Murray, P. W., Agard, B., & Barajas, M. A. (2018b). Forecast of individual customer's demand from a large and noisy dataset. *Computers & Industrial Engineering*, *118*, 33–34. https://doi.org/10.1016/j.cie.2018.02.007

Nikolopoulos, K. (2021). We need to talk about intermittent demand forecasting. *European Journal of Operational Research*, *291*(2), 549–559. https://doi.org/10.1016/j.ejor.2019.12.046

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. https://doi.org/10.1057/jors.2010.32

Nikolopoulos, K. I., Babai, M. Z., & Bozos, K. (2016). Forecasting supply chain sporadic demand with nearest neighbor approaches. *International Journal of Production Economics*, *177*, 139–148. https://doi.org/10.1016/j.ijpe.2016.04.013

Petropoulos, F., & Kourentzes, N. (2015). Forecast combinations for intermittent demand. *Journal of the Operational Research Society*, *66*(6), 914–924. https://doi.org/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. https://doi.org/10.1016/j.ijpe.2016.04.017

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. https://doi.org/10.1109/TASSP.1978.1163055

Sarda-Espinosa, A. (2018). dtwclust: Time series clustering along with optimizations for the dynamic time warping distance. In *R: A language and environment for statistical computing, version 5.5.0*.

Särndal, C.-E., Swensson, B., & Wretman, J. (2003). *Model assisted survey sampling*. Springer Science & Business Media.

Serban, N., & Wasserman, L. (2005). CATS: Clustering after transformation and smoothing. *Journal of the American Statistical Association*, *100*(471), 990–999. https://doi.org/10.1198/016214504000001574

Shale, E., Boylan, J., & Johnston, F. (2006). Forecasting for intermittent demand: The estimation of an unbiased average. *Journal of the Operational Research Society*, *57*(5), 588–592. https://doi.org/10.1057/palgrave.jors.2602031

Shenstone, L., & Hyndman, R. J. (2005). Stochastic models underlying Croston's method for intermittent demand forecasting. *Journal of Forecasting*, *24*(6), 389–402. https://doi.org/10.1002/for.963

Snyder, R. D., Ord, J. K., & Beaumont, A. (2012). Forecasting the intermittent demand for slow-moving inventories: A modelling approach. *International Journal of Forecasting*, *28*(2), 485–496. https://doi.org/10.1016/j.ijforecast.2011.03.009

Sokal, R. R., & Michener, C. D. (1958). A statistical method for evaluating systematic relationships. *University of Kansas Science Bulletin*, *38*(22), 1409–1438.

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., Babai, Z., Boylan, J. E., Kolassa, S., & Nikolopoulos, K. (2016). Supply chain forecasting: Theory, practice, their gap and the future. *European Journal of Operational Research*, *252*(1), 1–26. https://doi.org/10.1016/j.ejor.2015.11.010

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

Syntetos, A. A., Boylan, J. E., & Croston, J. (2005). On the categorization of demand patterns. *Journal of the Operational Research Society*, *56*(5), 495–503. https://doi.org/10.1057/palgrave.jors.2601841

Syntetos, A. A., Zied Babai, M., & Gardner, E. S. (2015). Forecasting intermittent inventory demands: Simple parametric methods vs. bootstrapping. *Journal of Business Research*, *68*(8), 1746–1752. https://doi.org/10.1016/j.jbusres.2015.03.034

van Steenbergen, R. M., & Mes, M. R. K. (2020). Forecasting demand profiles of new products. *Decision Support Systems*, *139*, 113401. https://doi.org/10.1016/j.dss.2020.113401

Vigtil, A. (2007). Information exchange in vendor managed inventory. *International Journal of Physical Distribution & Logistics Management*, *37*(2), 131–147. https://doi.org/10.1108/09600030710734848

Waller, M., Johnson, M. E., & Davis, T. (1999). Vendor-managed inventory in the retail supply chain. *Journal of Business Logistics*, *20*, 183–204.

Wallström, P., & Segerstedt, A. (2010). Evaluation of forecasting error measurements and techniques for intermittent demand. *International Journal of Production Economics*, *128*(2), 625–636. https://doi.org/10.1016/j.ijpe.2010.07.013

Wang, X., Smith, K., & Hyndman, R. (2006). Characteristic-based clustering for time series data. *Data Mining and Knowledge Discovery*, *13*(3), 335–364. https://doi.org/10.1007/s10618-005-0039-x

Wang, X., Yu, F., Pedrycz, W., & Wang, J. (2019). Hierarchical clustering of unequal-length time series with area-based shape distance. *Soft Computing*, *23*(15), 6331–6343. https://doi.org/10.1007/s00500-018-3287-6

Wang, Z., Ye, F., & Tan, K. H. (2014). Effects of managerial ties and trust on supply chain information sharing and supplier opportunism. *International Journal of Production Research*, *52*(23), 7046–7061. https://doi.org/10.1080/00207543.2014.932931

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. https://doi.org/10.1016/S0169-2070(03)00013-X

Witkowski, K. (2017). Internet of things, big data, Industry 4.0—Innovative solutions in logistics and supply chains management. *Procedia Engineering*, *182*, 763–769. https://doi.org/10.1016/j.proeng.2017.03.197

Zhang, Z., Kaiqi, H., & Tieniu, T. (2006). Comparison of similarity measures for trajectory clustering in outdoor surveillance scenes. In *18th International Conference on Pattern Recognition (ICPR'06), Hong Kong* (pp. 1135–1138). https://doi.org/10.1109/ICPR.2006.392

Zhou, M., Dan, B., Ma, S., & Zhang, X. (2017). Supply chain coordination with information sharing: The informational advantage of GPOs. *European Journal of Operational Research*, *256*(3), 785–802. https://doi.org/10.1016/j.ejor.2016.06.045

AUTHOR BIOGRAPHIES

Corey Ducharme received his doctorate in industrial engineering from the École Polytechnique de Montréal in 2021. During his studies there, he was a member of the Laboratoire en Intelligence des Données (LID) and of the Interuniversity Research Centre on Enterprise Networks, Logistics and Transportation (CIRRELT).

Bruno Agard is Professor of Industrial Engineering at École Polytechnique de Montréal, Montreal, Quebec, Canada. He graduated in manufacturing from the École Normale Supérieure de Cachan (1998). He has a master's degree (1999) and a PhD (2002) in industrial engineering from the Institut National Polytechnique de Grenoble, France. He joined the Department of Mathematics and Industrial Engineering in 2003. He is interested in various applications of artificial intelligence in industrial engineering. He is the director of the Laboratoire en Intelligence des Données (LID) and a member of CIRRELT and IVADO.

Martin Trépanier is a civil engineer and professor at the Department of Mathematics and Industrial Engineering of École Polytechnique de Montréal, an engineering school affiliated with the Université de Montréal. He is the titular of the Chair in the Transformation of Transportation. He conducts many research projects about intelligent transport systems, urban mobility, transportation planning, operations logistics, data mining, and processing of large datasets. Since January 2019, he has been the director of the Interuniversity Research Centre on Enterprise Networks, Logistics and Transportation (CIRRELT).

How to cite this article: Ducharme, C., Agard, B., & Trépanier, M. (2024). Improving demand forecasting for customers with missing downstream data in intermittent demand supply chains with supervised multivariate clustering. *Journal of Forecasting*, *43*(5), 1661–1681. https://doi.org/10.1002/for.3095

APPENDIX A

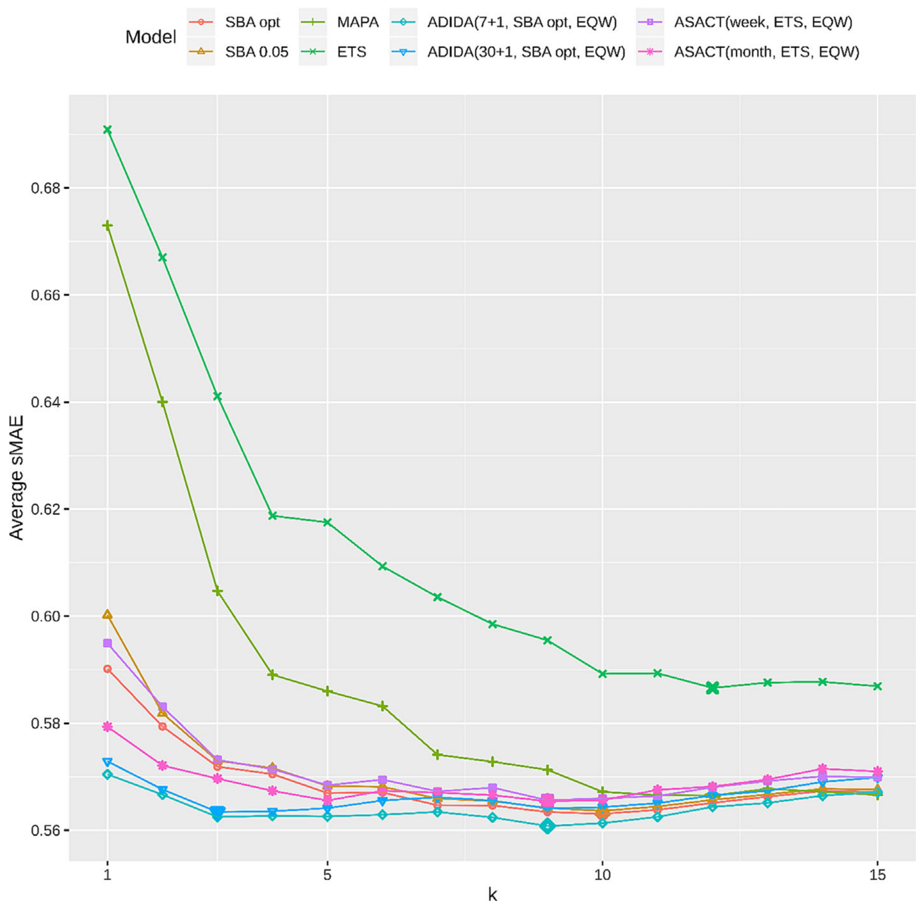

FIGURE A1 Error as a function of number of neighbors for nearest neighbor demand forecasting models. ADIDA, aggregate–disaggregate intermittent demand approach; ASACT, Aggregate, Smooth, Aggregate, Convert to Time-series; MAPA, Multiple Aggregation Prediction Algorithm; SBA, Syntetos–Boylan approximation; sMAE, scaled mean absolute error.

TABLE A1 Forecast error for nearest neighbor demand forecasts under mixed demand information strategy.

	h	SBA opt, $k = 10$	SBA 0.05, $k = 10$	MAPA, $k = 12$	ETS, $k = 12$	ADIDA (7 + 1, SBA opt, EQW), $k = 9$[a]	ADIDA (30 + 1, SBA opt, EQW), $k = 3$	ASACT (week, ETS, EQW), $k = 9$	ASACT (month, ETS, EQW), $k = 9$
sME	7	0.0214	0.0131	0.0162	0.0254	−0.0247	−0.0481	**−0.0025**	−0.0352
	14	0.0535	0.0452	0.0476	0.0483	0.0067	−0.0204	0.0321	**−0.0038**
	28	0.0928	0.0845	0.0852	0.0690	0.0462	**0.0190**	0.0774	0.0357
sMAE	7	**0.4816**	0.4839	0.4841	0.4984	0.4852	0.5036	0.4853	0.4915
	14	0.4906	0.4917	**0.4871**	0.5019	0.4927	0.5044	0.4932	0.4990
	28	0.5046	0.5046	**0.5039**	0.5242	0.5061	0.5150	0.5084	0.5104
sMSE	7	0.5772	0.5801	**0.5666**	0.5790	0.5685	0.6033	0.5714	0.5716
	14	0.6753	0.6734	0.6572	0.6670	**0.6547**	0.6582	0.6652	0.6573
	28	0.8699	0.8668	0.8489	0.8700	0.8468	0.8963	0.8589	**0.8461**
MAAPE	7	**0.4987**	0.4999	0.5042	0.5133	0.5078	0.5236	0.5053	0.5143
	14	**0.4859**	0.4867	0.4864	0.4987	0.4960	0.5097	0.4916	0.5021
	28	**0.4705**	0.4708	0.4754	0.4916	0.4807	0.4926	0.4771	0.4850
sPIS	7	1.287	1.520	1.378	**1.090**	2.575	3.239	1.953	2.869
	14	−1.402	**−0.5301**	−0.9205	−1.474	3.468	6.102	0.9569	4.572
	28	−25.46	−22.09	−22.96	−20.65	−6.521	3.972	−17.76	**−2.251**

Note: The bolded number represents the best performing model for the error and time horizon.
Abbreviations: ADIDA, aggregate–disaggregate intermittent demand approach; ASACT, Aggregate, Smooth, Aggregate, Convert to Time-series; MAAPE, mean absolute arctangent percent error; MAPA, Multiple Aggregation Prediction Algorithm; SBA, Syntetos–Boylan approximation; sMAE, scaled mean absolute error; sME, scaled mean error; sMSE, scaled mean squared error; sPIS, scaled periods in stock.
[a]Best performing model in the training + validation dataset.