Navigating demand forecasting in make-to-order manufacturing: the role of global models and intermittent time-series

Jonatan Flyckt[1,2,*], Niklas Lavesson[2]

[1] Herenco AB, Skolgatan 24, 553 16, Jönköping, Sweden

[2] Department of Software Engineering, Blekinge Institute of Technology, Valhallavägen 10, 371 79, Karlskrona, Sweden

Abstract

Demand forecasting can optimise production and supply chain practices in manufacturing organisations. However, demand forecasting is not widely adopted among make-to-order (MTO) manufacturers with mass customisation offers. Building effective demand forecasting systems is challenging in such organisations due to the numerous unique manufactured articles and sparse demand patterns. This position paper argues that make-to-order manufacturers should employ demand forecasting to a larger extent, and that the forecasting community should address challenges related to the domain. Key challenges include creating models capable of predicting both demand size and timing of intermittent forecasts, as well as a deeper insight into the effects of global deep learning time-series models. We perform a pilot experiment using demand forecasting in a purchasing decision support system to validate the usefulness of demand forecasting for MTO manufacturing organisations with mass customisation offers. A research roadmap is proposed to address the identified challenges.

Keywords

Demand Forecasting, Intermittent Time-Series, Global Models, Machine Learning, Make-to-Order Manufacturing, Mass Customization

1. Introduction

Demand forecasting is an important tool in many domains, e.g., retailing and manufacturing, to help ensure that goods are available when customers need them. Demand forecasting involves training machine learning models on historical demand patterns for products or product groups. During inference, the trained models are supplied with recent demand data to predict the demand for a future period of interest [1]. The goal is to estimate when customers will purchase items, and how many items they will purchase.

1.1. Demand forecasting challenges

Accurate demand forecasts allow manufacturing organisations to plan production proactively to have goods ready when the customers need them and become more efficient at doing so.

SAIS2025: Swedish AI Society Workshop 2025, 16-17 June 2025, Halmstad, Sweden.

*Corresponding author.

jonatan.flyckt@bth.se (J. Flyckt); niklas.lavesson@bth.se (N. Lavesson)

0000-0003-1380-1408 (J. Flyckt); 0000-0002-0535-1761 (N. Lavesson)

Retail and make-to-stock (MTS) manufacturing organisations adopt demand forecasting into their supply chain planning to a larger extent than make-to-order (MTO) manufacturing organisations [2, 3]. MTO manufacturing organisations generally wait until the customer places an order until they produce the final goods and therefore cannot use demand forecasts to the same extent that MTS manufacturing organisations do.

Demand forecasts can help MTO organisations decide when to produce extra finished goods to stock, making production more efficient. However, due to the challenges involved, MTO organisations often avoid using demand forecasting in their workflows [2, 3]. MTO manufacturing organisations often produce many articles and need to know when demand will occur for each article and customer [4]. However, it is not feasible to train separate machine learning models for thousands of unique articles (local models). Instead, focus shifts to global models that train on many time-series and forecast for all of them with one model [5].

1.2. Aim and scope

This position paper highlights knowledge gaps in the system requirements for demand forecasting for MTO manufacturing organisations and specific challenges that need to be addressed to close them. We believe more attention should be given to training and assessing intermittent time-series that predict both demand size and timing, especially from a global perspective when dealing with many time-series. To support this position, we conduct a pilot experiment where demand forecasting is incorporated into a smart purchasing system used by an MTO manufacturer. The aim of the experiment is to evaluate the usefulness of demand forecasting for improving purchasing decisions in an MTO context.

1.3. Outline

In the sections that follow, we first present the background and relevant terminology for the paper. This is followed by a description of the use of demand forecasting both in general and from a manufacturing perspective, highlighting specific demand forecasting challenges that arise in MTO manufacturing organisations. The argumentation section of the paper argues three main points:

- The need for a better understanding of the requirements of demand forecasting systems in MTO manufacturing organisations.
- The need for models capable of forecasting intermittent time-series with both accurate demand timings and sizes.
- The need for a deeper insight into the effects of using global deep learning time-series models compared to local models, and how to improve global models.

Following this, we present results from a pilot experiment where demand forecasting is integrated into a purchasing system in a mass customisation purchasing task, demonstrating the usefulness and potential of demand forecasting in an MTO context. Lastly, we present a research roadmap of open, tangible demand forecasting problems which we believe should be addressed.

2. Background

Demand forecasting has been used for decades in MTS manufacturing to manage inventory of semi-finished and finished goods [6, 7]. However, demand forecasting continues to see limited use in the MTO and mass customisation manufacturing domain [2]. MTO organisations want to keep the stock levels low to avoid tying up capital or producing goods which become obsolete. In our experience, MTO organisations often use qualitative methods, relying on expert knowledge to forecast contract orders without historical data or macro trends. However, these approaches are time-consuming and difficult when there are many articles [8].

2.1. Terminology

A **make-to-stock** manufacturing strategy aims to keep adequate stock of manufactured goods in anticipation of future demand [2]. A **make-to-order** manufacturing strategy means that goods are only manufactured after an order has been placed, and is a common strategy among manufacturers who have **mass customisation** offers [9, 2]. The purpose of mass customisation is to provide customers with customised products tailored to their specific needs [9], e.g., unique cosmetic designs integrated into base products, or tools unique to certain customers. We use the term **article** to describe goods that have been customised for a specific customer. Organisations with repeat customers can use a hybrid manufacturing strategy and manufacture customised goods ahead of demand with a calculated risk to improve production efficiency.

The purpose of **demand forecasting** is to estimate when customers are likely to have a demand for goods which an organisation supplies. Customers can sometimes provide their own demand forecasts, but they are often not accurate enough to use for decision support. **Time-series analysis** can be used to estimate demand of goods by treating historical sales orders as time-series. These time-series, either **univariate** or **multivariate** with covariates, can inform the training of machine learning models to predict future demand for a specified time-period (**forecasting horizon**) [1, 10].

Mass customisation often leads to sparse demand due to many unique articles, making the time-series **intermittent** or **lumpy**. Intermittent and lumpy time-series consist of many zero values followed by a single demand or burst of demands [11]. Such time-series are difficult to forecast because most methods handle continuous values [1].

Deep learning time-series models can be local or global [5]. **Local time-series models** train on historical samples of a single time-series for forecasting that series only, whereas **global time-series models** train on multiple time-series and forecast future values for any series in the dataset. We use **semi-global model** to describe multiple models trained on subsets of data to leverage global model strengths while having a more homogeneous training data. Both global and local models handle one time-series (and covariates) at a time during inference.

2.2. Related work

2.2.1. Demand forecasting and applications

Demand forecasting is widely employed in the retail and MTS manufacturing domains [4, 12, 13]. The most common methods for forecasting univariate time-series with machine learning are statistical models like ARIMA [14] for local models aimed at forecasting one time-series, and deep learning models like LSTMs and transformers, which can be both local or global, i.e., forecasting one or multiple time-series using the same model [1, 10]. Intermittent demands are common in both retail and manufacturing domains [4]. A hybrid MTS/MTO approach: make-to-forecast (MTF) has been proposed which uses forecasts to manufacture or purchase partially completed products based on forecasts, modifying them when the orders are placed [2, 3]. However, there are few practical real-world applications of such approaches to date [8]. Recent work on time-series have attempted to create foundation models which would be capable of predicting on time-series from completely new domains [15, 16].

2.2.2. Intermittent forecasting

The most widely adopted method for forecasting intermittent demand is the Croston method [17] and its many derivatives [18, 19, 20, 21]. The Croston method divides intermittent time-series into period length and demand size, then smooths them separately before combining. This estimates the average demand rate for a future time-period [17]. Another statistical approach from the retail domain is to use a hierarchical forecasting structure with a greedy aggregation-decomposition to forecast stock-keeping units for intermittent demand products [13].

Deep learning models like LSTMs [22] or transformers [10, 23] often predict zeros when faced with intermittent demand. Several specialised approaches have been proposed to forecast intermittent demand with deep learning [1, 4]. One such approach, DeepAR, uses non-linear transformations combined with appropriate likelihoods to overcome the issue of predicting only zero values, creating a model architecture that works for both smooth and intermittent time-series [1]. Non-sequential deep learning models have also been shown to perform well on intermittent time-series by using the demands and intervals as separate inputs into a multi-layer perceptron with a single hidden layer [4]. Most intermittent forecasting methods predict an average demand rate, which is useful for domains which focus on stock-keeping efficiency. However, few studies focus on predicting both the timing and size of demands

2.2.3. Global models

Deep learning forecasting has mostly used local models historically [1, 5]. Many deep learning architectures train both local and global models [1, 10], but papers often don't specify which is used. Rožanec et al. [24] used anomaly detection and explainable AI technologies to enable users to determine when the output of forecasting models could be trusted. They noted that global models can perform better, but there's a gap in explaining why they sometimes produce bad forecasts. Montero-Manso and Hyndman [5] found that global models could be both more complex and better at generalisation than local models.

Semi-global models trained on groups of homogeneous time-series have been shown to perform better than fully global models trained on more heterogeneous data [25]. Strong models can also be produced by training global base models and using transfer learning for local or semi-global contexts [26]. Bandara et al. used clustering on a set of 18 time-series features proposed by Hyndman [27] to determine which time-series should be grouped together for semi-global training [25]. [28] used an iterative refinement approach to the clustering where they let forecasting accuracy guide cluster assignment. These studies show good results in clustering time-series for semi-global models, but it is unclear which features are most relevant. Is there a difference in feature importance for smooth, intermittent, and lumpy time-series? It has been noted that disparate time-series in the same training cluster are detrimental to model performance [25]. Further exploration is required into which characteristics determine good semi-global training clusters, and whether there is a "similarity cutoff" to aim for.

2.2.4. Metrics

Forecast success can be assessed with different metrics, the most common method being some variation of either the Mean Absolut Deviation (MAD), percentage error, or Mean Squared Error (MSE) calculated from the prediction error of each time-step [29]. MSE, MAD, and a multitude of other metrics, e.g., MASE [30] and RMASE [13] have been used to assess intermittent forecasts with a per-time-step error in various ways [29]. Although current metrics are effective in some contexts, more specialised metrics are called for, as the large number of zero values make absolute errors a blunt assessment of the actual forecasting performance [30].

Some metrics account for the timing of demands to address the issue of over-penalising near-misses: MAD_n and MSE_n calculate the error when a demand occurs by summing the forecasted values leading up to the demand point and measuring its agreement with the specific demand for that time-step [29]. Cumulative Forecast Error (CFE) measures positive and negative errors at each time-step for the forecasting horizon before adding the errors together, essentially forming a sum aggregate error [29]. This bypasses some of the issues with misjudging near-misses, but these metrics only help for near-misses which occur before the demand, not after. Additionally, they penalise forecasts across the forecasting horizon equally, meaning that the timing of the forecasts is not assessed.

There are also domain-specific metrics such as Periods In Stock (PIS) [29] or Stock-keeping-oriented Prediction Error Cost (SPEC) [31] which simulate the performance of a forecasting model in a stock-keeping scenario at retail organisations. Such metrics only work for a stock-keeping oriented approach, however, and are not suitable for use in an MTO manufacturing organisation or in domain-agnostic approaches.

3. Forecasting intermittent demand: global and local model approaches for an MTO manufacturing context

This section will explore the challenges and needs in demand forecasting for MTO manufacturing organisations. We will focus on understanding system requirements, forecasting intermittent time-series, and comparing global and local models.

3.1. MTO manufacturing

MTO manufacturers should forecast the demand of their goods more, but first, it is essential to understand the specific requirements of demand forecasting systems in MTO manufacturing. Effective forecasts would allow many mass customisation manufacturers to take calculated risks to occasionally produce extra finished goods to stock, or to purchase semi-finished goods before demand arrives. By doing so they could improve the efficiency of their inventory management, production planning, staff scheduling, and material purchasing [32]. They could also serve customers better by contacting them with "win-win" opportunities when a demand is forecasted. This could both help the manufacturing organisation be more efficient and give the customer a more stable delivery and potentially better prices.

In our experience from working in the domain, customer-supplied forecasts are the primary decision support used to gauge future demand. Unfortunately, these forecasts are often either inaccurate or not detailed enough to be useful beyond high-level budgeting. Manual forecasts are sometimes made by observing past sales and using expertise to estimate future demand. These manual estimates inform production planning and purchasing decisions. Unfortunately, this process is time-consuming and difficult to scale accurately. This highlights the need for automated forecasting, which, besides increasing forecasting performance, could also free up a lot of time for other tasks. It is difficult to know beforehand which level of forecasting performance is needed to support decisions in MTO manufacturing organisations, as there have not been many studies on this. To fully support decision-making in MTO organisations where demands are often intermittent, forecasts should predict both demand timing and size. These forecasts could help reduce setup times and increase machine up-times by adjusting production order timing.

3.2. Intermittent time-series forecasting

Forecasting intermittent time-series is particularly challenging due to the sporadic nature of demand. Intermittent time-series forecasts usually predict an average demand rate per time-step, rather than the size and timing of individual demand occurrences. It should be a priority to develop models capable of predicting both demand size and timing, as they would allow many organisations to use the forecasts in ways which are not possible with average demand rates. To accomplish this, the focus should be on modelling the time-series in a way which deep learning models can understand. Due to their non-linear nature, deep learning models can capture underlying patterns and detect subtle trends ranging across time-series, whereas statistical models often oversimplify time-series problems.

Of course, the effort to cater to every situation with a black box model may be too large, and the stability and consistency of the standard statistical models is one of the reasons that they have been used for a long time. The black box nature of deep learning models also makes them less interpretable than statistical models, which can be detrimental to user trust in the models.

Model performance metrics should relate to the real-world goal one is trying to accomplish. There are good domain-specific metrics that simulate stock-keeping in e.g., the retail domain [29, 4, 31], but such metrics are unfeasible for an MTO manufacturing setting. When training a deep learning model, the loss function should ideally correlate strongly to the metric, and simulation metrics may be too slow to use efficiently. The non-domain specific metrics often judge a sum aggregate error or the error per time-step in the forecasting horizon [29]. These metrics are insufficient with models that forecast both the timing and size of demands. Therefore, it is essential to develop domain-agnostic metrics capable of correctly judging how well both the timing and size of demands are predicted.

3.3. Global vs. local time-series models

Global models, which train on multiple time-series, offer the advantage of simplicity: There is no need to tune the hyper-parameters and to spend time and compute resources on training tens of thousands of different models. Previous work indicates that global models can outperform local models for time-series forecasting [5]. We argue that the forecasting community should focus more on understanding the strengths and drawbacks of global deep learning models compared to local models.

A large disparity among time-series in the training data can negatively affect global model performance. Several studies suggest that semi-global models could help address this issue [25, 28]. This way we would get the best of both worlds: Robust models which generalise well and are not too resource intensive to train, but that are still specific enough to yield performance on a par with, or better than local models. To accomplish this, however, more knowledge is needed about how to best group time-series together for training. It is not understood today which characteristics of time-series are relevant for semi-global training clusters, and which characteristics are detrimental to forecasting performance. It would also be interesting to examine whether there are cut-offs in the diversity of time-series or in the number of time-series needed in a training cluster to produce high-performing models.

A counterargument to using clustering for semi-global models is that forecasting systems are already complex. Adding a clustering step to the forecasting process can make model maintenance more challenging. This could lead to slower development of such systems, specifically in general upkeep of models, and the designing of automatic re-training regimen. Additionally, it may be challenging to leverage conditional models with exogenous variables in a global or semi-global context, because the effects of the exogenous variables could differ too much across the time-series. Seasonal components would also need to be accounted for, which adds additional complexities. However, the performance increase could very well outweigh these issues, but more work is needed to fully understand this.

In real-world situations, it is common that some time-series are more important to forecast well than others. For this reason, a purely global forecasting approach may not produce the best results. A hybrid mix of local and global models could work well to ensure satisfactory performance on the important time-series. However, this requires a clearer understanding of which time-series characteristics determine whether global or local models are more effective. Perhaps transfer learning from a global model to a fine-tuning on a single time-series can yield stronger performance than either a global or local model can? Our experience from working with forecasting models is that some time-series are more difficult to forecast than others, and that training local models on such time-series often does not yield any valuable results whatsoever. However, when a global model trained on other time-series is applied to the same time-series the forecasting performance is improved. Although this is an anecdotal observation, it highlights that it could be worthwhile to study whether certain inherent characteristics make some time-series more difficult to forecast using local models.

4. Pilot experiment: demand forecasting in smart purchasing system

To validate the usefulness of demand forecasting in an MTO context, we perform a pilot experiment integrating demand forecasting into an existing purchasing decision support system. We simulate the system for one year and compare the savings with and without demand forecasting. We first provide a high-level description of the system's purpose and implementation, before detailing the experiment and results.

4.1. Purchasing task description and system implementation

An AI-assisted purchasing decision support system has been built for a company with an MTO and mass customisation strategy. The system helps purchase material sub-components that are unique to each product and customer, and need to be bought frequently and in large quantities. The material sub-component is one of several components used to produce the articles sold to customers. There are 2000 - 3000 unique articles active at any time, and anything between 50 and 100 different components are purchased each week at varying quantities.

The supplier uses bulk pricing: Purchases can be combined to reach different price tiers, affecting the final price per component. The system helps taking calculated risks outside of the known requirements to get a lower price per component. Components become obsolete if customers change their product design, so the company risks any component bought above the order-based quantities (purchasing requirements). Additionally, components have a shelf-life of roughly 9 months, after which they degrade too much in quality and cannot be used to produce the finished articles.

The decision support system uses a machine learning model which can estimate the probability of obsolescence P_O for a component purchase at a specific quantity X. The target variable is based on historical purchases and component usage before obsolescence: For training instances with historical known data, if we are assessing a purchasing quantity of $X = 10{,}000$ components, and the historical data shows that 100,000 components were used up before the customer discontinued the product version, then the regression target becomes $\frac{X}{100{,}000} = 0.1$. The input features include factors like customer consistency in purchase quantities, purchase frequency, and component lifetime before design changes. Other features include time since design swap, volume since design swap, and price list specific features.

The system takes a list of known purchase requirements and returns purchasing suggestions to the purchaser, who makes the final decision. Because the price lists are available during use, we know the following for a potential quantity increase X: (1) The monetary risk of increasing the quantity; (2) The best-case monetary reward of increasing the quantity. With these components together with the probability of obsolescence P_O, we can compute the expected value $\mathbb{E}(X)$ of increasing the purchase quantity:

$$\mathbb{E}(X) = (1 - P_O) \cdot \text{reward}(X) - P_O \cdot \text{risk}(X) \tag{1}$$

The system considers several quantity increases per component and chooses the one with the highest expected value. After this the system must calculate the expected values for the other quantity increases again, because the conditions have changed due to the previous quantity increase. Quantity increases are made until there are no more positive expected values, or other stopping criteria such as a time limit are reached.

4.2. Experiment setup

The experiment compares three different purchasing approaches: (1) Human-based purchasing; (2) System-based purchasing; (3) System-based purchasing enhanced with demand forecasting. We simulate one year of system use and compare the purchasing outcomes to the actual human purchasing done for the same year. The simulation tracks excess stock from purchase increases and subtracts these quantities from future purchase requirements to simulate the correct requirements. Additionally, expired and obsolete stock are tracked to further correct the excess stock. Expired stock occurs when a component has lied unused in stock for more than 9 months, and obsolete stock occurs when the customer updates their component to a new version, rendering the old version useless.

Summary statistics are computed after the simulation: How much did we spend compared to the human-based purchasing, and how many extra components did we buy? What component quantities went unused either due to expiration, or due to becoming obsolete because customers changed versions? How much stock remained at the end of the simulation (which would not become obsolete), and how much would that have cost to purchase with human purchasing? We estimate the cost of excess components using human purchasing and add it to the baseline cost to compute an estimated ground truth. The system's total spending is compared to this ground truth to estimate the percentage of monetary savings.

Two different simulations are run: The standard system-based purchasing uses the implementation described in subsection 4.1. The second simulation builds on the system by including demand forecasting to improve the decision making. We use TSB [21] for intermittent and lumpy demands and Prophet [33] for smooth and erratic demands. These models are chosen because they are readily available and can produce decent forecasts without much effort. TSB applies separate single exponential smoothing to estimate the demand probability and average demand size, which are then multiplied to produce a demand rate forecast [21]. Prophet performs an automatic decomposition into trend, holiday, and seasonal components, which are forecasted separately and then combined additively to produce the final forecast [33]. For each purchasing occasion, demand forecasts are computed for all included components monthly. The demand forecasts are based on the sales orders of the finished article which contain the material sub-component. The forecasted demand for the coming 9 months is set as an upper limit of the purchasing quantity for that article's sub-component, with the hope that this will reduce expired components. In addition to this, we set a max quantity cap of five times the required quantity. For articles with fewer than 3 historical sales, the forecast is skipped and a much lower max cap ($2\cdot$ the required quantity) is set to reflect the relatively high uncertainty of these articles. These limit parameters were found through trial and error by observing forecasts and purchasing suggestions in different scenarios.

4.3. Results and analysis

Adding forecasts to the system improves on all metrics (Table 1): The system achieves a significantly higher savings across one year of purchasing, and manages to do so while taking smaller risks and over-purchasing a smaller quantity in total. Additionally, the system is more successful at taking its risks (Table 2), mainly due to components expiring less frequently. This shows that adding demand forecasts to the system achieves the desired results: Fewer components expire due to not being used within 9 months of being purchased. The system without demand forecasts still outperforms human purchasing, but does so by taking less-calculated risks and relying on a more brute-force approach. The non-demand forecasting approach takes significantly larger risks (5.19 % across one year) than the demand forecasting approach which spends roughly the same amount of money as a human purchaser for a much higher quantity, most of which ends up being used. This occurs because of the bulk pricing, which causes many small purchases to have a very high component price, whereas bundling several purchases together in one larger purchase can lower the cost despite increasing the quantity. Overall, the results show that including demand forecasts in the system increases the profit margin and also has the potential to reduce the environmental impact from unnecessary transports.

Table 1
Comparison of overall system performance with and without demand forecasting. All metrics are compared to the baseline human purchasing across one year.

Metric	System	System + Forecasting
Total estimated monetary savings (%)	5.33	9.02
Over-purchasing quantity (%)	28.60	17.78
Monetary risk taken (%)	5.19	~0.00

Table 2
Comparison of system performance in over-purchasing decisions (quantities) with and without demand forecasting.

Metric	System	System + Forecasting
Discarded due to obsolescence (%)	16.62	14.33
Discarded due to expiration (%)	39.09	22.55
Successful (%)	44.29	63.12

5. Summary and roadmap

Demand forecasting is an important challenge for many organisations. MTO manufacturers face added complexities, causing them to avoid demand forecasting. The large number of articles warrants using global models that can forecast many time-series without training each one explicitly. Additionally, the per-article demand in MTO organisations is often sparse, causing an intermittent nature in the time-series.

Global and semi-global models could be trained better if we had more insights into how performance is affected by different time-series characteristics. It is essential to develop intermittent forecasting models which do not only produce an average demand rate across time, but that predict when demand is likely to occur and how large it will be. To assess such forecasts, new metrics need to be developed that assess how well each actual demand is predicted both in timing and size, and not just on a macro level.

The pilot experiment demonstrates that it is possible to use demand forecasting to significantly improve processes in an MTO context, and that even simple forecasting models can yield valuable results when embedded into operational decision support systems. The system with forecasts achieved higher savings and took smaller risks compared to the system without forecasts, resulting in a higher profit margin and a lower environmental impact. These results motivate putting focus on addressing demand forecasting challenges from an MTO manufacturing perspective, both in developing forecasts better suited to the problems for such organisations, and in performing experiments to validate the approaches.

A research roadmap is proposed to address the knowledge gaps discussed in this position paper:

- **MTO forecasting challenges**
  - What are the challenges faced by MTO organisations when it comes to demand forecasting? What type of information and systems would be needed to make full use of demand forecasts in operations? What are the data and software requirements for such forecasting systems? Do explanatory methods need to be developed for the forecasts to be useful as decision support to humans? We are currently investigating this though a real-world case study.
  - In what areas of operation can demand forecasts be used in MTO organisations, and what performance is needed? An important next step is to perform empirical studies which examine actual uses of forecast in different scenarios to support production planning and purchasing. The pilot experiment in this paper is a promising first step, but future experiments should also study additional aspects such as the effects of using global models, or on incorporating intermittent demand forecasts which predict both demand timing and size.
- **Global time-series models**
  - How can the correct "likeness" or "diversity" be determined for data used to train semi-global models for producing better forecasts than local or global models? How do we construct hyperparameter searches on specific datasets to find the optimal training data diversity? Which time-series features are the most relevant to represent diversity in this regard?
  - What time-series characteristics determine whether time-series benefit from using global, semi-global, or local models?
- **Intermittent time-series models**
  - How can intermittent time-series data be processed, and how do we construct deep learning models that train on the data in order to forecast both demand size and demand timings accurately? Can these models be global instead of local, and how are the methods for building global intermittent models similar or different to global models aimed at smooth time-series?

* Are some intermittent time-series too difficult for deep learning models to learn? Should these time-series instead be left to proven statistical methods such as the Croston method? What are the characteristics of such time-series, and is it possible to identify them automatically?

- What is the best way to assess the strength intermittent time-series forecasts both in terms of demand size and timing? Which metrics best represent the usefulness in MTO organisations or other organisations where sparse demand is common?

Acknowledgments

This research was sponsored via the KKS project SERT (Software Engineering ReThought), www.rethought.se.

References

[1] David Salinas et al. "DeepAR: Probabilistic Forecasting with Autoregressive Recurrent Networks". en. In: *International Journal of Forecasting* 36.3 (July 2020), pp. 1181–1191. ISSN: 01692070. DOI: 10.1016/j.ijforecast.2019.07.001.

[2] Jack Meredith and Umit Akinc. "Characterizing and Structuring a New Make-to-Forecast Production Strategy". In: *Journal of Operations Management* 25.3 (2007), pp. 623–642. DOI: 10.1016/j.jom.2006.04.006.

[3] Umit Akinc and Jack R. Meredith. "Make-to-Forecast: Customization with Fast Delivery". en. In: *International Journal of Operations & Production Management* 35.5 (May 2015), pp. 728–750. ISSN: 0144-3577. DOI: 10.1108/IJOPM-12-2012-0567.

[4] Nikolaos Kourentzes. "Intermittent Demand Forecasts with Neural Networks". en. In: *International Journal of Production Economics* 143.1 (May 2013), pp. 198–206. ISSN: 09255273. DOI: 10.1016/j.ijpe.2013.01.009.

[5] Pablo Montero-Manso and Rob J. Hyndman. "Principles and Algorithms for Forecasting Groups of Time Series: Locality and Globality". en. In: *International Journal of Forecasting* 37.4 (Oct. 2021), pp. 1632–1653. ISSN: 01692070. DOI: 10.1016/j.ijforecast.2021.03.004.

[6] N.S. Adshead and D.H.R. Price. "Demand forecasting and cost performance in a model of a real manufacturing unit". In: *International Journal of Production Research* 25.9 (1987), pp. 1251–1265. DOI: 10.1080/00207548708919913.

[7] Shaw K. Chen and Maling Ebrahimpour. "Demand forecasting for a jewelry packaging company: pattern identification and assessment". In: *International Journal of Production Economics* 22.3 (1991), pp. 203–209. DOI: 10.1016/0925-5273(91)90096-C.

[8] Marius Syberg et al. "Framework for Predictive Sales and Demand Planning in Customer-Oriented Manufacturing Systems Using Data Enrichment and Machine Learning". en. In: *Procedia CIRP* 120 (2023), pp. 1107–1112. ISSN: 22128271. DOI: 10.1016/j.procir.2023.09.133.

[9] B. Joseph Pine II, Bart Victor, and Andrew C. Boynton. “Making Mass Customization Work”. In: *Harvard Business Review* 71 (1993).

[10] Bryan Lim et al. “Temporal Fusion Transformers for Interpretable Multi-Horizon Time Series Forecasting”. en. In: *International Journal of Forecasting* 37.4 (Oct. 2021), pp. 1748–1764. ISSN: 01692070. DOI: 10.1016/j.ijforecast.2021.03.012.

[11] Aris A. Syntetos, John E. Boylan, and J. D. Croston. “On the categorization of demand patterns”. en. In: *Journal of the Operational Research Society* 56.5 (May 2005), pp. 495–503. DOI: 10.1057/palgrave.jors.2601841.

[12] Fatima Zohra Benhamida et al. “Demand Forecasting Tool for Inventory Control Smart Systems”. en. In: *Journal of Communications Software and Systems* 17.2 (2021), pp. 185–196. ISSN: 18456421, 18466079. DOI: 10.24138/jcomss-2021-0068.

[13] Chongshou Li and Andrew Lim. “A Greedy Aggregation–Decomposition Method for Intermittent Demand Forecasting in Fashion Retailing”. en. In: *European Journal of Operational Research* 269.3 (Sept. 2018), pp. 860–869. ISSN: 03772217. DOI: 10.1016/j.ejor.2018.02.029.

[14] George E. P. Box and Gwilym M. Jenkins. *Time Series Analysis: Forecasting and Control*. San Francisco: Holden-Day, 1970.

[15] Mononito Goswami et al. “MOMENT: A Family of Open Time-series Foundation Models”. In: vol. 235. 2024, pp. 16115–16152.

[16] Yuxuan Liang et al. “Foundation Models for Time Series Analysis: A Tutorial and Survey”. In: 2024, pp. 6555–6565. DOI: 10.1145/3637528.3671451.

[17] J.D. Croston. “Forecasting and Stock Control for Intermittent Demands”. In: *Operational Research Quarterly* 23.3 (1972), pp. 289–303. DOI: 10.1057/jors.1972.50.

[18] Aris A. Syntetos and John E. Boylan. “On the Bias of Intermittent Demand Estimates”. en. In: *International Journal of Production Economics* 71.1-3 (May 2001), pp. 457–466. ISSN: 09255273. DOI: 10.1016/S0925-5273(00)00143-2.

[19] Aris A. Syntetos and John E. Boylan. “The Accuracy of Intermittent Demand Estimates”. en. In: *International Journal of Forecasting* 21.2 (Apr. 2005), pp. 303–314. ISSN: 01692070. DOI: 10.1016/j.ijforecast.2004.10.001.

[20] Estelle A. Shale, John E. Boylan, and F. R. Johnston. “Forecasting for Intermittent Demand: The Estimation of an Unbiased Average”. In: *Journal of the Operational Research Society* 57.5 (2006), pp. 588–592. DOI: 10.1057/palgrave.jors.2602031.

[21] Ruud H. Teunter, Aris A. Syntetos, and Mohamed Zied Babai. “Intermittent Demand: Linking Forecasting to Inventory Obsolescence”. In: *European Journal of Operational Research* 214.3 (2011), pp. 606–615. ISSN: 0377-2217. DOI: 10.1016/j.ejor.2011.05.018.

[22] Sepp Hochreiter and Jürgen Schmidhuber. “Long Short-Term Memory”. In: *Neural Computation* 9.8 (1997), pp. 1735–1780. DOI: 10.1162/neco.1997.9.8.1735.

[23] Ashish Vaswani et al. “Attention is All You Need”. In: *Advances in Neural Information Processing Systems*. Vol. 2017-December. 2017, pp. 5999–6009.

[24] Jože Rožanec et al. "Explaining Bad Forecasts in Global Time Series Models". en. In: *Applied Sciences* 11.19 (Oct. 2021), p. 9243. ISSN: 2076-3417. DOI: 10.3390/app11199243.

[25] Kasun Bandara, Christoph Bergmeir, and Slawek Smyl. "Forecasting Across Time Series Databases Using Recurrent Neural Networks on Groups of Similar Series: A Clustering Approach". en. In: *Expert Systems with Applications* 140 (Feb. 2020), p. 112896. ISSN: 09574174. DOI: 10.1016/j.eswa.2019.112896.

[26] Rui Ye and Qun Dai. "Implementing Transfer Learning Across Different Datasets for Time Series Forecasting". en. In: *Pattern Recognition* 109 (Jan. 2021), p. 107617. ISSN: 00313203. DOI: 10.1016/j.patcog.2020.107617.

[27] Rob J. Hyndman, Earo Wang, and Nikolay Laptev. "Large-Scale Unusual Time Series Detection". en. In: *2015 IEEE International Conference on Data Mining Workshop (ICDMW)*. Atlantic City, NJ, USA: IEEE, Nov. 2015, pp. 1616–1619. ISBN: 978-1-4673-8493-3. DOI: 10.1109/ICDMW.2015.104.

[28] Angel López Oriona, Pablo Montero Manso, and José Antonio Vilar Fernández. *Time Series Clustering Based on Prediction Accuracy of Global Forecasting Models*. en. arXiv:2305.00473 [cs, stat]. Apr. 2023.

[29] Peter Wallström. "Evaluation of Forecasting Techniques and Forecast Errors". Available at https://www.diva-portal.org/smash/get/diva2:990519/FULLTEXT01.pdf. Licentiate thesis. Luleå, Sweden: Luleå University of Technology, May 2009.

[30] Nikolaos Kourentzes. "On Intermittent Demand Model Optimisation and Selection". en. In: *International Journal of Production Economics* 156 (Oct. 2014), pp. 180–190. ISSN: 09255273. DOI: 10.1016/j.ijpe.2014.06.007.

[31] Dominik Martin, Philipp Spitzer, and Niklas Kühl. "A new metric for lumpy and intermittent demand forecasts: Stock-keeping-oriented Prediction Error Costs". In: *Proceedings of the Annual Hawaii International Conference on System Sciences*. Vol. 2020-January. 2020, pp. 974–983.

[32] Jamal Fattah et al. "Forecasting of demand using ARIMA model". In: *International Journal of Engineering Business Management* 10 (2018). DOI: 10.1177/1847979018808673.

[33] Sean J Taylor and Benjamin Letham. "Forecasting at scale". In: *PeerJ Preprints* 5 (Sept. 2017), e3190v2. DOI: 10.7287/peerj.preprints.3190v2.