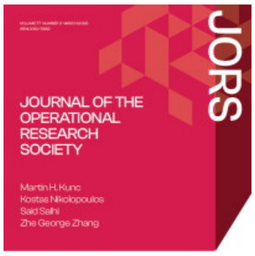

Journal of the Operational Research Society

ISSN: 0160-5682 (Print) 1476-9360 (Online) Journal homepage: www.tandfonline.com/journals/tjor20

Drug demand forecasting for hospital pharmacies using temporal hierarchies

Devon Barrow, Nikolaos Kourentzes, Martin Utley & Hazel Kirkland

To cite this article: Devon Barrow, Nikolaos Kourentzes, Martin Utley & Hazel Kirkland (06 Feb 2026): Drug demand forecasting for hospital pharmacies using temporal hierarchies, Journal of the Operational Research Society, DOI: 10.1080/01605682.2026.2620515

To link to this article: https://doi.org/10.1080/01605682.2026.2620515

JOURNAL OF THE OPERATIONAL RESEARCH SOCIETY
https://doi.org/10.1080/01605682.2026.2620515

Drug demand forecasting for hospital pharmacies using temporal hierarchies

Devon Barrow[a], Nikolaos Kourentzes[b], Martin Utley[c] and Hazel Kirkland[d]

[a]University of Birmingham, Birmingham, UK; [b]Institutionen för informationsteknologi, Högskolan i Skövde Högskolevägen, Skövde, Sweden; [c]Clinical Operational Research Unit, University College London, London, UK; [d]Applied Information, Digital Data Technology Service, Nottinghamshire Healthcare NHS Foundation Trust, Nottingham, UK

ABSTRACT

The provision of patient care in hospital pharmacies requires forecasting across a range of operations planning horizons to support adequate drug availability. This means that demand forecasts need to (1) be aligned across various planning horizons to support inventory management, (2) accommodate volatile drug delivery lead times and (3) be robust against erratic demand patterns including varying levels of intermittency. These are the challenges observed at the UK-based hospital pharmacy in the study. In response, we propose constructing forecasts that leverage the different time scales intrinsic to hospital pharmacies' inventory management. Using temporal hierarchies, we mitigate the challenge of intermittency and volatility in drug demand, while also enabling coherent decisions across planning time scales. Across the range of service requirements, lead times, and from 'Drug' to 'Drug by Dispensary' planning, the proposed approach, based on temporal hierarchies, ranks consistently well, reducing modelling risk and supporting automation. It ranks statistically best on pinball loss when planning for 95–99% service at the Drug by Dispensary level for all lead times from 1 to 20 days, a key challenge for the hospital. The approach is model-agnostic allowing hospital pharmacies to adopt either state-of-the-art forecasting models, or adjust to existing software and modelling capabilities.

ARTICLE HISTORY
Received 17 December 2024
Accepted 18 January 2026

KEYWORDS
Hospital; pharmacy; drug; forecasting; temporal hierarchies

1. Introduction

Within hospital pharmacies, drug demand forecasting is central to the effective management of drug inventory, improving efficiencies (Khalil Zadeh et al., 2014) and ensuring quality of patient care (Litvak & Long, 2000). Understanding the demand for each drug can reduce holding costs, optimise the frequency of wholesaler orders, and reduce the time pharmacy staff spend on managing inventory instead of valuable clinical work (Lega et al., 2013). Forecasting within hospital pharmacies however comes with several complexities. Often the supply chain is highly regulated with lead times of up to a week or more for external supply, and shorter lead times from one day to a few hours across the internal hospital network . The internal pharmacy network will also have multiple stockholding areas exhibiting high levels of demand variation (Beier, 1995; George & Elrashid, 2023). Additionally, to provide specialist patient care and to respond to emergencies, hospital pharmacies will stock costly specialist medicines not available from retail pharmacies and having low demand and high levels of intermittency. These characteristics mean that demand forecasts must:

1. be aligned across various planning horizons to support inventory management at various areas of the logistic network,
2. accommodate volatile drug delivery lead times, and
3. be robust against erratic demand patterns including varying levels of intermittency.

In the case of the hospital under study, these forecasts also need to be automated to allow pharmacist to focus on more value-adding activity. Forecasting therefore needs to be robust and reliable across challenges (1) to (3).

There are several forecasting challenges that a potential solution needs to address. Firstly, the desired forecast horizon at the aggregate planning level can be relatively long, especially on daily data granularity. For the chosen hospital pharmacy, this can be as long as up to 6 weeks ahead for some products. This implies a choice among different models performing well for short and long-term forecasts as is well documented in the literature (Makridakis et al., 2022; Makridakis & Hibon,

CONTACT Devon Barrow d.k.barrow@bham.ac.uk Department of Management, Birmingham Business School, University of Birmingham, Birmingham, B15 2TY, UK.

Supplemental data for this article can be accessed online at https://doi.org/10.1080/01605682.2026.2620515.

2000). Second, the delivery lead times that the hospital pharmacy faces are volatile. Longer lead times result in increased uncertainty and higher safety stocks, while shorter lead times reduce forecast uncertainty, resulting in lower safety stock and holding costs albeit at usually higher ordering costs (Sagaert et al., 2019). This creates a requirement for robust forecasting performance across multiple long-term horizons, and fine-tuning the performance for a specific target forecast horizon is of limited benefit (Van Belle et al., 2025). A less obvious modelling issue is that forecasts within the range of forecast horizons that span the volatile delivery lead times can be volatile themselves. This can manifest in a degradation of the forecasting performance of a given model past a given horizon, reducing the robustness of models across horizons and substantially complicating model selection in the presence of stochastic forecast horizons.

Assuming a reliable evaluation methodology, one can mix forecasts from different models across horizons to obtain minimal errors. However, this approach can introduce discontinuities in the forecasted demand trajectories, where for, instance, the forecast between two horizons may have substantial jumps as the forecasts originate from different models. This can reduce the trustworthiness of the forecasts, leading to unnecessary manual adjustments by demand planners (Spavound & Kourentzes, 2022), and distortions in replenishment orders. Finally, multiple time series have intermittent demand, while a minority have continuous or almost continuous demand. This may lead to the need to use diverse models to cover the various cases, further complicating model implementation and selection in practice. While these challenges and the potential solutions are not unique taken one at a time, together they present a challenge for the hospital pharmacy under study and will apply to other similar operational environments within and outside of healthcare where forecasting is required.

To address these challenges and facilitate automated forecasting, we propose forecasting with temporal hierarchies to mitigate modelling uncertainty, address the challenge of intermittency, and produce reconciled forecasts across multiple lead times (Athanasopoulos et al., 2017; Kourentzes & Athanasopoulos, 2019). Now, demand forecasting challenges with multiple stockholding levels are typically modelled with hierarchical forecasting (Athanasopoulos et al., 2024), however, in the case of the given hospital pharmacy, we show that the structure of the supply chain, is mirrored well in a temporal hierarchy. We leverage the hierarchical structure to address the aforementioned demand forecasting challenges and potential solution conflicts. This way, we (1) align forecasts across planning horizons, (2) provide consistent accuracy improvements across horizons, mitigating the complexity of forecasting for volatile lead times, and (3) simplify forecasting for drugs with intermittent demand. We demonstrate the benefits both in terms of forecast and quantile performance metrics, the latter acting as a proxy for inventory performance. Beyond mirroring decision requirements in the hospital pharmacies, temporal hierarchies have the advantage that they are forecasting model independent. This makes them easily applicable within the analytical constraints of organisations. For the case study hospital, this is a major advantage, given their forecasting capability are in the infancy stage. This flexibility, allows us to benchmark standard forecasting models for the creation of accurate base forecasts, providing tangible guidelines for improving the forecasting process in organisations.

In the following sections, we discuss the current state of demand forecasting within the wider pharmaceutical sector and specifically hospital pharmacies. We discuss the inventory decisions for which these forecasts are needed, the forecasting challenges faced and motivate the use of temporal hierarchies for addressing these challenges . Using a real dataset from a pharmacy serving one of the largest non-acute hospitals in the UK, we evaluatethe application of forecasting with temporal hierarchies for improving demand forecasting within hospital pharmacies .

2. Forecasting drug demand for inventory management

For hospital pharmacies, demand forecasting to support the management of drug inventory is critical. A mismatch between demand and supply will impact patient care, and being at a critical point of care, can lead to loss of life. This means that forecasting to facilitate availability of drugs is of vital importance. This forecast is generally needed within the short to medium term from 1 to 8 weeks ahead (Cook, 2016) and will vary based on supplier, drug, and patient needs, and the location of a given drug within the internal supply network of the hospital. Hence the drug delivery lead time will be volatile, and planning horizons will vary across the internal supply network of the hospital with longer horizons for strategic planning and budgeting. As such this literature review will focus on the current approaches to forecasting available to hospital pharmacies to facilitate planning for drugs considering the wider pharmaceutical sector. It will review the challenge presented in accommodating volatile delivery lead times and what forecasting has to offer in dealing with the resulting erratic demand patterns arising from drugs having varied demand characteristics across the hospital internal supply network.

2.1. Multiple planning levels

A key challenge for hospital pharmacies is obtaining forecasts of drug demand which are aligned across the various planning levels and horizons required to support inventory management at the various points of their internal logistic network from wards to pharmacy stores. What we observe in the literature however, is that developments in forecasting methodology have in the main reflected the needs of the actors at various levels of the supply network (Zhu et al., 2021). Focusing on inventory planning, we observe that in the upper levels of the supply network decision makers are often interested in forecasts to support new drug development, manufacturer production needs, and market strategies. Typical data granularities are monthly and annual, reflecting that even for inventory management, these actors consider much longer time scales than actors closer to patients, as is the case with hospital pharmacies. As a result of the data granularity, most of these studies utilise univariate models capable of tracking market potential (Nikolopoulos et al., 2016; Restyana et al., 2021; Siddiqui et al., 2022; Sousa et al., 2019). For retail pharmacies the planning horizon is typically weeks to months as the demand for drugs is prescription-based not being a critical point of care as are hospital pharmacies. The focus of many forecasting studies is on the inclusion of various types of exogenous variables to capture market movements and promotions in response to consumer needs and patient perceptions (Abolghasemi et al., 2020; Anusha et al., 2014; Mousa & Al-Khateeb, 2023; Pall et al., 2023; Papanagnou & Matthews-Amune, 2018). Some studies have investigated machine learning methods (Hussein et al., 2019; Ribeiro et al., 2017) but again with no focus on aligning forecasts across planning horizons.

For settings close to the point of care, such as hospital pharmacies, there is forecasting research at multiple levels of decision-making. At the patient level, Fruggiero et al. (2012) propose to forecast drug demand at pharmacies using hospitalised patient diagnostic data. In their scenario, forecasts are patient-specific and begin when the patient is hospitalised. Vila-Parrish et al. (2012) develop an inventory policy for the inpatient hospital pharmacy based on a Markov Decision Process (MDP) used to represent and estimate drug demand over a lead time of one day, as a function of patient condition. There are other studies that focus on patient-level management of drug inventory, using for example MDPs and knowledge of patient health conditions (e.g., Biagi et al., 2018; Saha & Ray, 2019). Referring to Figure 2, in our case hospital, the lead time between dispensaries and wards often matches the 1-day ahead focus of the research, however that is too short for the overall hospital pharmacy forecasting requirements (see, Figure 3). There is forecasting at higher levels within hospital pharmacies but again these are not focused on the quality of forecasts and appropriateness for aligning inventory decisions across planning horizons.

Wettermark et al. (2010) apply linear regression analysis to annual hospital sales and drug data distributed in ambulatory care to predict prescription and hospital drug expenditure for early warning systems and horizon scanning. Silva-Aravena et al. (2020) develop a decision-aid tool for supporting purchasing and inventory decisions for hospital-level controlled drug pharmacies. They consider a planning period of one year and forecast monthly data for 34 drugs using polynomial regression. Both Ramirez et al. (2014) and Varghese et al. (2012) evaluate exponential smoothing and ARIMA for forecasting drug demand at a hospital pharmacy. They find that no method uniquely provided the best results, while other studies, e.g., Varghese et al. (2012), evaluate exponential smoothing and ARIMA to forecast weekly drug demand for hospital pharmacies and find potential for significant cost savings based on an (r, Q) inventory policy.

Several researchers have examined the impact of information sharing for improving forecasting across planning levels. Merkuryeva et al. (2019) study a supply chain in which distributors forecast product demand using sales data and stocks, while the wholesaler additionally uses market sales data from the distributors. Sohrabi et al. (2021) use a dataset of pharmacies (point of sale) dealing with a specific pharmaceutical company in Iran to predict drug sales at the manufacturer level. Utilising decisions trees and neural networks to investigate links between physicians and pharmacy sales, they found the most significant variables to be the number of pharmacy visits by either physicians or reps, the number of different physicians visiting the pharmacy, and the time between visits by a physician. More recently, Zhu et al. (2021) proposed a new forecasting framework to support decision-making in a study of a major pharmaceutical manufacturer and its supply chain. The approach leverages cross-drug information on historical demand and non-demand characteristics including downstream inventory data and supply chain structure information. Their work claims to be the first to empirically capture the value of downstream inventory information in a cross-series demand forecasting environment.

Other similar work (e.g., Khalil Zadeh et al., 2014; Kim et al., 2015; Van Belle et al., 2021) have tried to incorporate supply chain information in forecasting for drug manufacturers and distributors, with none focused on the critical care level of hospital pharmacies or their forecast inventory needs. Even with information, we observe that the aim of these studies isn't to produce a forecast which is aligned across multiple planning levels or reconciled to keep forecasts consistent across horizons.

2.2. Supply lead times and forecast horizon

The supply lead time faced by hospital pharmacies, being the time between when an order is placed and when it is available to satisfy patient demand, plays a crucial part in the ability of pharmacy to serve patients. It determines the (maximum) forecast horizon to be used by your forecasting model, which is lengthened for example, by any imposed review period. While in some cases this lead time can be controlled, it is the case that the delivery lead times for the hospital pharmacy under study are volatile. However, we observe that many of the research studies on forecasting for drug demand do not consider the issue of volatile supply lead times and many assume fixed lead times. Anusha et al. (2014) who study a retail pharmacy and Fruggiero et al. (2012) who produce forecasts for a local hospital pharmacy each assume constant supply lead time. Vila-Parrish et al. (2012) assume same-day production lead time in a study of perishable drugs for inpatient hospital pharmacies. Many more studies do not consider supply lead time at all and therefore do not address the challenge of robustness across multiple long-term forecast horizons.

Because the case study hospital has lead times which change constantly based on patient need, supply source, and availability of a drug within or outside their internal logistics network, forecasts of demand need to be robust across multiple horizons. The literature does offer up some solutions, but these are not without their challenges. One approach is to use multiple models, each forecasting a different planning horizon, with some models performing well for short horizons and some for long. This is known as the direct or independent strategy (Kline, 2004; Taieb et al., 2012; Tiao & Tsay, 1994). This approach requires greater computational effort especially at daily data granularity due to the number of models being equal to the horizon. It also reduces accuracy where there are dependencies in the time series across horizons, and so there is no guarantee of coherency across horizons. Also forecasts from one period to the next can potentially disagree substantially being from different models and can show large jumps which introduce new challenges for planning. A more common approach is the recursive or iterated strategy where a single model is used to produce forecasts across all required horizons (Kline, 2004; Taieb et al., 2012; Tiao & Tsay, 1994). A weakness of this strategy is that errors made in the past will propagate through the forecast and worsen at longer horizons. If several models are considered based on the iterative approach, then a challenge becomes selecting among models performing differently across different forecast horizons; that is, when no model is consistent across all horizons. We also observe that none of these strategies take into consideration the data granularity or inventory decision. For the case study hospital, data are available at the daily level with inventory decisions made at the daily and weekly level at different parts of the hospital pharmacy network. It also means that forecasts are required at both daily and weekly levels for inventory, and at multiple horizons of each. For a given planning decision and horizon (leadtime), it means that the original data e.g., daily can be used to generate the forecast, which can then be aggregated to the appropriate leadtime, or alternatively, the data can first be aggregated to the leadtime level and then the forecast produced (Boylan & Babai, 2016; Goodwin, 2018). It is here that temporal aggregation (TA) becomes a candidate for addressing the multiple challenges presented, in this section and the previous Section 2.1. Temporal aggregation (TA) transforms a time series from higher (e.g., daily) into lower frequencies (e.g., weekly, monthly), strengthening or attenuating different information contained within the time series across data granularity.

2.3. Drug demand patterns and intermittency

The varying characteristics of drugs and resulting patient needs, as well as the data granularity, can result in varied set of time series demand patterns. Many drugs exhibit intermittency being specialist drugs prescribed only occasionally, and therefore having long sequences of zeros followed by some demand occurrence. Some will show lumpiness being ordered irregularly for example for special patient cases, while others will have seasonal demand being linked to time of the year. Other factors such as changes in treatment guidelines and dosage will also result in changes in demand patterns over time. Patterns will also differ based on data granularity with many of the characteristics manifesting themselves differently at the daily, weekly and monthly level. Additionally, as is the case for the hospital pharmacy under study, demand will vary by location, for example, ward versus dispensary. For hospital pharmacies, drug demand is also linked to patient arrival and composition, and unlike retailers, hospitals may have to hold very high stocks of certain drugs to ensure high service level. To support good inventory planning, a forecasting methodology needs to address the intermittent nature of demand due to the characteristics of the drug and patient need, and the varying characteristics of time series data due to data granularity and multi-location demand. This may lead to the need to use diverse models to cover the various cases, further complicating implementation and model selection in practice. Again, temporal aggregation and hierarchical reconciliation are candidates for addressing these challenges, while keeping computational cost to a minimum and being model independent.

3. Drug demand in hospital pharmacies

A simplified representation of a drug supply chain network is provided in Figure 1. In the typical case, drugs flow from manufacturers to trade partners, such as wholesalers, through centralised distribution hubs and then to the final points of consumption. This is served by retail and hospital pharmacies (Chen et al., 2013; Nguyen et al., 2022), the latter being the focus of this study.

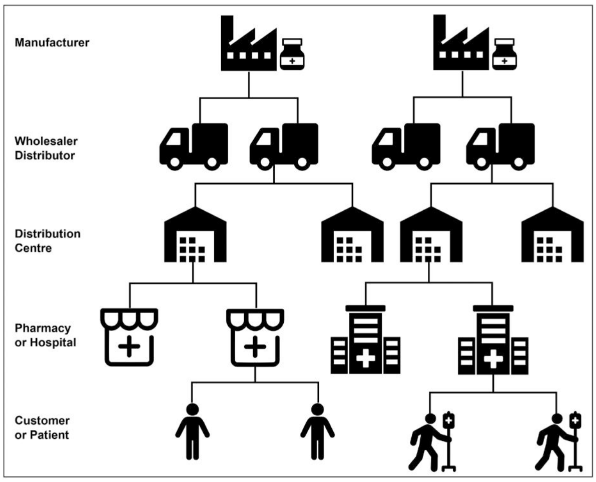

Figure 1. A drug supply network from manufacturer to customer/patient.

While the upstream part of the network remains relatively stable with pharmaceutical companies being the main developers and manufacturers of drugs, the remaining distribution network can vary extensively in structure and relationships.

3.1. The internal supply network in hospital pharmacies in the UK

In this work we focus on the case of a major non-acute hospital pharmacy in the UK. Within UK National Health Service (NHS) hospital pharmacies, the downstream (internal) network is comprised by (1) a pharmacy store, (2) a main dispensary, and (3) smaller satellite dispensaries which are located either within the same hospital building, or nearer to geographically remote sites, wards/outpatient clinics which serve patients, and (4) small stores of "stock medicines" housed at patient wards. Figure 2 shows the downstream internal supply network for the case study hospital pharmacy network.

The pharmacy store orders from wholesalers and in some cases directly from the manufacturers. The pharmacy store supplies medication to the dispensaries and stocks medication directly to wards and outpatient clinics. Dispensaries provide and label medications for individual patients according to their prescriptions. This medication can be supplied directly to the patient or delivered to the outpatient clinic or ward. During the study period, the layout of the network changed. Before the change, the pharmacy store and the main dispensary were combined. The current architecture, with the separated roles, is the more typical configuration within UK National Health Service (NHS) hospital pharmacies.

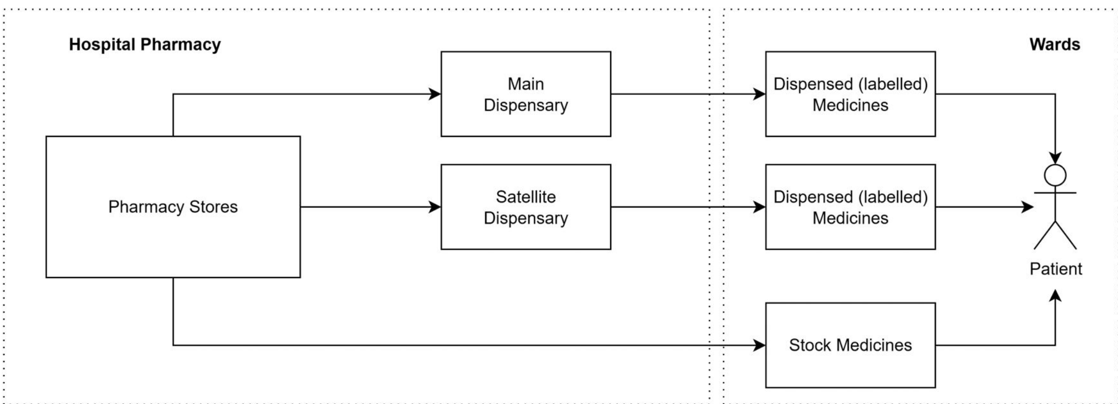

Figure 2. Configuration of the downstream supply chain within the case study hospital pharmacy.

3.2. Drug order processing and inventory planning

Pharmacy stores typically place an order with a specific wholesaler once a week, based on a preset reorder level for stocked drugs. While this allows the hospital pharmacy to schedule workload and group orders with each wholesaler, it runs the risk of potential stockouts during the weekly review period. As such interim orders need to be placed for stocked drugs where the actual stock level is insufficient for current demand as determined by procurement, and for non-stock drugs as required, for example, in response to a specific prescription. These orders can often disrupt the flow of operations taking up the time of the pharmacist which is better spent dispensing patient drugs rather than managing emergency order requests. The case hospital pharmacy deals with over 1,100 different drugs. One such example is paracetamol, a low-grade painkiller which has a long shelf life, is ordered in large batches and takes up limited storage space.

Dispensaries place daily orders to the stores for each stocked drug, based on its current stock level and established re-order point. Ad-hoc items are added to the order if a non-stock drug is required. For our case, the order quantity for most drugs is calculated to last approximately 2 wk. While a particular drug will be present across the pharmacy network, inventory management occurs primarily at the drug-by-dispensary level.

The hospital pharmacy in this study operates on a 2-week ordering schedule. The delivery lead time is different across drugs. Although it is typically a few days, it can reach up to two weeks. Figure 3 illustrates the standard case, where the demand is forecasted for sufficient days to cover the re-ordering cycle and the maximum delivery lead time. A wholesaler may impose a minimum order quantity affecting the re-ordering schedule and can bring the total forecast horizon up to 6 wk ahead. Although there is an organisational preference for adherence to the ordering schedule, this is not always feasible. Adding to the complexity, delivery lead times can be volatile. These suggest that any forecasts must perform well across a range of horizons and planning levels (Drug Store and Drug by Dispensaries) to reduce stockouts, emergency orders and ensure good inventory performance.

3.3. Drug demand data and forecasting needs

The hospital pharmacy retains demand data by drug by dispensary by ward. From the recorded data, we can construct demand measurements at various points of the internal supply network; however, not all are of decision-making interest. This is due to the structure of the network and the geographical proximity of the dispensaries and the wards, resulting in very short lead times for satisfying ward demand. Likewise, various potential data aggregations do not match relevant decisions. For example, demand by dispensary is uninformative as it is the drug information that is relevant for placing orders. The operational decision-relevant data aggregations are highlighted in Figure 4. Note that we use the term demand as per the usual convention in the forecasting literature, and data are of observed usage, excluding unmet needs.

Orders to manufacturers and wholesalers are placed per *Drug*, making it a forecasting objective, for the forecast horizons discussed before (see, Figure 3). Likewise, the pharmacy store needs to supply the dispensaries with the required drugs, making the *Drug by Dispensary* level decision-making relevant; this is the level where external inventory orders decisions are made. Although there is overlap between the types of drugs offered by each dispensary, there are also unique allocations. This blends the information content of the two

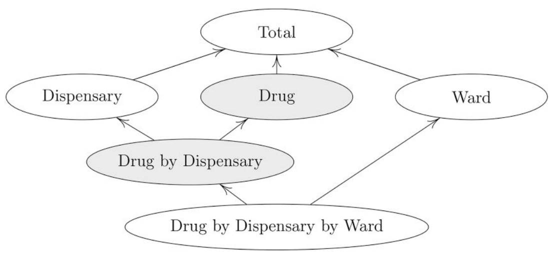

Figure 4. The data construction of the demand series, sampled from the Drug by Dispensary by Ward level. The shaded nodes are the focus of this work. Not all combinations of factors are visualised.

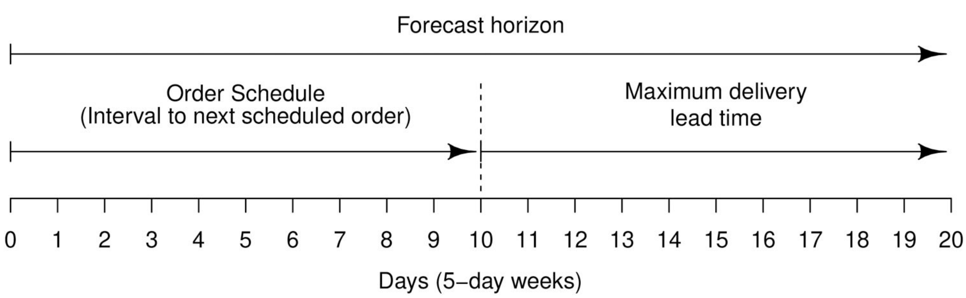

Figure 3. Order horizon.

levels. As there are limited drug substitutions, there is no benefit in looking at demand by drug category as this is beyond the current operations of the hospital pharmacy, and the scope of this study. Such information could be useful to wholesalers and manufacturers, as discussed in the literature overview.

All planning currently happens on data of daily granularity, and forecasts are produced at that level. The provided data span from 2nd April 2018 to 31st March 2022 and weekends are excluded, resulting in up to 1,044 observations per time series (series from one dispensary are only 433 days long). With the advice of experts at the hospital pharmacy, from all the available drugs we focus on 34 drug series. To establish a degree of generalisability for other hospital pharmacies, in addition to choosing series displaying different patterns of consumption, the drug series we focused on comprise a wide range of pharmaceutical use cases in terms of the forms of the drugs, whether a delay in administration of the drug could result in serious harm or death (criticality) and whether the drug is required at all points within the internal distribution network or just some (see Table 1).

When we look at the data as Drug by Dispensary this results in 120 time series. Many time series are highly intermittent. Figure 5 provides scatter plots of the percentage of zero demand periods against the coefficient of variation of the non-zero demand periods of the daily time series for the Drug and Drug by Dispensary series. It should be noted that observed zero values correspond to true absence of consumption, and they were validated as not being due to data issues. The Drug by Dispensary series being at a lower aggregate level exhibit greater intermittency. The plots show the relationship between demand intermittency and demand volatility. For both Drug and Drug by Dispensary series, highly intermittent time series are highly volatile.

In addition to modelling the daily demand data, we also model temporally aggregate data of weekly buckets. Given the required forecast horizons the weekly data have the advantage that 4 wk ahead are only 4-step ahead forecasts, rather than 20-steps ahead (5-day weeks), accumulating significantly fewer forecast errors. Moreover, temporal aggregation is a moving average operation, reducing the volatility of the time series as well as their intermittency, which decreases the average percentage of zero demand periods from 52.71 and 77.30% to 30.92 and 56.09% for the Drug and the Drug by Dispensary series respectively.

Table 1. Characteristics of drugs included in sample.

Dosage form	*Critical drugs	*Non-critical drugs
Tablets	11	10
Capsules	2	2
Prepared injections	4	3
Liquids	–	1
Sprays	–	1
Required at all dispensaries	13	14
Required at some only	4	3

*As per classification obtained from the case study hospital.

4. Temporal hierarchy forecasting for hospital pharmacies

4.1. The application of temporal hierarchies

In forecasting for hospital pharmacies, there are several forecasting challenges (see Section 1 and 2) that a potential solution needs to address. These require that demand forecasts (1) be aligned across various planning horizons to support inventory management, (2) accommodate volatile drug delivery lead times and (3) be robust against erratic demand patterns including varying levels of intermittency.To

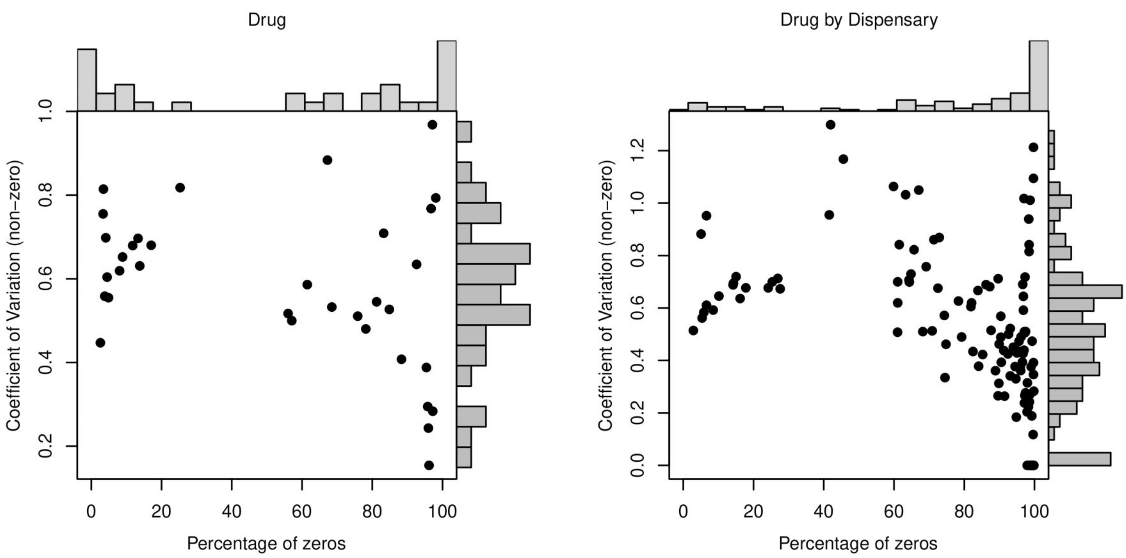

Figure 5. Visualisation of the intermittency of the daily time series. The percentage of zeros is plotted against the coefficient of variation of non-zero demand periods.

overcome these challenges, we propose using temporal hierarchies. The methodological underpinnings of Temporal Hierarchy Forecasting (THieF) share the same foundations as standard cross-sectional hierarchical forecasting (Athanasopoulos et al., 2024). The notion of using multiple levels of temporal aggregation for time series forecasting was introduced by Kourentzes et al. (2014) and embedded within the same methodological framework as hierarchical forecasting by Athanasopoulos et al. (2017). Although historically this was seen as a choice between bottom-up and top-down, i.e., forecasting at the bottom level of the hierarchy and aggregating, or forecasting at the most aggregate level of the hierarchy and disaggregating, nowadays, it is seen as a question of how to best reconcile forecasts from different hierarchical levels. This suggests that forecasts are produced at all levels of the hierarchy, independently, and then appropriately combined to generate a set of reconciled coherent forecasts. When forecasts are not required at different temporal aggregation levels, THieF uses the additional levels as statistical devices to improve the forecasting performance and reduce modelling uncertainty (Athanasopoulos & Kourentzes, 2023).

In relation to the case study, THieF combines forecasts on both the high frequency drug data with that of the low frequency. The targeted 20-day ahead drug forecast is merely 1-step ahead forecast on monthly data. Through the forecast combination, the accumulation of forecast errors across steps is mitigated. Kourentzes et al. (2014) and Athanasopoulos et al. (2017) show that THieF and similar approaches perform particularly well for longer-term forecasts, which can be explained by this effect and by the ability to model long-term trends better at temporally aggregate data views. Moreover, the forecast combination can reduce the modelling risk, where we do not have to rely on a single model for the most disaggregate data which may be misspecified, nor on a single modelling family. Panagiotelis et al. (2021) provide a proof that hierarchical forecasts will always have lower mean squared errors across the hierarchy than the base forecasts, which describes a similar effect. Given this, we can anticipate that THieF will exhibit similar forecasting performance for neighbouring forecast horizons, lessening the concern of volatile forecasting performance across the stochastic delivery lead times.

4.2. Implementation of temporal hierarchies

Given a time series y_t for periods $t = 1, \ldots, n$ and at sampling frequency m, we construct a non-overlapping temporally aggregate version of it, $y_j^{[k]}$, where k is a factor of m and $j = 1, \ldots, \lfloor n/k \rfloor$,

$$y_j^{[k]} = \sum_{t=t^*+(j-1)*k}^{t^*+jk-1} y_t, \tag{1}$$

where $m_k = m/k$. As (1) requires complete blocks of k observations, the summation starts from period $t^* = n - \lfloor n/m_k \rfloor m_k + 1$. Different levels of temporal aggregation filter some parts of the time series, while strengthening others. For example, at the annual level all seasonality has been filtered, and the long-term trend is easier to model. This suggests that with THieF we typically aggregate up to the annual level. Table 2 shows the temporal aggregation levels for the case of daily and weekly data, using a 5-day week for the daily series. Note that only the levels with integer multiples are considered, as these are factors of m. Figure 6 visualises the effect of temporal aggregation of the daily demand for a drug from our case study. At the original sampling frequency, the drug exhibits some intermittency, with no clear discernible patterns. Temporal aggregation reduces the intermittency but also makes it increasingly obvious that the demand has been increasing over the years. This reduction of intermittency motivated Kourentzes and Athanasopoulos (2021) to investigate the use of THieF for intermittent demand forecasting and found gains in mean and quantile forecast accuracy.

Furthermore, observe that in Table 2 the corresponding aggregation levels for a weekly time series are included in the daily temporal hierarchy. Therefore, by enforcing coherency across the temporal hierarchy, i.e., that the daily forecasts sum up to the weekly ones, and so on, any decisions done independently on the two planning levels will be conditioned on the same information. Naturally, this is connected with a requirement to plan in the face of longer lead times. The 1-step ahead prediction at the weekly level corresponds to 5-steps ahead at the daily level. Therefore, one can borrow predictive power for longer-term forecasts from temporally aggregate levels, instead of relying on multiple steps ahead forecasts at disaggregte levels that will accumulate forecast errors from each step-ahead. Even if one retains the daily and weekly forecasting processes separately, using THieF will provide both the more aggregate information from the levels represented in Figure 6, which due to their commonality will help in bringing the outcome of the otherwise disjointed forecasting processes close. This provides various implementation options that we expand upon in the discussion. Note that the prescribed temporal aggregation exclude calendar months. As these would have a different number of days per month, they can introduce an artificial seasonality in the data that can be complex

Table 2. Temporal aggregation levels for daily and weekly sampled series.

	Aggregation periods (k)											
Starting sampling frequency	Daily			Weekly	Bi-weekly		Monthly			Quarterly	Semi-annual	Annual
Weekly				1	2		4			13	26	52
Daily	1	2	4	5	10	13	20	26	52	65	130	260

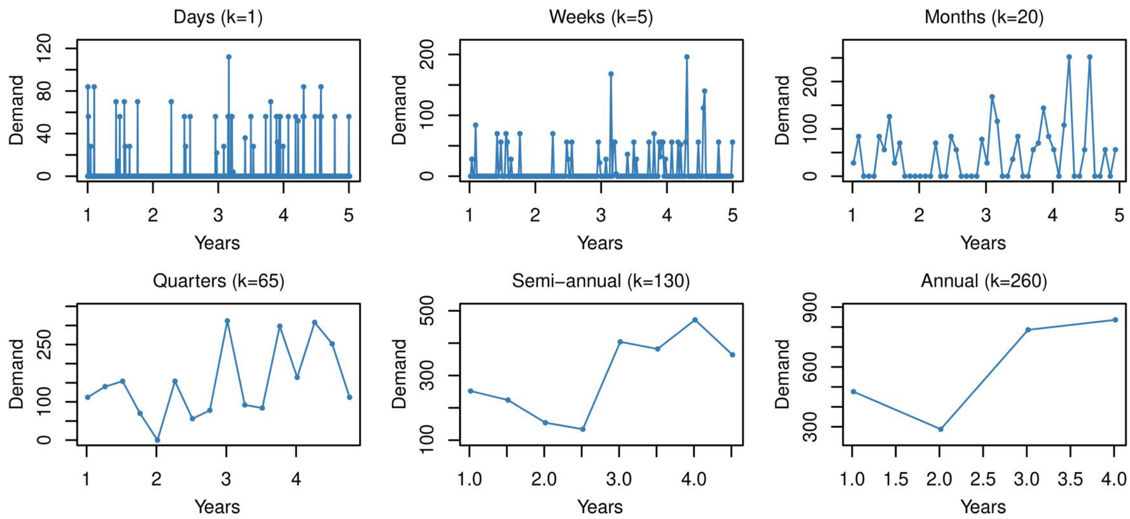

Figure 6. Examples of temporally aggregate views of the daily demand for a drug.

to model (Makridakis et al., 2008, Chapter 4). Instead, if the user requires forecasts at that level, we recommend to aggregate the coherent daily predictions as needed. This is not required in our case, and therefore omitted.

To achieve coherency across the temporal hierarchy, we leverage the well-studied hierarchical forecasting methodology (Athanasopoulos et al., 2017; Wickramasuriya et al., 2019). Figure 7 represents the information from Table 2 as a hierarchical structure for a single year, making the connection between THieF and hierarchical forecasting apparent. In Eq. (1) the index j is specific for each aggregation level. We keep track of timing for the complete hierarchy at the most aggregate level ($k = m$), using index i. Observations at any level can be written as

$$y^{[k]}_{m_k(i-1)+z}, \tag{2}$$

with $z = 1, \ldots, m_k$. Observations of any given level can be collected in a column vector

$$\mathbf{y}^{[k]}_i = \left(y^{[k]}_{m_k(i-1)+1}, y^{[k]}_{m_k(i-1)+2}, \ldots, y^{[k]}_{m_k i}\right)' \tag{3}$$

and observations from all hierarchical levels for a single period i into

$$\mathbf{y}_i = \left(y^{[m]}_i, \ldots, \mathbf{y}^{[1]'}_i\right)', \tag{4}$$

which contains all observations visualised as nodes of the hierarchies in Figure 7.

We further need to define a summing matrix $\mathbf{S}$ that maps how to aggregate from the originally sampled series at the bottom level of the hierarchy, to the complete representation. Given $\mathbf{b}_i = \mathbf{y}^{[1]}_i$ the complete hierarchy is constructed as

$$\mathbf{y}_i = \mathbf{S}\mathbf{b}_i. \tag{5}$$

The $\mathbf{S}$ is a $\sum m_k \times m$ matrix with binary elements that resolves the necessary linear combination. For example, if we were dealing with a quarterly time series, then

$$\mathbf{S} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ & \mathbf{I}_m & & \end{bmatrix}, \tag{6}$$

with the annual aggregate observation corresponding to the first row, the two semi-annual to the second and third rows, and the diagonal $\mathbf{I}_m$ connecting each quarter to itself. Obviously, $\mathbf{S}$ is much larger for daily and quarterly data (588×260 and 98×52 respectively, using 5-day weeks).

Having the data structure organised, we replicate it for the forecasts. In analogy to $\mathbf{y}_i$ we can organise the forecasts to a collection $\hat{\mathbf{y}}_i$, where the forecasts at each temporal aggregation level are generated independently. The method is model agnostic, so the forecasts can be obtained from any source, forecasting models or methods, or include expert judgement. This also means that in the typical case $\hat{\mathbf{y}}_i \neq \mathbf{S}\hat{\mathbf{b}}_i$, i.e., that the base forecasts for each temporal aggregation level will be incoherent. Although these forecasts are incoherent, each is based on different information, having access to different raw data at each different temporal aggregation level, which highlights different information (see Figure 6). This also suggests that the resulting models at

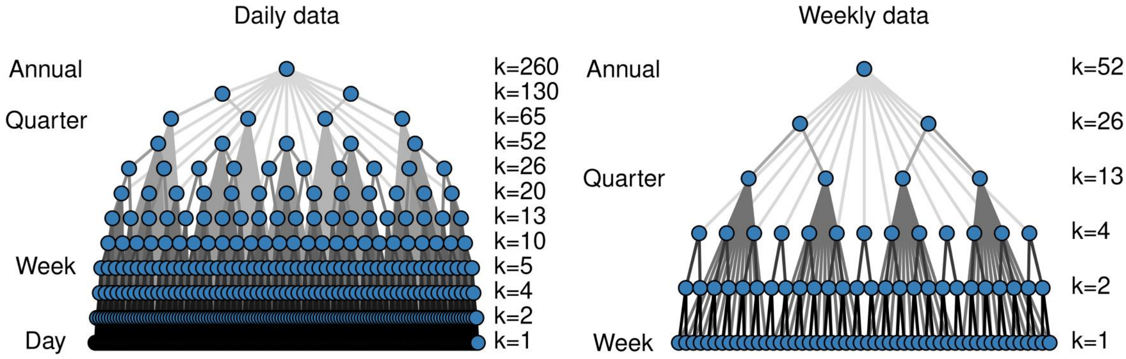

Figure 7. The temporal hierarchies for daily and weekly data.

each level will be different (Athanasopoulos et al., 2017; Kourentzes et al., 2014). To overcome these differences, in hierarchical forecasting, we combine all base forecasts at different levels in a coherent set of reconciled forecasts:

$$\tilde{\mathbf{y}}_i = \mathbf{SG}\hat{\mathbf{y}}_i, \tag{7}$$

where $\tilde{\mathbf{y}}_i$ is the vector containing the coherent forecasts, across all levels of the hierarchy. Equation 7 blends the information contained in all forecasts $\hat{\mathbf{y}}_i$, into a set of bottom-level forecasts, which are multiplied by $\mathbf{S}$ to generate forecasts for all levels (following from Eq. (5)). As all $\tilde{\mathbf{y}}_i$ are based on the new bottom-level forecasts, these are by construction coherent. Matrix $\mathbf{G}$ contains the necessary combination weights. Wickramasuriya et al. (2019) showed that

$$\mathbf{G} = (\mathbf{S}'\mathbf{W}^{-1}\mathbf{S})^{-1}\mathbf{S}'\mathbf{W}^{-1}, \tag{8}$$

where $\mathbf{W}^{-1}$ is the inverse of the variance-covariance matrix of the forecasts errors between the nodes in the hierarchy. $\mathbf{W}^{-1}$ is $\sum m_k \times \sum m_k$, and needs to be estimated on a number of observations prescribed by the most aggregate level, $\lfloor n/m \rfloor$. Therefore, a direct estimation is often impractical, and instead we rely on approximations.

In the literature, there are many alternatives (the reader is referred to Pritularga et al., 2021, for a discussion of various options). Here we rely on the proposal by Athanasopoulos et al. (2017), where $\mathbf{W}$ is a diagonal matrix with its elements being the variance of the forecasts, estimated as the in-sample mean squared errors. An important advantage of this option is that it does not require forecasts to be model-based, with the easy to calculate in-sample mean squared error being sufficient for $\mathbf{W}$. This allows the inclusion of method-based forecasts, as is the case for intermittent demand forecasting methods and machine learning methods, which may lack variance expressions for their forecasts.

5. Empirical evaluation

5.1. Empirical evaluation design

We conduct two experiments, one with daily and one with weekly observations. The last 65 days (13 wk) are kept as a test set. The corresponding forecast horizons are $h_{\text{day}} = \{1, 5, 10, 20\}$ days and $h_{\text{week}} = \{1, 2, 4\}$ weeks. The weekly horizons correspond to the daily ones as weekends are excluded. These horizons overlap various operational lead times. The longer test sets enable us to perform a rolling origin evaluation, with multiple forecasts being generated for each time series.

In terms of evaluation metrics we track the cumulative errors to construct the Root Mean scaled Squared cumulative Error (RMsSE) as:

$$\text{RMsSE}_{i,h} = \sqrt{\frac{1}{p}\sum_{o=1}^{p}\left(\frac{\left(\sum_{j=o+1}^{o+h} y_{i,j} - \sum_{j=o+1}^{o+h} \hat{y}_{i,j}\right)^2}{\sum_{m=2}^{n}(y_{i,m} - y_{i,m-1})^2}\right)}, \tag{9}$$

where p is the number of rolling origins (with index o), y_j is the observed value for period j of series i, and $\hat{y}_j$ its corresponding forecast. The denominator is a scaling factor used to make the errors scale independent and enable the calculation of summary metrics across time series, and corresponds to the 1-step ahead in-sample mean squared error of a random walk. The RMsSE is the quadratic and cumulative error counterpart to the MASE proposed by Hyndman and Koehler (2006). We use the quadratic to track the mean of the demand series and cumulative as we are interested in the over-the-lead time accuracy (Athanasopoulos & Kourentzes, 2023).

Accuracy is not necessarily strongly correlated with inventory outcomes (Goltsos et al., 2022; Guerrero et al., 2013; Kourentzes et al., 2020; Van der Laan et al., 2016). Therefore, we also track the scaled cumulative pinball loss (sPIN), which is constructed similarly to the RMsSE:

$$\text{sPIN}_{i,h} = \frac{1}{p}\sum_{o=1}^{p}\left(\frac{w_{i,o,h}}{\sum_{m=2}^{n}|y_{i,m} - y_{i,m-1}|}\right) \tag{10}$$

$$w_{i,o,h} = \begin{cases} \left(\sum_{j=o+1}^{o+h} y_j - \sum_{j=o+1}^{o+h} \hat{Q}_j \right) \alpha, & \text{if } \sum_{j=o+1}^{o+h} y_j \geq \sum_{j=o+1}^{o+h} \hat{Q}_j \\ \left(\sum_{j=o+1}^{o+h} \hat{Q}_j - \sum_{j=o+1}^{o+h} y_j \right) (1-\alpha), & \text{if } \sum_{j=o+1}^{o+h} y_j < \sum_{j=o+1}^{o+h} \hat{Q}_j \end{cases}, \tag{11}$$

where $\hat{Q}_i$ is the prediction for the α% quantile for period j. The scaling factor uses the absolute errors to match the loss order of $w_{i,o,h}$. The Pinball loss follows the newsvendor loss (Axsäter, 2015), and relates closely with inventory performance. For these experiments, we use an order-up-to policy with no lost sales, which makes the pinball loss a proper score, and therefore the best-performing method in terms of sPIN will also perform best in terms of average inventory costs.

5.2. Forecasting methods

From the exploration of the data, it becomes apparent that we need to consider methods that can deal with intermittent demand time series, either exclusively, or together with conventional demand series. We further need reliable benchmarks. We rely on a diverse selection of well-known time series models and methods: the (i) Naive (random walk); together with the (ii) arithmetic mean, for benchmarks; (iii) exponential smoothing and (iv) ARIMA (weekly data only, due to the reduced intermittency) families of models; and the (v) TSB method for intermittent demand forecasts. The TSB method is preferred over alternative standard intermittent demand methods, such as Croston's method and its variants, as the information that is used to produce the forecast is updated at every period, in contrast to only when demand occurs. This has been shown to result in better performance (Kourentzes, 2014; Teunter et al., 2011).

Temporal hierarchies are forecasting model-independent, and therefore one can substitute these with other readily available forecasting methods. As the various base forecasts are not critical, we provide additional implementation details and reasoning for our choices in the online supplement.

We implement two variants of THieF. For the first variant, we use ETS at all temporal aggregation levels. As the data properties change at each level the appropriate ETS model is automatically identified, using the Akaike Information Criterion corrected for sample size. The second variant mixes intermittent demand forecasts with conventional demand. If the intermittency (i.e., the percentage of zero-demand observations) is more than 40%, then TSB is used. Otherwise, ETS is used. This combination of intermittent and non-intermittent forecasts with THieF was shown to outperform intermittent demand forecasts (Kourentzes & Athanasopoulos, 2021) due to easier-to-model patterns that emerge in more aggregate less intermittent levels. The 40% threshold is used based on the recommendations by Kourentzes and Athanasopoulos (2021), who found limited sensitivity of THieF to the intermittency threshold. For both variants, the structural scaling outlined in Section 4 is used.

5.3. Estimation of quantiles

For inventory management, point (mean) forecasts are not adequate. We are interested in the demand over the range of potential lead times, and specifically in quantiles of demand. A challenge is that many forecasting approaches lack variance expressions, and more so expressions of variance as a function of forecast horizon. Saoud et al. (2022) showed that an additional challenge in obtaining the lead time variance is the estimation of the covariances between the predictive distribution of different forecast horizons. Analytical expressions for these are lacking even for well-studied families of models, such as the ARIMA. Instead, they propose to empirically estimate the variance based on the forecast errors of the cumulative demand over the lead time. This has the additional benefit that it is applicable irrespective of whether the forecasts are a product of a method or a model. Trapero et al. (2019) investigated the impact of non-normality and distributional asymmetries in the estimation of quantiles for inventory planning purposes. They find that using a kernel density estimate to obtain a smooth distribution of the empirical forecast errors is beneficial.

We combine both contributions to obtain quantiles for the various forecasts. First, we calculate the lead time demand errors as the difference between the total demand and the cumulative forecasts. The resulting collection of errors is smoothed using a kernel density estimate, assuming a Gaussian kernel and using the well-known Silverman's heuristic for obtaining the kernel bandwidth. Subsequently, the resulting density is sampled at the desired quantiles. The advantages of the approach are that it does not impose any distributional assumptions and accounts for all covariances over the forecast horizon, with minimal additional modelling complexity. The process is repeated for different lead times, to provide the relevant lead time errors.

6. Results

6.1. Forecast accuracy

Table 3 summarises the RMSsE results for the daily time series, for the Drug and Drug by Dispensary series. Each column corresponds to a different lead time, and the best performing method is highlighted in bold. As the RMSsE is a normalised error, to provide a feel for the importance in the reported differences, the range of errors (excluding the poorly performing Naive) is provided.

For the drug demand series THieF is consistently the best performing method, followed by THieF-I in all cases but the 1-day ahead forecast, where ETS is performing equally well to the best. Comparing ETS and TSB that is appropriate for intermittent demand series, on average ETS performs better, which explains the difference between THIeF and THieF-I. On average these series have 52.71% zero demand periods, with about 20% of the series having 5% or less intermittency. That makes the ETS still able to provide reasonably good forecasts. These also consistently outperform the remaining benchmarks. The Drug by Dispensary series are more intermittent, and although that results in a closer performance of TSB to ETS, the latter is still consistently better. If fact, in many cases it performs best overall, closely followed by THIeF-I.

We investigate the differences between the forecasts by testing whether the reported differences are statistically significant. We employ non-parametric tests to investigate whether different methods can be grouped together or whether there is sufficient evidence of differences. We rely on the combination of the non-parametric Friedman and post-hoc Nemenyi tests for this purpose. This choice has three advantages. First, as they are non-parametric there is no need for distributional assumptions of the errors, which is relevant given the intermittent nature of the data. This provides a complimentary view of the comparison, as only the ranking of the methods across series is retained and not the size of the errors. Second, since we are evaluating multiple methods we need to avoid doing repeated testing (i.e., multiple pairwise tests) as this will distort the corresponding outcome of the test, and third, as the statistic is based on ranks, we can use any metric

Table 3. RMsSE summary results (daily data).

Method	1 day	5 days	10 days	20 days
		Drug		
Naive	0.775	11.636	42.362	160.456
Mean	0.425	2.580	6.839	21.635
TSB	0.415	2.308	5.705	16.871
ETS	**0.391**	1.732	3.418	7.472
THieF-I	0.392	1.700	3.300	7.158
THieF	**0.391**	**1.695**	**3.269**	**6.882**
Range*	0.034	0.885	3.570	14.753
		Drug by dispensary		
Naive	0.574	8.567	30.762	116.766
Mean	0.306	1.758	4.310	12.786
TSB	0.304	1.686	4.027	11.577
ETS	**0.291**	**1.355**	**2.651**	6.136
THieF-I	**0.291**	1.363	2.707	**6.127**
THieF	0.298	1.437	2.826	6.275
Range*	0.016	0.404	1.659	6.659

*Range calculation excludes the Naïve.

and lead time. We use the implementation of the Friedman and Nemenyi tests available in the tsutils package (Kourentzes, 2023) for R.

Figure 8 visualises the results of the tests for the RMsSE for the different lead times. In each panel, the horizontal axis lists the methods as in Table 3. The vertical axis ranks the methods according to their mean rank which is provided next to the name of the method. The figure can be read by rows or columns, resulting in the same intuition. All methods with a coloured cell in a row/column belong to the same group. The cell coloured in black is matching the same method horizontally and vertically. For example, in the top-left panel, the second row means that THieF and THieF-I belong the same group, and that THieF ranks second according to the mean rank.

The Nemenyi test results provide a clear narrative. For one-day ahead forecasts ETS is best, while which variant of THieF is used does not result in different performance. For longer horizons, THieF and THieF-I begin to dominate. For 1-week ahead, all THieF, THieF-I, and ETS are grouped together, while for longer $t+10$ and $t+20$ (2 and 4 wk) THieF-I ranks first, significantly different from methods that do not rely on temporal hierarchies. This agrees with theory, where THieF is expected to perform well in long-term forecasts as it has access to more aggregate information.

On the weekly data, the results are similar, with THieF dominating in most cases, particularly for longer horizons. We find ARIMA to perform best for 1-step ahead forecasts for the Drug by Dispensary series, but otherwise does not change the remaining results significantly. For brevity, we provide a detailed discussion of the weekly results in the online supplement.

6.2. Quantile performance

Table 4 provides the results for the pinball loss. In our case, since there can be back-orders, this matches the inventory performance of the methods for the various lead times. The table is organised in a similar fashion to Table 3, providing the 90, 95, and 99% pinball losses for the Drug and Drug by Dispensary series.

We first draw attention to some overall observations. The range of values increases with the lead times and decreases with the target quantile. These results are consistent with our expectations, as longer lead times have to account for the more uncertain forecasts of longer horizons and the covariances between them (Saoud et al., 2022).

In contrast to the RMsSE results, the ranking of the competing methods is less consistent.

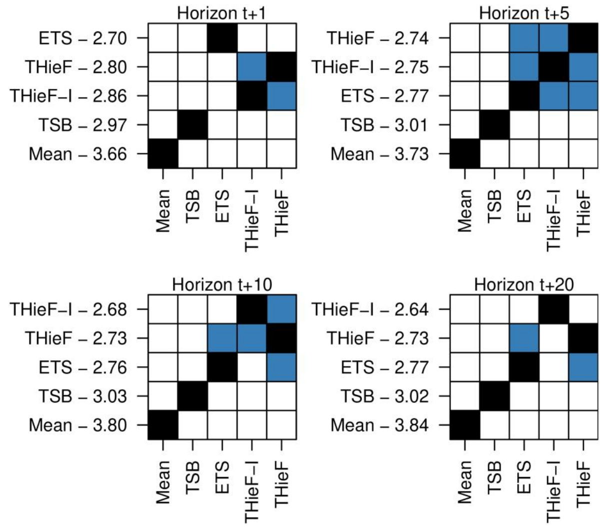

Figure 8. Plots of the Nemenyi test results at 5% significance for RMsSE across lead times.

Nonetheless, for the Drug series ETS performs well, often being the leading method, closely followed by the THieF variants. The triplet of ETS, THieF and THieF-I exhibit better performance compared to the Naive, Mean, and TSB. However, this difference becomes smaller as we look at the results for higher quantiles. For the Drug by Dispensary series THieF and ETS dominate, with the latter more for the lower quantiles and the former for the higher quantiles.

Figure 9 plots the pinball loss of the various forecasting methods, without the poorly performing Naive, between lead times $t+1$ and $t+20$, to provide an impression of how consistent the forecast accuracy remains across lead times. A method that is consistently performing well will be located at the bottom-left of the figure, while methods located either in the bottom-right or top-left perform very well only for a specific lead time. Therefore, this figure helps to see if a method dominates the rest across multiple lead times. To this end, we choose the shortest and longest lead times in the experiment. We only provide the results for 95% for brevity.

No method is dominant across all lead times. However, both for the Drug and Drug by Dispensary series the THieF is close to the bottom left corner, suggesting the most consistent good performance across lead times. For the 1-day ahead horizon most methods perform similarly, with the exception of ETS, which performs better on the 20-day lead time, ranking best in the case of the Drug series. Note that the difference in pinball is relatively small for $t+1$ and conversely large for $t+20$ when comparing results by Drug and Drug by Dispensary. This demonstrates that the inventory challenge is smaller when decisions are required by Drug.

Table 4. Pinball loss summary results (daily data).

	Drug				Drug by dispensary			
Method	1 day	5 days	10 days	20 days	1 day	5 days	10 days	20 days
				Pinball 90%				
Naive	0.279	0.845	1.394	2.540	0.377	1.069	1.704	2.753
Mean	**0.227**	0.570	0.788	1.351	**0.231**	0.775	1.199	1.715
TSB	0.228	0.579	0.815	1.398	0.236	0.770	1.173	1.757
ETS	0.234	**0.549**	**0.767**	**1.267**	0.255	0.764	**1.115**	**1.575**
THieF-I	0.228	0.563	0.785	1.321	0.232	0.770	1.183	1.664
THieF	**0.227**	0.555	0.768	1.279	0.237	**0.737**	1.136	1.582
Range*	0.008	0.030	0.048	0.131	0.024	0.038	0.085	0.182
				Pinball 95%				
Naive	0.190	0.498	0.883	1.536	0.274	0.723	1.053	1.690
Mean	**0.169**	0.348	0.545	0.893	**0.198**	0.543	0.741	1.295
TSB	0.170	0.356	0.555	0.918	0.199	0.547	0.754	1.303
ETS	0.172	**0.346**	**0.527**	**0.844**	0.211	0.547	0.710	1.201
THieF-I	0.170	0.349	0.543	0.877	**0.198**	0.543	0.726	1.266
THieF	0.170	0.347	0.530	0.850	0.201	**0.530**	**0.705**	**1.162**
Range*	0.003	0.010	0.027	0.075	0.013	0.017	0.049	0.141
				Pinball 99%				
Naive	0.063	0.169	0.240	0.384	0.117	0.258	0.339	0.490
Mean	**0.060**	0.132	0.159	0.233	0.102	0.227	0.275	0.363
TSB	**0.060**	0.132	0.161	0.238	0.102	0.226	0.277	0.367
ETS	**0.060**	**0.128**	**0.154**	0.223	0.w104	0.226	0.262	0.346
THieF-I	**0.060**	0.129	0.156	0.222	0.102	0.228	0.271	0.357
THieF	**0.060**	0.129	0.155	**0.220**	**0.101**	**0.214**	**0.258**	**0.344**
Range*	0.001	0.005	0.007	0.018	0.002	0.014	0.019	0.023

*Range calculation excludes the Naïve.

Figure 10 provides the results of the Nemenyi statistical test for different lead times and target quantiles. In many cases, THieF is the top-ranking method, and when it is not, it is part of the first group of methods, for which the performance is statistically indistinguishable. THieF-I typically ranks in the middle, in some cases being part of the first or the second group of methods. ETS has a similar performance. TSB and Mean tend to rank lower, though, again, the ranking is inconsistent across cases.

From the analysis of the pinball loss results for the daily time series we conclude that THieF has consistently good performance. THieF-I does not perform equally well, although it consistently outperforms its intermittent demand forecasting counterpart, the TSB method. ETS exhibits strong performance in cases, but inconsistently. We argue that the consistency of the performance of THieF is its major advantage. This observation is in agreement with the accuracy results.

The pinball results on the weekly data are largely in agreement with the daily ones. THieF performs overall well and fairly consistently. Although other methods perform competitively, they lack performance consistency, in agreement with the accuracy results investigation. We provide detailed results in the online supplement.

7. Discussion

The contributions of this work have been to evaluate and address three key challenges in forecasting

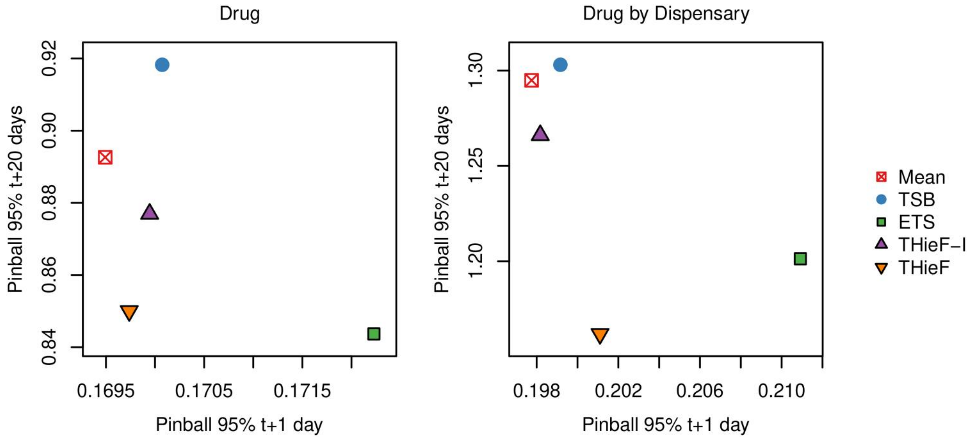

Figure 9. Scaled Pinball metric for the 95% quantile, for Drug and Drug by Dispensary (daily data).

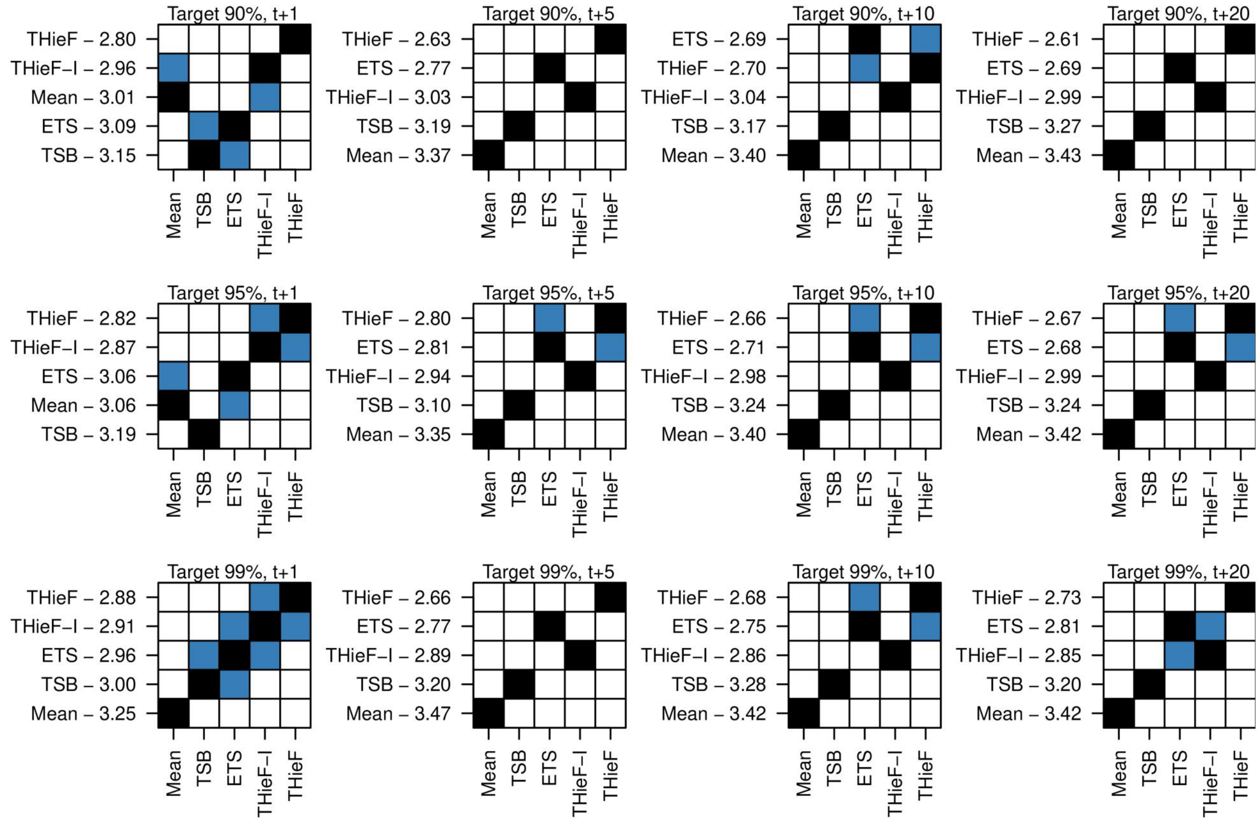

Figure 10. Nemenyi test results at 5% significance for pinball loss across lead times (daily series).

drug demand for hospital pharmacies. That demand forecasts (1) be aligned across various planning horizons to support inventory management, (2) accommodate volatile drug delivery lead times and (3) be robust against erratic demand patterns including varying levels of intermittency. In this section we draw on the idea of forecast trustworthiness to reflect on the work done across these areas. Spavound and Kourentzes (2022) identify four dimensions of trustworthy forecasting: (i) reliability; (ii) stability; (iii) intelligibility; (iv) alignment. In the section on forecast coherency and inventory planning we draw on the notion of forecast alignment and intelligibility to discuss how the proposed approach enhances planning across horizons while still being understandable for decision makers. We then discuss the need for automation and trust and how the proposed approach improves on the reliability and stability of forecasts in the face of volatile and intermittent data.

7.1. Forecast coherency and inventory planning

It is required that forecasts be aligned across various planning horizons to support inventory decisions at different levels of the supply network. In this work, we discussed two sides of the inventory decisions, orders to suppliers, and supply to wards. Although these problems can be considered independently, a misalignment between the two can increase supply chain friction. For example, the inventory for the wards may be optimally planned, but if there is insufficient stock ordered from the manufacturers, the supply to the wards will fail too. Spavound and Kourentzes (2022) refers to this as alignment, that is, the design of forecasts to match the objectives of supported decisions, in our case, to approximate well the drugs' demand over the lead time. THieF achieves this by directly modelling the cumulative demand in its temporally aggregate levels. Moreover, with THieF, analysts can intervene to include additional contextual information both at the day-by-day or the total lead time demand as it becomes available. THieF reconciles any such additional information across planning horizons and decisions.

This also brings operational benefits. Planning for the use of drugs at wards, or for ordering from suppliers, may be done by different experts, who may or may not be in communication, or using the same or different contextual information. The advantage of THieF is that it assumes that forecasts at different planning levels are built independently. Organisationally, this suggests that current practice at the hospital can continue, with some analysts performing the final reconciliation step. The reconciled forecasts will contain the blended information from all component forecasts, on which the decisions will be based. Therefore, on the one hand, minimal organisational and analytical changes are sufficient to implement THieF in practice. On the other hand, it brings managerial opportunities to the organisation, as it provides an "analytics way" to increase the information flow between analysts and teams.

The current focus at the case hospital pharmacy is daily forecasts of the Drug by Dispensary demand. We show that this can be questioned, with forecasting at a Drug level leading to better performance. However, such forecasts can only support ordering decisions upstream, Likewise, we find evidence that modelling at a weekly level can further simplify the forecasting challenge. However, this may be damaging to the ability of analysts to react quickly with ad-hoc orders when necessary. Given the overlap of information in the daily and weekly temporal hierarchies, one could operate on the weekly data and supervise using the daily data. In this scenario, the forecast is still relatively simple for decision makers and analyst to understand, given that details of the algorithm are not important, but rather how different pieces of information are incorporated into the forecast, through simple aggregation mechanism. This is referred to as the concept of intelligibility (Spavound & Kourentzes, 2022).

7.2. Forecast automation and trustworthiness

A further challenge presented when forecasting drug demand for hospital pharmacies are the potential erratic demand patterns with a high level of demand variation across the hospital demand/supply network. We observed for the case hospital pharmacy high levels of intermittency and volatility at both Drug and Drug by Dispensary level (see Figure 5). Additionally, pharmacists spend too much time placing emergency orders and managing inventory, suggesting the need for varying degrees of automation. This requires reliability and stability in the production of forecasts (Spavound & Kourentzes, 2022).

By reliability we are looking for forecasts that fail gracefully, as large and unexpected forecast errors can harm the trust of users (Dietvorst et al., 2015, 2018). THieF forecasts consistently performed well. They did not rank first in all cases, but they were always amongst the top contenders, across different inventory settings and metrics. Competing forecasts that ranked first in some cases demonstrate rather erratic ranking in other cases. This outcome follows from the theory of temporal hierarchies, where the use of multiple temporal aggregation levels helps to identify more structures from the data, while at the same time mitigating the modelling risk of misspecifying forecasting models. This is highlighted in the reported Nemenyi results, which focus on the ranking of individual forecasts.

By stability, Spavound & Kourentzes (2022) refers to the property of forecasts to not vary wildly over time. The notion of stability relates to the changes in the forecasts from one forecast origin to the next, or from a review period to the next in the inventory context. Temporal hierarchies utilise more temporally aggregate views of the data, resulting in less volatile forecasts as the aggregate data are smoothed. Forecasts at the more disaggregate levels may differ substantially as more data become available, but the more aggregate ones remain similar. For instance, on a daily level the data updates every day, changing the forecast origin. On the annual view of the data, the origin changes only by 1/260th (5-day weeks) and reasonably the forecasts remain very similar. Temporal hierarchy forecasts have increased stability by construction, irrespective of the forecasting model deployed. We evidenced that THieF forecasts perform consistently well across various forecast horizons. This contributes to simplifying the forecast generation when the hospital pharmacy is faced with stochastic lead times, as there is an expectation that THieF's performance will not differ substantially for the range of relevant lead times.

In the presence of volatile, intermittent data, the proposed forecasting setup greatly simplifies the forecast selection and the forecasting process supporting automation. Through the temporal aggregation mechanism, THieF produces useful forecasts for intermittent demand drugs, to the extent that there is no need to resort to specialised forecasting methods, like the TSB, or THieF-I. This results in a less cumbersome implementation. Likewise, there is no need to follow a separate forecasting process for continuous and intermittent demand drugs.

8. Conclusions

Motivated by the challenges faced in hospital pharmacy drug demand forecasting we propose the use of temporal hierarchies. We find that these are able to mitigate some of the more challenging aspects, such as intermittent demand, volatile delivery lead times, and relatively long forecast horizons. We show that they result in high-performing forecasts in terms of accuracy and inventory performance, the latter as measured with the pinball loss that is a proper score for simple inventory policies. Future research should investigate more thoroughly the inventory benefits of THieF, particularly under stochastic lead times. We also show that the consistent behaviour of THieF has the potential to simplify the forecasting process and any future implementation in practice.

In this work, motivated largely by the intelligibility of forecasts, we looked at established methods for the generation of base forecasts, all of which have been extensively researched when it comes to automatic specification and estimation. The use of THieF does not preclude the use of more recent advances in forecasting methods, for instance relying on machine learning and artificial intelligence. We argue that the data characteristics of our dataset, particularly the very high intermittency, are not conducive to a straightforward and automatic implementation of these methods. Notwithstanding, with more data, or for a different hospital pharmacy they may be more apt. THieF can take advantage of these, if they are found to perform better at a specific temporal aggregation level or across. This connects with one avenue for future research that relates to model selection across a temporal hierarchy. Although there are multiple methodologies to choose a forecasting model for a time series, they are inherently constrained by the model families considered in the model pool by the analyst. We argue that THieF brings a different aspect to this question, as the time series change across the temporal hierarchy in a de facto way. Silvestrini and Veredas (2008) provide how a specific ARIMA model changes with temporal aggregation analytically. Although this treatment is very challenging in the general case, it may be that temporal aggregation can inform model choice across levels of the temporal hierarchy, instead of considering these fully independently. We are unaware of any research in this direction, but we recognise that this could provide computational and accuracy benefits.

Disclosure statement

There are no conflicts of interest to declare.

Funding

This work was supported by The Health Foundation's Advancing Applied Analytics programme under Grant number 1910259.

ORCID

Devon Barrow http://orcid.org/0000-0001-8919-1701
Nikolaos Kourentzes http://orcid.org/0000-0003-0211-5218
Martin Utley http://orcid.org/0000-0001-9928-1516

References

Abolghasemi, M., Beh, E., Tarr, G., & Gerlach, R. (2020). Demand forecasting in supply chain: The impact of demand volatility in the presence of promotion. *Computers & Industrial Engineering*, *142*, 106380. https://doi.org/10.1016/j.cie.2020.106380

Anusha, S. L., Alok, S., & Shaik, A. (2014). Demand forecasting for the indian pharmaceutical retail: A case study. *Journal of Supply Chain Management Systems*, *3*(2), 1–8.

Athanasopoulos, G., Hyndman, R. J., Kourentzes, N., & Panagiotelis, A. (2024). Forecast reconciliation: A review. *International Journal of Forecasting*, *40*(2), 430–456. https://doi.org/10.1016/j.ijforecast.2023.10.010

Athanasopoulos, G., Hyndman, R. J., Kourentzes, N., & Petropoulos, F. (2017). Forecasting with temporal hierarchies. *European Journal of Operational Research*, *262*(1), 60–74. https://doi.org/10.1016/j.ejor.2017.02.046

Athanasopoulos, G., & Kourentzes, N. (2023). On the evaluation of hierarchical forecasts. *International Journal of Forecasting*, *39*(4), 1502–1511. https://doi.org/10.1016/j.ijforecast.2022.08.003

Axsäter, S. (2015). *Inventory control*. Vol. 225 Springer.

Beier, F. J. (1995). The management of the supply chain for hospital pharmacies: A focus on inventory management practices. *Journal of Business Logistics*, *16*(2), 153.

Biagi, M., Carnevali, L., Santoni, F., Vicario, E. (2018). Hospital inventory management through markov decision processes@ runtime. In: *Quantitative Evaluation of Systems: 15th International Conference, QEST 2018*, Beijing, China, September 4-7, 2018, (pp. 87–103). Springer.

Boylan, J. E., & Babai, M. Z. (2016). On the performance of overlapping and non-overlapping temporal demand aggregation approaches. *International Journal of Production Economics*, *181*, 136–144. https://doi.org/10.1016/j.ijpe.2016.04.003

Chen, D. Q., Preston, D. S., & Xia, W. (2013). Enhancing hospital supply chain performance: A relational view and empirical test. *Journal of Operations Management*, *31*(6), 391–408. https://doi.org/10.1016/j.jom.2013.07.012

Cook, A. G. (2016). *Forecasting for the pharmaceutical industry: Models for new product and in-market forecasting and how to use them*. CRC Press.

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology. General*, *144*(1), 114–126. https://doi.org/10.1037/xge0000033

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2018). Overcoming algorithm aversion: People will use imperfect algorithms if they can (even slightly) modify them. *Management Science*, *64*(3), 1155–1170. https://doi.org/10.1287/mnsc.2016.2643

Fruggiero, F., Iannone, R., Martino, G., Miranda, S., Riemma, S. (2012). A forecast model for pharmaceutical requirements based on an artificial neural network. In *Proceedings of 2012 IEEE International Conference on Service Operations and Logistics, and Informatics*. (pp. 263–268) IEEE.

George, S., & Elrashid, S. (2023). Inventory management and pharmaceutical supply chain performance of hospital pharmacies in bahrain: A structural equation modeling approach. *Sage Open*, *13*(1), 21582440221149717. https://doi.org/10.1177/21582440221149717

Goltsos, T. E., Syntetos, A. A., Glock, C. H., & Ioannou, G. (2022). Inventory–forecasting: Mind the gap. *European Journal of Operational Research*, *299*(2), 397–419. https://doi.org/10.1016/j.ejor.2021.07.040

Goodwin, P. (2018). *Profit from your forecasting software: A best practice guide for sales forecasters*. John Wiley & Sons.

Guerrero, W. J., Yeung, T. G., & Guéret, C. (2013). Joint-optimization of inventory policies on a multi-product multi-echelon pharmaceutical system with batching and ordering constraints. *European Journal of Operational Research*, *231*(1), 98–108. https://doi.org/10.1016/j.ejor.2013.05.030

Hussein, B. R., Kasem, A., Omar, S., Siau, N. Z. (2019). A data mining approach for inventory forecasting: a case study of a medical store. In Computational Intelligence in Information Systems: Proceedings of the Computational Intelligence in Information Systems Conference (CIIS 2018). (pp. 178–188). Springer. https://doi.org/10.1007/978-3-030-03302-6_16

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

Khalil Zadeh, N., Sepehri, M. M., & Farvaresh, H. (2014). Intelligent sales prediction for pharmaceutical distribution companies: A data mining based approach. *Mathematical Problems in Engineering*, *2014*(1). https://doi.org/10.1155/2014/420310

Kim, W., Won, J. H., Park, S., & Kang, J. (2015). Demand forecasting models for medicines through wireless sensor networks data and topic trend analysis. *International Journal of Distributed Sensor Networks*, *11*(9), 907169. https://doi.org/10.1155/2015/907169

Kline, D. M. (2004). Methods for multi-step time series forecasting neural networks. In *Neural networks in business forecasting*. (pp. 226–250). IGI Global Scientific Publishing.

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. (2023). tsutils: Time Series Exploration, Modelling and Forecasting. R package version 0.9.4. https://github.com/trnnick/tsutils/

Kourentzes, N., & Athanasopoulos, G. (2019). Cross-temporal coherent forecasts for Australian tourism. *Annals of Tourism Research*, *75*, 393–409. https://doi.org/10.1016/j.annals.2019.02.001

Kourentzes, N., & Athanasopoulos, G. (2021). Elucidate structure in intermittent demand series. *European Journal of Operational Research*, *288*(1), 141–152. https://doi.org/10.1016/j.ejor.2020.05.046

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

Kourentzes, N., Trapero, J. R., & Barrow, D. K. (2020). Optimising forecasting models for inventory planning. *International Journal of Production Economics*, *225*, 107597. https://doi.org/10.1016/j.ijpe.2019.107597

Lega, F., Prenestini, A., & Spurgeon, P. (2013). Is management essential to improving the performance and sustainability of health care systems and organizations? a systematic review and a roadmap for future studies. *Value in Health: The Journal of the International Society for Pharmacoeconomics and Outcomes Research*, *16*(1 Suppl), S46–S51. https://doi.org/10.1016/j.jval.2012.10.004

Litvak, E., & Long, M. C. (2000). Cost and quality under managed care: Irreconcilable differences. *The American Journal of Managed Care*, *6*(3), 305–312.

Makridakis, S., & Hibon, M. (2000). The m3-competition: Results, conclusions and implications. *International Journal of Forecasting*, *16*(4), 451–476. https://doi.org/10.1016/S0169-2070(00)00057-1

Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2022). M5 accuracy competition: Results, findings, and conclusions. *International Journal of Forecasting*, *38*(4), 1346–1364. https://doi.org/10.1016/j.ijforecast.2021.11.013

Makridakis, S., Wheelwright, S. C., & Hyndman, R. J. (2008). *Forecasting methods and applications*. John wiley & sons.

Merkuryeva, G., Valberga, A., & Smirnov, A. (2019). Demand forecasting in pharmaceutical supply chains: A case study. *Procedia Computer Science*, *149*, 3–10. https://doi.org/10.1016/j.procs.2019.01.100

Mousa, B. A., & Al-Khateeb, B. (2023). Predicting medicine demand using deep learning techniques: A review. *Journal of Intelligent Systems*, *32*(1), 20220297. https://doi.org/10.1515/jisys-2022-0297

Nguyen, A., Lamouri, S., Pellerin, R., Tamayo, S., & Lekens, B. (2022). Data analytics in pharmaceutical supply chains: State of the art, opportunities, and challenges. *International Journal of Production Research*, *60*(22), 6888–6907. https://doi.org/10.1080/00207543.2021.1950937

Nikolopoulos, K., Buxton, S., Khammash, M., & Stern, P. (2016). Forecasting branded and generic pharmaceuticals. *International Journal of Forecasting*, *32*(2), 344–357. https://doi.org/10.1016/j.ijforecast.2015.08.001

Pall, R., Gauthier, Y., Auer, S., & Mowaswes, W. (2023). Predicting drug shortages using pharmacy data and machine learning. *Health Care Management Science*, *26*(3), 395–411. https://doi.org/10.1007/s10729-022-09627-y

Panagiotelis, A., Athanasopoulos, G., Gamakumara, P., & Hyndman, R. J. (2021). Forecast reconciliation: A geometric view with new insights on bias correction. *International Journal of Forecasting*, *37*(1), 343–359. https://doi.org/10.1016/j.ijforecast.2020.06.004

Papanagnou, C. I., & Matthews-Amune, O. (2018). Coping with demand volatility in retail pharmacies with the aid of big data exploration. *Computers & Operations Research*, *98*, 343–354. https://doi.org/10.1016/j.cor.2017.08.009

Pritularga, K. F., Svetunkov, I., & Kourentzes, N. (2021). Stochastic coherency in forecast reconciliation. *International Journal of Production Economics*, *240*, 108221. https://doi.org/10.1016/j.ijpe.2021.108221

Ramirez, A. L., Jurado, I., Garcia, M. F., Tejera, B. I., Llergo, J. D. P., Torreblanca, J. M. (2014). Optimization of the demand estimation in hospital pharmacy. In *Proceedings of the 2014 IEEE Emerging Technology and Factory Automation (ETFA)*. (pp. 1–6). IEEE. https://doi.org/10.1109/ETFA.2014.7005057

Restyana, A., Savitri, L., Laili, N. F., & Probosiwi, N. (2021). Analysis of drug forecasting with single moving average and single exponential smoothing approach (case study in jombang regency 2017-2019). *Journal of Physics: Conference Series*, *1899*, 012100. https://doi.org/10.1088/1742-6596/1899/1/012100

Ribeiro, A., Seruca, I., & Durão, N. (2017). Improving organizational decision support: Detection of outliers and sales prediction for a pharmaceutical distribution company. *Procedia Computer Science*, *121*, 282–290. https://doi.org/10.1016/j.procs.2017.11.039

Sagaert, Y. R., Kourentzes, N., De Vuyst, S., Aghezzaf, E.-H., & Desmet, B. (2019). Incorporating macroeconomic leading indicators in tactical capacity planning. *International Journal of Production Economics*, *209*, 12–19. https://doi.org/10.1016/j.ijpe.2018.06.016

Saha, E., & Ray, P. K. (2019). Modelling and analysis of inventory management systems in healthcare: A review and reflections. *Computers & Industrial Engineering*, *137*, 106051. https://doi.org/10.1016/j.cie.2019.106051

Saoud, P., Kourentzes, N., & Boylan, J. E. (2022). Approximations for the lead time variance: A forecasting and inventory evaluation. *Omega*, *110*, 102614. https://doi.org/10.1016/j.omega.2022.102614

Siddiqui, R., Azmat, M., Ahmed, S., & Kummer, S. (2022). A hybrid demand forecasting model for greater forecasting accuracy: The case of the pharmaceutical industry. In: *Supply Chain Forum: An International Journal*. (Vol. 23, pp.124–134). Taylor & Francis. https://doi.org/10.1080/16258312.2021.1967081

Silva-Aravena, F., Ceballos-Fuentealba, I., & Álvarez-Miranda, E. (2020). Inventory management at a chilean hospital pharmacy: Case study of a dynamic decision-aid tool. *Mathematics*, *8*(11), 1962. https://doi.org/10.3390/math8111962

Silvestrini, A., & Veredas, D. (2008). Temporal aggregation of univariate and multivariate time series models: A survey. *Journal of Economic Surveys*, *22*(3), 458–497. https://doi.org/10.1111/j.1467-6419.2007.00538.x

Sohrabi, B., Raeesi Vanani, I., Nikaein, N., & Kakavand, S. (2021). A predictive analytics of physicians prescription and pharmacies sales correlation using data mining. *International Journal of Pharmaceutical and Healthcare Marketing*, *13*(3), 346–363. https://doi.org/10.1108/IJPHM-11-2017-0066

Sousa, R. M., Hannachi, S., & Ramos, G. N. (2019). Statistical and deep learning models for forecasting drug distribution in the brazilian public health system. In: *2019 8th Brazilian Conference on Intelligent Systems (BRACIS)*. (pp. 723–728). IEEE. https://doi.org/10.1109/BRACIS.2019.00130

Spavound, S., & Kourentzes, N. (2022). Making forecasts more trustworthy. *Foresight: The International Journal of Applied Forecasting*, (66), 21–25.

Taieb, S. B., Bontempi, G., Atiya, A. F., & Sorjamaa, A. (2012). A review and comparison of strategies for multi-step ahead time series forecasting based on the nn5 forecasting competition. *Expert Systems with Applications*, *39*(8), 7067–7083.

Teunter, R. H., Syntetos, A. A., & Babai, M. Z. (2011). Intermittent demand: Linking forecasting to inventory obsolescence. *European Journal of Operational Research*, *214*(3), 606–615. https://doi.org/10.1016/j.ejor.2011.05.018

Tiao, G. C., & Tsay, R. S. (1994). Some advances in non-linear and adaptive modelling in time-series. *Journal of Forecasting*, *13*(2), 109–131. https://doi.org/10.1002/for.3980130206

Trapero, J. R., Cardós, M., & Kourentzes, N., (2019). Empirical safety stock estimation based on kernel and garch models. *Omega*, *84*, 199–211. https://doi.org/10.1016/j.omega.2018.05.004

Van Belle, J., Crevits, R., Guns, T., & Verbeke, W. (2025). To change or not to change the lead time in a dynamic order-up-to inventory system? *Journal of the Operational Research Society*, *76*(9), 1765–1776. https://doi.org/10.1080/01605682.2024.2444409

Van Belle, J., Guns, T., & Verbeke, W. (2021). Using shared sell-through data to forecast wholesaler demand in multi-echelon supply chains. *European Journal of Operational Research*, *288*(2), 466–479. https://doi.org/10.1016/j.ejor.2020.05.059

Van der Laan, E., van Dalen, J., Rohrmoser, M., & Simpson, R. (2016). Demand forecasting and order planning for humanitarian logistics: An empirical assessment. *Journal of Operations Management*, *45*(1), 114–122. https://doi.org/10.1016/j.jom.2016.05.004

Varghese, V., Rossetti, M., Pohl, E., Apras, S., & Marek, D. (2012). Applying actual usage inventory management best practice in a health care supply chain. *International Journal of Supply Chain Management*, *1*(2), 1–10.

Vila-Parrish, A. R., Ivy, J. S., King, R. E., & Abel, S. R. (2012). Patient-based pharmaceutical inventory management: A two-stage inventory and production model for perishable products with markovian demand. *Health Systems*, *1*(1), 69–83. https://doi.org/10.1057/hs.2012.2

Wettermark, B., Persson, M. E., Wilking, N., Kalin, M., Korkmaz, S., Hjemdahl, P., Godman, B., Petzold, M., & Gustafsson, L. L. (2010). Forecasting drug utilization and expenditure in a metropolitan health region. *BMC Health Services Research*, *10*(1), 1–14.

Wickramasuriya, S. L., Athanasopoulos, G., & Hyndman, R. J. (2019). Optimal forecast reconciliation for hierarchical and grouped time series through trace minimization. *Journal of the American Statistical Association*, *114*(526), 804–819. https://doi.org/10.1080/01621459.2018.1448825

Zhu, X., Ninh, A., Zhao, H., & Liu, Z. (2021). Demand forecasting with supply-chain information and machine learning: Evidence in the pharmaceutical industry. *Production and Operations Management*, *30*(9), 3231–3252. https://doi.org/10.1111/poms.13426