CIGI QUALITA MOSIM 2023
On comparing forecasting methods for postal delivery service

NADINE KAFA, M. ZIED BABAI , WALID KLIBI

The Centre of Excellence in Supply Chain (CESIT), KEDGE BUSINESS SCHOOL
Bordeaux, France
nadine.kafa@kedgebs.com, mohamed-zied.babai@kedgebs.com, walid.klibi@kedgebs.com

***Résumé* –Ce travail fait partie d'un projet de recherche lié au domaine de la prévision pour le bien social (FSG) visant à améliorer les prévisions de la demande de courrier pour une grande organisation postale française. L'objectif de l'article est de proposer une nouvelle approche d'aide à la décision basée sur les prévisions des flux de courrier à différents niveaux de l'organisation. L'approche proposée étudie la prévision hiérarchique de séries temporelles en utilisant les approches Top-down (TD), Bottom-up (BU) et la combinaison optimale (OC). Elle intègre une approche en quatre phases comprenant la préparation des données, le schéma hiérarchique, les méthodes de prévision et les mesures d'évaluation. Elle évalue différentes méthodes de prévision communes comme ARIMA and ETS selon des indicateurs de précision et des indicateurs sociaux. Cette approche montre des prévisions de la demande plus précises et performant par rapport à la méthode actuelle utilisée par l'organisation.**

***Abstract* – This work is part of a research project related to the field of Forecasting for Social Good (FSG) to improve mail demand forecasts for a major French postal organization. The aim of the paper is to propose a new forecast-based decision-making approach that predicts the mail demand flow at different levels of the organization to better plan the delivery activities. It includes a four-phase approach including data preparation, a hierarchy scheme, forecasting methods, and evaluation measures. The proposed approach investigates hierarchical time series forecasting using the Top-down (TD), Bottom-up (BU), and Optimal combination (OC) approaches. It also evaluates different common forecasting methods like the autoregressive integrated moving average (ARIMA) method and the exponential smoothing with Error, Trend, and Seasonality (ETS) method based on accuracy and social measures. The results show more accurate mail demand forecasts using an efficient hierarchical approach compared to the current method used by the organization.**

***Mots clés* - prévision hiérarchique des séries temporelles, demande de courrier, service de livraison par voie postale.**
***Keywords* – hierarchical time series forecasting, mail demand flow, postal delivery service.**

1 INTRODUCTION AND BACKGROUND

Forecasting for Social Good (FSG) is a forecasting process that aims to support decisions that prioritize the thriving of humanity over the thriving of economies by improving the social foundation and ecological ceilings that impact the public as a whole on both local and global scales (Rostami-Tabar et al., 2022). Based on this definition, research has to satisfy four main attributes to be considered as an FSG study. First, the forecasting process is related to a real business context. Second, the studied problem is driven by humanity thriving rather than economies thriving. Third, the proposed solution improves the social foundation and ecological ceiling. Finally, the study impacts the public as a whole. These attributes represent the essential motivation for this work as this study deals with a real problem to forecast mail demand flows as a social good in order to ensure mail distribution and to support decisions related to planning delivery activities.
However, the FSG research area is new and few studies have been conducted in this area with limited applications (Rostami-Tabar et al., 2022). (Gross & Sohl, 1990) studied forecasting crime for policy and planning decisions to support the tactical deployment of police resources. (van der Laan et al., 2016) provided an empirical analysis of the demand planning and distribution operations at the Operational Center Amsterdam of Médecins sans frontières (MSF-OCA) supported by forecasting methods. (Litsiou et al., 2022) applied judgmental methods for forecasting the success of megaprojects. (Wicke et al., 2022) used scenario generation to forecast the outcomes of a refugee crisis. (Rostami-Tabar & Ziel, 2022) proposed a forecasting model to generate both point and probabilistic daily forecasts of Emergency Department (ED) attendance with the aim of allocating available resources more effectively.
Furthermore, most studies in the postal sector discuss the network design problem or vehicle routing problem (Ducret, 2014; Janjevic & Winkenbach, 2020) rather than the mail demand flow forecasting. For instance, (Chlosta & Froberg, 1977) is a basic reference that proposes a multivariate linear regression model for mail forecasts to support production planning. (Munkhdalai et al., 2020) improve the mail forecast accuracy for Korea Post using a deep learning-based time series forecasting model.
On the other hand, many forecasting problems could be organized in a hierarchical structure in accordance with the hierarchal setting of the decisional context. In the literature, hierarchical time series forecasting approaches are proposed to deal with such setting (Babai et al., 2022). A hierarchical forecasting approach is proposed by (Villegas & Pedregal, 2018) to efficiently tackle the case of Spanish grocery retail.
(Punia et al., 2020) propose an empirical study to investigate the suitability of hierarchical forecasting approaches in the context

of online and offline retail. This study shows that the forecasts provided by the hierarchical approaches are more accurate than the direct forecasts for all levels of decision-making for the retailer. (Abolghasemi et al., 2022) apply the three principal approaches in hierarchical forecasting Top-down (TD), Bottom-up (BU), and optimal linear combination (OC) using machine learning methods to investigate the impact of promotions using FMCG products data from a food manufacturing company in Australia. There were debates in the literature regarding the best performant hierarchical approach (Punia et al., 2020). However, the optimal linear combination (OC) outperforms or at least gives equal accuracy as the TD and BU forecasting approaches (Hyndman et al., 2011). In addition to forecast accuracy, (Babai et al., 2022) recommend proposing new measures related to utility functions in order to evaluate the different forecasting approaches. To the best of our knowledge, no existing studies have offered a hierarchical forecasting approach for the mail demand flow to support delivery activities applied to a real case study with a social-oriented context.

With respect to the case under study, this work aims to improve the mail demand forecasts, which impact the planning and delivery service for a major French postal organization. The organization network has a hierarchy with two levels composed of 19 platforms (PFs) and 87 distribution centers (DCs), deployed over the territory as shown in (Figure 1). Each Platform PF distributes the mail flow to one or more distribution centers DCs, which collect, sort, and distribute the mail. The majority of these activities are accomplished by temporary employees. Thus, after forecasting the month's demand, the labor workload for each month is determined for each platform (P) and distribution center (DC) to ensure high delivery service. Inaccurate mail flow forecasts lead to either over- or under-workload that generates extra costs or poor service levels, respectively.

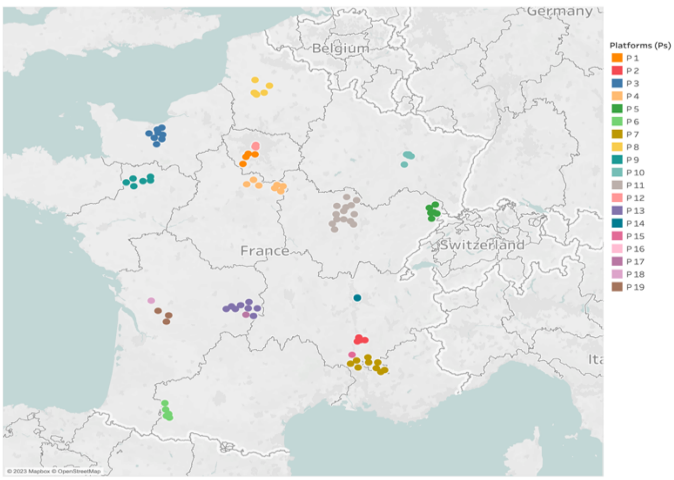

Figure 1. Mapping of 87 DCs and 19 PFs

Moreover, the current forecasting process used in the organization is time-consuming. Therefore, the decision-makers aim to get more accurate mail forecasts in order to better plan the distribution activities and to decrease the time calculation and effort. Most distribution activities are done by temporary employees who are planned based on the forecasts. In the case, the forecasts consistently overestimate the real volumes, the organization release all or part of the planned temporary employees. The organization aims to obtain useful forecasts that could be integrated into a new social key performance indicator that reflects the variability of the release level. This issue constitutes one of the objectives of this research.

Thus, the primary contribution of this study is to compare direct and hierarchical time series approaches for forecasting the mail demand flow in the postal delivery service. The comparison of both approaches is based not only on accuracy measures but also on the social impact.

The remainder of this paper is structured as follows. Section 2 describes the proposed approach including the data preparation, hierarchy scheme, forecasting methods, and accuracy measures as sub-sections. The implementation, findings, and discussion are drawn in Section 3. Finally, concluding remarks and recommendations for future research are presented in Section 4.

2 FORECASTING APPROACH

In this section, we present the proposed approach from data preparation to forecasting as illustrated in Figure 2. Initially, the data preparation and cleaning process is first stated to obtain a complete and proper database for the monthly mail demand flow. Then, hierarchical forecasting is proposed using the Top-down (TD), Bottom-up (BU), and Optimal combination (OC) approaches. See (Hyndman & Athanasopoulos, 2018) for an introduction to hierarchical forecasting and (Babai et al., 2022) for a review of aggregation and hierarchical approaches in the supply chain. For each approach, the baseline forecasting method currently used by the organization, as well as the proposed forecasting methods which are the autoregressive integrated moving average (ARIMA) method, and the exponential smoothing with Error, Trend, and Seasonality (ETS) method are presented. These methods are compared and evaluated based on accuracy measures and a new social indicator, namely, the Mean Absolute Percentage Error (MAPE) and the Root Mean Squared Scaled Error (RMSSE) commonly used in the literature, and a new social indicator reflects a dismissing level DL. This indicator represents the probability that the organization planned more employees than it needed as the forecast load is higher than the real volume load.

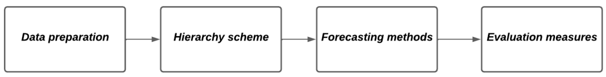

Figure 2. Approach

2.1 Data preparation

First of all, empirical data is required to analyze the performance of the current forecasting method used in the postal organization and the proposed ones. Therefore, we started by preparing, cleaning, and defining data with the corresponding historical period. As expected, even if the data were existing in the different databases of the organization, the data cleaning phase was a complicated task for many reasons. First, there was no integrated database including the relevant monthly mail flow at the platforms "PFs" and the distribution centers "DCs". Besides, each level of the organization had its own information system and databases which are differently structured. Thus, we combined several databases and corrected the historical data to construct a complete database for the mail volume during the period.

2.2 Hierarchy scheme

The structure of the postal organization can be presented as a hierarchy diagram as it is grouped into platforms having their distribution centers (Figure 3). We recognize the following two levels of the hierarchy that can be mapped to the time series:

1. Platform (PF) flow is the total mail flow at a specific top-level PF.
2. Distribution center (DC) flow is the total mail flow at a specific down-level DC.

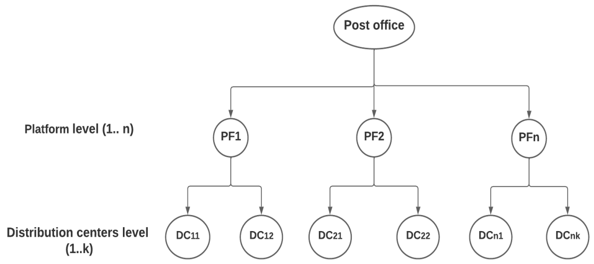

Figure 3. Hierarchical organization structure

This schema helps to take advantage of the hierarchical approaches in order to improve the forecast accuracy at the P and the DC levels as those are the relevant levels for supporting the mail delivery planning, and also improving the service quality.

The common approaches for forecasting the hierarchical time series are the Top-down (TD), Bottom-up (BU) and Optimal combination (OC) approaches (Athanasopoulos et al., 2009; Hyndman & Athanasopoulos, 2018).

The different hierarchical forecasting approaches can be presented with a unified structure:

$$\tilde{Y}_n(\mathrm{h}) = \mathbf{SG}\hat{Y}_n(\mathrm{h}), \tag{1}$$

where $\tilde{Y}_n(\mathrm{h})$ is the final reconciled forecasts of all series, n is the number of the observations in each series, **S** is a summing matrix of order $\mathrm{m} \times \mathrm{m}_k$ that aggregates the bottom level series, m is the total number of series in the hierarchy, k is the total number of levels in the hierarchy, **G** is a matrix of order $\mathrm{m} \times \mathrm{m}_k$ which elements depend on the type of the HF method used, $\hat{Y}_n(\mathrm{h})$ is an h-step-ahead independent base forecast of all series based on "n" observations (Hyndman & Athanasopoulos, 2018).

The Bottom-up (BU) approach generates forecasts for the bottom level and aggregates them to compute the forecasts for top-level series in the hierarchy. Some researchers suggest using this approach for forecasts related to operational decisions like production planning and logistics (Abolghasemi et al., 2022; Kahn, 1998).

In this approach, the **G** in equation (1) is $\mathbf{G} = \left[0_{mk\times(m-m_k)}|\mathrm{I}_{mk}\right]'$, where $0_{i\times j}$ is an $i \times j$ null matrix at levels i, j. Thus, **G** obtains the bottom-level forecasts and combines them with the summing matrix **S** to generate the final forecasts of the hierarchy.

The Top-down (TD) approach produces forecasts at the top level and disaggregates them to lower levels using some disaggregation feature. Here, researchers suggest using this approach for forecasts related to strategic decisions like defining budgets (Abolghasemi et al., 2022; Kahn, 1998). According to the TD approach, the vector $[P_j]$ of all disaggregation proportions corresponding to the series in the bottom level j=1,.., m_k is defined using two TD principal approaches known in the literature. These two approaches are the average historical proportions and the proportions of the historical averages (Gross & Sohl, 1990). According to the average historical proportions, each proportion p_j reflects the average of the historical proportions of the bottom level series y_j:

$$p_j = \frac{1}{n}\sum_{t=1}^{n}\frac{y_{j,t}}{y_t} \tag{2}$$

According to the proportions of the historical averages, each proportion p_j captures the average historical value of the bottom-level series y_{jt} , relative to the average value of the total aggregate y_t:

$$p_j = \frac{\sum_{t=1}^{n} y_{j,t}}{\sum_{t=1}^{n} y_t} \tag{3}$$

These proportions can be used to form the vector $\mathbf{g} = [\mathrm{p}_1, \mathrm{p}_2, \mathrm{p}, \dots, \mathrm{p}_{mk}]$ so that $\mathbf{G} = \left[\mathrm{g}\,|0_{mk\times(m-1)}\right]'$. On this point, **G** disaggregates the forecast at the top level to the lower levels.

The Optimal combination (OC) creates base forecasts for all series across all hierarchical levels and then combines them with a linear model to obtain the reconciled forecasts (Hyndman et al., 2016; Wickramasuriya et al., 2019). In the OC approach, the basic formula for calculating all final h-step-ahead forecast $\tilde{Y}_n(\mathrm{h})$ is presented as:

$$\tilde{Y}_n(\mathrm{h}) = S(S'W_h^{-1}S)^{-1}S'W_h^{-1}\,\hat{Y}(\mathrm{h}), \tag{4}$$

where W_h is the variance-covariance matrix of the base forecast errors which could be estimated using several ways like ordinary least square, weighted least squares, structural scaling, and the minimum trace. The minimum trace is used in this study as it gives the most accurate forecasting results (Mircetic et al., 2022).

2.3 Forecasting methods and accuracy measures

We propose two alternative common forecasting methods for producing the base forecasts used in the hierarchical forecasting approaches, namely: the autoregressive integrated moving average (ARIMA) method (Box, 1970; Silvestrini & Veredas, 2008), and the exponential smoothing with Error, Trend, and Seasonality (ETS) based on innovations state space method (Holt, 1957; Hyndman et al., 2008). Then, the forecasting results of both methods are compared to the results generated by the current forecasting method used in the organization as a baseline.

ARIMA combines three statistical models, the autoregressive (AR) to forecast a time series using a linear combination of its past p-values, the integrated (I) to make the time series stationary by removing (or mitigating) trend or seasonality differencing d times, and the moving average (MA) to forecast future values using a linear combination of previous forecast q-errors (Silvestrini & Veredas, 2008). The ARIMA model (p,d,q) can be written as follows:

$$y_t = \varphi_0 + \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \cdots + \varphi_p y_{t-p} + \tau_1\varepsilon_{t-1} + \tau_2\varepsilon_{t-2} + \cdots + \tau_p\varepsilon_{t-q}\,, \tag{5}$$

where y_t and ε_t represent the time series value and the error at *th* time instance, respectively., φ_0 is a constant coefficient, ε_t (i = 1, 2,...,q) are moving average (MA) coefficients and φ_j (j = 1, 2,...,p) are autoregressive (AR) coefficients.

Besides, the exponential smoothing methods provide future forecasts based on the weighted averages of past observations. The state space models for the exponential smoothing with Error, Trend, and Seasonality (ETS) (Hyndman et al., 2008) can be presented as:

$$y_t = \mathrm{w}(x_{t-1}) + \mathrm{r}(x_{t-1})\varepsilon_t \tag{6}$$

$$x_t = \mathrm{f}(x_{t-1}) + \mathrm{g}(x_{t-1})\varepsilon_t \tag{7}$$

Equation (6) represents the relationship between the observation y_t and the unobserved state x_{t-1}.
Equation (7) is the transition equation that represents the evolution of states over time, where f, g, and w, are coefficients and ε_t stands for the random errors.

2.4 Evaluation measures

To compare the performance of the forecasting methods, many measures are considered in the literature. For instance, In the inventory management literature, there are two main categories of performance measures: forecast accuracy measures and inventory performance measures (Pinçe et al., 2021).
In this study, the Mean Absolute Percentage Error (MAPE) (Makridakis et al., 1982) and the Root Mean Squared Scaled Error (RMSSE) (Hyndman & Koehler, 2006) are the main accuracy measures to evaluate the performance of the current forecasting method used by the organization, ARIMA, and ETS methods.
MAPE is the most commonly used error measure (Hong et al., 2019) which represents the average of the total percentage of errors between the real values and the forecasted values. A low MAPE value shows that the forecasted value is approaching its current value. Here is the expression of the MAPE:

$$\mathrm{MAPE} = 100.\frac{1}{n}\sum_{i=1}^{t}\left|\frac{Y_t - F_t}{Y_t}\right| \tag{8}$$

RMSSE is a scaled error measure independent of the data scale based on a training mean squared error of the Naïve method and it is recommended for evaluating the performance of the hierarchical forecasts (Makridakis et al., 2022). Here is the expression of the RMSSE:

$$\mathrm{RMSSE} = \sqrt{\frac{\frac{1}{h}\ \sum_{t=n+1}^{n+h}(Y_t - F_t)^2}{\frac{1}{n-1}\ \sum_{t=2}^{n}(Y_t - Y_{t-1})^2}} \tag{9}$$

Furthermore, in the same vein, as demand forecasting accuracy measures (Pinçe et al., 2021), we propose a new social indicator representing a dismissing level DL as it reflects the average time that the organization planned more employees than it needed based on the forecasts:

$$\mathrm{DL} = \frac{1}{n}\sum_{i=1}^{t}(Y_t \ \leq F_t)\,, \tag{10}$$

where F_t is the forecast value for period t, Y_t is the real observation in period t, n is the length of the within-sample and h is the length of the evaluation horizon.

3 Implementation, Findings, and Discussion

In order to implement the proposed hierarchical approach, the historical cleaned data for the monthly mail demand over 4 years is used, examples in (Figure 4) and (Figure 5) present strong trends and seasonality in the platform and distribution center levels. The data set is divided into a training set (from January 2015 to December 2017) and a test set (from January 2018 to December 2018). The number of time series is 19 at the platform (PF) level and 87 at the distribution center (DC) level.

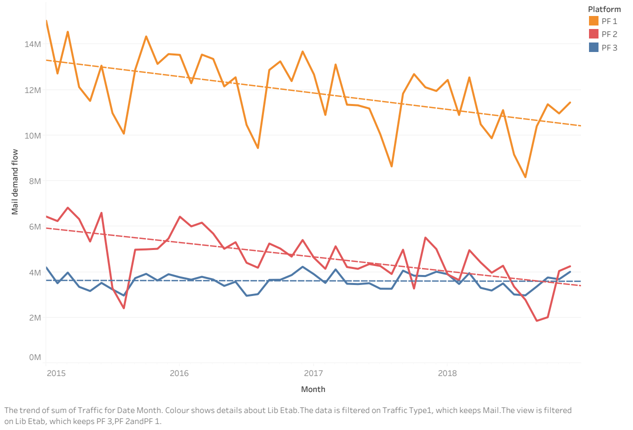

Figure 4. Mail demand flow from 2015 to 2018 at PF_1, PF_2, and PF_3

We apply the proposed forecast methods ARIMA and ETS at both levels using the direct approach first and then the hierarchical approach.

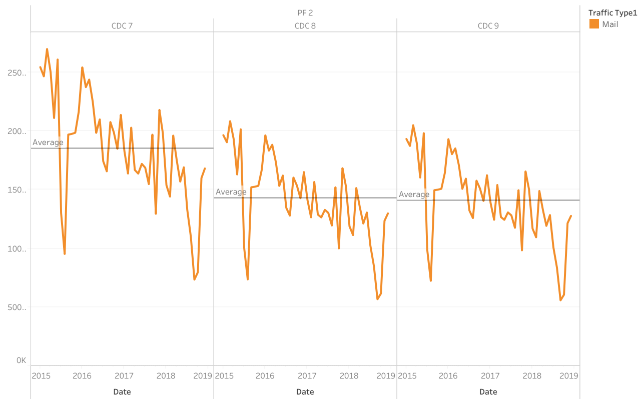

Figure 5. Mail demand flow from 2015 to 2018 at DC_1, DC_2, DC_3 associated to the platform PF_1

The automatic ARIMA; ARIMA () that integrates SARIMA and the automatic exponential smoothing; ETS() function in R fable package (O'Hara-Wild et al., 2020) are used to generate the forecasts of monthly mail in both levels directly. The best of all possible models was chosen according to Akaike information criterion (AIC) (Hyndman & Athanasopoulos, 2018).
For the hierarchical approaches the reconcile () function and forecast () function in the "fable" package were also used (O'Hara-Wild et al., 2020).
First, we use ARIMA, and ETS methods directly without considering hierarchical approaches. The current method used by the organization is considered the baseline to be compared to the proposed methods. The results shown in (Table 1) indicate that the method ETS outperforms the baseline and ARIMA methods according to both accuracy measures MAPE and RMSSE in both hierarchical levels of the organization. However, the proposed ARIMA and ETS methods are much better than the baseline.

Table 1. Accuracy measures using the direct methods

Level	Baseline		ARIMA		ETS	
	MAPE	RMSSE	MAPE	RMSSE	MAPE	RMSSE
PF	11.3	1.32	5.69	0.524	**4.91**	**0.466**
DC	12.3	1.41	5,88	0.562	**5.30**	**0.513**

Then, we study hierarchical forecasting using the Top-down (TD), Bottom-up (BU), and Optimal combination (OC) approaches. Consequently, we calculate the hierarchical forecasts based on the forecast combinations introduced in Section 2.2. The proposed hierarchical approaches TD, BU, and OC are evaluated according to the same accuracy measures MAPE and RMSSE by applying the baseline, ARIMA, and ETS methods (Table 2), (Table 3), and (Table 4).
For the TD forecasts, we subtract the allocation proportions at the distribution center (DC) down-level based on the observations in the training set at the platform (PF) top-level. Using this approach, the forecasts at the top-level PF are slightly improved using both methods ARIMA, and ETS compared to the direct approach. In the DC down-level, ETS outperforms the baseline and ARIMA methods, (Table 2).

Table 2. Accuracy measures using the TD hierarchical approach

Level	Baseline		ARIMA		ETS	
	MAPE	RMSSE	MAPE	RMSSE	MAPE	RMSSE
PF	11.9	1.37	5.65	0.519	**4.91**	**0.462**
DC	13	1.47	5,78	0.561	**5.37**	**0.519**

The BU forecasts for the top-level PF are the sum of the forecasts at the down-level DC. As shown in Table 3, based on this approach, the ETS method is the best method according to MAPE and RMSSE with improved accuracy measures at the top level compared to the direct approach.

Table 3. Accuracy measures using the BU hierarchical approach

Level	Baseline		ARIMA		ETS	
	MAPE	RMSSE	MAPE	RMSSE	MAPE	RMSSE
PF	11.3	1.32	5.72	0.527	**4.78**	**0.454**
DC	12.3	1.41	5,88	0.562	**5.30**	**0.513**

For the OC approach, a regression model is used to explore the information available across the platform and distribution center levels. First of all, the OC approach provides the best forecast accuracy measures and outperforms the TD and BU approaches, which agrees with the findings in the literature (Hyndman & Athanasopoulos, 2018; Mircetic et al., 2022).
Then, the EST is still the most efficient method for both top-level PF and down-level DC, (Table 4).

Table 4. Accuracy measures using the OC hierarchical approach

Level	Baseline		ARIMA		ETS	
	MAPE	RMSSE	MAPE	RMSSE	MAPE	RMSSE
PF	11.3	1.31	5.64	0.528	**4.56**	**0.427**
DC	12.1	1.39	5,88	0.565	**4.93**	**0.474**

Regarding the DL indicator, the results are different. First of all, ARIMA is the best method using the three hierarchical approaches at both levels PF, and DC. Then, the BU approach provided better results than TD, and OC at both levels, which agrees with the suggestion to use the BU approach for forecasts related to operational decisions (Abolghasemi et al., 2022; Kahn, 1998).

Table 5. DL in % using the hierarchical approaches

Level	Baseline			ARIMA			ETS		
	BU	TD	OC	BU	TD	OC	BU	TD	OC
PF	74%	76%	74%	**57%**	**59%**	**58%**	65%	60%	60%
DC	75%	79%	75%	**52%**	**52%**	**55%**	60%	55%	56%

Thus, the present study shows that the new hierarchical forecast-based decision-making approach is more accurate than the direct approach for all levels in the hierarchical structure. It provides the forecasts simultaneously for the platform (PF) level, and the distribution center (DC) level instead of calculating them separately using a direct approach. The forecast methods ETS, and ARIMA also provide more accurate forecasts for the mail flow than the current baseline method. Besides, based on accuracy measures ETS is more performant than ARIMA. On the other hand, ARIMA is better than ETS based on the new social indicator.
In summary, the proposed forecasting approach for the postal delivery service is more accurate and better than the current approach based on both accuracy and social measures. It helps the decision-makers to better plan delivery activities, and save time and effort as a managerial implication.

4 CONCLUSION

This study presents a new hierarchical approach for decision-makers at a French postal organization to forecast the mail flow in a real-life problem. The proposed approach gives more accurate forecasts to better plan the delivery activities at the different levels of the organization promoting the social facet. It incorporates a four-phase approach including data preparation, hierarchy scheme, forecasting methods, and evaluation measures. In the first phase, a structured process for preparing data is provided. The result of this phase is a clean dataset that can be used to generate forecasts for mail demand. The second phase designs the organization structure as a hierarchy scheme including the platforms (PFs) at the top level and the distribution centers (DCs) at the down level. The aim is to take advantage of the hierarchical time series approaches for the mail flow, namely, Top-down (TD), Bottom-up (BU), and Optimal combination (OC) approaches. In the third phase, ARIMA, and ETS methods are applied using direct and hierarchical approaches. Finally, these methods are compared according to MAPE and RMSSE as accuracy measures and dismissing level DL as a social indicator to evaluate their performance. The implementation of the proposed approach is presented and some findings are carried out to provide insights. By applying the different methods using the proposed approach, several evaluations and comparisons are obtained.
The forecast accuracy results indicate that the proposed methods ARIMA and ETS are both better than the current method used by the organization, regardless of the approach that is applied. The ETS is the most performant forecasting method in this case. In addition, the optimal combination (OC) hierarchical approach outperforms the Top-down (TD), Bottom-up (BU) approaches. By applying this approach, decision-makers can save time and improve postal delivery service by improving mail forecasts.
Based on the findings, future research could be conducted in the following ways. The suggested approach can be applied to other organizations and cases that could be designed as a hierarchical structure to solve different forecasting decision-making problems. It would be also interesting to consider in depth the

forecast utility (Babai et al., 2022) in addition to accuracy measures by proposing other new social measures in link with workforce and workload planning. Furthermore, an ultimate direction for future work is to forecast the hierarchical time series using machine learning methods like XGBoost or LightGBM, which have already been shown in the literature to be associated with a good hierarchical forecasting performance in the retail context.

5 REFERENCES

Abolghasemi, M., Tarr, G., & Bergmeir, C. (2022). Machine learning applications in hierarchical time series forecasting : Investigating the impact of promotions. *International Journal of Forecasting*. https://doi.org/10.1016/j.ijforecast.2022.07.004

Athanasopoulos, G., Ahmed, R. A., & Hyndman, R. J. (2009). Hierarchical forecasts for Australian domestic tourism. *International Journal of Forecasting*, *25*(1), 146-166. https://doi.org/10.1016/j.ijforecast.2008.07.004

Babai, M. Z., Boylan, J. E., & Rostami-Tabar, B. (2022). Demand forecasting in supply chains : A review of aggregation and hierarchical approaches. *International Journal of Production Research*, *60*(1), 324-348. https://doi.org/10.1080/00207543.2021.2005268

Box, G. E. (1970). GM Jenkins Time Series Analysis : Forecasting and Control. *San Francisco, Holdan-Day*.

Chlosta, N. W., & Froberg, L. (1977). Forecasting mail volume for production planning in a post office. *Computers & Operations Research*, *4*(4), 247-255. https://doi.org/10.1016/0305-0548(77)90023-5

Ducret, R. (2014). Parcel deliveries and urban logistics : Changes and challenges in the courier express and parcel sector in Europe — The French case. *Research in Transportation Business & Management*, *11*, 15-22. https://doi.org/10.1016/j.rtbm.2014.06.009

Gross, C. W., & Sohl, J. E. (1990). Disaggregation methods to expedite product line forecasting. *Journal of Forecasting*, *9*(3), 233-254. https://doi.org/10.1002/for.3980090304

Holt, C. E. (1957). Forecasting seasonals and trends by exponentially weighted averages (ONR Memorandum No. 52). *Carnegie Institute of Technology, Pittsburgh USA*, *10*.

Hong, T., Xie, J., & Black, J. (2019). Global energy forecasting competition 2017 : Hierarchical probabilistic load forecasting. *International Journal of Forecasting*, *35*(4), 1389-1399. https://doi.org/10.1016/j.ijforecast.2019.02.006

Hyndman, R. J., Ahmed, R. A., Athanasopoulos, G., & Shang, H. L. (2011). Optimal combination forecasts for hierarchical time series. *Computational Statistics & Data Analysis*, *55*(9), 2579-2589. https://doi.org/10.1016/j.csda.2011.03.006

Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting : Principles and practice*. OTexts.

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

Hyndman, R. J., Koehler, A. B., Snyder, R. D., & Grose, S. (2002). A state space framework for automatic forecasting using exponential smoothing methods. *International Journal of Forecasting*, *18*(3), 439-454. https://doi.org/10.1016/S0169-2070(01)00110-8

Hyndman, R. J., Lee, A. J., & Wang, E. (2016). Fast computation of reconciled forecasts for hierarchical and grouped time series. *Computational Statistics & Data Analysis*, *97*, 16-32. https://doi.org/10.1016/j.csda.2015.11.007

Hyndman, R., Koehler, A. B., Ord, J. K., & Snyder, R. D. (2008). *Forecasting with Exponential Smoothing : The State Space Approach*. Springer Science & Business Media.

Janjevic, M., & Winkenbach, M. (2020). Characterizing urban last-mile distribution strategies in mature and emerging e-commerce markets. *Transportation Research Part A: Policy and Practice*, *133*, 164-196. https://doi.org/10.1016/j.tra.2020.01.003

Kahn, K. B. (1998). Revisiting Top-Down versus Bottom-Up Forecasting. *The Journal of Business Forecasting Methods & Systems*.

Litsiou, K., Polychronakis, Y., Karami, A., & Nikolopoulos, K. (2022). Relative performance of judgmental methods for forecasting the success of megaprojects. *International Journal of Forecasting*, *38*(3), 1185-1196. https://doi.org/10.1016/j.ijforecast.2019.05.018

Makridakis, S., Andersen, A., Carbone, R., Fildes, R., Hibon, M., Lewandowski, R., Newton, J., Parzen, E., & Winkler, R. (1982). The accuracy of extrapolation (time series) methods : Results of a forecasting competition. *Journal of Forecasting*, *1*(2), 111-153. https://doi.org/10.1002/for.3980010202

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

Mircetic, D., Rostami-Tabar, B., Nikolicic, S., & Maslaric, M. (2022). Forecasting hierarchical time series in supply chains : An empirical investigation. *International Journal of Production Research*, *60*(8), 2514-2533. https://doi.org/10.1080/00207543.2021.1896817

Munkhdalai, L., Park, K. H., Batbaatar, E., Theera-Umpon, N., & Ryu, K. H. (2020). Deep Learning-Based Demand Forecasting for Korean Postal Delivery Service. *IEEE* https://doi.org/10.1109/ACCESS.2020.3030938

O'Hara-Wild, M., Hyndman, R., & Wang, E. (2020). fable : Forecasting models for tidy time series. *R package version 0.2, 1*.

Pinçe, Ç., Turrini, L., & Meissner, J. (2021). Intermittent demand forecasting for spare parts : A Critical review. *Omega*, *105*, 102513. https://doi.org/10.1016/j.omega.2021.102513

Punia, S., Singh, S. P., & Madaan, J. K. (2020). A cross-temporal hierarchical framework and deep learning for supply chain forecasting. *Computers & Industrial Engineering*, *149*, 106796. https://doi.org/10.1016/j.cie.2020.106796

Rostami-Tabar, B., Ali, M. M., Hong, T., Hyndman, R. J., Porter, M. D., & Syntetos, A. (2022). Forecasting for social good. *International Journal of Forecasting*, *38*(3), 1245-1257. https://doi.org/10.1016/j.ijforecast.2021.02.010

Rostami-Tabar, B., & Ziel, F. (2022). Anticipating special events in Emergency Department forecasting. *International Journal of Forecasting*, *38*(3), 1197-1213. https://doi.org/10.1016/j.ijforecast.2020.01.001

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

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*, 114-122. https://doi.org/10.1016/j.jom.2016.05.004

Villegas, M. A., & Pedregal, D. J. (2018). Supply chain decision support systems based on a novel hierarchical forecasting approach. *Decision Support Systems*, *114*, 29-36. https://doi.org/10.1016/j.dss.2018.08.003

Wicke, L., Dhami, M. K., Onkal, D., & Belton, I. (2022). Using Scenarios to Forecast Outcomes of a Refugee Crisis : BPS Defence and Security Psychology in the 21st Century. *International Journal of Forecasting*, *38*(3), 1175-1184. https://doi.org/10.1016/j.ijforecast.2019.05.017

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