Forecasting the number of firemen interventions using Exponential Smoothing methods: a case study

Roxane Elias Mallouhy, Christophe Guyeux, Chady Abou Jaoude and Abdallah Makhoul

Abstract Predicting the number of firemen interventions to size the appropriate workload of firefighters to the appropriate need is vital for reducing material and human resources. Therefore, it will have a great impact on reducing the financial crisis resulting from global warming and population growth. The database in this research includes interventions recorded hourly from ”1 January, 2015 00:00:00” to ”31 December, 2019 23:00:00” in Doubs, France. The data were processed, decomposed, outliers were detected and replaced. Thenceforth, optimal smoothing values were selected and then three different models of Exponential Smoothing were deployed. Experiments have shown that Holt-Winters' method has the best accuracy comparing to the baseline and other Exponential Smoothing techniques. The results are promising and would optimize the number of firefighters' resources.

Roxane Elias Mallouhy
Prince Mohammad Bin Fahd University, Khobar, Kingdom of Saudi Arabia
e-mail: `reliasmallouhy@pmu.edu.sa`

Christophe Guyeux, Abdallah Makhoul
Univ. Bourgogne Franche-Comté, Belfort, France
e-mail: `christophe.guyeux@univ-fcomte.fr`, `abdallah.makhoul@univ-fcomte.fr`

Chady Abou Jaoude
ntonine University, Baabda, Lebanon.
e-mail: `chady.aboujaoude@ua.edu.lb`

Acknowledgement: This work has been supported by the EIPHI Graduate School (contract ANR-17-EURE-0002) and is partially funded with support from the Hubert Curien CEDRE programme n° 46543ZD.

1 Introduction

Time series forecasting is a data science analysis used in a variety of fields. It has a substantial impact on future decisions with a generous range of complexity and applications. In agriculture, predicting the weather of the coming days helps to plan sowing or harvesting. In retail, forecasting daily sales gives guidance for inventory decisions. In finance, more successful investments are possible after predicting stock prices. Also, forecasting the number of firemen brigades would help optimize human and material resources.

Over time, the number of firefighters interventions in France has steadily increased, reaching more than 4.65 million operations in 2017 [1]. Predicting the future number of firefighting missions would help the emergency response system to estimate the possible flow of events in the coming hours and days, and then be able to optimize human and material resources. This would have a remarkable impact on protecting people, the environment, and property from damage, critical incidents, and disasters. Consequently, this would increase the efficiency of emergency operations while reducing operational costs.

On top of that, the use of machine learning to predict the number of firefighters operations seems to be efficient as it can be assumed that the number of operations is affected by climate, time, and other events such as New Year's Eve, holidays. Following this principle, it will be possible to analyze past observations using historical values, and associated patterns to predict future deployments. These properties are well represented in time series forecasting approaches that consider the trend, level, and seasonality of a time-ordered series. They analyze and forecast data observations and the result can lead to better decisions. To date, several models of time series forecasting models have been created, one of which is Exponential Smoothing, which gives more weight to recent values but gives less exponential importance to older observations. This is particularly the case with the firefighter dataset, as it is reasonable that the number of deployments in the upcoming hour tends to be influenced by the number of deployments in the previous hour.

The objective of this study is to predict the number of firemen brigades by applying the techniques of forecasting Simple Exponential Smoothing, Holt and Holt-Winters to a concrete dataset over the period 2015-2019. In this paper, we will show how we prepare the data and how we build each model to realistically predict the future. This paper is organized as follows: Section 2 presents an overview of Exponential Smoothing techniques and related work to our research; Section 3 demonstrates how the data was prepared; Section 4 provides the optimal choice of values of each model and experiments. Results are interpreted in Section 5 and last section provides conclusions and future work.

2 State of the art

Exponential Smoothing is a machine learning technique that can support level, trend, and seasonality unlike other time series forecasting methods used in a previous study [2] for the firefighter dataset, such as AutoRegression, Moving Average, and Auto Regressive Integrated Moving Average.

2.1 Exponential Smoothing methods

There are mainly three types of Exponential Smoothing: single (does not treat systematic structure), double (treats trend), and triple (both trend and seasonality).

2.1.1 Simple Exponential Smoothing (SES)

SES was developed by Robert G. Brown (1956) [3] and is naturally called Single Exponential Smoothing. It is a time series forecasting method for data with no clear trend or seasonal pattern. Forecasts are calculated by using the weighted average of the previous level and the current observations. SES associates more weightage to recent observations and fewer weights to older ones [4] and essentially requires a ”level” component called alpha (α). The formula of Simple Exponential Smoothing is as follows: $S_t = \alpha y_t + (1-\alpha)S_{t-1}, t > 0$, where: α is the smoothing coefficient between 0 and 1; S_t is the forecast value for period t; y_t is the refinement constant for the whole data.

2.1.2 Double Exponential Smoothing (Holt)

Charles C. Holt (1957) extended Single Exponential Smoothing, to allow prediction of data with trend [5]. Holt's method assumes that datasets have a trend and do not have seasonality, so it uses two components ”level” and ”trend”. This method is mainly used for linear trends with short/medium forecast periods. The two smoothing parameters for level and trend, alpha (α) and beta (β) respectively, are ranged between 0 and 1. A second equation is added to handle the trend aspect: $b_t = \beta(S_t - S_{t-1}) + (1-\beta)b_{t-1}$, wherein: β is the refinement constant for trends; b_t is the trend for period t.

2.1.3 Triple Exponential Smoothing (Holt-Winters)

Holt-Winters (1960) is an extension of Holt's method by Holt's student, Peter Winters, who assumes the existence of seasonal and trend variations for data observation [6]. It considers level, trend, and seasonality with the corresponding parameters alpha (α), beta (β), and gamma (γ). Triple Exponential Smoothing adds a third equation to the single and double smoothing as follows: $I_t = \gamma y_t + (1-\gamma)I_{t-l} + m$, where: γ is the seasonal smoothing constant; I_t is the seasonal index for period t; l is the size of the season.

2.2 Related Work

This section reviews various literature studies based on Exponential Smoothing methods in many fields. A large number of researches have been done on the powerful ES methods for time series prediction in many fields such as tourism, weather, COVID-19 pandemic, finance, economics, and medicine.

Singh, K. et al. [7] implemented in his research Exponential Smoothing method to predict the number of tourists for 2018 of an Indian state using Java programming language for the years 2008 to 2017. He applied different values of smoothing constant to find the best accuracy of the model. Zafar S. et al. [8] analyzed and studied temperature data and variability of two major regions using the Simple Smoothing Technique and concluded that SES gives the best predictive values compared to other models.
Moreover, Argawu, A. [9] predicted the number of COVID -19 new cases in the 10 most infected African countries by applying regression, ARIMA, and Exponential Smoothing Models. Yasar, H. and Kilimci, Z.H. [10] emphasized how to mix Time Series Forecasting methods with Financial Sentiment Analysis data collected from Twitter, Instagram, and Facebook. In their case study, they employed ARIMA, Holt, and Holt-Winters to provide a more consistent exchange rate prediction to any user wishing to exchange Turkish Lira /US dollars, and ended their study with the best-observed performance belonging to the Holt-Winters' method.
Anggrainingsih, R. et al. [11] analyzed time series data of website visitors using the Triple Exponential Smoothing method. Their results showed the optimal alpha, beta, and gamma for the best prediction accuracy. Lai, K.K. et al. [12] proposed a hybrid methodology by integrating Neural Network with Exponential Smoothing for financial time series prediction. Their experimental results considered the accuracy and directional predictions and showed that the hybrid-integrated method performs better than the two benchmark models. Jones, S.S. et al. [13] examined and evaluated the use of SARIMA, Exponential Smoothing, time series regression, and ANN to predict daily patient volume in the emergency department, compared the results with the multiple linear regression model previously performed and concluded that the regression-based model provided the most consistent accuracy.

Incidentally, few studies on time series forecasting using skills in Artificial Intelligence have considered the topic of our research about fire department operations. Nahuis, S.L.C. et al. used Long Short-Term Memory and demonstrated the possibility to build a neural network from scratch and the ability to predict the number of deployments in 2017 from those from 2012 to 2016 [14]. Couchot, J.F. et al. [15] worked on a learning process based on real but anonymous data and conducted the study using the Extreme Gradient Boosting technique. In [16], Guyeux C. et al. applied an ad-hoc Multi-Layer Perceptron in which hyper-parameters were selected using a supercomputer, and their work gave a reasonable prediction of firefighters' operations. Furthermore, S. Cerna et al. [17] compared between XG-Boost, Gradient Boosting, and AdaBoost, as these techniques are considered very effective in modeling nonlinear systems. Arcolezi, H.H. et al. [18] focused on local-differential privacy-based data in their study. Their approach was to predict the number of firefighters in specific locations by applying differential privacy along with XG-Boost techniques. None of these methods are classified as exponential smoothing techniques, which is the added value of this research compared to the aforementioned previous studies.

3 Data preparation

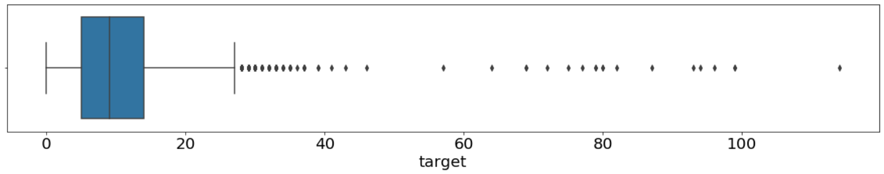
3.1 Data Acquisition

The data considered in this study contains the number of firemen interventions collected from "January 1, 2015, 00:00:00" to "December 31, 2019, 23:00:00" by the Fire and Rescue Department SDIS25, in Doubs, France. In this paper, two datasets are analyzed: the first one (hourly-dataset) contains the number of interventions per hour, while the second one (daily-dataset) contains the average of interventions per day. Therefore, the hourly dataset consists of 43824 interventions, while the daily dataset carries 1826 interventions.

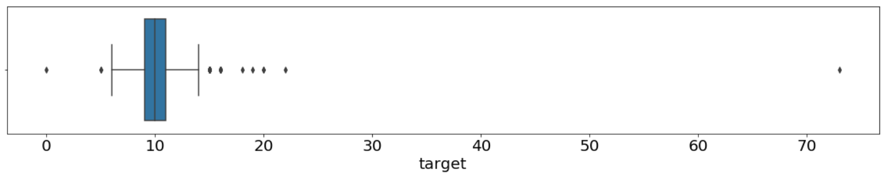
3.2 Outliers Detection

As can be seen in Figure 1, there are black dots outside the blue box: these are the outliers.

We consider everything above 25 for the hourly-dataset and everything below 6 and above 14 for the daily-dataset as anomalies. The boxplot detects 327 outliers for the hourly-dataset and 39 outliers for the daily-dataset.
To remove what can be considered as anomalies, we replaced the outliers with the lower or upper whisker, and consequently, the datasets have been updated, as can be viewed in Figure 2.

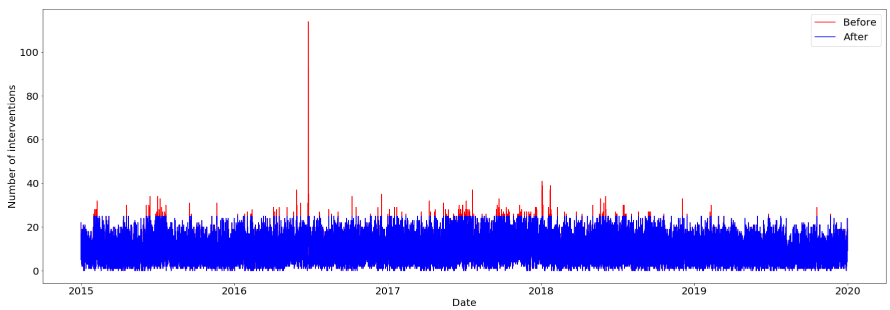


(a)

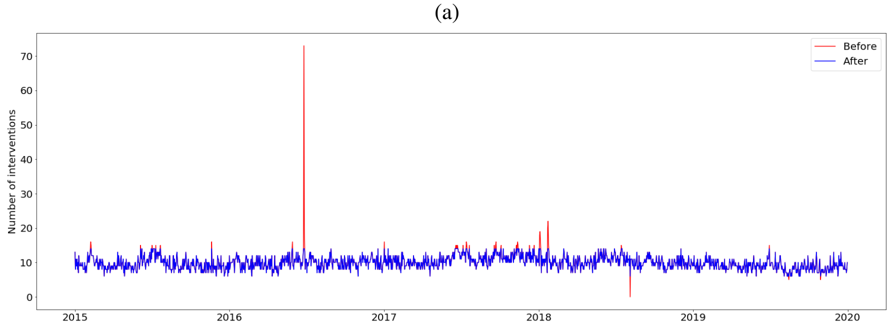


(b)

Fig. 1: Graphical visualization for outliers detection for (a) hourly-dataset and (b) daily-dataset



(a)



(b)

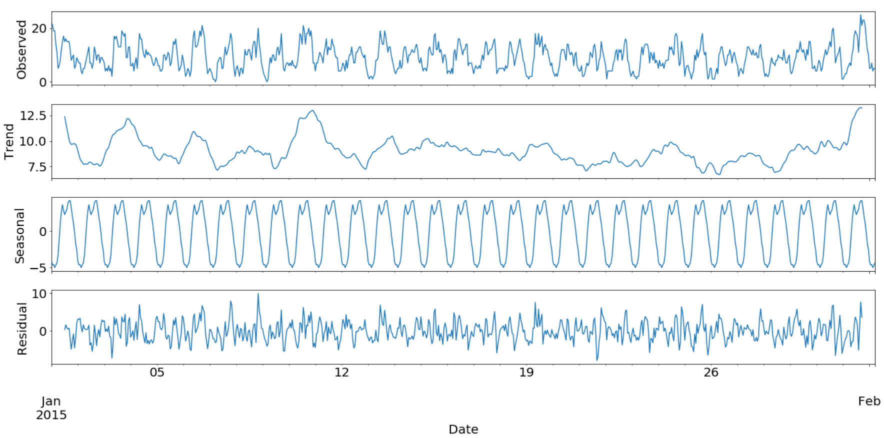
Fig. 2: Number of firefighters' interventions before and after replacing outliers for (a) hourly-dataset and (b) daily-dataset

3.3 Datasets Decomposition

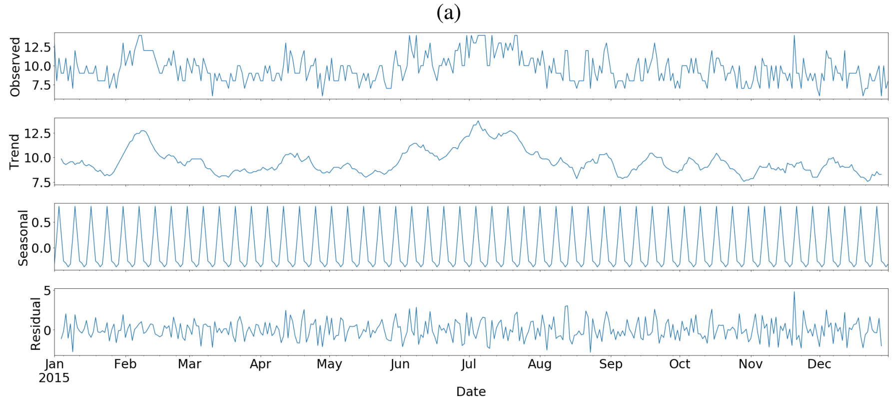
It is important to perform the decomposition of the datasets to get a structured view about the components used in the Exponential Smoothing methods, such as:

- trend: increasing/decreasing tendencies of firefighters' interventions;
- seasonality: repeating cycle;
- residual: random variation of the dataset.

In this study, both the daily and hourly datasets show an interesting seasonality and trend: the cycle repeats every day/24 hours for the hourly-dataset and every week/7 days for the daily-dataset, as shown in Figure 3. Moreover, the residuals are also reasonable and show different variability over time.



(a)



(b)

Fig. 3: Decomposition charts for (a) hourly-dataset and (b) daily-dataset

4 Experimental methods and results

4.1 Datasets

Each dataset is divided into training and testing, and to backtest the three Exponential Smoothing methods, this study uses the walk-forward validation technique by repeating the following steps:

1. training the model using the minimum number of samples in the window;
2. prediction for the unique next step;
3. evaluation of the predicted value against the real value;
4. expansion of the window to include the known value.

4.2 Selection of appropriate smoothing constants

The sensitivity of the predictions depends on the smoothing constants. Larger values of alpha (α) form a forecast that is more sensitive to recent observations, while in contrast, smaller values give a dampening result. The same concept applies to beta (β), which emphasizes recent trends over older observed values.
In this study, the smoothing parameters were selected as a function of the minimum values of mean absolute error (MAE) and root mean squared error (RMSE) so that the forecasts are more accurate. Different values of α and β were tried on the datasets. The concept is to repeat the loop 99 times for simple Exponential Smoothing, whereas for Holt's method the loop is repeated 99*99 times because it has two smoothing parameters. α and β are ranged from 0 to 1 and the loops increase the value of α and β by 0.01 at each iteration. On the other hand, the seasonality used for the Holt-Winters' method is picked up depending on the seasonal curve presented in section 3.3.
The obtained optimal constants for the hourly-dataset are alpha=0.9, beta=0.05 and the seasonality is 24 hours. On the other side, the optimal values for the daily-dataset are alpha=0.1, beta=0.05 and the seasonality is 7 days. These results are reported in Table 1 and Table 2

Daily-dataset		Hourly-dataset	
alpha	RMSE	alpha	RMSE
0.1	1.409	0.1	4.775
0.5	1.464	0.5	3.476
0.9	1.624	0.9	3.172

Table 1: The RMSE measures using different value of smoothing constant (α)

Daily-dataset			Hourly-dataset		
alpha	beta	RMSE	alpha	beta	RMSE
0.1	0.05	1.408	0.9	0.05	3.175
0.5	0.05	1.451	0.9	0.5	3.246
0.1	0.5	1.656	0.4	0.5	3.961

Table 2: The RMSE measures using different value of smoothing constants (α) and (β)

4.3 Forecasting Results

To check the efficiency of the models used, forecasting was made for different time intervals in the future, and to compare the obtained results with a reference, the persistence model is considered as the baseline. MAE and RMSE are calculated for each selected period for hourly-dataset and daily-dataset and the results are outlined in Table 3 and Table 4 respectively. Moreover, the prediction of the number of firemen for 300 hours in 2019 for the hourly-dataset is shown in Figure 4 and the result of prediction for the daily-dataset in the whole year 2019 is displayed in Figure 5.

Period	SES		Holt		Holt-Winters		Persistence	
	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
1 h	0.47	0.47	0.605	0.605	0.631	0.631	1	1
2 h	1.485	1.259	1.55	1.356	0.947	0.906	1.581	1.5
3 h	2.821	2.241	2.73	2.3	2.15	1.762	2.646	2.333
5 h	3.072	2.76	3.088	2.801	2.727	2.401	3.033	2.8
7 h	2.842	2.458	2.851	2.481	2.394	1.983	2.803	2.429
12 h	3.218	2.79	3.237	2.826	3.44	2.76	3.215	2.833
1 day	3.05	2.543	3.062	2.566	3.17	2.525	3.021	2.542
2 days	3.339	2.616	3.348	2.63	3.347	2.488	3.189	2.5
3 days	3.251	2.476	3.251	2.481	3.1	2.383	3.238	2.458
4 days	3.172	2.362	3.175	2.368	3.06	2.322	3.135	2.326
5 days	3.136	2.367	3.138	2.369	3.046	2.33	3.129	2.342
10 days	3.237	2.514	3.312	2.575	3.149	2.46	3.206	2.471
15 days	3.157	2.425	3.23	2.484	2.991	2.336	3.121	2.376
1 month	3.223	2.499	3.298	2.564	2.957	2.306	3.179	2.455
2 months	3.219	2.481	3.293	2.542	2.926	2.23	3.179	2.428

Table 3: RMSE and MAE for different prediction models for hourly-dataset over various time period

To show the results for multiple periods and since seasonality is evident in this study for both the daily and hourly datasets, we calculated the average number of firefighters' interventions on an hourly basis from "00:00:00" to "23:00:00" for the hourly-dataset, and on weekly basis from Monday to Sunday for the daily-dataset. Subsequently, the same concept is processed by applying Simple Expo-

Period	SES		Holt		Holt-Winters		Persistence	
	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
1 day	0.402	0.402	0.414	0.414	0.75	0.75	1	1
2 days	0.533	0.52	0.531	0.52	0.55	0.479	1	1
3 days	1.467	1.155	1.473	1.16	1.571	1.189	1.915	1.667
5 days	1.238	1.004	1.241	1.005	1.481	1.247	2	1.6
7 days	1.234	1.025	1.236	1.026	1.394	1.128	1.773	1.429
2 week	1.421	1.203	1.42	1.202	1.442	1.191	1.964	1.571
4 week	1.218	0.929	1.218	0.928	1.201	0.95	1.69	1.289
8 week	1.323	1.017	1.322	1.016	1.304	1.03	1.778	1.411
16 weeks	1.204	0.946	1.203	0.945	1.197	0.94	1.573	1.205
32 weeks	1.357	1.075	1.356	1.075	1.332	1.07	1.654	1.246
48 weeks	1.405	1.117	1.404	1.118	1.353	1.09	1.659	1.265
1 year	1.409	1.114	1.408	1.114	1.37	1.093	1.682	1.274

Table 4: RMSE and MAE for different prediction models for daily-dataset over various time period

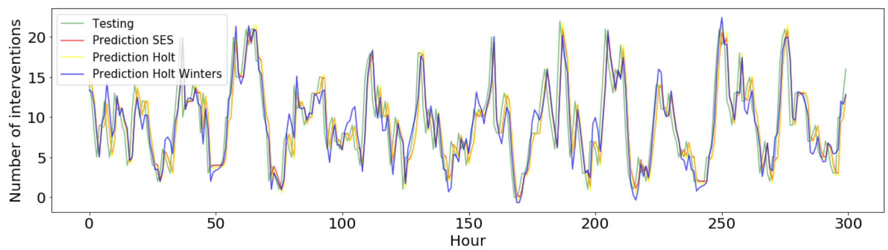


Fig. 4: Various models to predict the number of firefighters' interventions during 300 hours in January 2019 for hourly-dataset

nential Smoothing, Holt's, Holt-Winters', and persistence models. The results for the different methods used are illustrated in Figure 6 and Figure 7.

5 Discussion

This research aimed to develop three time forecasting methods using hourly and daily data from 2015 to 2019, split them into training/testing, implement walk-forward validation to backtest datasets, and select optimal values of smoothing parameters and seasonalities.

From Table 1 and Table 2, it can be seen that the value of RMSE for hourly-dataset is increased with the increase of alpha. However, the opposite is observed for the daily-dataset as the RMSE decreases with the increase of alpha. The selected optimal values resulted from the minimal RMSE. These results reflect that Exponential Smoothing assigns larger weights to the recent observations in the hourly-dataset and fewer weights in the daily-dataset. This means that α has a smaller effect and

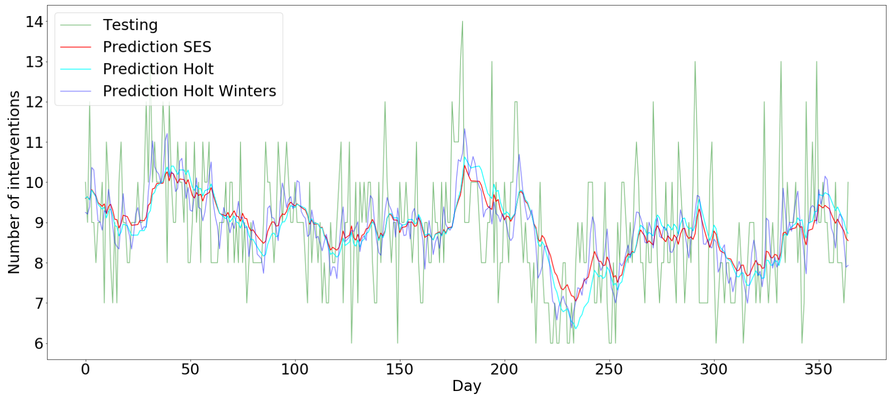


Fig. 5: Various models to predict the number of firefighters' interventions during the whole year 2019 for daily-dataset

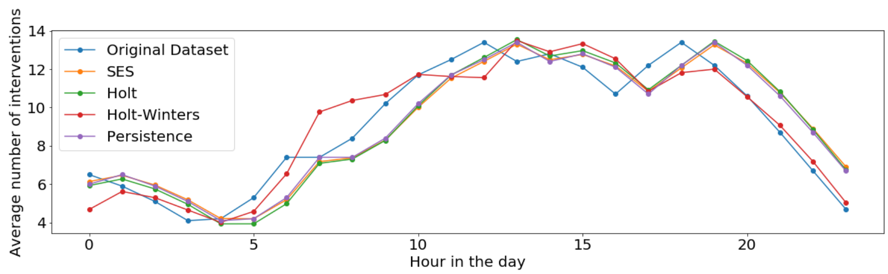


Fig. 6: The average number of firefighters' interventions over the hours of a day for hourly-dataset

gives more importance to the recent observations in hourly-dataset, while the data in the daily-dataset is less sensitive to the recent changes. Furthermore, the optimal value of beta is 0.05, which is very close to zero. This means that more weight has been given to the past trends in the estimation of current trends.
After selecting the smoothing constants that produce less error, MAE and RMSE are calculated for each prediction period for the hourly and daily datasets to measure the analysis performed during the forecasting process, as shown in Table 3 and Table 4. These tables indicate that the use of single and double Exponential Smoothing in this work is not effective compared to the Holt-Winters' method, which gives the lowest prediction error over time. In other words, as can be seen in Figure 8 and Figure 9, when Holt-Winters' method is used, the RMSE decreases as the prediction period increases. This result reveals that Triple Exponential Smoothing is a feasible technique used in this research because both the daily and hourly datasets are seasonal.
Finally, the compilation of the average number of firefighters operations during the

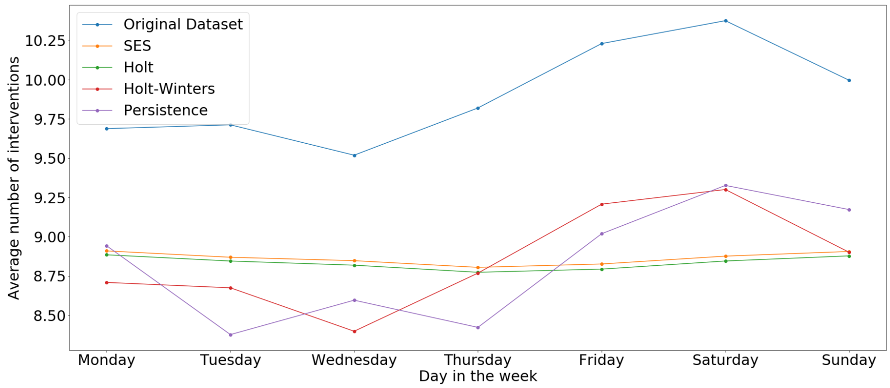


Fig. 7: The average number of firefighters' interventions over the days of the week for daily-dataset

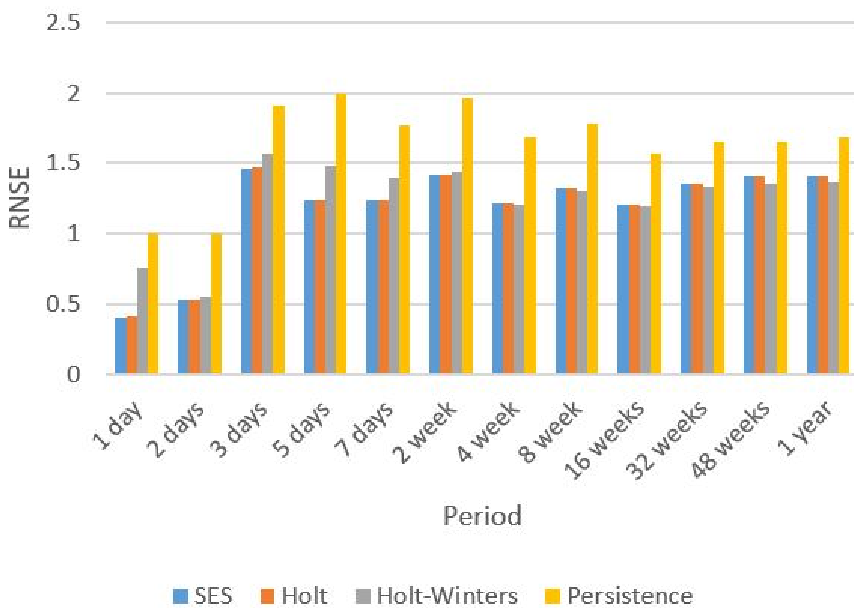


Fig. 8: Prediction results on hourly-dataset

days of the week for the hourly-dataset and during each hour of the day for the daily-dataset expose many relevant facts. Figure 6 and Figure 7 show that Holt-Winters' method has the most accurate values of prediction comparing to the original values of interventions. Additionally, it is observed in Figure 6 that the firemen services increase from 5:00 am, tended to peak, and remain broadly stable throughout the day before gradually decreasing at 7:00 pm. This is very reasonable because the

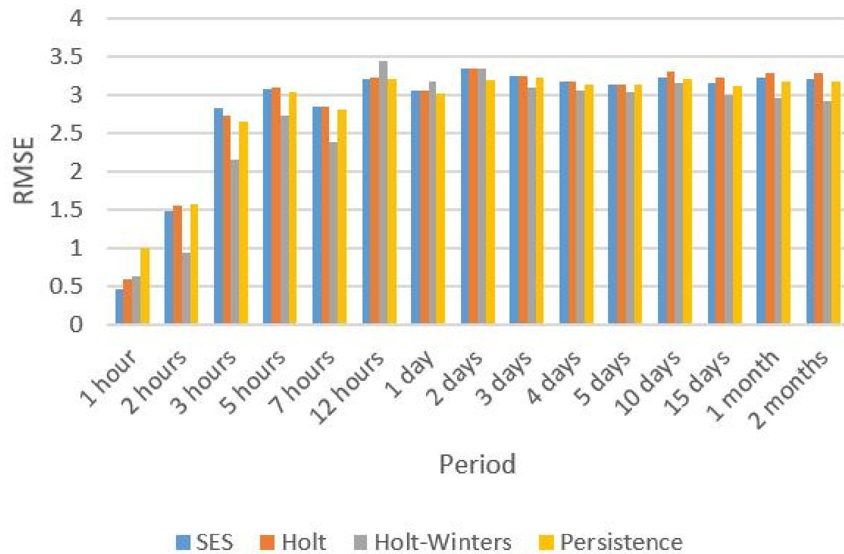


Fig. 9: Prediction results on daily-dataset

number of vehicles and flow of the people out of their homes is higher during the day causing rush hours and thus more risk of damage and incidents. On the other hand, it can be seen in Figure 7 that the firefighters have the highest number of interventions from Friday to Sunday with less fluctuations on weekdays. Unsurprisingly, weekends are the riskiest days for fatal crashes.

6 Conclusion

It is very discernible that the prediction of the number of firefighters' interventions is not a simple random process, but is influenced by hourly and weekly changes related to human activities. Exponential Smoothing is a prominent tool to use in such a study that provides reliable forecasting. Statistics characteristics and graphical exploration of data have been presented to find the best Exponential Smoothing technique and three models were developed and then compared to each other and the baseline. Based on the prediction error, the less variation measures were found for the method Holt-Winters' since it takes into account seasonality, which is the main component of hourly and daily datasets.
For future work, new machine learning techniques will be explored on this dataset. Testing new and larger time steps is planned, such as ambulatory transportation. Finally, adding integrated explanatory inventories to the dataset is a possible idea to predict not only the number of interventions but also the type of interventions.

References

1. Statista Research Department. Number of firefighting operations in france between 2007 and 2017.
2. Roxane Mallouhy, Christophe Guyeux, Chady Abou Jaoude, and Abdallah Makhoul. Time series forecasting for the number of firefighters interventions. In *The 35-th International Conference on Advanced Information Networking and Applications*, pages ***–***. Springer Series "Advances in Intelligent Systems and Computing", May 2021.
3. Robert Goodell Brown. Exponential smoothing for predicting demand. cambridge, mass., arthur d. little. *Book Exponential Smoothing for Predicting Demand*, 1956.
4. Everette S Gardner Jr. Exponential smoothing: The state of the art. *Journal of forecasting*, 4(1):1–28, 1985.
5. Charles C Holt. Forecasting seasonals and trends by exponentially weighted moving averages. *International journal of forecasting*, 20(1):5–10, 2004.
6. Peter R Winters. Forecasting sales by exponentially weighted moving averages. *Management science*, 6(3):324–342, 1960.
7. Kuljeet Singh, Sourabh Shastri, Arun Singh Bhadwal, Paramjit Kour, Monika Kumari, Anand Sharma, and Vibhakar Mansotra. Implementation of exponential smoothing for forecasting time series data. *Int. J. Sci. Res. Comput. Sci. Appl. Manag. Stud*, 8, 2019.
8. Muhammad Kashif Saba Zafar, Muhammad Imran Khan, and Hafiza Nida. Application of simple exponential smoothing method for temperature forecasting in two major cities of the punjab, pakistan.
9. Alemayehu Argawu. Modeling and forecasting of covid-19 new cases in the top 10 infected african countries using regression and time series models. *medRxiv*, 2020.
10. Harun Yasar and Zeynep Hilal Kilimci. Us dollar/turkish lira exchange rate forecasting model based on deep learning methodologies and time series analysis. *Symmetry*, 12(9):1553, 2020.
11. Rim Anggrainingsih, Gilang Romadhon Aprianto, and Sari Widya Sihwi. Time series forecasting using exponential smoothing to predict the number of website visitor of sebelas maret university. In *2015 2nd international conference on information technology, computer, and electrical engineering (ICITACEE)*, pages 14–19. IEEE, 2015.
12. Kin Keung Lai, Lean Yu, Shouyang Wang, and Wei Huang. Hybridizing exponential smoothing and neural network for financial time series predication. In *International Conference on Computational Science*, pages 493–500. Springer, 2006.
13. Spencer S Jones, Alun Thomas, R Scott Evans, Shari J Welch, Peter J Haug, and Gregory L Snow. Forecasting daily patient volumes in the emergency department. *Academic Emergency Medicine*, 15(2):159–170, 2008.
14. Selene Leya Cerna Nahuis, Christophe Guyeux, Héber Hwang Arcolezi, Raphaël Couturier, Guillaume Royer, and Anna Diva Plasencia Lotufo. Long short-term memory for predicting firemen interventions. In *2019 6th International Conference on Control, Decision and Information Technologies (CoDIT)*, pages 1132–1137. Ieee, 2019.
15. Jean-François Couchot, Christophe Guyeux, and Guillaume Royer. Anonymously forecasting the number and nature of firefighting operations. In *Proceedings of the 23rd International Database Applications & Engineering Symposium*, pages 1–8, 2019.
16. Christophe Guyeux, Jean-Marc Nicod, Christophe Varnier, Zeina Al Masry, Nourredine Zerhouny, Nabil Omri, and Guillaume Royer. Firemen prediction by using neural networks: A real case study. In *Proceedings of SAI Intelligent Systems Conference*, pages 541–552. Springer, 2019.
17. Selene Cerna, Christophe Guyeux, Héber H Arcolezi, Raphaël Couturier, and Guillaume Royer. A comparison of lstm and xgboost for predicting firemen interventions. In *World Conference on Information Systems and Technologies*, pages 424–434. Springer, 2020.
18. Heber H Arcolezi, Jean-Francois Couchot, Selene Cerna, Christophe Guyeux, Guillaume Royer, Bechara Al Bouna, and Xiaokui Xiao. Forecasting the number of firefighter interventions per region with local-differential-privacy-based data. *Computers & Security*, 96:101888, 2020.