Time Series Forecasting: A Study on Local Urban Waste Management in a Portuguese City

A. Manuela Gonçalves[1,3](✉) and Vítor Hugo Silva[2,3]

[1] Center of Mathematics, University of Minho, Guimarães, Portugal
mneves@math.uminho.pt
[2] VITRUS AMBIENTE, EM, S.A., Guimarães, Portugal
vitor.silva@vitrusambiente.pt
[3] Department of Mathematics, University of Minho, Guimarães, Portugal

Abstract. Due to economic and social development in general and to population growth, the volume of waste, particularly municipal waste, has been significantly increasing in recent years. VITRUS AMBIENTE, EM, S.A. is a municipal company that operates at various levels in local business management, namely in Urban Waste Management (UWM), and ensures the collection of waste in the Municipality of Guimarçes (located in the northwest of Portugal). A pioneer project termed "Pay-As-You-Throw" (PAYT) was implemented in the Historic Center (HC) of this municipality and is managed by the VITRUS AMBIENTE. VITRUS is the managing entity and the Urban Hygiene Service is responsible for implementing the necessary measures to ensure the success of this project. This work focuses specifically on Urban Waste Management and aims at modeling and forecasting the behavior of urban waste production within the company's PAYT area. Time series forecasting models – Box-Jenkins SARIMA models – are applied both for the estimation and forecasting of short-term organic and recyclable intramural waste collection in the Historic Centre of Guimarçes in the PAYT system implementation pilot zone. The dataset consists of weekly measurements from 2016 to 2019. The methodologies used support the company's management and decision-making process regarding Urban Waste Management, aiming at improving the services provided to the population and having always as its cornerstone the preservation of the environment.

Keywords: Waste management · Recycling · PAYT · Modeling · Forecasting · Time series

1 Introduction

In this work, we address the issue of urban waste in the city of Guimarães located in the northwest of Portugal. An appropriate solid urban waste management

A. Manuela Gonçalves was partially financed by Portuguese Funds through FCT (Fundação para a Ciência e a Tecnologia) within the Projects UIDB/00013/2020 and UIDP/00013/2020 of CMAT-UM.

O. Gervasi et al. (Eds.): ICCSA 2021, LNCS 12952, pp. 527–538, 2021.
https://doi.org/10.1007/978-3-030-86973-1_37

becomes a local and national challenge regarding the development of legislative plans and standards set in the "Plano Estratégico da Gestão de Resíduos Urbanos" (PERSU). This plan encourages the maximization of waste potential and suggests the introduction of a tax to promote waste separation at source and collection rate increase: the Pay-As-You-Throw (PAYT) system. This system is based on the "polluter-payer" principle and on the concept of shared responsibility, according to which those who generate less waste pay less [1]. This check is carried out based on the number of bags sold, that is, citizens buy the bags of organic waste according to the capacity requested, and the bags for recycling are free of cost. This is one of the system's strengths in adopting good environmental practices. Guimarães was a pioneering city in the implementation of this system, and VITRUS AMBIENTE is a local public company that manages all the environments of the system: from waste collection to rising public awareness to this issue.

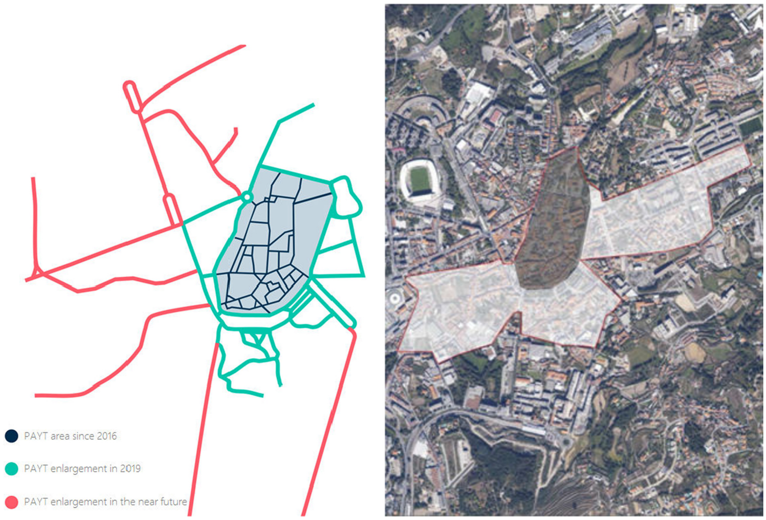

Fig. 1. Map with the PAYT system's expansion plans (left); PAYT system's current area of operation (right).

The main objective of this research is to analyze the modeling and forecasting processes of recyclable and organic waste production in the system's operation area, i.e., in the Historic Center of Guimarães, which is the PAYT system's implementation pilot zone (Fig. 1).

Predicting and forecasting environment processes - in this case, urban waste production - has always been a difficult field of research analysis, with very slow progress rate over the years. How to best model and forecast these patterns has been a long-standing issue in time series analysis. This research investigates the forecasting performances of the forecasting methods of Box-Jenkins SARIMA models [2]. These models are chosen because of their ability to model seasonal fluctuations present in this type of data, time series with seasonal patterns. Thus, statistical models are developed in this context of time series analysis to estimate and forecast, in the period from 2016 to 2019, the weekly production of waste (recyclable and organic) in the PAYT area. The validity of the modeling and forecasting models performed is judged by comparing its estimate and forecast error of the models. The smaller the error, the better the estimate and the forecast value produced by the model.

2 Methodology

2.1 SARIMA Model

A time series is an ordered sequence of values of a variable at equally spaced time intervals, in this case weekly organic and recyclable intramural waste collection in the Historic Centre of Guimarães, which is the PAYT system's implementation zone. Time series forecasting is an important area in which past observations of the same variable are collected and analyzed to develop a model describing the underlying relationship. The model is then used to extrapolate the time series into the future. Forecasting methods are a key tool in decision-making processes in many areas, such Urban Waste Management, as is the case studied in this work [3]. There are several approaches to modeling time series, but we decided to study and to compare the accuracy of the seasonal autoregressive integrated moving average models denominated by SARIMA models, because these models can increase the chance of capturing the proprieties and the dynamics in the data, in particular seasonality, and improving forecast accuracy.

The SARIMA $(p, d, q)(P, D, Q)_s$ is a short memory model and is a very flexible model, given that it accounts for stochastic seasonality, and is one of the most versatile models for forecasting seasonal time series. Such seasonality is present when the seasonal pattern of a time series changes over time. The theory of SARIMA models has been developed by many researchers and its wide application the result of the work by Box et al. [4], who developed a systematic and practical model-building method. Through an iterative three-step model-building process, model identification, parameter estimation and model diagnosis, the Box-Jenkins methodology has proven to be an effective practical time series modeling approach.

The SARIMA model has the following form

$$\mathit{\Phi}_p(B)N_P(B^s)(1-B)^d(1-B^s)^D Y_t = \mathit{\Theta}_q(B)H_Q(B^s)\epsilon_t,$$

where Y_t is the time series, with
$\mathit{\Phi}_p(B) = 1 - \phi_1 B - \cdots - \phi_p B^p,$
$N_P(B^s) = 1 - \nu_1 B - \cdots - \nu_P P_s,$
$\mathit{\Theta}_q(B) = 1 + \theta_1 B + \cdots + \theta_q B^q,$
$H_Q(B^s) = 1 + \eta_1 B^s + \cdots + \eta_Q Q_s,$

where s is the seasonal length, B is the backshift operator defined by $B^k Y_t = Y_{t-k}, \Phi_p(B)$ and $\Theta_q(B)$ are the regular autoregressive and moving average polynomials of orders p and q, respectively, $N_P(B^s)$ and $H_Q(B^s)$ are the seasonal autoregressive and moving average polynomials of orders P and Q, respectively, and $epsilon_t$ is a sequence of white noises with zero mean and constant variance σ^2. $(1-B)^d$ and $(1-B^s)^D$ are the nonseasonal and seasonal differencing operators, respectively.

The model with the minimum value of the AIC (Akaike's Information Criteria) is often the best model for forecasting [5]. We investigated the required transformations for variance stabilization and decided to apply logarithms to the time series under study.

Once the model has been specified, its autoregressive, moving average, and seasonal parameters (SARIMA model) need to be estimated. The parameters of SARIMA models are usually estimated by maximizing the likelihood of the model (for more details about this procedure, see [6]). The estimation is carried out in R.

2.2 Forecast Error Measures

Let's denote the actual observation for time period t by Y_t and the estimated or forecasted value for the same period by $\hat{Y}_t$ and n is the total number of observations. The most commonly used forecast error measures are the mean squared error (MSE), the root mean squared error (RMSE), the mean absolute percentage error (MAPE), the mean absolute scaled error (MASE) [7]. They are defined by the following formulas, respectively:

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{t=1}^{n} e_t^2} = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(Y_t - \hat{Y}_t)^2},$$

$$\mathrm{MAPE} = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{Y_t - \hat{Y}_t}{Y_t}\right| \times 100 \quad (\%)\,,$$

$$\mathrm{MASE} = \frac{1}{n}\sum_{t=1}^{n}|q_t| = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{Y_t - \hat{Y}_t}{\frac{1}{n-1}\sum_{t=2}^{n}|Y_t - Y_{t-1}|}\right|.$$

When comparing the performance of forecast methods on a single dataset, there is no absolute criterion for a "good" value of the error measure: it depends on the units in which the variable is measured and on the degree of forecasting, as measured in these units, which is sought in a particular application. Frequently, different accuracy measures will lead to different results regarding the forecast method is best.

3 Solid Waste Forecasting in the PAYT Area

3.1 DATA

The urban waste data is collected weekly by the VITRUS AMBIENTE and for this study only the data collected in the city's Historic Center, in the PAYT area, is considered. The observation period was between the 9th week of 2016 and the 34th week of 2019. The time series of recyclable waste (REC) and organic waste (ORG) were analysed, in tonnes. The recyclable waste is the sum of paper, plastic/metal, and glass waste.

Table 1. Descriptive statistics of solid waste - organic, and recyclable (paper, plastic/metal, and glass) waste - in PAYT area (monthly).

	Minimum	Maximum	1st quartile	Median	Mean	3rd quartile	Standard deviation
ORG	29.34	79.44	44.88	50.06	51.63	58.64	10.21
REC	16.08	36.92	21.08	24.40	24.92	28.14	5.31
PAP	3.02	11.24	5.00	6.44	6.58	8.00	2.15
PLA	3.24	6.72	3.72	4.26	4.53	5.30	0.93
GLA	7.80	21.70	10.44	12.88	13.81	16.74	3.86

Table 1 presents descriptive statistics for the five types of waste during the observed period by month in the PAYT system. As expected, organic waste has higher statistics; in particular, the standard deviations of 10.21 tons indicate a larger variability during the observed period. Regarding the total amount of waste produced, in this system pilot zone the average amount of waste produced is equal to 76.55 tons per month and with a minimum and maximum amount equal to 45.42 tons and 103.84 tons, respectively.

In the analysis of recyclable waste, paper waste has an average production of 6.58 tons per month and, during the observed period, a production between 3.02 tons and 11.24 tons. Plastic/metal waste presents an average production of 4.53 tonnes, and during the observed period, plastic/metal production was between 3.24 tonnes and 6.72 tonnes. Finally, glass waste has an average production of 13.81 tonnes, with glass waste representing a large part of the recyclable collection, and during the observed period, glass waste production presented values between 7.80 tons and 21.70 tons monthly.

Figure 2 presents the box-plots of organic waste (ORG), and the recyclable waste (REC) (paper, plastic/metal, and glass waste), by month. The box-plots present asymmetric distributions (positive), but paper waste distribution and recyclable waste distribution are almost symmetric. It should be noted that there is a moderate outlier on the data of organic waste, which corresponds to the month of June 2018.

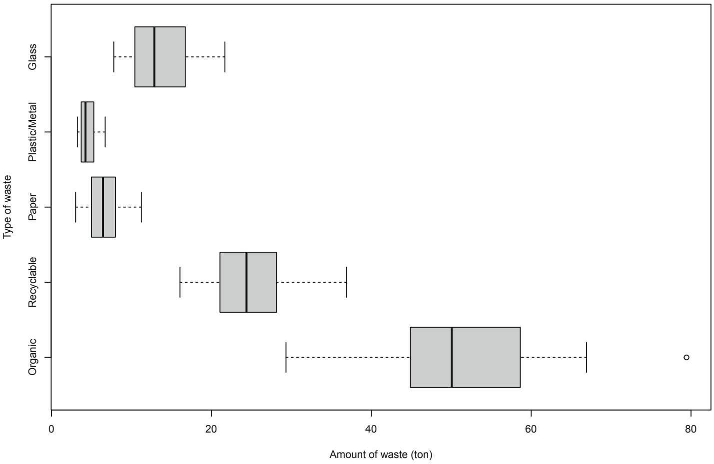

Fig. 2. Box-plots of solid waste - organic, and recyclable (paper, plastic/metal, and glass) - in PAYT area (monthly).

3.2 Results

Two time series are considered in the modeling process: recyclable waste (REC) and organic waste (ORG). The methods considered in this study are applied to two sets: training data (in-sample data) and testing data (out-of-sample data) in order to testify the accuracy of the proposed forecasting models.

For the REC and ORG waste time series, the selected training period was from the 9th week of 2016 to the 38th week of 2018 (first 134 observations) and was used to fit the models to data, and the testing period with the last 15 observations (the period from the 39th week of 2018 to the 1st week of 2019) was used to forecast. This approach allows the effectiveness of different methods of prediction.

The main task in SARIMA forecasting is to select an appropriate model order, i.e., the p, d, q, P, D, Q and s values. The modeling process follows the several steps to identify the model. Plotting the time series and choosing the proper variance-stabilizing transformation (in this case study, because the series present nonconstant variance, it was applied a logarithm transformation). Than computing and examining the ACF sample (Autocorrelation Function) and the PACF sample (Partial Autocorrelation Function) of the transformed data to further confirm a necessary degree of differencing, starting with the seasonal differencing. For seasonal modeling after analyzing the FAC, it appears that seasonality is weekly and, therefore, $s = \frac{365.25}{7} \approx 52.18 \approx 52$. This period of 52 weeks is due to the fact that there are leap years (366 d) and/or 53 weeks. In this way, this seasonality adjustment will allow a better formulation of the SARIMA model. Finally, computing and examining the ACF sample and PACF sample, we identify the properly transformed and differenced series of p, q, P, and Q by matching the patterns in the ACF sample and the PACF sample with the theoretical patterns of known models and via AICs.

After identifying an appropriate SARIMA model we have to check whether the model assumptions are satisfied. The basic assumption for both models is that ϵ_t is a zero mean Gaussian white noise process. It is considered the usual significance level of 5%. These residual analyses are not presented in this paper, but were performed. For the two final SARIMA models (recyclable waste and organic waste) all the model assumptions were verified; therefore, the diagnostics for these models leads to the conclusion that the models are adequate.

Recyclable Waste Forecasting. For the recyclable waste time series modeling process, the final SARIMA model was selected according to the procedure described earlier. Fitting the several models suggested by these observations and computing AIC for each, we obtain the results presented in Table 2. Based on the AICs, the selected model is the SARIMA $(1, 0, 1)(1, 1, 0)_{52}$.

Table 2. Parameters estimates of SARIMA models for the recyclable waste time series, and the AICs.

Model	$\hat{\phi}_1$	$\hat{\phi}_2$	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\nu}_1$	AIC
SARIMA(1,0,0)$(1,1,0)_{52}$	–0.393	–	–	–	–0.548	126.80
SARIMA(0,0,1)$(1,1,0)_{52}$	–	–	-0.377	–	–0.547	126.05
SARIMA(1,0,0)$(0,1,0)_{52}$	–0.354	–	–	–	–	135.66
SARIMA(2,0,1)$(1,1,0)_{52}$	–1.045	–0.234*	0.663*	–	–0.550	129.95
SARIMA(1,0,2)$(1,1,0)_{52}$	–0.734	–	0.328	–0.182	–0.557	129.31

* The parameter is not significant for a significant level of $\alpha = 0.10$.

The estimation results of the selected model can be consulted in Table 3.

Table 3. Results for the final SARIMA model for the recyclable waste time series in the historic center.

Final model: SARIMA $(0, 0, 1)(1, 1, 0)_{52}$	AIC $= 126.05$	$\hat{\sigma}^2 = 0.204$	
	θ_1	ν_1	
Estimate	–0.377	–0.547	
Standard error	0.095	0.120	

In Fig. 3 are represented the original values of recyclable waste, the estimates in the modeling period (training period), the forecasts in the forecasting period (testing period) and the forecast intervals for a confidence level of 90% by applying the SARIMA model.

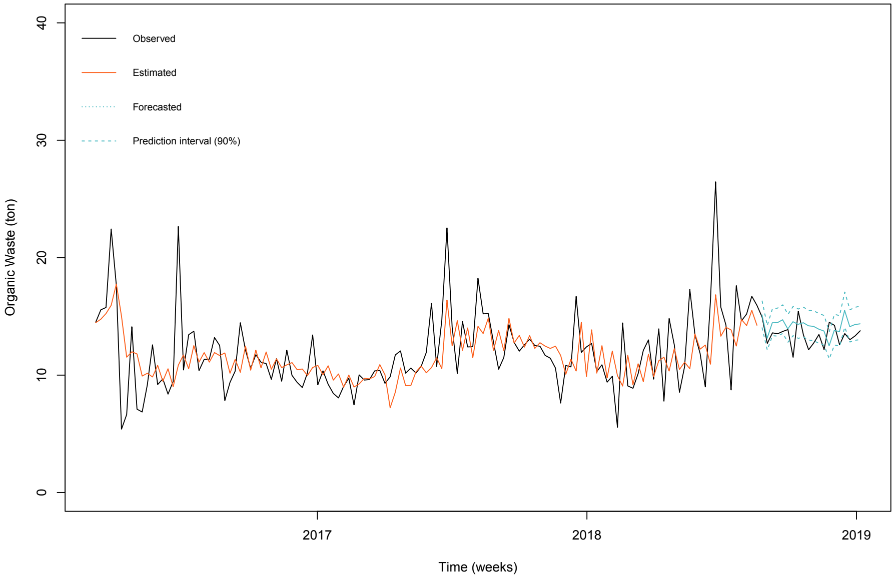

Fig. 3. Observed values, estimates and forecasts (with 90% confidence bounds) for recyclable waste time series using the final SARIMA model.

Figure 3 suggests that the model's predictive quality is better in the training series than in the testing series, since there is an atypical decrease in the amounts of recyclable waste in the original series that the model would not be able to predict given the observations of the past.

The coverage rate of empirical confidence of corrected forecasts is 40%, not similar to the confidence intervals with a confidence of 90% (6 observations of the testing series belong to the confidence interval).

Organic Waste Forecasting. The same modeling process was applied to the organic waste time series. Fitting the several models suggested by these observations and computing the AIC for each, we obtain the results presented in Table 4.

On the bases of the AICs, it is preferred the $SARIMA(1,1,2)(1,0,0)_{52}$ model. The estimation results of this model can be consulted, in more detail in Table 5.

In Fig. 4 are represented the original values of organic waste, the estimates in the modeling period (training period), the forecasts in the forecasting period (testing period) and the forecast intervals for a confidence level of 90% by applying the SARIMA model.

The confidence intervals with a 90% confidence have a coverage rate of 60% (less than 90%), since only 9 observations in the testing series belong to them.

Table 4. Parameters estimates of SARIMA models for the organic waste time series, and the AICs.

Model	$\hat{\phi}_1$	$\hat{\phi}_2$	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\nu}_1$	$\hat{\eta}_1$	AIC
SARIMA(0,1,1)$(1,0,1)_{52}$	–	–	–0.800	–	0.908	-0.741^*	10.45
SARIMA(2,1,0)$(1,0,1)_{52}$	–0.563	–0.381	–	–	0.994	–0.931	24.41
SARIMA(0,1,0)$(1,0,1)_{52}$	–	–	–	–	0.949	–0.848	64.66
SARIMA(0,1,2)$(1,0,1)_{52}$	–	–	–0.721	-0.116^*	0.875^*	-0.701^*	11.25
SARIMA(1,1,2)$(1,0,1)_{52}$	0.911	–	–1.719	0.719	0.896	–0.721	11.61
SARIMA(0,1,1)$(0,0,1)_{52}$	–	–	–0.810	–	–	0.221	12.24
SARIMA(0,1,1)$(1,0,0)_{52}$	–	–	–0.812	–	0.291	–	10.54
SARIMA(1,1,1)$(0,0,1)_{52}$	0.147^*	–	–0.876	–	–	0.217^*	12.62
SARIMA(1,1,2)$(1,0,0)_{52}$	–0.815	–	0.145	–0.822	0.280	–	4.61

* The parameter is not significant for a significant level of $\alpha = 0.10$.

Table 5. Results for the final SARIMA model for the organic waste time series in the Historic Center.

Final model: SARIMA $(1,1,2)(1,0,0)_{52}$			AIC $= 4.61$	$\hat{\sigma}^2 = 0.054$
	$\hat{\phi}_1$	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\nu}_1$
Estimate	–0.815	0.145	–0.822	0.280
Standard error	0.066	0.070	0.064	0.114

3.3 Forecasting Models Evaluation

Once established the forecasting models to the data provided, it is necessary to identify the models that best fit and (more accurately) forecast the urban waste time series. For comparative purposes, four evaluation measures are used: MSE and its corresponding on the same scale as the data, RMSE, MAPE, and MASE. In addition to these, measurements having been calculated for the testing time series, for the respective 15 observations in each time series, they are also determined for the training time series. The results can be found in Table 6.

The results in Table 6 show that the overall estimation and forecasting performance of the SARIMA models evaluated via MSE, RMSE, MAPE, and MASE are quite similar in both the training and testing periods (with the exception of the MASE measurement in each REC and ORG time series). From the analysis of Table 6, the model for recyclable waste time series in the PAYT implementation zone forecasts more accurately in the training period than in testing period. When considering the organic waste time series modeling process in the PAYT implementation area, it can be said that it has the best model adjusted to the testing time series than to the training time series.

It may be conclude that, in fact, a model that better fits the data does not necessarily forecast better, and the fit error measures should not be used to select a model for forecast.

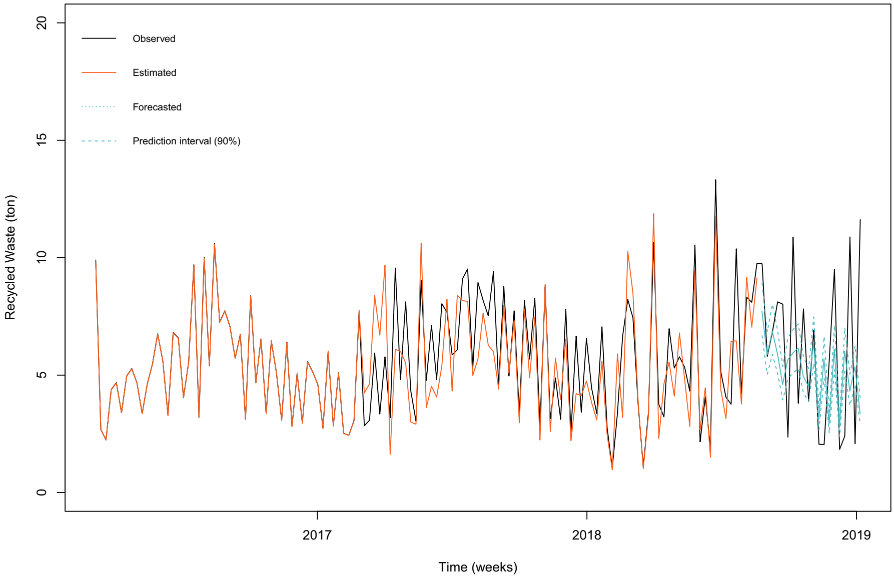

Fig. 4. Observed values, estimates and forecasts (with 90% confidence bounds) for organic waste time series using the final SARIMA model.

Table 6. Forecasting performance evaluation of the method for the training period and the testing period.

Time series	Training time series				Testing time series			
	MSE	RMSE	MAPE	MASE	MSE	RMSE	MAPE	MASE
REC	**3.202**	**1.789**	**23.473**	0.414	13.633	3.692	69.687	**0.363**
ORG	7.804	2.793	16.488	**0.579**	**1.654**	**1.286**	**7.965**	0.723

Once the forecasts' accuracy (punctual) is evaluated, it is essential to understand the effectiveness of the forecast intervals. Theoretically, the forecast intervals are calculated at 90% confidence, which means that 90% of the intervals must include the observed (real) observation. That is, it is considered that the most effective interval forecasts are those whose effective coverage rate is closer to 90%. Note that the forecast intervals are obtained based on the testing series, for each distinct series where, in this study, they contain only 15 observations and, therefore, the analysis of coverage rates must be taken care of.

In the two time series under study, REC waste and ORG waste, coverage rates of 40% and 60%, respectively, are calculated. It is notorious that the model formulated for the time series corresponding to the production of organic waste in the PAYT system's area of implementation presents better results (Table 7).

Table 7. 90% forecast confidence intervals (CI), forecasts, and observed value by using the final SARIMA for recyclable and organic waste.

REC			ORG		
Forecast CI	Forecast	Observed value	Forecast CI	Forecast	Observed value
(2.300 ; 6.667)	4.484	2.360	(12.376 ; 14.463)	13.419	13.880
(4.357 ; 8.725)	6.541	10.880	(13.072 ; 15.159)	14.115	11.520
(3.757 ; 8.124)	5.940	3.800	(12.782 ; 14.869)	13.826	15.440
(2.931 ; 7.299)	5.115	7.820	(12.972 ; 15.059)	14.015	13.420
(2.236 ; 6.604)	4.420	3.880	(12.666 ; 14.753)	13.710	12.160
(4.320 ; 8.688)	6.504	6.920	(12.649 ; 14.737)	13.693	12.740
(0.921 ; 5.288)	3.105	2.060	(12.384 ; 14.471)	13.428	13.460
(3.480 ; 7.848)	5.664	2.040	(12.249 ; 14.336)	13.293	12.180
(0.768 ; 5.135)	2.951	5.640	(10.967 ; 13.054)	12.011	14.500
(3.982 ; 8.350)	6.166	9.500	(12.303 ; 14.390)	13.347	14.240
(0.588 ; 4.956)	2.772	1.840	(12.181 ; 14.268)	13.225	12.580
(3.861 ; 8.228)	6.044	2.400	(14.008 ; 16.095)	15.052	13.540
(2.092 ; 6.460)	4.276	10.880	(12.605 ; 14.693)	13.649	13.020
(3.216 ; 7.584)	5.400	2.080	(12.796 ; 14.883)	13.839	13.340
(1.224 ; 5.591)	3.408	11.620	(12.851 ; 14.939)	13.895	13.780

In general, we find that both SARIMA models have the capability to estimate and forecast fairly well the behavior and seasonal fluctuations of the waste time series.

4 Conclusions

Due to economic and social development in general and to population growth, the amount of waste, particularly urban waste, has been significantly increasing in recent years. It is one of the major problems both at a national and global level, and action is urgently needed to ensure that waste is recovered and its volume reduced. Therefore, accurate (recyclable and organic) waste volume can have a great impact on effective local Urban Waste Management (UWM). Both established SARIMA models in this study have capabilities to enhance forecasting accuracy. Though forecasts are never totally accurate, they are the essential starting point of decision-making for any organization. Determining the expected values well ahead in time helps in fulfilling solid urban waste management orders in a municipal company as VITRUS AMBIENTE. This study will serve to draw preliminary conclusions, in order to improve management of the new PAYT system.

References

1. Reichenbach, J.: Status and prospects of pay-as-you-throw in Europe - a review of pilot research and implementation studies. Waste Manage. **28**(12), 2809–2814 (2008)
2. Navarro-Esbrí, J., Diamadopoulos, E., Ginestar, D.: Time series analysis and forecasting techniques for municipal solid waste management. Resour. Conserv. Recycl. **35**, 201–214 (2002)
3. Rimaitytė, I., Ruzgas, T., Denafas, G., Racys, V., Martuzevicius, D.: Application and evaluation of forecasting methods for municipal solid waste generation in an eastern-European city. Waste Manage. Res. **30**(1), 89–98 (2012)
4. Box, G., Jenkins, G., Reinsel, G.: Time Series Analysis, 4th edn. Wiley, Hoboken (2008)
5. Hyndman, R.J., Athanasopoulos, G.: Forecasting: Principles and Practice. Online Open-Acess Texbooks. http://otexts.com/fpp. Accessed 5 Oct 2020
6. Hyndman, R.J.: Forecast: Forecasting Functions for Time Series. R Package Version 4.06 http://cran.rstudio.com. Accessed 5 Oct 2020
7. Shumway, R.H., Stoffer, D.F.: Time series analysis and its applications: with R examples. Springer, New York (2017)