ORIGINAL ARTICLE

Evaluation of time-series models for forecasting demand for emergency health care services

José DÍAZ-HIERRO[1], José Jesús MARTÍN MARTÍN[2], Ángel VILCHES ARENAS[3], María Puerto LÓPEZ DEL AMO GONZÁLEZ[2], José María PATÓN ARÉVALO[1], Clara VARO GONZÁLEZ[4]

[1]Empresa Pública de Emergencias Sanitarias, Málaga, Spain. [2]Universidad de Granada, España. [3]Universidad de Sevilla, Spain. [4]Investigador independiente, Spain.

CORRESPONDENCE:
José Díaz Hierro
Avda. Benyamina
Urb. Playamar, Torre 12, 2ºA
29620 Torremolinos.
Málaga, Spain
E-mail: jdiaz@epes.es

RECEIVED:
15-12-2011

ACCEPTED:
10-2-2012

CONFLICT OF INTEREST:
None

Objective: To evaluate a combined set of 6 time-series models for improving the management of short-term calls for emergency health services.
Methods: The demand for emergency health services in the province of Malaga was analyzed between 2004 and 2008. Using standard software, we constructed and evaluated 3 decomposition models and 3 econometric models. The models considered summer months and atypical values, influenza cases, and number of overnight admissions as the exogenous inputs. We compared the models using the usual econometric tests, such as the root mean square error (RMSE), the mean absolute percentage error (MAPE), and the maximum absolute percentage error (MaxAPE) among others.
Results: The models had MAPEs under 5%. Autoregressive integrated moving average (ARIMA) modeling with intervention had the lowest RMSE. Harmonic analysis had the smallest difference between the MAPE and MaxAPE. In the validation phase, ARIMA with intervention had the poorest fit, and harmonic analysis and ARIMA with exogenous input had the best fits.
Conclusions: A forecast of the demand for emergency calls can be generated using 2 models simultaneously to improve short-term planning. Decomposition models and ARIMA with intervention warn of unexpected changes, whereas ARIMA or other models with exogenous inputs and harmony component analysis can introduce alternative planning scenarios, improve our understanding of demand, and facilitate decision-making. Implementing these models with standard software decreases the cost of this approach in emergency services. [Emergencias 2012;24:181-188]

Key words: Emergency care, prehospital. Forecasting. Trends. Organization and administration.

Introduction

Call centers for health emergencies have become the third gateway to health systems, which influences primary and specialty care. Emergency calls can be resolved with health advice or mobilization of a medical resource. Managing an emergency call center involves fast response and balancing costs. The workload is determined by time to treatment and the volume of demand. This study focuses on forecasting such demand. Improved efficiency in the management of emergency call centers acquires special significance in periods of economic crisis such as the current one. The instruments used to forecast calls for help are a key element for sustained improvement and quality service.

In academic circles, the analysis of time series[1-4] has been the focus of attention mainly because it is particularly suitable for short-term forecasts, allowing the detection of behavioral changes in the series which in turn allows rapid intervention[5]. There are two relevant approaches: the classical methods of analysis based on decomposition of the series, and econometric models that can incorporate exogenous variables which enable the study of causality[2-4,6-12]. The comparative analysis of different models, although somewhat limited[13-15],

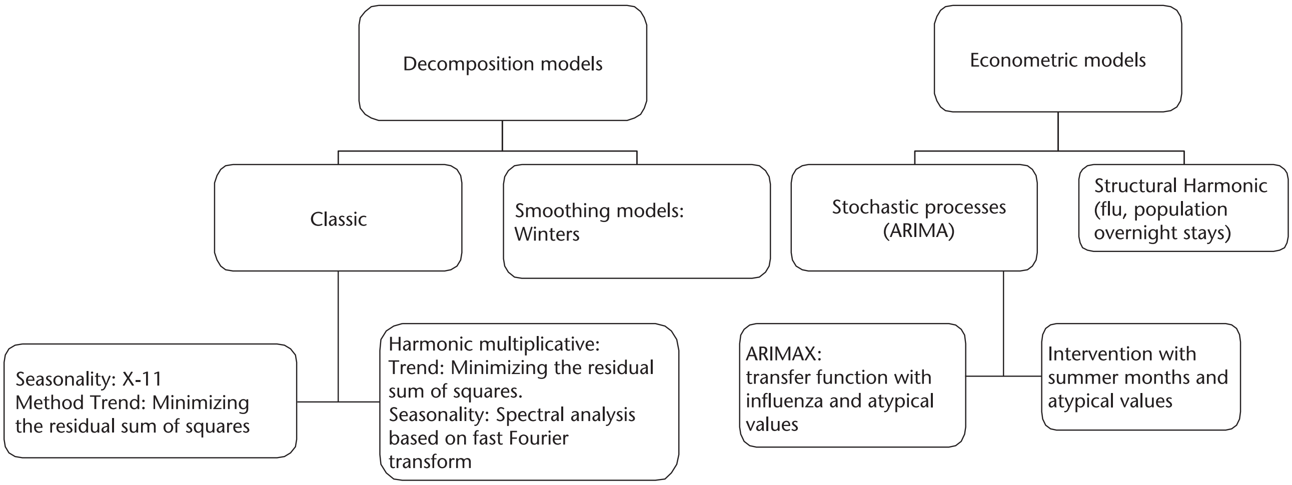

Figure 1. Time series models. Source: authors.

suggests that greater complexity is not necessarily associated with significant improvements, and indicates the usefulness of incorporating exogenous variables to adjust forecasts.

The aim of this study was to evaluate and compare the models that best predict the number and type of calls to an emergency center in Malaga and assess the potential synergistic effects of combined use. This is the first study to combine 6 methods of time series, including exogenous variables and using standard software, which analyzes the comparative advantages of each method, with the added value of assessing their combined use.

Method

We performed a descriptive observational study of the time series of calls received monthly from January 2004 to December 2008 at the Public Health Emergency call center (EPES in Spanish) in Malaga. The study period was 2004-2007, and 2008 as the validation period to assess the predictive power of the models. The study setting was out-of-hospital emergency attention in the province of Malaga, with a population of 1,600,000 inhabitants. The sources of information used were the public health emergencies information system (EPES) and, from the National Institute of Statistics, estimates of the population of Malaga[16], the overnight stays register[17] and the notifiable diseases register[18].

The selection of methods combined two criteria: the adequacy of data structure for the models and the existence of standard software for application. The 6 models chosen were: a harmonic multiplicative model, X-11 method of seasonal adjustment, Winters smoothing, autoregressive integrated moving average (ARIMA) by intervention analysis of summer months and atypical values, ARIMA with the transfer function of confirmed cases of influenza (ARIMAX), and a harmonic structure model, with population, influenza cases and number of overnight admissions as exogenous variables (Figure 1). The techniques and statistics used for each of the models are presented in Table 1[19-23].

The models were compared using stationay R^2, the root mean square error (RMSE), the mean absolute percent error (MAPE), which must be less than 5% to adjust the forecast to the budgetary objectives, the maximum absolute mean percent error (MaxAPE)[24], and standard Bayesian Information Criteria (BIC)[25].

In hypothesis testing, differences with a p value of < 0.05 were considered statistically significant, with 95% confidence intervals (CI). The analysis was conducted with the statistical packages SPSS v.18 and Eviews 5.1.

Results

The number of calls received showed a continuous increase during the study period, from 4% in 2004 to 13% in 2007, with a total of 572,674. We observed a trend to both growth and seasonality, with peaks in the demand for December-January and July-August each year.

We identified a dependency structure for close values, decreasing for distant values, and six-month seasonality (1/0.16), which can be explained by the double component seasonality: the influenza effect and the vacation effect. The great-

Table 1. Techniques and statistical tests applied

<table>
<tr><th rowspan="3">Models</th><th rowspan="3">Trend</th><th rowspan="3">Seasonality</th><th colspan="5">General tests</th><th rowspan="3">Specific tests</th><th rowspan="3">Parsimony</th><th rowspan="3">Goodness</th></tr>
<tr><th colspan="3">Residuals test</th><th colspan="2">Significance parameters</th></tr>
<tr><th>Normality</th><th>Heterocedasticity</th><th>Independence</th><th>Individual</th><th>Joint</th></tr>
<tr><td>General analyses</td><td>Linear and quadratic function fit</td><td>Correlograms (ACF & ACFP). Tukey-Hamming smoothing windowed quadratic periodogram</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>Multiplicative Harmonic</td><td rowspan="2">Ordinary quadratic minimum fit</td><td>Fast Fourier transform with 3 harmonics: 16 maximums and 1/4 and 1/8 of each wave.</td><td rowspan="5">Shapiro-Wilks</td><td rowspan="6">White</td><td rowspan="5">Ljung-Box</td><td>T</td><td>F</td><td></td><td rowspan="6">BIC criterion</td><td rowspan="6">Seasonal R^2, RMSE, MAPE, MaxAPE</td></tr>
<tr><td>X-11 Method</td><td>General indices of seasonal variation by X-11 method</td><td rowspan="2">T</td><td rowspan="2">F</td><td rowspan="2"></td></tr>
<tr><td>Winters smoothing</td><td colspan="2">Level, trend and seasonality fit; Winters equations</td></tr>
<tr><td>Arima intervention</td><td colspan="2">Box-Jenkins method with intervention analysis</td><td>T</td><td>F</td><td>Augmented Dickey-Fuller test</td></tr>
<tr><td>Arimax</td><td colspan="2">Box-Jenkins method with intervention analysis and pre-whitening transfer function</td><td>T</td><td>F</td><td>Augmented Dickey-Fuller test
Grange's causality test</td></tr>
<tr><td>Structural harmonic</td><td>Stepwise multiple regression</td><td>Fast Fourier transform with 3 harmonics</td><td>Jarque-Bera</td><td>Ljun-Box Durbin-Watson, Breusch-Godfrey</td><td>T</td><td>F</td><td>y. Multicolinearit Variance inflation factor. Condition index and Farrar-Glauber. test. Chow's structural stability test and CUSUM test. Functional specification. Ramsey's RESET test. Exogenicity. Hausman contrast.</td></tr>
</table>

The additive or multiplicative hypothesis of components was applied by calculating the Pearson coefficient of variation of standard transformations applied to the series. Source: authors. T: Student's t test, F: ANOVA test.

est explanatory range of variability occurred in the frequency range 0.1 to 0.2, with 0.1 being very close to a seasonal component of one month. Annex I shows estimates generated by the 6 models and Figure 2 shows the adjusted models.

In the multiplicative harmonic model the adjustment line stayed close to the real values for 2008 with an estimated total number of calls 2% lower than the real number.

A similar tendency was observed in the X-11 seasonal adjustment model, and the line of observed values was near the lower limit of the CI in 2008 (Figure 2), showing monthly seasonal variation indexes greater than 1 month in January, March, July, August and December and 5% higher estimate of total number of calls than the real number of calls received.

In the Winters smoothing model, the line of values observed in the first 6 months was slightly below the lower limit of the CI for the year 2008, with 7% higher estimate of the total number of calls than the real number of calls received.

In the ARIMA model with intervention analysis, May 2005 showed an increase in level of the se-

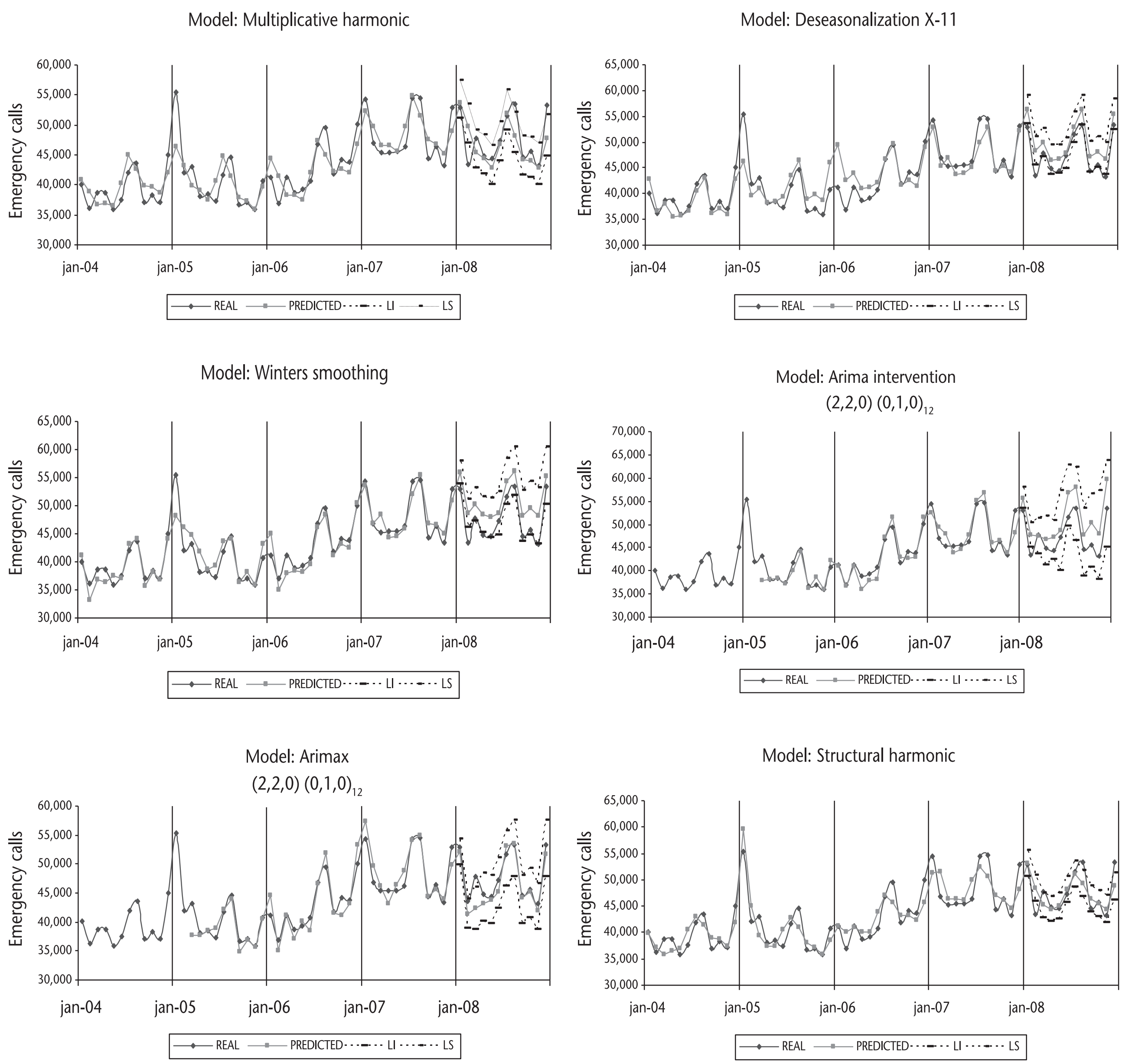

Figure 2. Predictions and fit of the 6 models regarding emergency calls in Malaga during the period 2004-2008. LL: lower limit, UL: upper limit. Source: the authors, using internal information of the "Empresa Pública de Emergencias Sanitarias".

ries which affected the nearest subsequent values and then decreased. The month of January 2006 showed a value below the mean for the same period in subsequent years.

As from March 2006 there was a sustained increase in the level of the series and the values were within the CI. In 2008, the line of observed values was within the limits of CI, but predicted values exceeded real values by 7.21%.

In the ARIMAX model, the number of calls received increased due to reported cases of influenza in the previous and current months, as well as cases reported in the same month of the year before. May 2005 showed an increase in the level of the series which affected the closest subsequent values and then decreased. As shown in Figure 2, the line of observed values was within the limits of the CI in 2008, with real values being higher than predicted values. The predicted total number of calls was 2.62% lower than the real number.

The structural harmonic multiple regression model showed a correlation coefficient of 0.78 between the variables. The goodness of fit showed a corrected R2 of 0.60. As shown in Figure 2, the value for January 2005 (outbreak of epidemic influenza) was adequately predicted. In

Table 2. Comparison of the goodness of fit of the model predictions regarding emergency calls in Malaga in 2004-2007 and in the validation period (2008)

Comparison of models						
Statistics	**Multiplicative Harmonic**	**Deseasonalization X-11**	**Winter's smoothing**	**ARIMA intervention**	**ARIMAX (2,2,0) $(0,1,0)_{12}$**	**Structural Harmonic**
Seasonal R^2	0.73	0.73	0.73	0.84	0.83	0.73
RMSE	2.633.47	2.709.37	2.058.76	1.988.46	2.128.89	2.393.98
MAPE	4.8%	4.6%	3.5%	3.1%	3.2%	4.4%
MaxAPE	16.4%	20.2%	12.8%	12.2%	12.8%	9.9%
Standardized BIC			15.5	15.7	15.9	
Ljung-Box	Independence	Independence	Independence	Independence	Independence	Independence
Values in the validation period (2008)						
Statistics	**Multiplicative Harmonic**	**Deseasonalization X-11**	**Winter's smoothing**	**ARIMA intervention**	**ARIMAX (2,2,0) $(0,1,0)_{12}$**	**Structural Harmonic**
Seasonal R^2	0.72	0.74	0.7	0.48	0.84	0.74
RMSE	3,039.17	2,673.34	3,403.00	3,643.39	2,024.78	2,503.27
MAPE	4.4%	5.2%	6.9%	6.9%	3.2%	3.8%
MaxAPE	14.2%	11.1%	11.9%	11.9%	11.4%	11.3%
Standardized BIC			16.2	16.9	15.5	
Annual estimate error	1.9%	–5.1%	–6.8%	–7.2%	2.6%	0.5%

Source: the authors, using internal information of the "Empresa Pública de Emergencias Sanitarias".

2008, the line of the observed values was within the limits of the CI, except for February, March, August and December. The predicted total number of calls was 0.5% less than the real number.

Synthetic comparison models

Table 2 compares the goodness of fit of the 6 prediction methods, and in all cases shows a MAPE less than 5%. In the estimation phase, the model with the lowest RMSE was ARIMA with intervention, and that with least difference between MAPE and MaxAPE was the structural harmonic model.

In the validation phase, the ARIMA model with intervention presented the worst fit, compared to the ARIMAX model that presented the best results. The models based on historical results (ARIMA, Winters, X-11) predicted higher values than those observed, which was to be expected given the annual systematic growth.

The multiplicative harmonic model showed the same trend as the time series models. However, the influence of the seasonal component decreased over the years and the estimate was lower than the observed values (1.95%).

The ARIMAX and harmonic structural models with exogenous variables showed intermediate results in terms of RMSE, but they improved in the validation period. Their error in annual estimates were 2.62% and 0.51% respectively, and they showed the best results in RMSE.

The ARIMAX model presented the best results. The historical influence of the series along with the decline in reported cases of influenza led to a decreased prediction of the number of calls, which proved greater than that observed.

Discussion

This study evaluated the combined use of 6 time series methods, developed with standard software, to improve the predicted demand for health emergencies. Some of their results are consistent with the available empirical evidence; on the one hand, greater complexity is not necessarily associated with improved predictive capacity, and on the other, the incorporation of exogenous variables is important to improve the results[5,15].

There was considerable variability in the predictions made by the models; the X-11, Winters smoothing and ARIMA with intervention, predicted more calls than were observed, while the multiplicative harmonic, the ARIMAX and structural harmonic predicted fewer calls. The former are useful as alert models of unusual behavior and the latter provide greater understanding of the variations in demand. For example, the ARIMAX model provided a significant association between number of calls and reported cases of influenza but not with population and overnight admissions, possibly due to the autoregressive component, which captures developments in the same sense of these two variables.

The results obtained by the structural harmonic model showed that people with influenza were responsible for the highest number of calls to emergency centers and people admitted overnight the lowest.

The approach used allowed the inclusion of other exogenous variables (level of pollution, rates of traffic accidents, cardiovascular disease incidence rate, temperature, etc.) that could increase the explanatory power of the structural harmonic model and ARIMAX[26].

The joint vision of the models offered a range of estimated annual emergency service calls in 2008, from 557,672 to 613,960, when in reality there were 572,674 calls. This enriches the information on considering scenarios with potential decreases, a fact confirmed in 2008, which allows planning alternative courses of action for the available resources and establishes predictions that incorporate knowledge of alternative quantitative models.

The main limitation of the present study is the uncertain extrapolation of these results to other geographical areas which may be more affected by the exogenous variables mentioned above.

However, application of the models presented here is still appropriate and recommended for any health emergency call center. In conclusion, the combined use of different methods of analysis generates synergies in knowledge of scenarios with emergency service demand, contributing to greater efficiency in the short term planning of human resources, reducing costs and simultaneously improving quality. The use of standard software lowers the costs of learning and allows stable periodic application in emergency centers.

References

1 Mandelbaum A. Call Centers. Research Bibliography with Abstracts. Version 7, May, 2006. [revista electrónica]. (Consultado 10 Diciembre 2011) Disponible en: http://ie.technion.ac.il/serveng/References/US7_CC_avi.pdf.

2 Baker JR, Fitzpatrick KE. Determination of an optimal forecast model for ambulance demand using goal programming. J Oper Res Soc. 1986;37:1047-59.

3 Mabert, V. Short interval forecasting of emergency phone call (911) work loads. J Oper Manag. 1985;5:259-71.

4 Channouf N, L'Ecuyer P, Ingolfsson A, Avramidis. A. The application of forecasting techniques to modeling emergency medical system calls in Calgary, Alberta. Health Care Manage Sci. 2007;10:25-45.

5 Peña D. Análisis de series temporales. Madrid: Alianza S.A., 2010.

6 Kamenetsky R, Shuman L, Wolfe H. Estimating need and demand for prehospital care. Oper Res. 1982;30:1148-67.

7 McConnell CE, Wilson RW. The demand for prehospital emergency services in an aging society. Soc Sci Med. 1998;46:1027-31.

8 Matteson D, McLean M, Woodard D, Henderson SG. Forecasting Emergency Medical Service Call Arrival Rates. Ann Appl Stat. 2011;5:1379-406.

9 Zhu Z, McKnew MA, Lee J. Effects of time-varied arrival rates: an investigation in emergency ambulance service systems. In: Proceedings of the 1992 winter simulation conference. Piscataway: IEEE Press; 1992. pp. 1180.

10 Tych W, Pedregal DJ, Young PC, Davies J. An unobserved component model for multi-rate forecasting of telephone call demand: the design of a forecasting support system. Int J Forecast. 2002;18:673-95.

11 Weinberg J, Brown LD, Stroud R. Bayesian forecasting of an inhomonegeous poisson process with applications to call center data. J Am Stat Assoc. 2007;102:1185-99.

12 Shen H, Huang JZ. Interday forecasting and intraday updating of call centers arrivals. Manufacturing and service operations managements. 2008;10:391-410.

13 Millán-Ruiz D, Pacheco J, Hidalgo JI, Vélez JL. Forecasting in a Multiskill Call Centre. ICAISC'10 Proceedings of the 10th international conference on Artificial intelligence and soft computing: Part II. Heidelberg: Springer-Verlag Berlin; 2010.

14 Setzler H, Saydam C, Park S. EMS Call Volume Predictions: A Comparative Study. Comput Oper Res. 2009;36:1843-51.

15 Taylor J. A comparison of univariate time series methods for forecasting intradia arrivals at a call center. Manage Sci. 2008;54:253-65.

16 Instituto Nacional de Estadística (2004 a 2008). Cifras oficiales de población: Revisión del Padrón Municipal. Madrid: Instituto Nacional de Estadística. (Consultado 11 Diciembre 2012). Disponible en: http://www.ine.es/jaxi/menu.do?type=pcaxis&path=%2Ft20%2Fe260&file=inebase&L=0

17 Instituto Nacional de Estadística (2004 a 2008). Encuesta de ocupación hotelera. Madrid: Instituto Nacional de Estadística. (Consultado 11 Diciembre 2012). Disponible en: http://www.ine.es/jaxi/menu.do?type=pcaxis&path=%2Ft11%2Fe162eoh&file=inebase&L=0

18 Instituto Nacional de Estadística (2004 a 2008). Enfermedades de Declaración Obligatoria. Madrid: Instituto Nacional de Estadística. (Consultado 11 Diciembre 2012). Disponible en: http://www.ine.es/jaxi/menu.do?type=pcaxis&path=/t15/p063&file=inebase&L=0

19 Casas Sánchez JM. Introducción a la estadística para economía. Madrid: Centro de Estudios Ramón Areces, S.A.; 1999.

20 Cáceres Hernández JJ, Martín Rodríguez G, Martín Álvarez FJ. Introducción al análisis univariante de series temporales económicas. Madrid: Delta Publicaciones Universitarias; 2008.

21 Jenkins G y Watts D. Spectral Analysis and Its Applications. Londres: Holden-Day; 1969.

22 Pérez López C. Econometría de las series temporales. Madrid: Pearson Educación S.A.; 2006.

23 Box G, Jenkins G. Time series analysis: Forecasting and control. San Francisco: Holden-Day; 1970.

24 Aguirre JA. Introducción al tratamiento de series temporales. Aplicación a las ciencias de la salud. Madrid: Ediciones Díaz de Santos, S.A.; 1994.

25 Hernández Alonso J. Análisis de series temporales económicas II. Madrid: ESIC Editorial; 2009.

26 Carracedo Martínez E, Sánchez C, Taracido M, Saenz M, Jato V, Figueiras A. Effect of short-term exposure to air pollution and pollen on medical emergency calls: a case-crossover study in Spain. Allergy. 2008;63:347-53.

Annex I. Equations of time series models adjusted for predicted emergency calls

1. **Multiplicative harmonic model: trend (Y_t) and seasonality (Y_t')**

$Y_t = 37{,}312.19 + 232.32 * t + e_t \qquad e_t = Y_t'$

$$Y_t' = 0.99 - 0.015 * \cos\left(\frac{2\pi}{48} * 16 * t\right) + 0.029 * \text{sen}\left(\frac{2\pi}{48} * 16 * t\right) + 0.023 * \cos\left(\frac{2\pi}{48} * 4 * t\right) +$$
$$+ 0.081 * \text{sen}\left(\frac{2\pi}{48} * 4 * t\right) - 0.019 * \cos\left(\frac{2\pi}{48} * 0.5 * t\right) + 0.066 * \text{sen}\left(\frac{2\pi}{48} * 0.5 * t\right)$$

2. **Deseasonalization X-11 model: trend (Y_t) and seasonality (indices)**

$Y_t = 36{,}833.45 + 245.07 * t$

SEASONAL VARIATION INDICES

JAN	1.15	JUL	1.05
FEB	0.98	AUG	1.11
MAR	1.01	SEP	0.93
APR	0.94	OCT	0.94
MAY	0.94	NOV	0.91
JUN	0.96	DEC	1.08

3. **Winters smoothing model:**

Parameters of exponential smoothing model

Model			Estimate	ET	t	Sig.
Calls-Model_1	Without transformation	Alpha (Level)	.707	.142	4.996	.000
		Gamma (Trend)	9.44E-006	.024	.000	1.000
		Delta (Season)	.001	.329	.003	.998

4. **ARIMA model with intervention analysis:** (Calls = C)

$$(1 + 1.021B + 0.611B^2)\nabla^2\nabla_{12}Ln(C)_t = a_t + 0.203\nabla^2\nabla_{12}I_{2005.5} - 0.185\nabla^2\nabla^{12}I_{2006.1} + 0.122^2{}_{12}E_{2006.3}$$

5. **ARIMA model with transfer function (ARIMAX):** (Calls = C and Flu = F)

$$(1 + 1.015B+0.586B^2)\nabla^2\nabla_{12}Ln(C)_t = a_t + (-0.66\text{-}0.66B\text{-}1.358B^{12}\nabla Ln(F)_t) + 0.201\nabla^2\nabla_{12}I_{2005.5}$$

6. **Structural Harmonic model:**

Calls = –59,869.6 + 0.67 * Population + 0.822 * Flu + 0.099 * Stays + e_t
Residual seasonality: $e_t = Y_t$

$$Y_t' = 507.95 - 138.22 * \cos\left(\frac{2\pi}{48} * 16 * t\right) + 749.65 * \text{sen}\left(\frac{2\pi}{48} * 16 * t\right) + 1{,}347.02 * \cos\left(\frac{2\pi}{48} * 4 * t\right) +$$
$$+ 1{,}822.56 * \text{sen}\left(\frac{2\pi}{48} * 4 * t\right) - 198.41 * \cos\left(\frac{2\pi}{48} * 0{,}5 * t\right) + 2{,}420.90 * \text{sen}\left(\frac{2\pi}{48} * 0.5 * t\right)$$

Evaluación de modelos de series temporales para la previsión de la demanda de emergencias sanitarias

Díaz-Hierro J, Martín Martín JJ, Vilches Arenas A, López del Amo González MP, Patón Arévalo JM, Varo González C

Objetivo: Evaluar las ventajas de la utilización conjunta de 6 modelos de series temporales para mejorar la gestión de la demanda a corto plazo de llamadas de emergencias sanitarias.
Método: Se ha analizado la demanda de emergencias sanitarias en el Servicio Provincial de Málaga entre 2004 y 2008 mediante 6 modelos desarrollados con *software* estándar, tres modelos de descomposición y tres econométricos, que consideran meses estivales y valores atípicos, casos de gripe y número de pernoctaciones como variables exógenas. La comparación de modelos se ha realizado mediante test econométricos habituales: la raíz cuadrada del error cuadrático medio (RMSE), el error absoluto porcentual medio (MAPE) y el máximo del error absoluto porcentual medio (MaxAPE) entre otros.
Resultados: Los modelos presentan un MAPE inferior al 5%. En la fase de estimación, el modelo ARIMA con intervención presenta la menor RMSE. El modelo estructural armónico obtiene el menor recorrido entre el MAPE y MaxAPE. En la fase de validación, el modelo ARIMA con intervención muestra el peor ajuste, y el modelo estructural armónico y ARIMAX los mejores.
Conclusiones: El empleo simultáneo de los modelos genera un intervalo de pronósticos de demanda de emergencias que mejora la planificación a corto plazo. Los modelos de descomposición y ARIMA con intervención alertan ante cambios inesperados, mientras que los modelos que incorporan variables exógenas, ARIMAX y estructural armónico, introducen escenarios alternativos de planificación, mejoran el conocimiento de la demanda y apoyan la toma de decisiones. Su implementación con *software* estándar disminuye los costes de aplicación en centros de emergencias. [Emergencias 2012;24:181-188]

Palabras clave: Emergencias sanitarias. Previsión. Tendencia. Organización y administración.