FORECASTING ^{7}BE CONCENTRATIONS IN SURFACE AIR USING TIME SERIES ANALYSIS

María del Carmen Bas, Ph.D* , Josefina Ortiz, Luisa Ballesteros, Sebastian Martorell, Ph.D

Laboratorio de Radiactividad Ambiental, Grupo MEDASEGI, Universitat Politecnica de Valencia, Spain

Abstract

^{7}Be is a cosmogenic radionuclide widely used as an atmospheric tracer, whose evaluation and forecasting can provide valuable information on changes in the atmospheric behavior. In this study, measurements of ^{7}Be concentrations were made each month during the period 2007-2015 from samples of atmospheric aerosols filtered from the air. The aim was to propose a Seasonal Autoregressive Integrated Moving Average (SARIMA) model to develop an explanatory and predictive model of ^{7}Be air concentrations. The Root Mean Square Error (RMSE) and the Adapted Mean Absolute Percentage Error (AMAPE) were selected to measure forecasting accuracy in identifying the best historical data time window to explain ^{7}Be concentrations. A measure based on the variance of forecast errors was calculated to determine the impact of the model uncertainty on forecasts. We concluded that the SARIMA method is a powerful explanatory and predictive technique for explaining ^{7}Be air concentrations in a long-term series of at least eight years of historical data to forecast ^{7}Be concentration trends up to one year in advance.

Keywords: ^{7}Be, time series, Forecasting, SARIMA model.

1. Introduction

^{7}Be is widely used as an atmospheric radiotracer due to its relatively short life ($T_{1/2} = 53.3$ days) and ease of measurement by γ-spectrometry, which provides important information on atmospheric air mass motions. A better understanding of its distribution would facilitate refinement and validation of global atmospheric circulation models (Dueñas et al. 2015). ^{7}Be forecasting can thus be adopted as a target value in analyzing fluctuations or deviations that could imply important atmospheric changes.

^{7}Be is a cosmogenic radionuclide formed by spallation reactions of light atmospheric nuclei (such as carbon, nitrogen and oxygen) with very high-energy protons and neutrons of the primary cosmic rays (Lal et al., 1958; Bruninx, 1961). Most ^{7}Be production (~70%) occurs in the stratosphere and the remainder (~30%) is produced in the troposphere, so that the ^{7}Be production rate is altitude-dependent (Feely et al., 1989; Baeza et al., 1996; Kotsopoulou & Ioannidou, 2012).

It is generally accepted that the ^{7}Be production rate depends on a number of atmospheric factors. Several studies have pointed out that the intensity of galactic cosmic rays in the Earth's orbit is affected by solar activity and the geomagnetic field, which is under constant cosmic ray bombardment from space (O'Brien, 1979; Vogt et al., 1990; Hötzl et al., 1991; Ioannidou & Papastefanou, 1994). In particular, an increase in solar activity and geomagnetic field reduce the galactic cosmic ray flux, which is followed by reduced ^{7}Be production.

In addition to the above-mentioned sources of variability, ^{7}Be concentrations in the lower layers of the atmosphere present temporal variations caused by solar radiation and meteorological parameters that can affect regional weather patterns (temperature, relative humidity, precipitations, wind speed and wind direction) (Feely et al., 1989; Baeza et al., 1996).

A recent study applied a decomposition of the ^{7}Be time series into a trend-cycle, a seasonal and an irregular component in order to separate the inter- and intra-annual patterns of ^{7}Be variability (Bas et al, 2016). The

results of this study showed the need to apply time series analysis to correlated data in order to separate the different sources of variability of ^{7}Be concentrations and to develop a forecasting model.

Many research studies have applied Multiple Linear Regression (MLR) analysis to develop a forecasting model for ^{7}Be air concentrations using atmospheric and meteorological variables as predictors (e.g.Azahra et al, 2004a, 2004b; Piñero-García et al., 2012, 2013; Dueñas et al., 1999, 2015). However, the disadvantage of the MLR technique is that it requires forecast meteorological parameters to predict^{7}Be air concentrations. Several authors recommend the use of time series modeling techniques when monitoring correlated process data (Alwan & Roberts 1988; Harris & Ross 1991; Wardell et al. 1994).

The objective of this study is to propose a seasonal Autoregressive Integrated Moving Average (SARIMA) model to develop a powerful explanatory and forecasting model of ^{7}Be air concentrations. For this, different data ranges of historical data are proposed to identify and validate the number of periods which best fit to ^{7}Be data. The optimal range of historical data is identified by means of the Root Mean Square Error (RMSE) and the Adapted Mean Absolute Percentage Error (AMAPE) as forecasting accuracy measures.The impact of the model uncertainty on forecastsis measured by the variance of the forecast errors.

2. Material and Methods

2.1. Study area and sampling

Airborne particulate samples were collected weekly on the campus of the Universitat Politècnica de Valencia from January 2007 to December 2015. Valencia is situated on the east coast of Spain (15m above sea level) in the western Mediterranean Basin (39°28′50″ N, 0°21′59″ W) and has a relatively dry subtropical Mediterranean climate with very mild winters and long hot summers. The sampling point was located approximately 2 km away from the coastline.

Aerosol samples were collected using Eberlyne G21DX and Saic AVS28A air samplers placed approximately 1 m above ground level. The aerosol particles were retained on a cellulose filter of 4.2 $x10^{-2}$ m effective diameter and 0.8 μm pore size. The filters were changed weekly and the average volume ranged from 300 to 400 m^3 per week. Each filter was put inside a plastic box and kept in a desiccator until it was measured.

2.2. ^{7}Be activity measurements

A monthly composite sample containing 4-5 filters was measured by γ-spectrometry to determine specific ^{7}Be activities using an HPGe detector (ORTEC Industries, USA) n-type with relative efficiency of 18% for 60Co gamma-ray. A certificated standard containing radionuclides with energies ranging from 59 to 1836.1 keV was used for preparing the calibrated filters, which were placed inside their plastic boxeson the top of the detector. The counting time was 60000s and the γ-line 477.7 KeV was used to calculate the activity. ORTEC Gamma-Vision software was used for acquisition and analysis. Concentration activities were corrected for the radioactive decay to the mid-collection period. The mean measured uncertainties (K=2) were around 10 %.

2.3. Statistical analysis

A time series is a sequence of observations taken sequentially in time and influenced by four separate components: (i) a *trend component* or long-term movement, (ii) a *cycle component* or fluctuations about the trend of greater or lesser regularity (iii) a *seasonal component* reflecting seasonality, and (iv) a random or *irregular component.* Several techniques are available for separating the trend component from oscillating fluctuations and random variations in a seasonal time series. One of the most frequently used methods for estimating a time series is the method of the Seasonal Autoregressive Integrated Moving Average (SARIMA) model (Box & Jenkins 1976).

The SARIMA model building process is designed to take advantage of the association in the sequentially lagged relationships that usually exist in data collected periodically. If the time series has more than one seasonal behavior, for instance two, s and s^1, the model will be composed of more parameters in order to model the other seasonal period, e.g. SARIMA$(p,d,q)(P,D,Q)_s(P^1,D^1,Q^1)_{s^1}$.

According to Bas et al. (2016), ^{7}Be concentration variability can be decomposed in terms of three main components: 1) trend-cycle, 2) seasonal and 3) irregular variations. Therefore, in principle, two periods could be considered, i.e. trend-cycle (11 years solar cycle) and seasonal (12 months).The trend-cycle component could not be totally modeled because the evaluated period was too short (2007-2015). In order to explicitly model the three parameters P^1, D^1, Q^1 of the trend-cycle component we needed more representative sample data. However, the model was able to capture part of the trend in the data with the d parameter, as explained below. We could therefore only model one seasonal period: annual periodicity.

A time series $\{z_t, t = 1, \dots, N\}$ is generated by a SARIMA$(p,d,q)(P,D,Q)_s$model for only one seasonal period if:

$$\phi_p(B)\Phi_P(B^s)(1-B)^d(1-B^s)^D z_t = \theta_q(B)\Theta_Q(B^s)a_t (eq.1)$$

where N is the number of observations; p, d, q, P, D, Q are integers; B is the lag operator (e.g. $(1-B)z_t = z_t - z_{t-1}$; $(1-B^{12})z_t = z_t - z_{t-12}$); s is the seasonal period length; d is the number of regular differences ($d \leq 2$); D is the number of seasonal differences, anda_t is the random event or estimated residual at time t, which is a usual Gaussian white noise process (WN).

$\phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p$; $(eq.2)$,is the regular autoregressive operator (AR) of order p,
$\theta_q(B) = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q$; $(eq.3)$,is the regular moving average operator (MA) of order q,
$\Phi_P(B^s) = 1 - \Phi_1 B^s - \Phi_2 B^{s2} - \cdots - \Phi_P B^{sP}$; $(eq.4)$, is the seasonal autoregressive operator (SAR) of order P,
$\Theta_Q(B^s) = 1 - \Theta_1 B^s - \Theta_2 B^{s2} - \cdots - \Theta_Q B^{sQ}$; $(eq.5)$,is the seasonal moving average operator (SMA) of order Q.

The parameters $(p,d,q)(P,D,Q)_s$try to model the time series behavior in the period evaluated. The first part of the SARIMA defined by the (p,d,q) parameters is related to the regular part of the time series and the $(P,D,Q)_s$ with seasonal variations.

Considering the annual periodicity observed in Bas et al. (2016) and the ^{7}Be values measured monthly, the parameters of the SARIMA model have the following interpretation:

p: determines the influence of the previous months of the same year on the forecasting month. It is known as the time series inertia. *p* is represented by $\phi_p(B)$ in Eq.1, which is defined in Eq. 2.
d: associated with the influence of the trend on the time series. The parameter *d* represents the times that the time series should be differenced in order to eliminate the trend. *d* is represented by $(1-B)^d$ in Eq. 1.
q: determines the influence of random events produced by external factors, which affected previous months of the same year, on the forecasting month. *q* is represented by $\theta_q(B)$in Eq.1, which is defined in Eq. 3.
P: determines the influence of the months of past years on the forecasting month. *P* is represented by $\Phi_P(B^s)$in Eq.1, which is defined in Eq. 4.
D: associated with the influence of the seasonal behavior on the time series. The parameter *D* represents the times that the time series should be differenced in order to eliminate the visual part of the seasonality. If the time series has a seasonal period, it is necessary to eliminate this seasonality in order to identify real relationships between the values of the time series. *D* is represented by $(1-B^s)^D$ in Eq. 1.

Q: determines the influence of random events produced by external factors, which affected months of past years, on the forecasting month. Q is represented by $\Theta_Q(B^s)$in Eq.1, which is defined in Eq. 5.

As reported by Box & Jenkins (1976) and Shumway & Stoffer (2006), the SARIMA model consists of three main steps:

Identification and estimation step

First, the periodogram technique was applied to identify the periodic cycle in the time series (Schuster, 1898). The periodogram plot should have clear peaks at points corresponding to the periodic cycle in the cyclic model.

The time series should then be differenced in order to be stationary in mean and variance (identifying d and D parameters). Differencing is a technique that can also be used to remove trends. Trends are usually detected by inspecting the plot of the ^{7}Be data over the period considered. However, they are also characterized by the autocorrelation function.

After differencing the time series, a tentative autoregressive moving average (ARMA) process is carried out based on the estimated autocorrelation function (ACF) and the estimated partial autocorrelation function (PACF). The shape of the ACF and PACF of the real time series is compared with the shape of the theoretical model to identify possible different parameters p, q, P and Q of the SARIMA model (Peña, 2010; Shumway & Stoffer, 2006). Having specified tentative models in the identification step, the parameters of the candidate models are estimated by a maximum likelihood function (Shine & Lee, 2000).

After trying several combinations for parameters p, q, P and Q, the best and mostparsimonious model was selected, considering the minimum AMAPE and RMSE (defined in the section on the Forecasting Step) for the forecasting data as accuracy measures of the predictive power.

Validation step

In this step, the below statistical testswere used to check the adequacy of the identified models for each time period. An essential part of the procedure is to examine the residuals of the SARIMA model, which should be considered, if the model is satisfactory, as White Noise (WN). We examine some simple ~~tools~~testsfor checking the hypothesis that the residuals are WN and the model is valid. If the fit model passes the following tests, it can be used to make a forecast.

- *t-ratio test* to evaluate the significance of the parameters estimated in each model. The parameters are considered significant with a 95% of confidence level if p-values<0.05.
- *Kolmogorov-Smirnov (K-S) test* applying Lilliefors correction of the residual series to check that the noise process is Gaussian. The residual series is Gaussian if p-values>0.05.
- *Q* Ljung-Box test* to check the condition that the residuals can be considered as a WN. The statistic proposed is

$$Q^* = n(n+2)\sum_{K=1}^{m}(n-k)^{-1}\, r_k(a)\,(eq.6)$$

where $r_k(\varepsilon)$ is the sample autocorrelation f order k of the residual, n is the length of residual series and m is the number of lags considered, $Q^* \approx \mathcal{X}^2_{m-n}$, $n = p+q+P+Q$. The model is considered valid if $P(\mathcal{X}^2(m-n) > Q^*) = p-value > 0.05$. In this study, Q* Ljung-Box statistic is calculated for a large m in each model, as suggested by Peña (2010).

Forecasting step

To assess the forecasting performance of different models each data set is divided into two samples for training and testing. This procedure is known as an out-of-sample technique, which means that the training data used in model fitting are different to the test sample (out-of-sample) used to evaluate the established model.

Several measurement statistics can be used to examine the forecast accuracy of different models. Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE) are the most frequently used criteria to evaluate the performance of the forecasting models. One of the disadvantages of the MAPE criteria is the adverse effect of small actual values, in which case MAPE criteria will contribute large terms to the MAPE coefficient, even if the difference between the actual and forecast values is small. It is therefore better to use an adapted MAPE (AMAPE), as defined in various studies (Tsay, 2005; Wu &Shahidehpour, 2010):

$$RMSE = \sqrt{\frac{\sum_{t=1}^{n}(\hat{z}_t - z_t)^2}{n}} (eq.7)$$

$$AMAPE = \frac{1}{n}\sum_{t=1}^{n}\left(\frac{|\hat{z}_t - z_t|}{\frac{1}{n}\sum_{t=1}^{n} z_t}\right) * 100\% \ (eq.8)$$

where t represents the time and n is the sample size for forecasts; $\hat{z}_t$ is the forecast at t from any mentioned model and z_t is the actual value at t. The RMSE statistic depends on the scale of the variables and measures the absolute errors. The AMPAE statistic measures the relative errors. The smaller the RMSE and AMAPE the better the accuracy of the model.

However, the model could give good results in the accuracy measurements for three-months-ahead, but poor results for one-year-ahead, for instance, which means the impact of the model uncertainty on forecasts needs to be measured (Chatfield, 2000). In this study, a measure was used based on the variance of forecast errors (difference between the actual and forecast value) to quantify this uncertainty. The smaller the variance the less uncertain the model or the more accurate the forecast results. The variance of error for sample size *n*for forecasts is defined as:

$$\sigma_{\varepsilon}^2 = \frac{1}{n}\sum_{t=1}^{n}\left(\left[\frac{|\hat{z}_t - z_t|}{\frac{1}{n}\sum_{t=1}^{n} z_t}\right] - \frac{1}{n}\sum_{t=1}^{n}\left(\frac{|\hat{z}_t - z_t|}{\frac{1}{n}\sum_{t=1}^{n} z_t}\right)\right)^2 (eq.9)$$

3. Results and discussion

Figure 1 shows the evolution of the actual ^{7}Be air concentrations and its measurement uncertainty during the entire study period from 2007 to 2015. A seasonal pattern with a sinusoidal trend can be clearly seen.

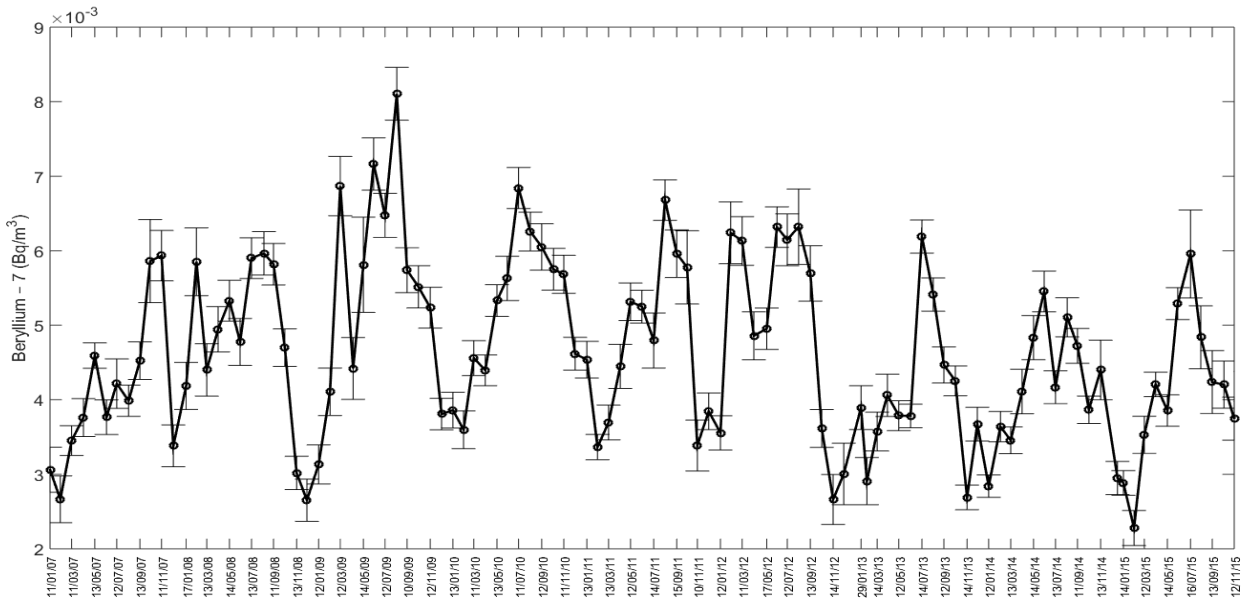

Fig 1. Temporal evolution of ^{7}Be air concentration over the period 2007-2015.

The steps involved in developing a SARIMA model are applied to different data ranges of historical data in order to identify and validate the number of periods which best fit actual ^{7}Be data.

First of all, results in Table 1 show the identification of a SARIMA model for each data range proposed, which were between two and eight years of historical data, considering 2007 as the initial year.

In the identification and estimation step, the periodogram analysis was applied to all the time series considered All these results identified a relevant peak of a period of 12 months (annual periodicity) ($s = 12$). For instance, Figure 2 shows the periodogram plot for the ^{7}Be time series over period 2007-2014.In this periodogram, one relevant peak is observed, corresponding to a cyclical period of 12 months ($1/0.083333$)., which indicates an annual periodicity.

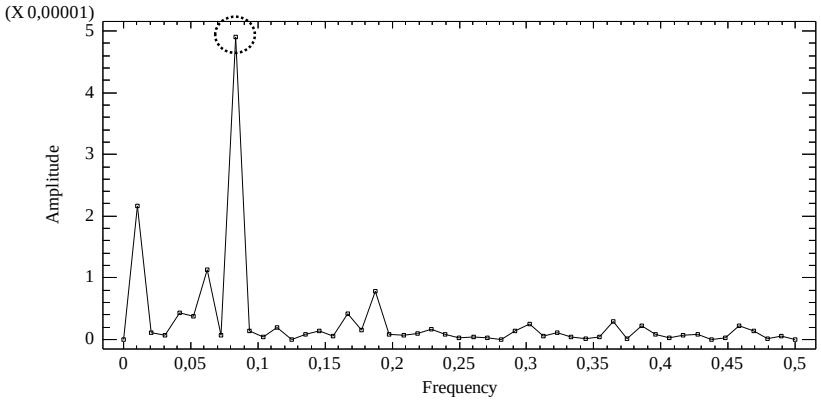

Fig 2. Periodogram of the ^{7}Be time series over the period 2007-2014

After differencing the time seriesand trying several combinations for parameters p, q, P and Q, based on the shape of the ACF and PACF, the best and mostparsimonious model was selected, considering the minimum

AMAPE and RMSE for the forecasting data as accuracy measures of the predictive power. Tables 1 reports the results of the parametersp, d, q, P, D and Q identifiedfor each selected model over the time period specified.

In the validation step, the t-ratio test, the Kolmogorov-Smirnov test and the Q* Ljung-Box test introduced in Section 2.3 were applied to validate the adequacy of the identified models for each time period. Tables 1show that, since they pass the above tests (p-values for *t-ratio test* are less than 0.05 and p-values for *K-S* and *Q* tests* are greater than 0.05) all the models selected to explain ^{7}Be air concentrations in a specific time period could be used to make forecasts.

The training data used for model fitting in the forecasting step was the data from the time period specified in Table 1, and the test sample used to evaluate the established model was the data of the year following the time period analyzed.

Figures 4-5show the RMSE, AMAPE for the different models proposed. Note that RMSE and AMAPE values were calculated considering the sample sizes for out-of-sample forecasts of 1,3,6,9, and 12 months. As can be observed, in general, the RMSE value for 1 month is very different to that of more than 1 month, suggesting that predictions for one month period are uncertain. The selection model criteria are thus based on forecasts of at least three months.

The models estimated with only a few years of historical data have higher values for the accuracy measurements in the forecasting sample sizes proposed (Figures 4-5), which means these models have a high degree of uncertainty and are therefore not useful for predicting ^{7}Be concentrations. As an exception, the model proposed with a time window of three years (2007-2009) provides good results in the RMSE and AMAPE coefficients for a forecasting sample of more than three months. However, the accuracy measurements for three-months-ahead are somewhat higher, indicating that the forecasting errors are not constant and the model is uncertain for short-term forecasting.

Figures 4-5show that the SARIMA$(0,1,1)x(1,1,3)_{12}$model provides good results for the RMSE and AMAPE accuracy measurements in a time window of seven (2007-2013) or eight (2007-2014) years. However, a time series with an 8-year time window appears to be better, because the measurement based on the variance of the forecast errors defined in eq. 9, is lower, σ_{ε}^{2} =0.0067 (Table 1), indicating that the model is less uncertain and the errors are more constant and stable across a one-year-ahead forecast. This result shows it is important to control the quality of the forecasting data. Note that the errors are minimal with a forecast window of six months.

In order to confirm these results and identify the time window that best fits ^{7}Be air concentrations, the analysis was repeated remaining the out-of sample fixed to the year 2015. Results are showed in Table 2.

The steps applied to identify and validate the best model in each period of time proposed in Table 2 are the same as the steps followed and explained above for Table 1.

Again, the RMSE, AMAPE and the low value in σ_{ε}^{2} (Figure 4-5 and Table 2 respectively) suggest that the model with 8 years of historical data (2007-2014) is the most suitable for monitoring and forecasting ^{7}Be data.

Model Identification and Estimation		Model Validation				Model Forecasting
Period	**Parameters**		**t-ratio statistic (p-value)**	**K-S statistic (p-value)**	**Q* statistic (p-value) m=nº of lags**	σ_ε^2
2007-2008	$(0,1,0)(1,0,1)_{12}$	Φ_1 Θ_1	6.74 (0.000001) 9.45 (<0.000001)	0.11073 (0.6571)	17.541 (0.0632) m=12	0.0493
2007-2009	$(0,1,1)(1,1,2)_{12}$	θ_1 Φ_1 Θ_1 Θ_2	3.20 (0.004637) -4.03 (0.000701) -2.28 (0.033760) 4.94 (0.000090)	0.092656 (0.8756)	10.468 (0.2337) m=12	0.0143
2007-2010	$(0,1,1)(0,1,2)_{12}$	θ_1 Θ_1 Θ_2	3.50 (0.001381) 13.16 (<0.000001) -7.06 (<0.000001)	0.06604 (0.9623)	29.061 (0.1125) m=24	0.0236
2007-2011	$(0,1,1)(2,1,2)_{12}$	θ_1 Φ_1 Φ_2 Θ_1 Θ_2	7.18 (<0.000001) -4.84 (0.000017) -18.06 (<0.000001) 7.13 (<0.000001) -3.55 (0.000951)	0.12184 (0.078)	18.998 (0.9549) m=36	0.0224
2007-2012	$(0,1,1)(2,1,2)_{12}$	θ_1 Φ_1 Φ_2 Θ_1 Θ_2	8.71 (<0.000001) -5.14 (0.000004) -14.91 (<0.000001) 10.30 (<0.000001) -5.08 (0.000005)	0.08602 (0.3415)	39.543 (0.622) m=48	0.0599
2007-2013	$(0,1,1)(1,1,3)_{12}$	θ_1 Φ_1 Θ_1 Θ_2 Θ_3	7.08 (<0.000001) -6.84 (<0.000001) 6.59 (<0.000001) 14.74 (<0.000001) -12.12 (<0.000001)	0.074491 (0.426)	48.092 (0.2742) m=48	0.0172
2007-2014	$(0,1,1)(1,1,3)_{12}$	θ_1 Φ_1 Θ_1 Θ_2 Θ_3	8.04 (<0.000001) -9.49 (<0.000001) 8.08 (<0.000001) 19.50 (<0.000001) -13.70 (<0.000001)	0.079983 (0.2119)	47.809 (0.2837) m=48	0.0067

Table 1. SARIMA models proposed for different time window data.

Model Identification and Estimation		Model Validation				Model Forecasting
Period	**Parameters**		**t-ratio statistic (p-value)**	**K-S statistic (p-value)**	**Q* statistic (p-value) m=nº of lags**	σ_ε^2
2013-2014	$(1,1,0)(1,0,0)_{12}$	ϕ_1 Φ_1	4.02 (0.000609) -30.87 (<0.000001)	0.17806 (0.05674)	9.2619 (0.5074) m=12	0.1290
2012-2014	$(0,1,1)(0,1,2)_{12}$	θ_1 Θ_1 Θ_2	6.13 (0.000005) 8.90 (<0.000001) -4.61 (0.000169)	0.088324 (0.913)	10.972 (0.277) m=12	0.0219
2011-2014	$(0,1,1)(0,1,3)_{12}$	θ_1 Θ_1 Θ_2 Θ_3	7.80 (<0.000001) 15.48 (<0.000001) -8.49 (<0.000001) 5.07 (0.000018)	0.10949 (0.3577)	27.277 (0.1276) m=24	0.0363
2010-2014	$(0,1,1)(2,1,2)_{12}$	θ_1 Φ_1 Φ_2 Θ_1 Θ_2	5.36 (0.000003) -9.53 (<0.000001) -14.02 (<0.000001) 7.35 (<0.000001) -4.22 (0.000124)	0.093043 (0.3909)	28.889 (0.575) m=36	0.0359
2009-2014	$(0,1,1)(2,1,2)_{12}$	θ_1 Φ_1 Φ_2 Θ_1 Θ_2	11.09 (<0.000001) -2.54 (0.013958) -7.54 (<0.000001) 13.04 (<0.000001) -6.67 (<0.000001)	0.068811 (0.696)	45.291 (0.3766) m=48	0.0167
2008-2014	$(0,1,1)(3,1,2)_{12}$	θ_1 Φ_1 Φ_2 Φ_3 Θ_1 Θ_2	9.79 (<0.000001) -4.29 (0.000059) -9.99 (<0.000001) -5.42 (0.000001) 13.77 (<0.000001) -6.92 (<0.000001)	0.069582 (0.5373)	34.671 (0.7816) m=48	0.0146
2007-2014	$(0,1,1)(1,1,3)_{12}$	θ_1 Φ_1 Θ_1 Θ_2 Θ_3	8.04 (<0.000001) -9.49 (<0.000001) 8.08 (<0.000001) 19.50 (<0.000001) -13.70 (<0.000001)	0.079983 (0.2119)	47.809 (0.2837) m=48	0.0067

Table 2. SARIMA models proposed for different time window data with common out-of-sample.

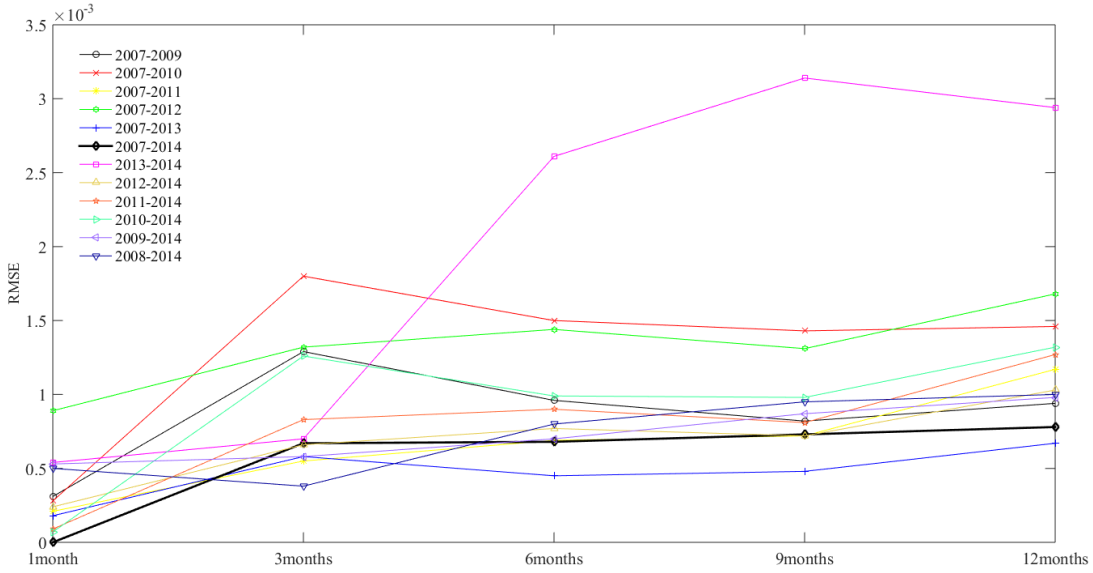

Fig 4. RMSE (eq. 7) for models evaluated in table 1 and 2.

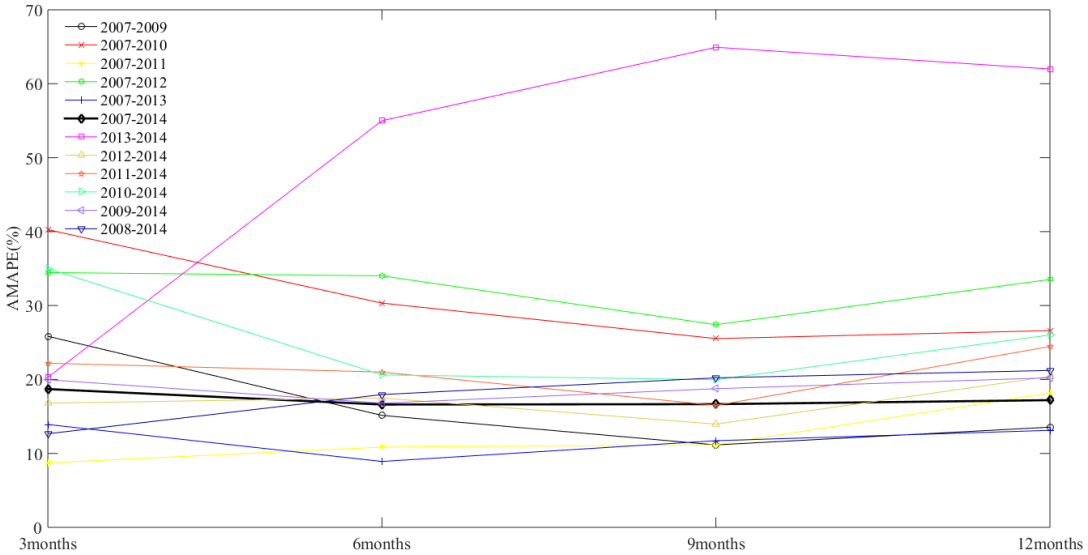

Fig 5. AMAPE (Eq. 8) for models evaluated in Tables 1 and 2.

The estimated and validated SARIMA$(0,1,1)x(1,1,3)_{12}$ model proposedfor a time window of eight years (2007-2014) is:

$$\Phi_1(B^{12})(1-B)^1(1-B^{12})^1 z_t = \theta_1(B)\Theta_3(B^{12})a_t \qquad (eq.10)$$

The coefficients are estimated by a maximum likelihood function, obtaining the following values:

$\theta_1 = 0.665$
$\Phi_1 = -0.814$
$\Theta_1 = 0.555$
$\Theta_2 = 0.932$
$\Theta_3 = -0.687$

Considering Equations 2 to 5 and the estimated parameters above, Equation 10 can be expressed as follows:

$$(1+0.814B^{12})(1-B)(1-B^{12})z_t = (1-0.665B)(1-0.555B^{12}-0.932B^{24}+0.687B^{36})a_t \quad (eq.11)$$

where $a_t \approx WN(0, 6.6\mathrm{E}-07)$ and B is the lag operator.

According to Bas et al. (2016), ^{7}Be concentration variability can be decomposed into terms of three main components: 1) trend-cycle, 2) seasonal and 3) irregular variations. Solar activity is a cosmogenic factor with a high influence on the trend-cycle component of ^{7}Be variability. Solar radiation, temperature and relative humidity are influential factors in seasonal ^{7}Be variations and, finally, precipitations and wind speed influence the irregular part of the ^{7}Be time series decomposition.

Considering the results obtained in Bas et al. (2016) and the model proposed in this paper (Eq. 11), the following relation between the SARIMA parameters and the atmospheric factors could be interpreted.

In the model proposed, the parameter *p* is zero; this result means that no significant influence of the ^{7}Be for the previous months of the same year was observed on the ^{7}Be for the forecasting month, which means that the time series has no inertia.

The coefficient $(1+0.814B^{12})$ is associated with the parameter *P=1*, which determines the influence of the months of previous years on the forecasting month; for instance, the forecasting value for ^{7}Be activity in January

2015 is influenced by the ^{7}Be activity observed in January 2014. The dependence observed between the seasonal observations could be influenced by solar radiation, temperature and relative humidity, which are regular atmospheric variables that affect seasonal ^{7}Be variations (Bas et al., 2016).

The coefficient $(1 - B)$ is associated with the parameter *d=1*, which recognize the presence of a trend in the time series. This trend could be affected by solar activity considering the results of Bas et al. (2016). The model proposed is able to detect a trend influenced by a part of the solar cycle, but is unable to model the cycle component due to solar activity because the evaluated period is too short.

The parameter *D=1* recognize the presence of an annual seasonality and is associated with the coefficient $(1 - B^{12})$ in the model. This coefficient eliminates the visual part of the seasonality in order to capture the real dependencies between the months in different years.

The coefficient $(1 - 0.665B)$ is associated with the parameter *q=1*. This coefficient identifies the influence of external factors, which affected the previous month, on the forecasting month. According to the results of Bas et al. (2006), these external factors could be precipitation and wind speed, among others, due to their irregular and random behavior. This result means that the ^{7}Be activity obtained in the forecasting months through the model is affected by the irregular factors that happened in the previous month.

Finally, the coefficient $(1 - 0.555B^{12} - 0.932B^{24} + 0.687B^{36})$ is associated with the parameter Q=3 and also captures the influence of random factors such as precipitation and wind speed, which affected months of the previous three years, on the forecasting month. In general, this parameter captures the long-term influence of external factors on the forecasting month.

Figure 6 shows the comparison between measured and forecast values using a SARIMA$(0,1,1)x(1,1,3)_{12}$ in a training sample 2007-2014.

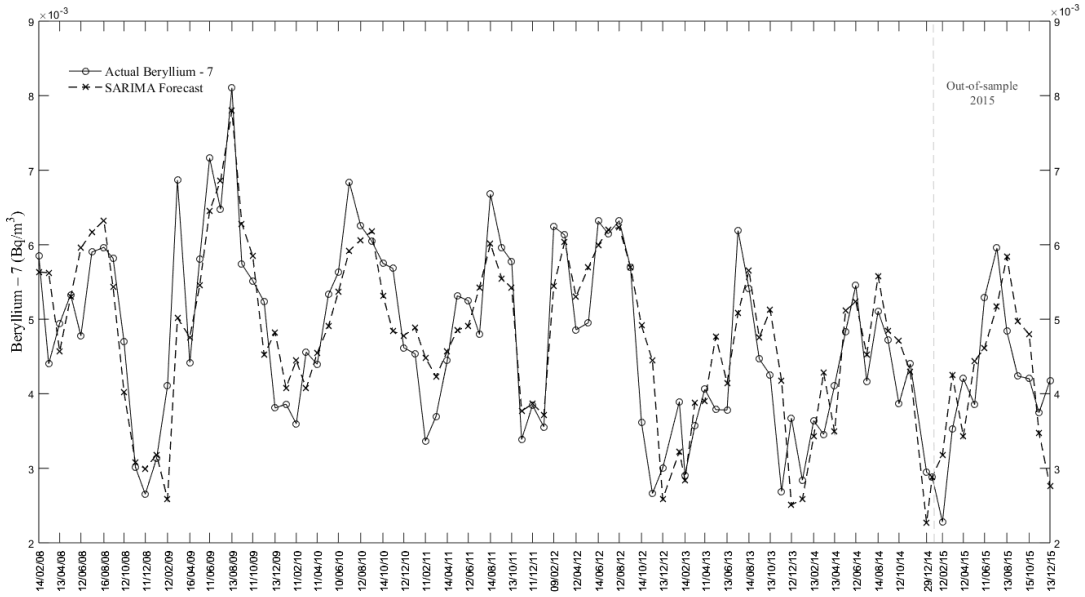

Fig 6. Comparison between measured and forecast (SARIMA) power

4. Conclusions

A Seasonal Autoregressive Integrated Moving Average (SARIMA) model was proposed in different year ranges to forecast ^{7}Be air concentrations in Valencia, in particular a SARIMA$(p,d,q)(P,D,Q)_s$. The forecasting results for the different models were compared and analyzed for the subsequent 12 months by out-of-sample tests.

The results show that the best time series models are based on a time window of at least eight years of data when forecasting ^{7}Be concentrations. Considering the forecasting power measured by the RMSE and AMAPE accuracy coefficients and the impact of the model uncertainty measured by the variance of the errors, a SARIMA$(0,1,1)x(1,1,3)_{12}$ model was proposed to best fit ^{7}Be concentration data for a time window of eight years (2007-2014). The prediction results for the out-of-sample year are appropriate and the errors observed are constant, with a minimum uncertainty of σ_ε^2 =0.0067. The results show also that the optimal forecasting time range is six months, since the errors are higher for longer prediction periods.

The time series models proposed in this paper have the advantage of not requiring any forecast meteorological parameters to develop the model, as is required for a Multiple Linear Regression (MLR) model. In this case, ^{7}Be forecasting can thus be adopted as a target value in analyzing deviations that could imply important atmospheric changes. The forecasting values obtained with SARIMA models do not explicitly capture an anomaly in specific atmospheric variables, which is the advantage of the MLR model. However, the SARIMA model can detect that a deviation occurred and was produced by an atmospheric factor. The SARIMA models combined with the influence of exogenous factors could cope with this problem and would be an interesting subject for future research. Despite this limitation, the model proposed in this study, in addition to the results of Bas et al. (2016), point to an interesting relationship between the model and the atmospheric parameters.

The availability of further data measurements will make it possible to adjust a time series with a wider time window period to submit the minimum period of years that best fit ^{7}Be air concentrations to a further analysis. For instance, the availability of more complete solar cycles could provide enough information to explicitly model the trend-cycle component. With regard to the application, one could envisage that it will be possible to develop the proposed forecasting models not only for ^{7}Be air concentrations, but also for monitoring and forecasting a range of different radionuclides.

5. Acknowledgements

This study has been supported partially by the REM program of the Nuclear Safety Council of Spain (SRA/2071/2015/227.06).

Bibliography

Alwan, L.C., Roberts, H.V.1988. Time series modelling for statistical process control. Journal of Business and Economic Statistics, 6, 87-95.

Azahra, M., López-Peñalver, J.J., Camacho García, C., González-Gómez, C., El Bardouni T., Boukhal, H. 2004a. Atmospheric concentrations of ^{7}Be and ^{210}Pb in Granada, Spain. Journal of Radioanalytical and Nuclear Chemistry, 261,401-405.

Azahra, M., González-Gómez, C., López-Peñalver, J.J., El Bardouni, T., Camacho García, A., Boukhal, H., El Moussaoui, F., Chakir, E., Erradi, L., Kamili, A., Sekaki, A. 2004b.The seasonal variations of ^{7}Be and ^{210}Pbconcentrations in air. Radiation Physics and Chemistry, 71, 789–790.

Bas, M.C., Ortiz, J., Ballesteros, L., Martorell, S. 2016. Analysis of the influence of solar activity and atmospheric factors on ^{7}Be air concentration by seasonal-trend decomposition. Atmospheric Environment, 145, 147-157.

Baeza, A., Del Río, L.M., Jiménez, A., Miró, C., Paniagua, J.M., Rufo, M., 1996. Analysis of the temporal evolution of atmospheric ^{7}Be as a vector of the behavior of other radionuclides in the atmosphere. Journal of Radioanalytical and Nuclear Chemistry, 207, 331-344.

Box, G.E.P., Jenkins, G.M.1976. Time series analysis: forecasting andcontrol. San Francisco: Holden Day.

Bruninx, E. 1961. High-energy nuclear reaction cross-sections. III Report CERN 64-17, Geneva, Switzerland.

Chatfield, C. 2000. Time-Series Forecasting. London, U.K.: Chapman & Hall/CRC.

Dueñas, C., Fernández, M.C., Cabello, M., Gordo, E., Liger, E., Cañete, S., Pérez, M. 2015.Study of the cosmogenic factors influence on temporal variation of ^{7}Be air concentration during the 23rd solar cycle in Málaga. Journal of Radioanalytical and Nuclear Chemistry, 303, 2151-2158.

Dueñas, C., Fernández, M.C., Liger, E., Carretero, J. 1999. Gross alpha, gross beta activities and ^{7}Be concentrations in surface air: analysis of their variations and prediction model. Atmospheric Environment, 33, 3705-3715.

Feely, H.W., Larsen, R.J., Sanderson, C.G. 1989. Factors that cause seasonal variations in Beryllium-7 concentrations in surface air. Journal of Environmental Radioactivity, 9, 223-249.

Harris, T.J., Ross, W.H. 1991. Statistical process control procedures for correlated observations. Canadian Journal of Chemical Engineering, 69, 48–57.

Hötzl, H., Rosner, G., Winkler, R., 1991. Correlation of ^{7}Be concentrations in surface air and precipitation with the solar cycle. Naturwissenschaften, 78, 215-217.

Ioannidou, A., Papastefanou, C., 1994. Atmospheric Beryllium-7 concentrations and sun spots. Nuclear Geophysics, 8, 539-543.

Kotsopoulou, E., Ioannidou A., 2012. ^{7}Be atmospheric concentration at mid latitudes (40°N) during a year of solar minimum. In Proceedings EPJ Web of Conferences. Vol 24. http://dx.doi.org/10.1051/epjconf/20122405005.

Lal, D., Malhotra, P.K., Peters, B., 1958. On the production of radioisotopes in the atmosphere by cosmic radiation and their application to meteorology. Journal of Atmospheric and Solar-Terrestrial Physics, 12, 306-328.

O'Brien, K., 1979. Secular variations in the production of cosmogenic isotopes in the earth's atmosphere. Journal of Geophysical Research, 84, 423-431.

Peña, D. (2010). Análisis de series temporales. Madrid: Alianza Editorial.

Piñero-García, F., Ferro-García, M.A., Azahra, M. 2012. ^{7}Be behaviour in the atmosphere of the city of Granada January 2005 to December 2009. Atmospheric Environment, 47, 84-91.

Piñero-García, F., Ferro-García, M.A. 2013. Evolution and solar modulation of ^{7}Be during the solar cycle 23. Journal of Radioanalytical and Nuclear Chemistry, 296, 1193-1204.

515 Schuster, A.1898. On the investigation of hidden periodicities with application to a supposed 26 day period of meteorological
516 phenomena, Terrestrial Magnetism, 3, 13-41.

517 Shine, D.W., Lee, J.H., 2000. Consistency of the maximum likelihood estimators for nonstationary ARIMA regressions with time
518 trends. Journal of Statistical Planning and Inference, 87, 55-68.

519 Shumway, R.H., Stoffer, D.S., 2006. Time series analysis and its applications: With R Exemples. Springer Texts in Statistics. New
520 York: Springer-Verlag.

521 Tsay, R. S. 2005. Analysis of Financial Time Series, 2nd ed. New York: Wiley.

522 Vogt, S., Herzog, G.F., Reedy, R.C., 1990. Cosmogenic nuclides in extraterrestrial materials. Reviews of Geophysics, 28, 253-275.

523 Wardell, D.G., Moskowitz, H., Plante, R.D.1994. Run length distributions of special-cause control charts for correlated processes.
524 Technometrics, 36, 3-17.

525 Wu, L., Shahidehpour, M. 2010. A Hybrid Model for Day-Ahead Price Forecasting. IEEE Transactions on power systems, 25, 1519-
526 1530.