Environmental Modelling & Software 51 (2014) 286–295

Contents lists available at ScienceDirect

Environmental Modelling & Software

journal homepage: www.elsevier.com/locate/envsoft

Modeling and forecasting daily average PM_{10} concentrations by a seasonal long-memory model with volatility

Valdério Anselmo Reisen [a,b,*], Alessandro José Queiroz Sarnaglia [a], Neyval Costa Reis Jr. [b], Céline Lévy-Leduc [c], Jane Méri Santos [b]

[a] Graduate Program in Statistics, Federal University of Minas Gerais, Brazil
[b] Graduate Program in Environmental Engineering, Federal University of Espírito Santo, Brazil
[c] AgroParisTech/INRA MIA 518, France

ARTICLE INFO

Article history:
Received 15 December 2012
Received in revised form
27 September 2013
Accepted 30 September 2013
Available online 7 November 2013

Keywords:
Fractional differencing
Long-memory
ARFIMA
Seasonality
Heteroscedasticity
PM_{10} contaminant

ABSTRACT

This paper considers the possibility that the daily average Particulate Matter (PM_{10}) concentration is a seasonal fractionally integrated process with time-dependent variance (volatility). In this context, one convenient extension is to consider the SARFIMA model (Reisen et al., 2006a,b) with GARCH type innovations. The model is theoretically justified and its usefulness is corroborated with the application to PM_{10} concentration in the city of Cariacica, ES (Brazil). The fractional estimates evidenced that the series is stationary in the mean level and it has long-memory phenomenon in the long-run and, also, in the seasonal periods. A non-constant variance property was also found in the data. These interesting features observed in the PM_{10} concentration supports the use of a more sophisticated time series model structure, that is, a model that encompasses both time series properties seasonal long-memory and conditional variance. The adjusted model well captured the dynamics in the series. The out-of-sample forecast intervals were improved by considering heteroscedastic errors and they were able to capture the periods of more volatility.

Software availability

All the model estimates were processed in the R environment (www.r-project.org). The subroutines and the data set are available upon request.

1. Introduction

The issue of airborne ambient Particulate Matter (PM) has become a well-recognized research topic in environmental sciences. Epidemiological studies have reported strong associations between PM_{10} concentrations (PM with an aerodynamic diameter of less than or equal to 10 μm) and several adverse health effects. Some of the most commonly health diseases are respiratory problems in children, death and increased hospital admissions for cardiopulmonary and respiratory conditions among others as found in, for example, Touloumi et al. (2004), Pérez et al. (2007), Zelm et al. (2008) and references therein.

In literature, several modeling strategies have been developed or optimized for the study and forecast of PM concentration in urban areas, such as Díaz Robles et al. (2008), Konovalov et al. (2009) and others. Among these modeling efforts, statistical models based on multiple regression (Stadlober et al., 2008) and time series tools, such as the Box–Jenkins time series Autoregressive Integrated Moving Average (ARIMA) model, have been widely used for this class of problems (Goyal et al., 2006; Liu, 2009).

Models which adequately describe the physical behavior of the data are essential for accurate forecasting in any area of application. In this paper, a Seasonal Autoregressive Fractionally Integrated Moving Average (SARFIMA) model with more than one fractional parameter and a non-constant conditional error variance (heteroscedastic errors) is used to illustrate how it can be useful to fit and forecast series with seasonality, volatility and long-range dependency (or long-memory) features. These time series phenomena are quite common characteristics found in data in many areas of interest. For example, Windsor and Toumi (2001) analyzed the variability of the pollutants ozone and PM with long-memory technique which was also the methodology applied by Baillie et al. (1996) to model and forecast temperature series. Karlaftis and Vlahogianni (2009) studied the memory and volatility properties in transportation time series Kumar and Ridder (2010)

* Corresponding author. Statistics Department, Federal University of Espírito Santo, 29075-910 Vitoria, ES, Brazil. Tel.: +55 2730192609.
E-mail address: valderioanselmoreisen@gmail.com (V.A. Reisen).

1364-8152/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.envsoft.2013.09.027

focused on modeling and forecasting ozone episodes through heteroscedastic processes (GARCH) associated with ARIMA model.

Roughly speaking, seasonality is a phenomenon where the observation in the instant t is highly correlated with the one in the time $t - s$. In this case, s is called season length. It is important to consider statistical tools which take into account the seasonality effect. However, some studies focusing on the forecast of daily PM_{10} concentrations do not regard for the seasonal influence of weather patterns (Goyal et al., 2006). Other studies, such as Stadlober et al. (2008) try to control the seasonal component by using dummy variables which is suitable just in the case when seasonality is present in the mean structure only.

Time series with volatility is characterized by a non-constant conditional variance, i.e., the error variance changes as a function of time. This fact contrasts with the usual assumption, namely the variance of the process is assumed to be constant. However, if the variance is time-varying, the forecast variance can be reduced by accommodating the conditional variance which will lead to more accurate forecast confidence intervals. A systematic structure for modeling volatility in a time series is the Autoregressive Conditional Heteroscedastic (ARCH) model proposed by Engle (1982). An extension of this model, the Generalized Autoregressive Conditional Heteroscedastic (GARCH), was proposed by Bollerslev (1986). See also Bollerslev et al. (1992) for a more complete review on this subject. Due to the high temporal variability of PM_{10} concentration, it is usually found to have a time-varying conditional variance (see Chelani and Devotta (2005) among others). Volatility models are popular tools in financial literature, however, only recently, these have caught the attention of many researchers interested in modeling time-varying variance in time series of environmental sciences, e.g. McAleer and Chan (2006).

Recently, time series analysis with long-term dependency have been studied by several authors in different areas of applications. In the time-domain, long-range dependency is usually characterized by a significant autocorrelation even for those observations separated by a relatively long time period. The ARFIMA model (Granger and Joyeux, 1980; Hosking, 1981) is a time series model that well accommodates the long-memory feature. As discussed in the next section, this model has the parameter d, which governs the memory of the process. Several estimation methods for the long-memory parameter have been proposed. The most popular semiparametric estimator is due to Geweke and Porter-Hudak (1983), Reisen (1994) among others. The usefulness of modeling time series with the long-memory characteristic by ARFIMA processes has been extensively studied, theoretically and empirically, in many areas, such as mathematics, economics among others. For a recent review of this subject, see Palma (2007). The characteristics of the long-memory parameter estimators have been extensively investigated under various model situations, such as the presence of non-Gaussian errors and outliers, e.g. Sena et al. (2006), Fajardo et al. (2009) among others.

However, in environmental science, more specifically, in the air pollution area, the use of the ARFIMA model has still not been well explored. Today, there is a lot of software that makes using this model less difficult in applied works. Due to the important model features of the ARFIMA process, this model is certain to motivate much research in the near future in the environmental science area. Iglesias et al. (2006) is an example of applied work with long-memory process in the air pollution area. The authors have investigated the use of an ARFIMA model to handle time series of $PM_{2.5}$, PM_{10} concentrations and other pollutants.

A natural extension of the ARFIMA model to accommodate seasonal features is the seasonal ARFIMA model. Since the early 90's, this model has caught the attention of researchers that are interested in studying long-memory time series with seasonal fractional parameters. Porter Hudak (1990) among others proposed the use of Geweke and Porter-Hudak (1983) method for the estimation of seasonal ARFIMA processes. A generalization of these seasonal long-memory models are the ARUMA and the GARMA models, which were originally proposed by Giraitis and Leipus (1995) and Woodward et al. (1998), respectively. A new class of models, in the frequency domain, called generalized exponential (GEXP) models for modeling seasonal long memory was recently proposed by Hsu and Tsai (2009). Reisen et al. (2006a,b) presented studies regarding the seasonal ARFIMA model, which is a particular case of the ARUMA/GARMA models, and suggested long-memory estimators. Empirical studies, performed by the authors, indicate the efficiency of the estimators when compared to other existing methods. Seasonality and long-memory properties have been explored theoretically and empirically by many authors see, for example, Reisen et al. (2006a,b), Arteche and Robinson (2000) among others.

For a series that presents seasonal long-memory features with conditional variance (or volatility), one convenient extension is to consider the SARFIMA model with GARCH type innovations. This model can provide a useful way of analyzing a process exhibiting seasonal long-memory with volatility. This is the main purpose of this paper, which proposes the use of a SARFIMA model with one non-seasonal and one seasonal fractional parameter and GARCH errors. The model is theoretically justified and its usefulness is corroborated with the application to PM_{10} ambient concentrations.

The rest of this paper is organized as follows. Section 2 introduces the model and discusses its properties. The section also summarizes the estimation method of the parameters. Section 3 deals with the analysis and modeling of the PM_{10} contaminant and forecasting issues. Some conclusions are given in Section 4.

2. The model and parameter estimation

A process $X_t \equiv \{X_t\}_{t\in\mathbb{Z}}$ is defined as a zero-mean SARFIMA$(p,d,q) \times (P,D,Q)_s$ model with non-seasonal orders p and q, seasonal orders P and Q, difference parameters d and D, and season length $s \in \mathbb{N}^* = \mathbb{N} - \{0\}$ if

$$U_t = \nabla^{\mathbf{d}} X_t \tag{1}$$

is a SARMA $(p,q) \times (P,Q)_s$ process. That is, the process $\{U_t\}_{t\in\mathbb{Z}}$ satisfies

$$\Phi(B^s)\phi(B)U_t = \Theta(B^s)\theta(B)\varepsilon_t, \tag{2}$$

where $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ is a white noise with $\mathbb{E}(\varepsilon_t) = 0$ and $\mathrm{Var}(\varepsilon_t) = \sigma_\varepsilon^2$ and B is the backward operator satisfying $BY_t = Y_{t-1}$ for any process $\{Y_t\}_{t\in\mathbb{Z}}$.

In (1), the operator $\nabla^{\mathbf{d}}$ is defined by:

$$\nabla^{\mathbf{d}} = (1-B)^d(1-B^s)^D, \tag{3}$$

where $\mathbf{d} = (d, D) \in \mathbb{R}^2$ is the memory vector parameter, d and D are the fractionally parameters at the zero (or long-run) and seasonal frequencies, respectively. Also, the fractional filters are

$$\left(1-B^k\right)^x = \sum_{j=0}^{\infty} \binom{x}{j} \left(-B^k\right)^j, k = 1, s, \quad \text{and} \quad x = d, D,$$

where

$$\binom{x}{j} = \frac{\Gamma(x+1)}{\Gamma(j+1)\Gamma(x-j+1)},$$

and $\Gamma(\cdot)$ is the well-known gamma function.