Mathematical Biosciences and Engineering

http://www.aimspress.com/journal/MBE

MBE, 18(1): 888–903.
DOI: 10.3934/mbe.2021047
Received: 05 October 2020
Accepted: 13 December 2020
Published: 31 December 2020

Research article

Detecting time-changes in PM_{10} during Covid pandemic by means of an Ornstein Uhlenbeck type process

Giuseppina Albano*

Dipartimento di Studi Politici e Sociali, Università degli Studi di Salerno, Via Giovanni Paolo II, 84084 Fisciano (SA), Italy

* **Correspondence:** Email: pialbano@unisa.it; Tel: +39089962645.

Abstract: Particulate matter with 10 micrometers or less in diameter (PM_{10}) from several italian cities is modeled by means of a non homogeneous Ornstein Uhlenbeck process. Such model includes two deterministic time dependent functions in the infinitesimal moments to describe the presence of exogeneous terms in the typical dynamics of the phenomenon. An iterative estimating procedure combining the maximum likelihood estimation and a generalized method of moments is provided. A Quandt Likelihood Ratio test for detecting structural breaks in PM_{10} data, in the period from 1st January 2020 to 8th July 2020 which includes the first lockdown due to Covid pandemic, confirms the presence of time-changes. These results show that the lockdown made the air once again cleaner. It is then shown that our model and the associated estimation procedure, while not explicitly contemplating the presence of structural breaks in the time series, implicitly incorporates them in the time dependence of the functions in the infinitesimal moments of the underlying process.

Keywords: non homogeneous Ornstein Uhlenbeck processes; estimating procedure; PM_{10}; structural breaks

1. Introduction

COVID-19 was initially found in December 2019 in Wuhan (China) and it then spread all over the world. The World Health Organization declared COVID-19 a Public Health Emergency of International concern in April 2020. Anyway, the rate of spread is remarkably different in different countries of the word. Such difference is also evident in regions of the same country. Important questions related to the influence of atmospheric factors, such as atmospheric pollution, on the spread of COVID-19 have been then raised.

It has been argued that significantly more infected cases have been observed in more polluted areas than in areas where the presence of pollutants is lower. Furthermore, the lockdown made us witness a situation in which the air has returned to being cleaner (see, for example, [1–3]).

This work does not intend to investigate the cause-effect link between the number of infections and air quality, but to investigate the trend and variability of one of the significant indexes of air quality during the pandemic period in Italy. In particular, we want to investigate the presence of time-changes in the observed time series of the particulate matter with 10 micrometers or less in diameter (PM_{10}), both as a result of the lockdown and because it is hoped that there has been a greater awareness of the importance of the environment by the population. The main focus of the paper is to provide an estimating procedure for the PM_{10} time series that is able to model also the dynamics when some change-points are present. Further, a sufficient condition for detecting the presence of structural breaks is given.

PM_{10} dynamics has been modeled by means of a non homogeneous Ornstein Uhlenbeck (OU) process in [4]. Such process, in its homogeneous version, was originally introduced to describe the velocity of a particle moving in a fluid, and then was generalized to model loan interest rates (see, for example, [5, 6]). In biological context it is able to model the membrane potential between two consecutive spikes (see, for example, [7, 8]). The wide applicability of this process can be documented by the vast literature in this regard (see, for example, [9–12]). In [4], in order to capture non linear trends in real phenomena, a generalization the OU process was considered and an iterative procedure for fitting the time dependent functions present in the drift and in the infinitesimal variance and the constant parameter in the drift term was provided. This approach seemed to work well since the sample paths obtained by plugging the estimated terms reproduce the observed PM_{10} time series quite well.

Here we make use of the non homogeneous OU model to model PM_{10} time series during Covid-19 pandemic as we argue that the drift and infinitesimal variance of the process depends on time. Furthermore, we show an iterative procedure along the line of that one proposed in [4] which makes use of the sample covariance between two consecutive observations instead of the sample variance. The use of the covariance between two successive observations has the advantage to take into account the dependence between two subsequent observations in addition to the variability between observations at the same time instant. The variance and the covariance functions have been compared in the context of neuronal activity modeling in [13]. There it is shown, by simulations, that the procedure implementing the sample covariance function, while showing similar performance to the procedure using the sample variance, better fits the conditional variance of the process. Further, looking at the PM_{10} we show that the procedure, while not explicitly contemplating the presence of structural breaks in the time series, implicitly incorporates them in the time dependence of the functions in the infinitesimal moments of the underlying process.

For our analysis we consider PM_{10} time series from 3 Italian cities, Milano, Torino and Bologna, which are among the 10 cities most affected by Covid-19 during the period 1st January 2020 to 8th July 2020.

The layout of the paper is the following. In Section 2 we introduce the methodology including the estimating procedure and a brief description of the idea underlying the tests for structural breaks in regression models and in time series. In Section 3 we describe the data, showing descriptive statistics and the estimates in the OU-type model. In Section 4 the outbreak detection in PM_{10} time series is investigated. Some concluding remarks close the paper.

2. The modeling strategy

OU process is a time homogeneous diffusion process described by the following stochastic differential equation:

$$dX(t) = [-aX(t) + b]dt + \sigma dB(t), \qquad X(t_0) = x_0, \tag{2.1}$$

where $B(t)$ is a standard Brownian motion. In [4] a generalization of the process (2.1), by including in the infinitesimal moments suitable deterministic time dependent functions, was considered.

Let $\{X(t),\ t \in [t_0, T]\}$ be a stochastic process in $\mathbb{R}$ described via the SDE:

$$dX(t) = [-aX(t) + b(t)]dt + \sigma(t)dB(t), \qquad \mathbb{P}[X(t_0) = x_0] = 1 \tag{2.2}$$

where $a \in \mathbb{R}$, $b(t)$ and $\sigma(t)$ are continuous deterministic functions with $\sigma(t) > 0$ for all $t \in [t_0, T]$.

The transition probability density function (pdf) of $X(t)$ $f(x, t|y, \tau)$ is normal with mean and variance

$$M(t \mid y, \tau) = ye^{-a(t-\tau)} + \int_\tau^t b(\theta)e^{-a(t-\theta)}d\theta, \qquad V(t \mid \tau) = \int_\tau^t \sigma^2(\theta)e^{-2a(t-\theta)}d\theta, \tag{2.3}$$

respectively. Further, the covariance function is (see [13]):

$$\begin{aligned} c(\tau, t) = \mathrm{cov}[X(\tau), X(t)] &= e^{-a(t-\tau)} \int_0^\tau \sigma^2(\xi)e^{-2a(\tau-\xi)}d\xi \\ &= e^{-a(t-\tau)}V(\tau|0), \end{aligned} \tag{2.4}$$

with $0 < \tau < t$.

From Eqs (2.3) and (2.4), it is easy to see that the functions $b(t)$ and $\sigma^2(t)$ satisfy the following relations:

$$b(t) = a\, M(t \mid y, \tau) + \frac{dM(t \mid y, \tau)}{dt}, \tag{2.5}$$

$$\sigma^2(t) = e^{-a(t-\tau)}\Big\{a\ c(\tau, t) + \frac{d\,c(\tau, t)}{d\tau}\Big\}. \tag{2.6}$$

In a discrete sampling, set $\tau = t - \Delta$ in (2.6) and Δ the step between two consecutive observations, we obtain:

$$\sigma^2(t) = e^{-a\Delta}\Big\{ac(t - \Delta, t) + \frac{d\,c(\tau, t)}{d\tau}\ |_{\tau=t-\Delta}\ \Big\}. \tag{2.7}$$

In the following section we provide an estimating procedure for the process $X(t)$ in (2.2) that is able of simultaneously estimating the parameter a and fitting the functions $b(\cdot)$ and $\sigma^2(\cdot)$. The proposed procedure is in line with that one proposed in [4]. It uses the sample covariance between two consecutive observations instead of the sample variance. In this way the procedure has the advantage of being able to capture the dependence between subsequent observations. The following assumption has to be required:

Assumption. *The functions $b(t)$ and $\sigma^2(t)$ in (2.2) are continuous and bounded in* $[t_0, T]$.

Under such assumption, the quantities in (2.5) and (2.6) are well-defined, so we can implement the sample versions of the functions involved in the estimation procedure.

In order to introduce the procedure in the following section, let us consider a discrete sampling of the process (2.2) based on d sample paths for the times t_j, with $j = 0, 1, \ldots, n$. Let Δ be the time between two consecutive observations, i.e., $\Delta = t_j - t_{j-1}$. Let $x_{i,j}$ be the observed values at times t_j, $j = 0, \ldots, n$ and $i = 1, \ldots, d$, i.e., $x_{i,j}$ is the observation of the i–th sample path at the time t_j. Clearly, $x_{i,0} = x_0 \ \forall i = 1, \ldots, d$.

2.1. The iterative procedure

Given an initial value $\widehat{a}_0$ to the estimate of the parameter a in (2.2), the idea is to estimate the functions $b(\cdot)$ and $\sigma^2(\cdot)$ by using (2.5) and (2.6) obtaining $\widehat{b}_1(\cdot)$ and $\widehat{\sigma}^2_1(\cdot)$; then, by using these estimates, we apply the MLE to obtain the estimate $\widehat{a}_1$ and so on until some form of convergence is reached.

In the following $\widehat{a}_k$, $\widehat{b}_k(\cdot)$ and $\widehat{\sigma}^2_k(\cdot)$ are the estimates of a, $b(\cdot)$ and $\sigma^2(\cdot)$, respectively, obtained at k-th iteration of the procedure. The initial value $\widehat{a}_0$ is fixed as the MLE obtained by the homogeneous OU process (2.1). Chosen a wanted precision level ε, the procedure works as follows:

- Step 1. From the observed sample $\{x_{i,j}\}$, with $i = 1, \ldots, d$, and $j = 1, \ldots, n$, obtain $\widehat{a}_0$ as MLE of the parameter a in Eq (2.1);
- Step 2. Obtain the sample mean μ_j and the sample covariance c_j as follows:

$$\mu_j = \frac{1}{d}\sum_{i=1}^{d} x_{i,j}, \qquad c_j = \frac{1}{d-1}\sum_{i=1}^{d}(x_{i,j-1} - \mu_{j-1})(x_{i,j} - \mu_j). \tag{2.8}$$

- Step 3. Interpolate the values μ_j and c_j. The obtained functions $\widehat{M}(t)$ and $\widehat{V}(t)$ provide estimates of $M(t|x_0, 0) \equiv M(t)$ and $c(t - \Delta, t) \equiv c(t)$ in (2.3) respectively.
- Step 4. Obtain the derivatives of $\widehat{M}(t)$ and $\widehat{c}(t)$.
- Step 5. Set $k = 1$, $C_3 = d(n-1)\Delta$, $toll = \varepsilon + 1$.
- Step 6. **while** ($toll > \varepsilon$)
  - Obtain the estimate of $b(t)$ and $\sigma^2(t)$ as follows:

$$\widehat{b}_k(t) = \widehat{a}_{k-1}\,\widehat{M}(t) + \frac{d\widehat{M}(t)}{dt}, \qquad \widehat{\sigma}^2_k(t) = e^{-\widehat{a}_{k-1}\Delta}\Big\{\widehat{a}_{k-1}\widehat{c}(t) - \frac{d\widehat{c}(t)}{dt}\Big\}. \tag{2.9}$$

  - From $\{x_{ij}\}$, with $i = 1, ..., d$, and $j = 1, ..., n$, obtain (see [4])

$$C_{1k} = \sum_{i=1}^{d}\sum_{j=2}^{n}\frac{x^2_{i\,j}}{\widehat{\sigma}^2_k(t_{j-1})}, \qquad C_{2k} = \sum_{i=1}^{d}\sum_{j=2}^{n}\frac{x_{i\,j}x_{i\,j-1} + x_{i\,j}\widehat{b}_k(t_{j-1})\Delta}{\widehat{\sigma}^2_k(t_{j-1})}.$$

  - Calculate

$$\alpha_k = \frac{C_{2k} + \sqrt{C^2_{2k} + 4C_{1k}C_3}}{2C_{1k}}.$$

  - Obtain the estimate of the parameter a as

$$\widehat{a}_k = \frac{\ln \alpha_k}{\Delta}.$$

  - $toll = |\widehat{a}_k - \widehat{a}_{k-1}|$.

– $k = k + 1$.

end

We point out that the proposed methodology differs from the methodology in [4] on the choice to estimate the covariance function $c(t-\Delta, t)$ rather than the variance function of the process $X(t)$. As in [4], the consistency of the estimators derives from the consistency of the ML and GMM estimators in addition to the uniform convergence of the interpolation method. Further, several simulation experiments show the consistency of the iterative method, since as the number of observations n and the number of sample paths d increase, the Mean Absolute Error decrease.

Finally, we point out that in our analysis the interpolation method used in Step 3 of the iterative procedure is the natural cubic spline interpolation. In such case the interpolating function $\widehat{M}(t)$ is:

$$\widehat{M}(t) = \begin{cases} \widehat{M}_1(t), & t_0 \le t \le t_1, \\ \widehat{M}_2(t), & t_1 < t \le t_2, \\ \vdots & \\ \widehat{M}_n(t), & t_{n-1} < t \le t_n. \end{cases} \tag{2.10}$$

with $\widehat{M}_j(t) = \alpha_j + \beta_j t + \gamma_j t^2 + \delta_j t^3$ $(\delta_j \neq 0)$, $j = 1, \ldots, n$. The coefficients $\alpha_j, \beta_j, \gamma_j$ and δ_j for each j are determined by the following boundary conditions on the functions, their prime and second derivatives:

$$\begin{aligned} &\widehat{M}_j(t_{j-1}) = \mu_{j-1}, \quad \widehat{M}_j(t_j) = \mu_j, \quad j = 1, \ldots, n, \\ &\widehat{M}'_j(t_j) = \widehat{M}'_{j+1}(t_j), \quad j = 1, \ldots, n-1, \\ &\widehat{M}''_j(t_j) = \widehat{M}''_{j+1}(t_j), \quad j = 1, \ldots, n-1, \end{aligned}$$

The expression of the interpolating function $\widehat{c}(t)$ can be obtained as $\widehat{M}(t)$ in (2.10), interpolating the points c_j, $j = 1, \ldots, n$.

2.2. Testing for structural changes: Quandt Likelihood Ratio test in a nutshell

A structural break is an unexpected change over time in the parameters of regression models. Preciselly, in the model

$$y_i = x_i^T \beta_i + u_i, \qquad i = 1, 2, \ldots, n, \tag{2.11}$$

where at time i, y_i is the observation of the dependent variable, $x_i = (1, x_{i2}, \ldots, x_{ik})$ is the vector of observations of the independent variables, u_i are *iid* $N(0, \sigma^2)$ and β_j is the $k \times 1$ vector of regression coefficients, the presence of time-changes can be tested through the hypothesis:

$$H_0 : \beta_i = \beta_0, \qquad i = 1, 2, ...n$$

against the alternative that at least one of the coefficients β_i depends on time.

Chow test is the classical test for structural change. Here the alternative is:

$$H_1 : \beta_i = \begin{cases} \beta_A & 1 \le i \le i_0, \\ \beta_B & i_0 < i \le n. \end{cases} \tag{2.12}$$

In it the sample is splitted into two sub-samples, estimates of the parameters are provided for the two sub-sample, and then a test on the equality of the two parameter vectors is performed (see, for a review, [14]). Chow test computes the test statistics

$$F_{i_0} = \frac{\hat{u}^T \hat{u} - \hat{e}^T \hat{e}}{\hat{e}^T \hat{e}/(n-k)}, \tag{2.13}$$

where $\hat{e} = (\hat{u}_A, \hat{u}_B)^T$ are the residuals under the alternative H_1, where the coefficients are estimates separately in the subsamples, and $\hat{u}$ are the residuals from the model under H_0, where the coefficients are estimated on the whole sample. If F_{i_0} is too large, H_0 is rejected and the presence of a change point in i_0 is confirmed. The main drawback of the Chow test is that the break-date must be known a priori, so a candidate break-date is generally fixed by looking at the data. Anyway, the results can be highly sensitive to these choices. A natural generalization of the Chow test, in which the breakdate is unknown, is the Quandt Likelihood Ratio (QLR) test. It computes the F statistics (2.13) for all possible breakdates in a fixed range $[\tau_0, \tau_1]$. Usually, $\tau_0 = \lfloor 0.15\, T \rfloor$ and $\tau_1 = \lfloor 0.85\, T \rfloor$. Further, to aggregate the series of F statistics into one, the following test statistics can be considered:

$$\begin{aligned} supF &= \sup_{\tau_0 \le i \le \tau_1} F_i, \\ aveF &= \frac{1}{\tau_1 - \tau_0 + 1} \sum_{i=\tau_0}^{\tau_1} F_i, \\ expF &= \log\Big(\frac{1}{\tau_1 - \tau_0 + 1} \sum_{i=\tau_0}^{\tau_1} \exp(0.5\, F_i) \Big). \end{aligned}$$

The distribution of such statistics is not exact, anyway in [15] a code computing p values for the F statistics was provided. In this way, under the null hypothesis of no structural change, boundaries can be computed such that the asymptotic probability that the F test statistics exceeds is α.

In our case, in which we observe PM_{10} concentrations, structural breaks detection is based on the Euler's discretization (2.15) and the regressor is the lag 1-delayed observation.
Preciselly, from (2.2) we have

$$X(t_j) = (1 - a\Delta) X(t_{j-1}) + b(t_{j-1})\Delta + \sigma(t_{j-1}) \sqrt{\Delta}\, Z_j, \qquad j = 2, \ldots, n. \tag{2.14}$$

where $Z_j \sim N(0, 1)$. Setting

$$Y_j = \frac{X(t_j)}{\sigma(t_{j-1}) \sqrt{\Delta}},$$

we can write the Eq (2.14) in the form of a linear regression model as in (2.11), i.e.,

$$Y_j = (1 - a\Delta) Y_{j-1} + \frac{b(t_{j-1})}{\sigma(t_{j-1})} \sqrt{\Delta} + Z_j. \tag{2.15}$$

We note that the Euler equation (2.15) presents a convergence order $1/2$ since the infinitesimal variance of the process $X(t)$ does not depend on x. Hence the Euler scheme coincides with Milstein scheme (which is based on the first order approximation of the diffusive term with respect to the variable x).

Finally we point out that the null hypothesis of no structural break in the regression (2.15) corresponds to testing that $1 - a\Delta$ and $\frac{b(t_{j-1})}{\sigma(t_{j-1})}\sqrt{\Delta}$ are both constant, i.e., a is constant and $b(t_{j-1})$ and $\sigma(t_{j-1})$ are proportional for all j, or better, the functions $b(\cdot)$ and $\sigma(\cdot)$ are both constant. Therefore the condition that at leat one of the functions $b(\cdot)$ and $\sigma(\cdot)$ is not constant constitutes a sufficient condition for the presence of structural breaks in the time series.

3. Data collection

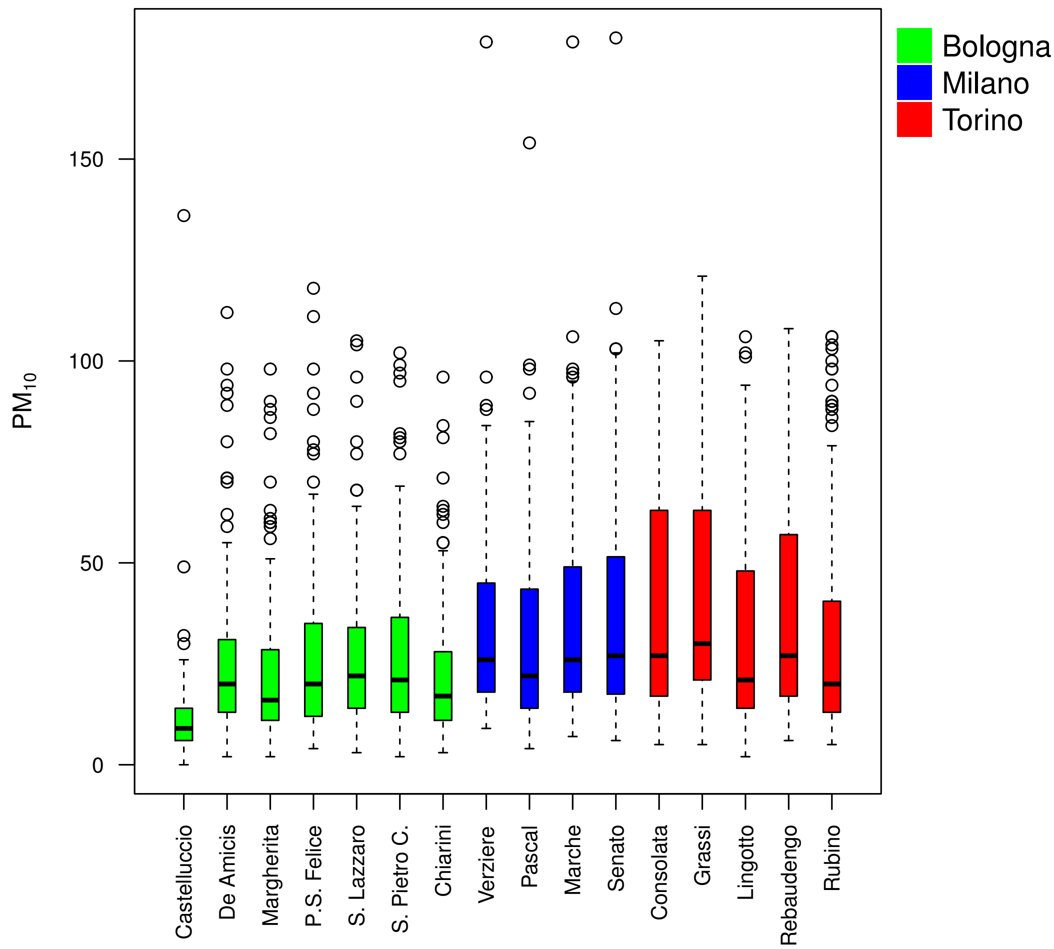

Figure 1. PM_{10} concentration distribution by station in Bologna, Milano and Torino.

We consider PM_{10} daily concentrations (in $\mu g/m^3$) measured from 1 January 2020 to 8 July 2020 in three italian cities severely affected by the pandemic that are Milano, Bologna and Torino. For them we consider several monitoring stations in the metropolitan areas (190 observations for each station). We have 7 monitoring stations in Bologna, that are Castelluccio, De Amicis, Giardini Margherita, Porta San Felice, San Lazzaro, San Pietro Capofiume and Via Chiarini; 4 in Milano, i.e., Verziere, Pascal, Viale Marche and Via Senato; 5 in Torino, that are Consolata, Grassi, Lingotto, Rebaudengo and Rubino. The data sets were provided by the regional agencies (Emilia-Romagna, Lombardia and Piemonte). Figure 1 shows the distribution of the PM_{10} concentration for each of the considered stations, showing that all the distributions are positively skewed, with many values exceeding the admitted concentration 50 $\mu g/m^3$, expecially for Milano and Torino. In particular Bologna seems the most "virtuous" city since PM_{10} values above the legal limit of 50 $\mu g/m^3$ turn out to be outliers in most cases. Further, the variability of Bologna PM_{10} data is the lowest among the three cities. Also for Milan, values exceeding 50 $\mu g/m^3$ are found only in 25% of the highest values of the data distribution. There are, however, a few outliers that even exceed 150 $\mu g/m^3$. Consolata and Grassi stations in Torino have PM_{10} values on average higher than the others, also presenting a greater variability with respect the other monitoring stations. Table 1 shows the descriptive statistics, included the number of missing values.

Table 1. Descriptive statistics on PM_{10} time series data.

Station	Min	Q_1	Median	Mean	Q_3	Max	NA's
			Bologna				
Castelluccio	0.00	6.00	9.00	10.86	14.00	136.00	13
De Amicis	2.00	13.00	20.00	25.34	30.75	112.00	4
Giardini Margherita	2.00	11.00	16.00	22.73	28.50	98.00	7
Porta San Felice	4.00	12.00	20.00	26.70	35.00	118.00	10
San Lazzaro	3.00	14.00	22.00	26.81	34.00	105.00	6
San Pietro Capofiume	2.00	13.00	21.00	27.51	36.50	102.00	7
Via Chiarini	3.00	11.00	17.00	22.02	28.00	96.00	12
			Milano				
Verziere	9.00	18.00	26.00	34.18	45.00	179.00	6
Pascal	4.00	14.00	22.00	31.96	43.25	154.00	5
Viale Marche	7.00	18.00	26.00	35.96	49.00	179.00	0
Via Senato	6.00	17.50	27.00	36.93	51.50	180.00	6
			Torino				
Consolata	5.00	17.00	27.00	39.85	62.75	105.00	64
Grassi	5.00	21.00	30.00	42.58	63.00	121.00	29
Lingotto	2.00	14.00	21.00	32.15	48.00	106.00	11
Rebaudengo	6.00	17.00	27.00	38.75	57.00	108.00	15
Rubino	5.00	13.00	20.00	31.45	40.25	106.00	14

Here, the missing values imputation is made by means of the procedure in [16], by looking at the geographical distances between the monitoring stations.

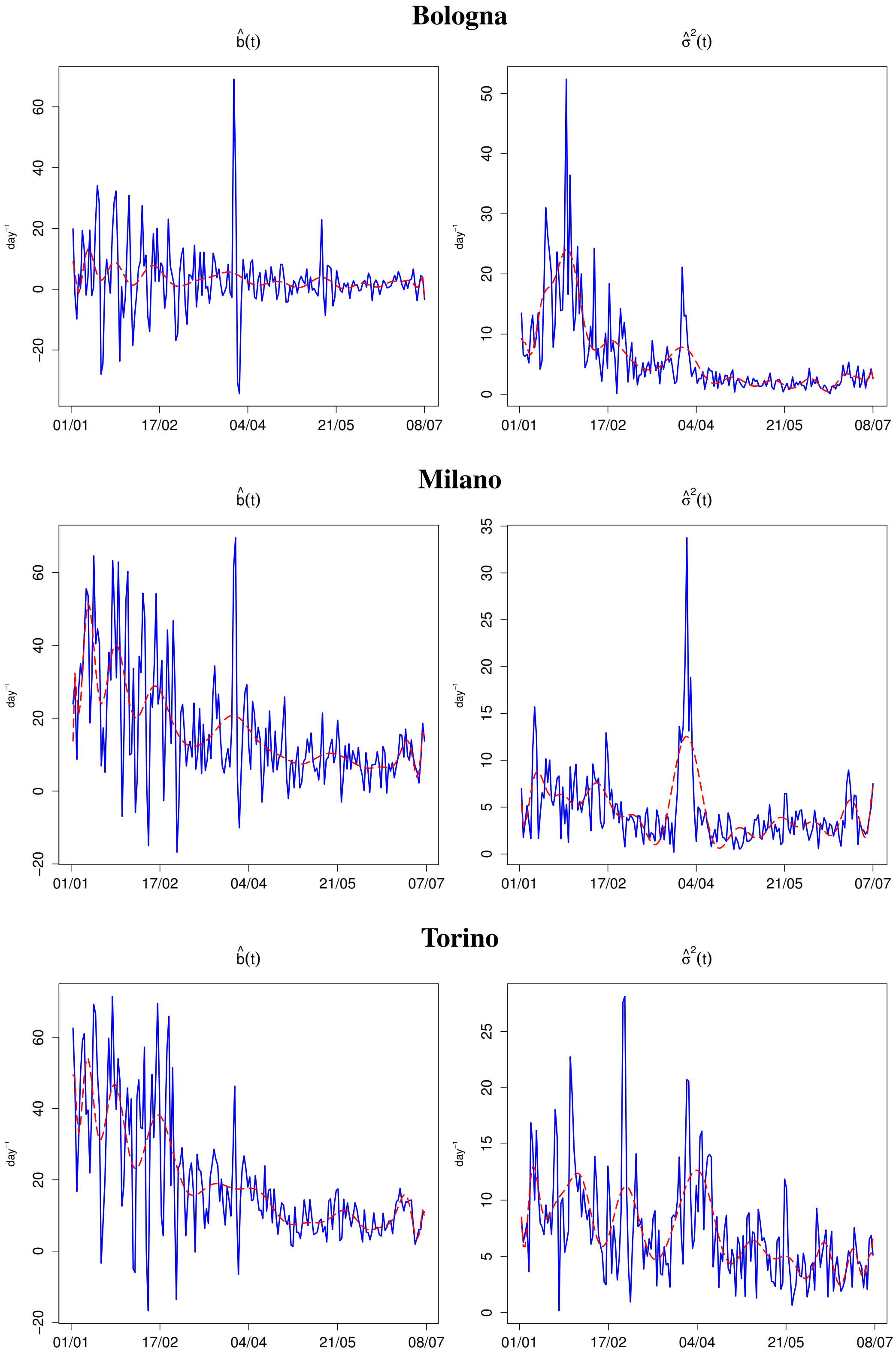

Figure 2. Fitted function of $b(t)$ (left) and of $\sigma^2(t)$ (right) for PM_{10} data in Bologna (top), Milano (middle) and Torino (bottom) along with their regular versions (dashed red line).

In the following analysis, we consider, for each city, each PM_{10} time series in that city as a sample path of a same diffusion diffusion process. Essentially, we consider three OU-type process, X_B, X_M and X_T, where B, M and T stands for Bologna, Milano and Torino and we observe 7 sample paths for $X_B(t)$, 4 for $X_M(t)$ and 5 for $X_T(t)$. The application of the iterative procedure provides the following estimates of the parameter a: $a_B = 0.6491973$ for the process $X_B(t)$; $a_M = 0.3301912$ for $X_M(t)$ and $a_T = 0.3873125$ for $X_T(t)$. The fitted functions $\widehat{b}(t)$ and $\widehat{\sigma}^2(t)$ are shown in Figure 2 along with their regular versions (in red). In all the cases, the fitted functions $\widehat{b}(t)$ and $\widehat{\sigma}^2(t)$ are far from constant in the period before April, about a month after the start of the lockdown due to the Covid pandemic. This observation leads to argue that the observed time series present some strucural breaks, since the lockdown period has somehow "regularized" both the trend and the variability of the process describing the PM_{10} dynamics. We point out that the constancy of the functions $b(t)$ and $\sigma(t)$ can be verified by means of a bootstrap test, in line with [17]. In our case, all the tests provide $p-$values of order of 10^{-16}, so the functions $b(t)$ and $\sigma(t)$ are not constant.

In the following section we test the null hypotesis of constant parameters in Eq (2.15).

4. PM_{10} outbreak detection

This section investigates the presence of structural breaks in the considered PM_{10} time series by means a well know fluctuation test, i.e., the QLR test. The linear regression on which the test is based is the Euler discretization (2.15). In terms of the model, if H_0 is not reject, PM_{10} time series can be modeled by means of an homogeneous OU process with Eq (2.1).

Table 2. Results of QLR estimates test on PM_{10} time series data.

Station	Breakdates	*supF*	*p*-value
Bologna			
Castelluccio	2020-04-05	27.041	0.000043
De Amicis	2020-02-03	15.490	0.009413
Giardini Margherita	2020-02-26	16.774	0.005224
Porta San Felice	2020-02-20	27.058	0.000042
San Lazzaro	2020-02-25	29.641	0.000012
San Pietro Capofiume	2020-01-30	19.450	0.001542
Via Chiarini	2020-02-19	20.418	0.000996
Milano			
Verziere	-	10.659	0.072890
Pascal	2020-04-27	17.001	0.004673
Viale Marche	2020-03-23	17.193	0.004283
Via Senato	2020-03-23	18.554	0.002304
Torino			
Consolata	2020-01-21	22.381	0.000511
Grassi	2020-01-21	15.593	0.011340
Lingotto	2020-01-21	18.914	0.002548
Rebaudengo	2020-01-21	20.366	0.001307
Rubino	2020-01-21	20.366	0.001307

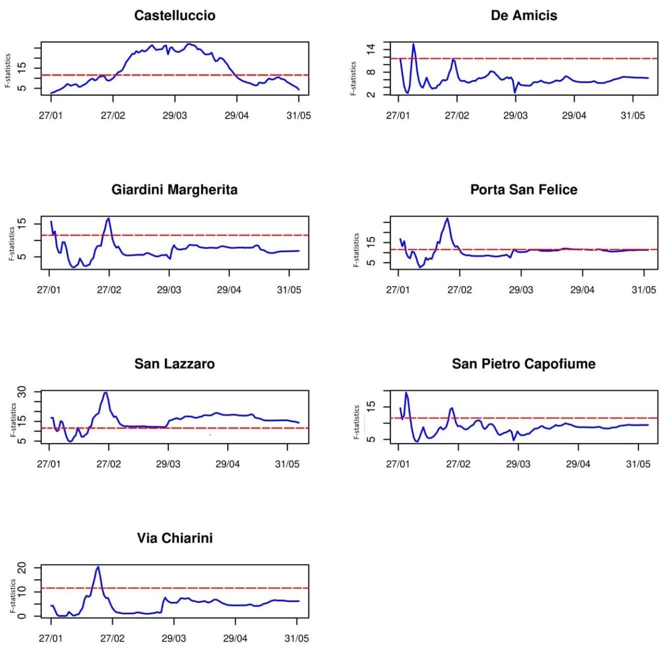

Figure 3. Process of F-statistics in the QLR test applied to the Euler discretization by considering data from the monitoring stations in Bologna. Horizontal red line is the corresponding boundary. The test reject the null hypothesis of no breaks for all the monitoring stations.

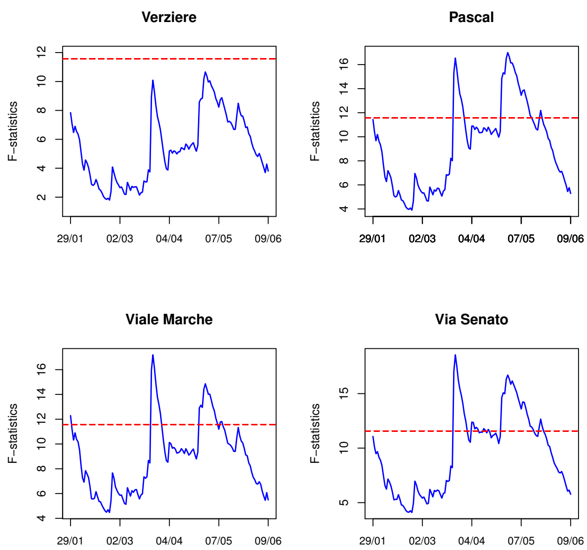

Figure 4. Process of F-statistics in the QLR test applied to the Euler discretization by considering data from the monitoring stations in Milano. Horizontal red line is the corresponding boundary. The test reject the null hypothesis of no breaks for the monitoring stations.

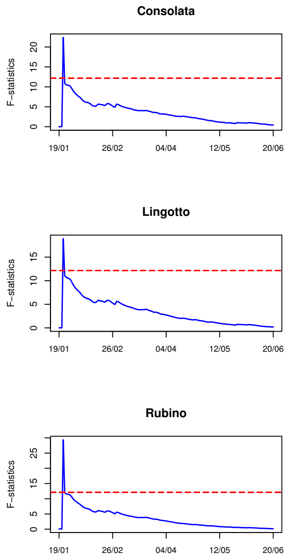

Figure 5. Process of F-statistics in the QLR test applied to the Euler discretization by considering data from the monitoring stations in Torino. Horizontal red line is the corresponding boundary. The test does not reject the null hypothesis of no breaks for Verziere monitoring station and reject it for the others.

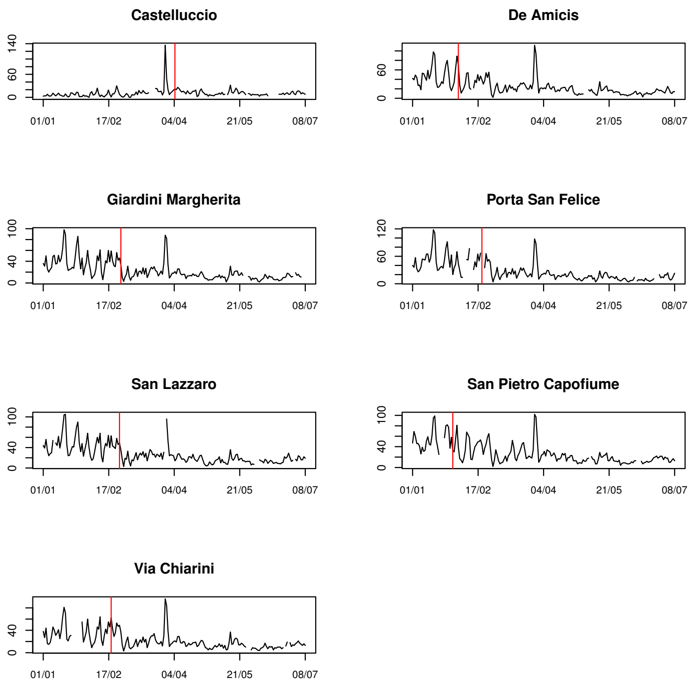

Figure 6. PM_{10} time series in Bologna from 1st January 2020 to 8th July 2020. Vertical red line indicates the break date identified by the QLR test.

In Figures 3–5 the processes of F-statistics, along with the corresponding boundaries at level

$\alpha = 0.05$, are shown for all the considered stations in Bologna, Milano and Torino, respectively. For this analysis the R-package *strucchange* has been used. It is evident that in almost all the cases the process exceeds the boundary with significance 0.05, so there is statistical evidence of a structural break in the considered period. It is also interesting to observe that the processes present very similar shapes for stations in the same town. In Table 2 the breakdates detected by the QLR test are shown for all the stations. The values of the test statistics *supF* and the corresponding *p*-values are reported. Only for Milano Verziere station there is no evidence of a break change point, although the process shows a shape that is very to the other stations in Milano, still remaining below the boundary.

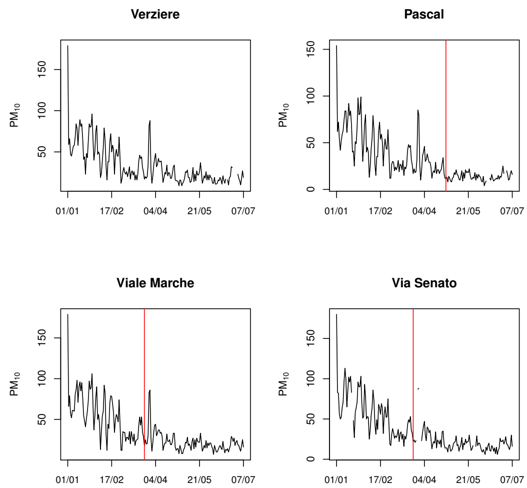

Figure 7. PM_{10} time series in Milano from 1st January 2020 to 8th July 2020. Vertical red line indicates the break date identified by the QLR test.

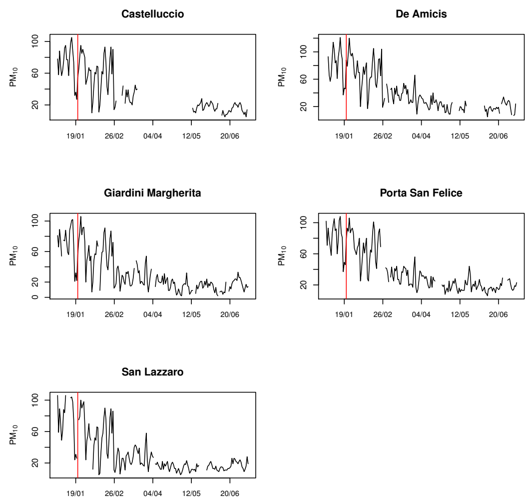

Figure 8. PM_{10} time series in Torino from 1st January 2020 to 8th July 2020. Vertical red line indicates the break date identified by the QLR test.

In Figures 6–8 the observed time series are shown along with the detected change points. We can see that, in all the stations, PM_{10} observations are characterized by a "flattening" of both the values and their variability, probably due to the new environmental conditions in the lockdown period.

In the following we use the estimates of a, $b(t)$ and $\sigma^2(t)$ provided by the proposed procedure and simulate 500 sample- paths of the process (2.2) in which such estimates are plugged in. In Figure 9 the observed sample-paths of $X_B(t)$, $X_M(t)$ and $X_T(t)$ are compared with the mean of the corresponding simulated sample-paths. The results show that the fitted processes via the proposed procedure are quite close to the real ones and they satisfactory capture the trend of the PM_{10} concentrations in the three analized cities. So the model (2.2), and consequently the estimating procedure, even if it does not explicitely include the presence of time-change points in the observed time series, is able to incorporate it in the dependece on time of the infinitesimal moments.

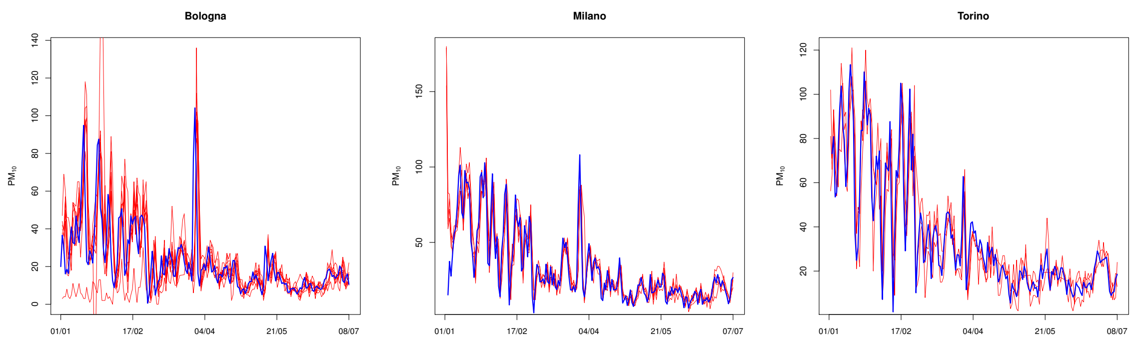

Figure 9. PM_{10} daily concentration in the metropolitan area in several monitoring stations in Bologna (top), Milano (middle) and Torino (bottom) (red lines) and the corresponding fitted values via the iterative procedure (blue line).

5. Conclusions

We modeled PM_{10} data from three italian cities by means of a non homogeneous OU process. An iterative estimatimg procedure combining the maximum likelihood estimation and a generalized method of moments is provided. Such procedure is able to fit the involved functions taking into account relations between subsequent observations and between observations at the same time.

QLR test for the considered time series during Covid pandemic show the presence of structural breaks during the period 1st January 2020–8th July 2020. In particular all the observed time series show a "flattening" of the PM_{10} values in terms of mean and variability. This is probably due to the lockdown imposed by the government or, we hope, to citizens' awareness of the environmental issue. This hope will be able to be verified in a later period and, in any case, when our life can return to how, and better, than before. Anyhow, the model, and consequently the proposed iterative procedure, is able to fit data in the whole considered period. Indeed, the non-constancy of the terms in the infinitesimal moments is a sufficient condition for detecting the presence of structural breaks in the data. Further, the time dependence of the involved functions is able to implicitly include time-changes in the data, so to model phenomena in which external conditions change the internal structure of the data.

Conflict of interest

The author declares no conflict of interest.

References

1. G. He, Y. Pan, T. Tanaka, The short-term impacts of COVID-19 lockdown on urban air pollution in China, *Nat. Sustain.*, **3** (2020).
2. X. Wu, R. C. Nethery, B. M. Sabath, D. Braun, F. Dominici, Air pollution and COVID-19 mortality in the United States: Strengths and limitations of an ecological regression analysis, *Sci. Adv.*, **6** (2020), 10051011.
3. Y. Han, J. C. Lam, V. O. Li, P. Guo, Q. Zhang, A. Wang, et al., The effects of outdoor air pollution concentrations and lockdowns on Covid-19 infections in Wuhan and other provincial capitals in China, preprints, (2020), 2020030364. Available from: https://www.preprints.org/manuscript/202003.0364/v1
4. G. Albano, V. Giorno, Inferring time non-homogeneous Ornstein Uhlenbeck type stochastic process, *Comput. Stat. Data Anal.*, **150** (2020), 107008–107008.
5. G. Albano, V. Giorno, On short-term loan interest rate models: a first passage time approach, *Mathematics*, **6** (2018), 70.
6. V. Linetsky, Computing hitting time densities for CIR and OU diffusions: Applications to mean reverting models, *J. Comput. Financ.*, **4** (2004), 1–22.
7. S. Ditlevsen, P. Lánský, Estimation of the input parameters in the Ornstein-Uhlenbeck neuronal model, *Phys. Rev. E*, **71** (2005), 011907.
8. H. Tuckwell, F. Wan, J. P. Rospars, A spatial stochastic neuronal model with OrnsteinUhlenbeck input current, *Biol. Cybern.*, **86** (2002), 137–145.
9. G. Albano, V. Giorno, P. Román-Pomán, F. Torres-Ruiz, On a non-homogeneous Gompertz-type diffusion process: inference and first passage time, in *LNCS 10672* (eds. R. Moreno-Díaz et al.), Springer (2018), 47–54.
10. A. Buonocore, L. Caputo, A. G. Nobile, E. Pirozzi, Restricted ornstein uhlenbeck process and applications in neuronal models with periodic input signals, *J. Comput. Appl. Math.*, **285** (2015), 59–71.
11. V. Giorno, S. Spina, On the return process with refractoriness for non-homogeneous Ornstein-Uhlenbeck neuronal model, *Math. Biosci. En.*, **11** (2014), 285–302.
12. R. Gutiérrez, R. Gutiérrez-Sánchez, A. Nafidi, A. Pascual, Detection, modelling and estimation of non linear trends by using a non-homogeneous Vasicek stochastic diffusion. Application to CO_2 emissions in Morocco, *Stoch. Environ. Res. Risk Assess.*, **26** (2012), 533–543.
13. G. Albano, V. Giorno, Inference on the effect of non homogeneous inputs in Ornstein Uhlenbeck neuronal modeling, *Math. Biosci. Eng.*, **17** (2020), 328–348.
14. D. W. K. Andrews, R. C. Fair, Inference in nonlinear econometric models with structural change, *Rev. Econ. Stud.*, **55** (1988), 615–639.

15. E. Hansen, Approximate asymptotic p values for structural-change tests, *J. Bus. Econ. Stat.*, **15** (1997), 60–67.
16. M. L. Parrella, G. Albano, M. La Rocca, C. Perna, Reconstructing missing data sequences in multivariate time series: an application to environmental data, *Stat. Methods Appl.*, **28** (2019), 359–383.
17. G. Albano, V. Giorno, P. Román-Pomán, S. Román-Pomán, J. J. Serrano-Pérez, F. Torres-Ruiz, Inference on an heteroscedastic Gompertz tumor growth model, *Math. Biosci.*, **328** (2020), 108428.