

Article

Functional Location-Scale Model to Forecast Bivariate Pollution Episodes

Manuel Oviedo-de La Fuente ¹, Celestino Ordóñez ^{2,*} and Javier Roca-Pardiñas ³

¹ Department of Statistics, Mathematical Analysis and Optimization, Universidad de Santiago de Compostela, 15782 Santiago de Compostela, Spain; manuel.oviedo@usc.es

² Department of Mining Exploitation and Propsecting, Universidad de Oviedo, Escuela Politécnica de Mieres, 33600 Mieres, Spain

³ Department of Statistics and Operation Research, Universidad de Vigo, 36310 Vigo, Spain; roca@uvigo.es

* Correspondence: ordonezcelestino@uniovi.es

Received: 7 May 2020; Accepted: 29 May 2020; Published: 8 June 2020



Abstract: Predicting anomalous emission of pollutants into the atmosphere well in advance is crucial for industries emitting such elements, since it allows them to take corrective measures aimed to avoid such emissions and their consequences. In this work, we propose a functional location-scale model to predict in advance pollution episodes where two pollutants are involved. Functional generalized additive models (FGAMs) are used to estimate the means and variances of the model, as well as the correlation between both pollutants. The method not only forecasts the concentrations of both pollutants, it also estimates an uncertainty region where the concentrations of both pollutants should be located, given a specific level of uncertainty. The performance of the model was evaluated using real data of SO₂ and NO_x emissions from a coal-fired power station, obtaining good results.

Keywords: pollution episodes; functional data; bivariate analysis; uncertainty region; generalized additive models

1. Introduction

Forecasting air quality and concentrations of pollutants in the atmosphere by means of statistical methods is an active area of research given the transcendence of the problem and the difficulty to find optimal solutions using deterministic mathematical models. Among the different methods that can be found in the literature to tackle this problem, models for time series analysis such as the integrated autoregressive moving average—ARIMA [1–3], multivariate regression [4–7], generalized linear or additive models (GAM) [8–11] and artificial neural networks (ANN) [12–19] are the most extended. Due to the increased access to continuous data over time, functional data analysis [20,21] was also proposed for air quality forecasting and outlier detection [22–24]. Parametric [25,26] and nonparametric [27–29] functional regression methods were tested. A functional framework allows considering the inherent correlation between observations, instead of considering them as independent realizations of an underlying stochastic process. Some functional approaches add related meteorological variables to the models [30–34], which can improve the result of the predictions and help to understand the process underlying the evolution of the pollutants.

Most of the documents in the literature propose solutions to predict the concentration of each pollutant individually, being much scarcer those focused on predicting more than one pollutant at a time. Vector autoregressive moving average (VARMA) [35,36] and vector autoregressive integrated moving average (VARIMA) [37] models were applied to reach this objective. In this work, we proposed a method for the simultaneous forecasting of pollution episodes when two pollutants, i.e., SO₂ and NO_x, are involved. Apart from transport, one of the main sources of these pollutants is public electricity

and heating. Their negative effects on human health are well known, and goes for mild (i.e., eyes irritated, nose or headache) to severe (i.e., lung damage or reduced oxygenation of tissues). They also have negative effects on animals and plants, as well as in other substances, such as water and soils. In addition, NO_x is a precursor of the tropospheric ozone. High levels of ozone contributes to climate change, cause adverse impacts on health and can damage vegetation.

Pollution episodes (incidents) are abnormally large emissions of one or more pollutants in short periods of time. Although the improvement of the chemical processes and particle filter systems have significantly reduced the amount and intensity of the pollution episodes, they are still of particular interest for the industries, as they may be subject to sanctions, or for other reasons, such as public health deterioration or industry discredit. Therefore, pollution industries, such as coal-fired power plants, are very interested in determining in advance when these episodes of excessive contamination might occur. Specifically, this is the purpose of our work: forecasting pollution episodes of SO_2 and NO_x early enough to allow corrective measures to be taken. Our approach uses a location-scale model [11,38,39] that treats the predictors, the concentrations of both pollutants over time, as functions, while the response is a scalar, the concentration of the pollutants some time in advance. The novelty of our approach is the combination of a bivariate location-scale model with functional additive models. This method combines the simplicity of the location and scale models with the capacity of functional data analysis to deal with data in the form of functions.

The document is structured as follows: In Section 2 we show the mathematical model proposed to solve the problem under analysis and the algorithm used to estimate a solution from the data. Section 3 is devoted to test the validity of the model using real data. Finally, a discussion of the results and the main conclusions of our work are exposed in Section 4.

2. Methodology

2.1. Mathematical Model

Let $\{\mathbf{X}_i, \mathbf{Y}_i\}_{i=1}^n$ be a set of observations of a stochastic process, $\mathbf{X} = (X^1(t), \dots, X^p(t))$, where $X^j(t) \in L_2[0, T], j = 1, \dots, p$, are predictor covariates and $\mathbf{Y} = (Y_1, Y_2)$, with $Y_j \in \mathbb{R}$, a response variable. In this context, the following bivariate location-scale model [40,41] is assumed

$$\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \begin{pmatrix} \mu_1(\mathbf{X}) \\ \mu_2(\mathbf{X}) \end{pmatrix} + \Sigma^{1/2}(\mathbf{X}) \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix} \tag{1}$$

where $\Sigma^{1/2}(\mathbf{X})$ represents the Cholesky decomposition of the variance-covariance matrix $\Sigma(\mathbf{X})$

$$\Sigma(\mathbf{X}) = \begin{pmatrix} \sigma_1^2(\mathbf{X}) & \sigma_{12}(\mathbf{X}) \\ \sigma_{12}(\mathbf{X}) & \sigma_2^2(\mathbf{X}) \end{pmatrix} \tag{2}$$

so that $\text{Var}(\mathbf{Y}|\mathbf{X}) = \Sigma(\mathbf{X}) = \Sigma^{1/2}(\mathbf{X}) \left(\Sigma^{1/2}(\mathbf{X}) \right)^T$. To guarantee the model identification in (1), the bivariate residuals $(\varepsilon_1, \varepsilon_2)$ are assumed to be independent of the covariates, with zero mean, unit variance, and zero correlation. Despite we do not assume any distribution for the error term, within the framework of functional data analysis this work might be addressed under the assumption of other structures for error distribution: generalized Gauss-Laplace distribution that relax the constrictive assumption of the normal distribution errors [42], generalized linear mixed models (GLMMs) [38] to estimate random effects and dependent (temporal or spatial) errors, and generalized additive models for location, scale and shape (GAMLSS) [43] to model the dynamically variable distribution, considering skewness and kurtosis.

We define the unconditionally probabilistic region for the errors $(\varepsilon_1, \varepsilon_2)$ as

$$\varepsilon_\tau(k) = \{(\varepsilon_1, \varepsilon_2) \in \mathbb{R}^2 | f(\varepsilon_1, \varepsilon_2) \geq k\}$$

f being the density function of the bivariate residuals $(\varepsilon_1, \varepsilon_2)$ and k the τ -quantile of $f(\varepsilon_1, \varepsilon_2)$. Then, for a given \mathbf{X} , we define a conditional τ^{th} - uncertainty region for (Y_1, Y_2) containing $\tau\%$ of the observations as:

$$R_\tau(\mathbf{X}) = \begin{pmatrix} \mu_1(\mathbf{X}) \\ \mu_2(\mathbf{X}) \end{pmatrix} + \Sigma^{1/2}(\mathbf{X})\varepsilon_\tau \tag{3}$$

2.2. Estimation Algorithm

To implement an algorithm that allows applying the mathematical model exposed in the previous section, we propose using a functional additive models to estimate the means, variances and covariances in (1). Given a sample of size n , $\{\mathbf{X}_i, (Y_{i1}, Y_{i2})\}_{i=1}^n$, where $\mathbf{X}_i = (X_i^1(t), \dots, X_i^p(t))$, the steps of the proposed estimation algorithm are the following:

Step 1: Perform a decomposition of each covariate $X^j(t)$ in basis functions of the form $X^j(t) \approx \sum_{k=1}^K \zeta_k^j \phi_k(t)$, where ϕ_k ($k = 1, \dots, K$) are K basis functions (i.e., B-splines, wavelets), and ζ_{il} are either the coefficients of an expansion in fixed basis or the principal component scores of the Karhunen-Loève expansion [44,45]. As a result, we obtain the transformed covariates

$$\tilde{\mathbf{X}}_i = \left((\zeta_{i1}^1, \dots, \zeta_{iK}^1); (\zeta_{i1}^2, \dots, \zeta_{iK}^2); \dots; (\zeta_{i1}^p, \dots, \zeta_{iK}^p) \right) \quad i = 1, \dots, n$$

Step 2: For $r = 1, 2$, fit an additive model to the sample $\{\tilde{\mathbf{X}}_i, Y_{i1}, Y_{i2}\}_{i=1}^n$ and obtain an estimation of the means

$$\hat{\mu}_r(\mathbf{X}_i) = \alpha_r + \sum_{j=1}^p \sum_{k=1}^K \hat{f}_{rk}^j(\zeta_{ik}^j) \tag{4}$$

and then estimate $\sigma_r^2(\mathbf{X})$ from the sample $\{\tilde{\mathbf{X}}_i, (Y_{ir} - \hat{\mu}_r(\mathbf{X}_i))^2\}_{i=1}^n$ as

$$\hat{\sigma}_r^2(\mathbf{X}_i) = \exp \left(\hat{\beta}_r + \sum_{j=1}^p \sum_{k=1}^K \hat{g}_{rk}^j(\zeta_{ik}^j) \right) \tag{5}$$

Then, compute the correlation $\rho(\mathbf{X})$, which is related to the covariance by $\sigma_{12}(\mathbf{X}) = \sigma_1(\mathbf{X})\sigma_2(\mathbf{X})\rho(\mathbf{X})$, using the sample $\{\mathbf{X}_i, \hat{\delta}_i\}_{i=1}^n$, as follows:

$$\hat{\rho}(\mathbf{X}_i) = \tanh \left(\hat{\gamma} + \sum_{j=1}^p \sum_{k=1}^K \hat{m}_k^j(\zeta_{ik}^j) \right)$$

being

$$\hat{\delta}_i = \frac{(Y_i^1 - \hat{\mu}_1(\mathbf{X}_i)) (Y_i^2 - \hat{\mu}_2(\mathbf{X}_i))}{\hat{\sigma}_1(\mathbf{X}_i)\hat{\sigma}_2(\mathbf{X}_i)}$$

where \hat{f}_{rk}^j , \hat{g}_{rk}^j and \hat{m}_k^j are smooth and unknown functions, α_r , β_r and γ are coefficients, p the number of predictors (covariates), and K the number of basis. Please note that the link functions $H_\sigma(\cdot) = \exp(\cdot)$ and $H_\rho(\cdot) = \tanh(\cdot)$ used in the variance and correlation structures, respectively, ensure that the restrictions on the parameter spaces ($\sigma_r^2(\mathbf{X}) \geq 0$ and $0 \leq \rho(\mathbf{X}) \leq 1$) are maintained. Moreover, in order to guarantee the identification of the model, we assume that all the means of functions f_j , g_j and m_j are zero.

Step 3: Compute the standardized residuals

$$\begin{pmatrix} \hat{\varepsilon}_{i1} \\ \hat{\varepsilon}_{i2} \end{pmatrix} = \hat{\Sigma}^{-1/2}(\mathbf{X}_i) \begin{pmatrix} Y_{i1} - \hat{\mu}_1(\mathbf{X}_i) \\ Y_{i2} - \hat{\mu}_2(\mathbf{X}_i) \end{pmatrix} \quad i = 1, \dots, n \tag{6}$$

where

$$\hat{\Sigma}(\mathbf{X}_i) = \begin{pmatrix} \hat{\sigma}_1^2(\mathbf{X}_i) & \hat{\sigma}_{12}(\mathbf{X}_i) \\ \hat{\sigma}_{12}(\mathbf{X}_i) & \hat{\sigma}_2^2(\mathbf{X}_i) \end{pmatrix} \tag{7}$$

with $\hat{\sigma}_{12}(\mathbf{X}_i) = \hat{\sigma}_1(\mathbf{X}_i)\hat{\sigma}_2(\mathbf{X}_i)\hat{\rho}(\mathbf{X}_i)$.

Step 4: Obtain the kernel estimation of the bivariate density $\hat{f}(\varepsilon_1, \varepsilon_2)$ given by

$$\hat{f}((\varepsilon_1, \varepsilon_2), \mathbf{H}) = \frac{1}{n} \sum_{i=1}^n K_{\mathbf{H}} \begin{pmatrix} \varepsilon_1 - \hat{\varepsilon}_{i1} \\ \varepsilon_2 - \hat{\varepsilon}_{i2} \end{pmatrix} \tag{8}$$

where $K(\cdot)$ is the kernel which is a symmetric probability density function and \mathbf{H} is a 2×2 positive definite matrix. Then, obtain the τ^{th} unconditional bivariate uncertainty region on the residual scale as

$$\hat{\varepsilon}_{\tau} = \{(\varepsilon_1, \varepsilon_2) \in \mathbb{R}^2 | \hat{f}(\varepsilon_1, \varepsilon_2) \geq \hat{k}\} \tag{9}$$

\hat{k} being the empirical τ quantile of the values $\hat{f}(\varepsilon_{11}, \varepsilon_{12}), \dots, \hat{f}(\varepsilon_{n1}, \varepsilon_{n2})$. Finally, for a given \mathbf{X} , the conditional bivariate uncertainty region $R_{\tau}(\mathbf{X})$ is estimated according to (3)

$$\hat{R}_{\tau}(\mathbf{X}) = \begin{pmatrix} \hat{\mu}_1(\mathbf{X}) \\ \hat{\mu}_2(\mathbf{X}) \end{pmatrix} + \hat{\Sigma}^{1/2}(\mathbf{X})\hat{\varepsilon}_{\tau} \tag{10}$$

3. Case Study: Joint Forecasting of (SO₂, NO_x) Pollution Episodes

The mathematical model exposed in the previous section was applied to the forecasting of pollution episodes registered at a coal-fired power station located in the northwest of Spain. SO₂ and NO_x are two of the main air pollutants generated by combustion processes, and both have harmful effects on human health. Moreover, it was proven that both pollutants are correlated [46], which is consistent with the model in (1). Fortunately, pollution episodes are not very frequent and the trend is that they will become scarcer as technology advances.

Let t_0 be the present time measured each five minutes, and $SO(t_0)$ and $NO(t_0)$ the concentrations obtained respectively by the series of bi-hourly SO₂ and NO_x means at instant t_0 . Being t_h the prediction horizon time, the interest is to predict

$$(Y_1, Y_2) = (SO(t_0 + t_h), NO(t_0 + t_h))$$

and provide an uncertainty region for these estimations given a specific value of τ , using the predictive covariates

$$\mathbf{X} = (X^1(t), X^2(t), X^3(t), X^4(t)) = (SO(t), NO(t), SO'(t), NO'(t)) \quad \text{with } t \in [t_0 - t_{lag}, t_0]$$

where $(NO'(t), SO'(t))$ represents the first derivatives of the functions that approximate the concentrations of both pollutants. These derivatives are obtained from the functional representation of the discrete data, according to Step 1 of the estimation algorithm. Please note that t_{lag} represents the lagged time used in the predictors. In particular,, we are interested in predicting an hour in advance, according to the requirements of current Spanish legislation and, therefore, we will consider $t_h = 12$ (60 min) from now on.

Most of the time, these concentrations times series are low, close to zero, and in order to obtain a reasonably large number of pollution incidents, we took as our sample a historical matrix $\{(\mathbf{X}_i, \mathbf{Y}_i)\}_{i=1}^N$ with pollution data of approximately 12 years, which includes a considerable number of pollution episodes (see [9] for a detailed description of the historical matrix construction). In summary, in the historical matrix not all the data are used, but only part of them, following a quantile-weighted criterion. This means that the larger the concentration, the greater the number of observations of that concentration in the sample. Figure 1 shows a sample of the historical matrix, on top are the curves of both pollutants ($NO(t), SO(t)$) measured in $t_{lag} = 20$ discretization points and evaluated in

5 B-spline basis functions of order $p = 4$. On the bottom, the first derivative of the B-spline curves ($NO'(t), SO'(t)$) with order $p - 1$ are represented.

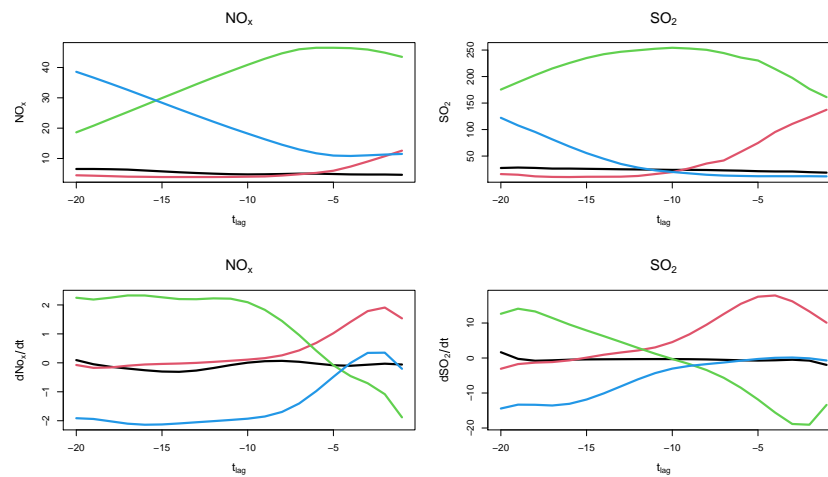


Figure 1. Five curves of the historical matrix for pollutants ($NO(t), SO(t)$) and their corresponding first derivatives ($NO'(t), SO'(t)$), observed in a period of time $t_{lag} = 20$.

In this paper, we tested four models using as predictors different combinations of the covariates ($X^1(t), X^2(t), X^3(t), X^4(t)$) that include the concentrations of both pollutants and their first derivatives. In particular, we will consider models given by:

$$\begin{pmatrix} SO(t_0 + 12) \\ NO(t_0 + 12) \end{pmatrix} = \begin{pmatrix} \mu_1(\mathbf{X}^1) \\ \mu_2(\mathbf{X}^2) \end{pmatrix} + \begin{pmatrix} \sigma_1^2(\mathbf{X}^1) & \sigma_{12}(\mathbf{X}^1, \mathbf{X}^2) \\ \sigma_{12}(\mathbf{X}^1, \mathbf{X}^2) & \sigma_2^2(\mathbf{X}^2) \end{pmatrix} \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix} \tag{11}$$

The four considered models, M_1, M_2, M_3 and M_4 , are configured in Table 1 where the cross X indicates the covariates included in each model.

Table 1. Selected models from equation in (11). Cross X indicates the covariates included in each of the four considered models. The derivatives of the functions are indicated with a single quote.

Model	X^1				X^2			
	$SO(t)$	$NO(t)$	$SO'(t)$	$NO'(t)$	$SO(t)$	$NO(t)$	$SO'(t)$	$NO'(t)$
M_1	X						X	
M_2	X	X			X	X		
M_3	X		X			X		X
M_4	X	X	X		X	X	X	X

To validate and compare the four proposed models, we randomly select from the full historical matrix a training set $\mathbf{M}^I = \{(\mathbf{X}_i^I, \mathbf{Y}_i^I)\}_{i=1}^{n_{train}}$ and a test set $\mathbf{M}^{II} = \{(\mathbf{X}_i^{II}, \mathbf{Y}_i^{II})\}_{i=n_{train}+1}^N$.

The estimates $\hat{\mu}_1, \hat{\mu}_2, \hat{\Sigma}$ were obtained from the samples in the first matrix \mathbf{M}_t^I . The bivariate uncertainty regions for the values of the covariates on the second matrix \mathbf{M}^{II} were obtained using (3).

The estimated coverage $\hat{\tau}$ is given by

$$\hat{\tau} = \frac{1}{n_{test}} \sum_{i=n_{train}+1}^N I\{\mathbf{Y}_i^{II} \in \hat{R}_\tau(\mathbf{X}_i^{II})\}; n_{test} = N - n_{train} \tag{12}$$

The performance of the proposed predictors was evaluated in two pollution incidents. A bivariate representation of these episodes is shown in Figure 2. The orientation of the points shows a clear correlation between both pollutants although the range of concentrations is quite different.

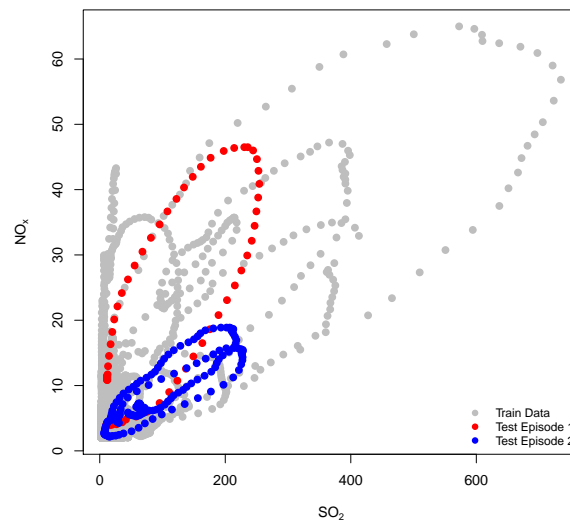


Figure 2. Observed and forecasted concentrations of SO₂ and NO_x for two pollution episodes.

The nominal and the estimated coverages for different time lags and training sample sizes are shown in Table 2. The coverages correspond to the bivariate solution, and were obtained for n_{test} consecutive observations that might or might not correspond to pollution incidents.

Table 2. Nominal τ and estimated $\hat{\tau}$ coverages for each of the four models under study. Two time lags, $t_{lag} = 10, t_{lag} = 20$, two sizes of the training sample $n_{train} = 10,000$ and $n_{train} = 4900$, and two numbers of principal components, $K = 3$ and $K = 5$, were considered. Results correspond to the test sample.

τ	t_{lag}	n_{train}^*	K	$\hat{\tau}$			
				M ₁	M ₂	M ₃	M ₄
0.50	10	20	3	0.43	0.47	0.45	0.51
			5	0.42	0.48	0.47	0.49
	10	49	3	0.51	0.52	0.52	0.52
			5	0.51	0.50	0.50	0.50
	20	20	3	0.45	0.49	0.44	0.49
			5	0.48	0.46	0.43	0.46
	20	49	3	0.50	0.54	0.51	0.50
			5	0.49	0.49	0.49	0.48
0.75	10	20	3	0.70	0.73	0.72	0.75
			5	0.69	0.74	0.73	0.74
	10	49	3	0.76	0.78	0.78	0.78
			5	0.77	0.76	0.76	0.76
	20	20	3	0.70	0.72	0.70	0.72
			5	0.70	0.71	0.69	0.69
	20	49	3	0.77	0.78	0.75	0.73
			5	0.75	0.74	0.72	0.72
0.90	10	20	3	0.88	0.87	0.87	0.89
			5	0.86	0.88	0.87	0.88
	10	49	3	0.91	0.90	0.90	0.90
			5	0.90	0.90	0.90	0.90
	20	20	3	0.87	0.87	0.85	0.86
			5	0.87	0.86	0.86	0.84
	20	49	3	0.90	0.89	0.90	0.86
			5	0.87	0.87	0.88	0.85
0.95	10	20	3	0.93	0.93	0.93	0.93
			5	0.93	0.93	0.93	0.93
	10	49	3	0.96	0.93	0.93	0.93
			5	0.95	0.94	0.94	0.94
	20	20	3	0.93	0.93	0.92	0.92
			5	0.92	0.92	0.93	0.90
	20	49	3	0.95	0.94	0.95	0.92
			5	0.92	0.92	0.93	0.90

RMSE values for each model are shown in Table 3, considering an expansion of the functions in three or five principal components, and lags $t_{lag} = 10$ and $t_{lag} = 20$. Please note that this table makes reference to the marginal distributions, and that the range of concentrations for each pollutant is very different, therefore the RMSE values are also different.

Table 3. RMSE values for two pollution episodes and the four models tested, considering curves with two different time lags, size of the training samples and number of principal components.

Response	t_{lag}	n_{train}^*	Episode 1				Episode 2				
			K	M ₁	M ₂	M ₃	M ₄	M ₁	M ₂	M ₃	M ₄
NO _x	10	20	3	20.7	25.2	20.0	23.0	1.9	1.7	1.4	1.4
			5	19.6	18.6	20.6	16.4	0.9	0.8	0.8	0.6
	100	49	3	20.9	24.7	18.2	23.8	1.8	1.9	1.4	1.4
			5	19.2	17.5	19.3	16.7	0.9	0.9	0.8	0.7
	20	100	3	34.0	24.7	19.9	19.6	3.3	5.5	1.7	3.1
			5	40.9	23.3	46.5	29.2	1.5	2.5	1.0	1.8
20	49	3	30.2	18.2	18.7	20.4	3.4	5.3	1.7	2.8	
		5	36.2	27.6	39.8	35.6	1.5	2.5	1.0	1.8	
SO ₂	10	100	3	505.0	841.0	407.5	837.3	544.8	531.9	419.8	419.1
			5	914.9	868.6	669.4	516.3	215.6	230.3	184.5	199.1
	100	49	3	686.0	685.3	515.8	518.3	481.4	484.4	338.0	361.1
			5	991.1	925.5	846.3	682.9	199.9	214.7	170.6	179.9
	20	100	3	1463.4	2172.7	825.4	1199.9	1154.5	1133.1	709.5	659.0
			5	1470.6	2531.2	1002.9	1428.5	525.5	482.1	352.0	341.3
20	49	3	1458.7	2485.4	768.4	698.4	1162.6	1125.8	644.8	628.7	
		5	1787.6	2811.0	1111.2	951.5	548.1	492.0	352.7	359.3	

Note: $n_{train} = 100 \cdot n_{train}^*$.

For the two episodes analyzed, Figure 3 shows the observed and the predicted values as well as the quantile for $\hat{\tau} = 0.95$, calculated for the test sample. The results correspond to curves observed in ten points ($t_{lag} = 10$) and represented in a basis expansion in three functional principal components. These univariate confidence intervals were respectively constructed from (11) as $\mu_1(\mathbf{X}^1) + \sigma_1(\mathbf{X}^1)\varepsilon_1^{0.95}$ and $\mu_2(\mathbf{X}^2) + \sigma_2(\mathbf{X}^2)\varepsilon_2^{0.95}$, $\varepsilon_1^{0.95}$ and $\varepsilon_2^{0.95}$ being the 0.95 quantile of the distributions of errors ε_1 and ε_2 , respectively.

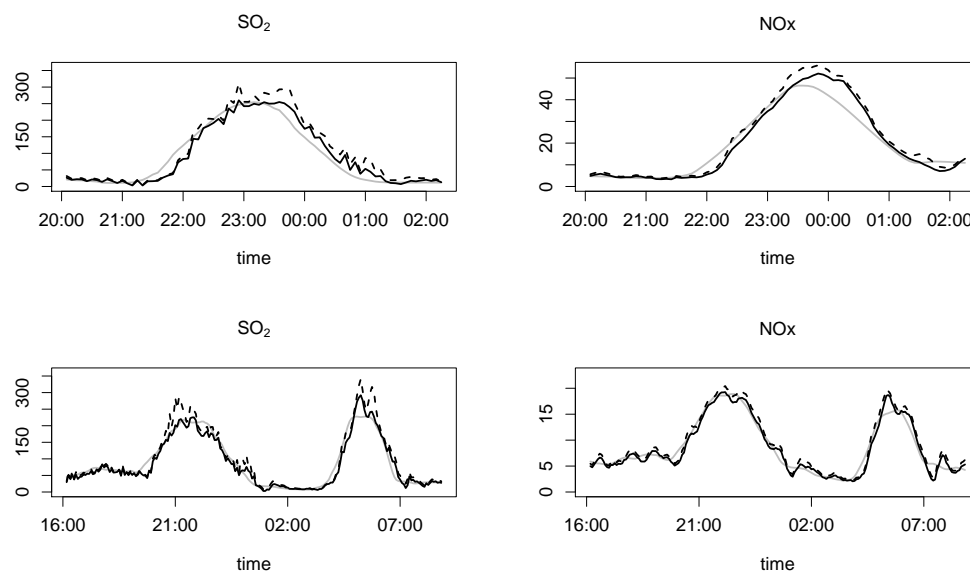


Figure 3. Observations (solid black line), mean (solid gray line) and 0.95th quantile estimations (discontinuous line) for both pollutants, SO₂ and NO_x and for two pollution episodes.

Table 4 shows the maximum consumed memory and the runtime (in seconds) for the four models tested and two different dimensions of the submatrices are executed in a Intel Core i7-2600K with 16 GB of RAM.

Table 4. Maximum memory consumption (MB) and computation time (seconds) for the four models tested, M_i , following different strategies concerning the time lag, t_{lag} , the size of the training sample, n_{train} , and the number of basis functions, K .

t_{lag}	n_{train}	K	Memory (MB)				Runtime (seconds)			
			M_1	M_2	M_3	M_4	M_1	M_2	M_3	M_4
10	20	3	652.70	1083.03	1319.57	2266.34	17.98	29.66	35.99	61.62
		5	1090.78	1855.39	2299.14	4075.88	35.55	55.29	63.15	124.7
	49	3	329.94	548.58	669.42	1153.38	10.53	17.42	21.38	37.01
		5	552.74	942.68	1171.67	2089.53	19.79	31.28	46.62	88.20
20	20	3	653.34	1084.15	1320.66	2268.16	18.11	29.63	35.97	61.21
		5	1091.90	1857.26	2300.99	4078.95	32.14	49.51	72.45	124.81
	49	3	330.10	548.84	669.69	1153.82	10.51	17.56	21.49	37.82
		5	553.01	943.13	1172.20	2090.36	17.80	36.44	39.03	76.11

Note: $n_{train} = 100 \cdot n_{train}^*$.

4. Discussion

We begin the discussion of the results analyzing Table 2 that show the estimated coverage for the bivariate prediction depending on the time lag, the size of the training sample, the number of principal components and the model used. It can be appreciated that the estimated coverages are generally lower than the theoretical coverages, although very close. Therefore, the mathematical models proposed show a good performance although there is a trend to underestimate the observed values. This effect can also be appreciated in Figure 3, where the mean tends to be under the observed values. Then, in order to be on the safe wide, it would be preferable to use the quantile $\tau = 0.95$, which provides greater guarantee of predicting the highest values of the pollution episodes. Regarding the rest of the parameters, it is not possible to establish a combination of them that provides the best results. However, in general they were obtained for the lowest training size, $n_{train} = 49,000$, and for models that includes one or two derivatives (models M_3 and M_4).

The prediction errors are shown in Table 1, where the best results (minimum RMSE) are marked in bold. As can be seen, they correspond to model M_4 for NO_x and model M_3 for SO_2 . In both cases, these models incorporate the derivatives of the original functions. Accordingly, we conclude that the derivatives contribute positively to improve the results, which reinforces the role of the functional approach. However, there is an asymmetry between both pollutants: using the concentrations of SO_2 and their derivatives improves the results for NO_x , but using the concentrations of NO_x and their derivatives is not an advantage in the estimation of SO_2 . When SO_2 and NO_x concentrations of both episodes are plotted against time (Figure 4), a slight advance can be seen on the first pollutant compared to the second, which would explain this asymmetry.

With respect to the time lag, the minimum RMSE values were obtained for the shorter period of time $t_{lag} = 10$, so it seems that using 20 observations to predict one our in advance introduced noise into the model instead of adding useful information. This result is in agreement with those obtained for the same data in previous studies of some of the authors, which indicated that only a few observations close to the time of prediction contribute to that prediction. Talking about the size of the training sample, simplifying the original data by removing small values of the concentrations improves the results in most of the cases, so this would be the advisable option.

When the effect of the number of principal components used as basis functions is analyzed, using $K = 5$ is always favorable for episode 2, for both SO_2 and NO_x , but not for episode 1, for which the trend is opposite.

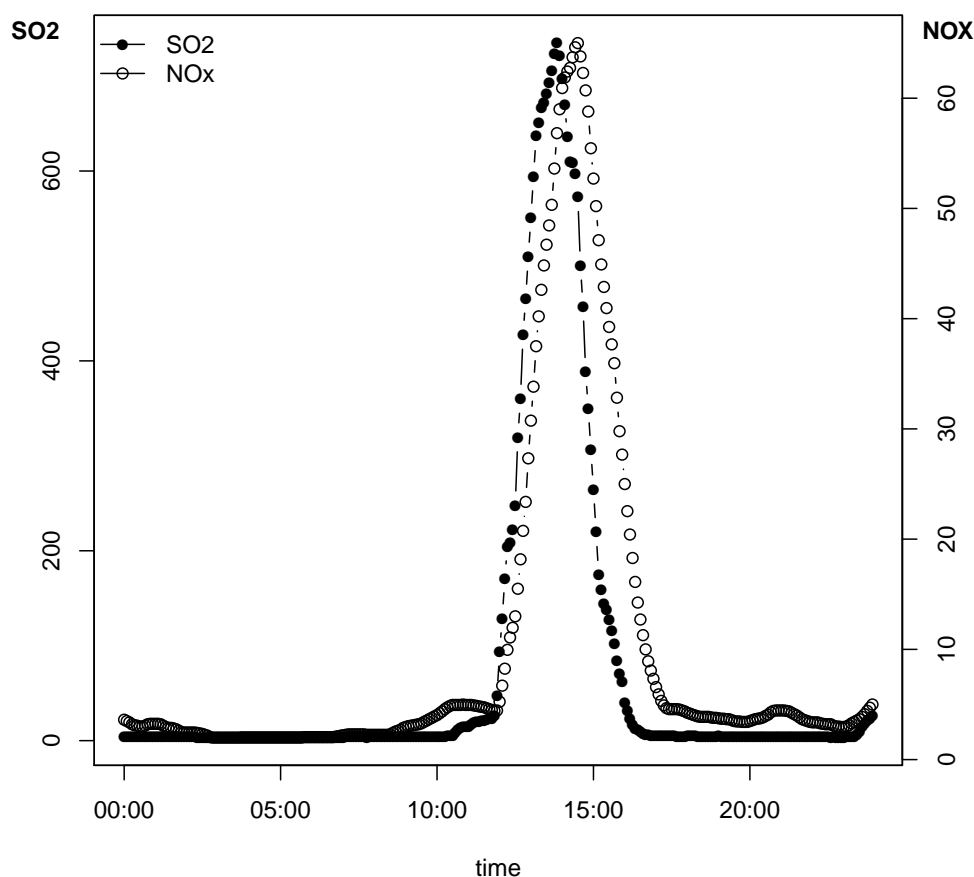


Figure 4. Example of a pollution incident showing SO₂ and NO_x concentrations versus time. Notice that there is an advance in the first pollutant compared to the second.

Although they are not shown in the article, so as not to overstretch it, a comparison of the estimated coverages using 3 or 5 principal components, or 5 B-splines basis functions, tell us that there are not substantial differences among them, so it seems that one or other base functions can be used interchangeably.

Finally, regarding memory consumption and runtime for model training, it is evident, from Table 4 that more complex models consume more resources and requires more computing time. For fixed values of the time lag (t_{lag}) and the size of the training sample (n_{train}), model M_4 is between 3 and 7 times more expensive than model M_1 in terms of memory consumption and runtime. Using time lags $t_{lag} = 10$ or $t_{lag} = 50$ has no effect in terms of computation requirements; and employing 5 principal components instead of 3 principal components implies an approximately double memory consumption and runtime.

To conclude, it is possible to establish that the functional location-scale model proposed were quite a good approach (in terms of coverage and prediction error) to forecast bivariate pollution episodes one hour in advance, as it is required by the Spanish legislation. The best results were obtained when the derivatives of functions adjusted to the observed data are included in the model, when the raw data are filtered and when the shorter period of time is used for the prediction. The size of the training data and the type and number of basis functions are, instead, parameters on which definitive conclusions could not be drawn.

Author Contributions: Conceptualization, J.R.-P. and C.O.; methodology, J.R.-P. and M.O.-d.L.F.; software, J.R.-P. and M.O.-d.L.F.; validation, J.R.-P., C.O. and M.O.-d.L.F.; writing—original draft preparation, J.R.-P. and C.O.; writing—review and editing, J.R.-P., C.O. and M.O.-d.L.F.; supervision, C.O. All authors have read and agreed to the published version of the manuscript.

Acknowledgments: The authors acknowledge financial support from: (1) UO-Proyecto Uni-Ovi (PAPI-18-GR-2014-0014), (2) Project MTM2016-76969-P from Ministerio de Economía y Competitividad—Agencia Estatal de Investigación and European Regional Development Fund (ERDF) and IAP network StUDyS from Belgian Science Policy, (3) Nuevos avances metodológicos y computacionales en estadística no-paramétrica y semiparamétrica—Ministerio de Ciencia e Investigación (MTM2017-89422-P).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Siew, L.Y.; Ching, L.Y.; Wee, P.M.J. ARIMA and integrated ARFIMA models for forecasting air pollution index in Shah Alam, Selangor. *Malay. J. Anal. Sci.* **2008**, *12*, 257–263.
2. Ibrahim, M.Z.; Roziah, Z.; Marzuki, I.; Muhd S.L. Forecasting and Time Series Analysis of Air Pollutants in Several Area of Malaysia. *Am. J. Environ. Sci.* **2009**, *5*, 625–632. [[CrossRef](#)]
3. Abhilash, M.S.K.; Thakur, A.; Gupta, D.; Sreevidya, B. Time Series Analysis of Air Pollution in Bengaluru Using ARIMA Model. In *Ambient Communications and Computer Systems; Advances in Intelligent Systems and Computing*; Perez, G., Tiwari, S., Trivedi, M., Mishra, K., Eds.; Springer: Singapore, 2018.
4. Liu, P.W.G. Simulation of the daily average PM10 concentrations at Ta-Liao with Box-Jenkins time series models and multivariate analysis. *Atmos. Environ.* **2009**, *43*, 2104–2113. [[CrossRef](#)]
5. Nazif, A.; Mohammed, N.I.; Malakahmad, A.; Abualqumboz, M.S. Regression and multivariate models for predicting particulate matter concentration level. *Environ. Sci. Pollut. Res. Int.* **2018**, *25*, 283–289. [[CrossRef](#)]
6. Zhao, R.; Gu, X.; Xue, B.; Zhang, J.; Ren, W. Short period PM2.5 prediction based on multivariate linear regression model. *PLoS ONE* **2018**, *13*, e0201011. [[CrossRef](#)]
7. Ng, K.Y.; Awang, N. Multiple linear regression and regression with time series error models in forecasting PM10 concentrations in Peninsular Malaysia. *Environ. Monit. Assess.* **2018**, *190*, 63. [[CrossRef](#)]
8. Roca-Pardiñas, J.; González Manteiga, W.; Febrero-Bande, M.; Prada-Sánchez, J.M.; Cadarso-Suárez, C. Predicting binary time series of SO₂ using generalized additive models with unknown link function. *Environmetrics* **2004**, *15*, 729–742. [[CrossRef](#)]
9. Martínez-Silva, I.; Roca-Pardiñas, J.; Ordóñez, C. Forecasting SO₂ pollution incidents by means of quantile curves based on additive models. *Environmetrics* **2016**, *27*, 147–157. [[CrossRef](#)]
10. Garcia, J.M.; Teodoro F.; Cerdeira, R.; Coelho, L.M.R.; Prashant, K.; Carvalho, M.G. Developing a methodology to predict PM10 concentrations in urban areas using generalized linear models. *Environ. Technol.* **2016**, *37*, 2316–2325. [[CrossRef](#)]
11. Roca-Pardiñas, J.; Ordóñez, C. Predicting pollution incidents through semiparametric quantile regression models. *Stoch. Environ. Res. Risk Assess.* **2019**, *33*, 673–685. [[CrossRef](#)]
12. Azid, I.A.; Ripin, Z.M.; Aris, M.S.; Ahmad, A.L.; Seetharamu, K.N.; Yusoff, R.M. Predicting combined-cycle natural gas power plant emissions by using artificial neural networks. In *Proceedings of the 2000 TENCON Proceedings, Intelligent Systems and Technologies for the New Millennium (Cat. No.00CH37119)*, Kuala Lumpur, Malaysia, 24–27 September 2000; Volume 3, pp. 512–517.
13. Perez, P.; Trier, A.; Reyes, J. Prediction of PM2.5 Concentrations Several Hours in Advance Using Neural Networks in Santiago, Chile. *Atmos. Environ.* **2000**, *34*, 1189–1196. [[CrossRef](#)]
14. Ferretti, G.; Piroddi, L. Estimation of NO_x Emissions in Thermal Power Plants Using Neural Networks. *J. Eng. Gas Turbines Power* **2001**, *132*, 465–471. [[CrossRef](#)]
15. Siwek, K.; Osowski, S. Improving the accuracy of prediction of PM10 pollution by the wavelet transformation and an ensemble of neural predictors. *Eng. Appl. Artif. Intell.* **2012**, *25*, 1246–1258. [[CrossRef](#)]
16. Muñoz, E.; Martín, M.L.; Turias, I.J.; Jimenez-Come, M.J.; Trujillo, F.J. Prediction of PM10 and SO₂ exceedances to control air pollution in the Bay of Algeciras, Spain. *Stoch. Environ. Res. Risk Assess.* **2014**, *28*, 1409–1420. [[CrossRef](#)]
17. He, H.D.; Lu, W.Z.; Xue, Y. Prediction of particulate matters at urban intersection by using multilayer perceptron model based on principal components. *Stoch Environ. Res. Risk. Assess.* **2015**, *29*, 2107–2114. [[CrossRef](#)]
18. Antanasijević, D.; Pocajt, V.; Perić-Grujić, A.; Ristić, M. Multiple-input–multiple-output general regression neural networks model for the simultaneous estimation of traffic-related air pollutants. *Atmos. Pollut. Res.* **2018**, *9*, 388–397. [[CrossRef](#)]

19. Gilson, M.; Dahmen, D.; Moreno-Bote, R.; Insabato, A.; Helias M. The covariance perceptron: A new paradigm for classification and processing of time series in recurrent neuronal networks. *BioRxiv* **2019**. [[CrossRef](#)]
20. Ramsay, J.O.; Silverman, B.W. *Applied Functional Data Analysis: Methods and Case Studies*; Springer: New York, USA, 2002.
21. Ferraty, F.; Vieu, P. *Nonparametric Functional Data Analysis*; Springer: New York, USA, 2006.
22. Febrero-Bande, M.; Galeano, P.; González-Manteiga, W. Outlier detection in functional data by depth measures with application to identify abnormal NOx levels. *Environmetrics* **2008**, *19*, 331–345. [[CrossRef](#)]
23. Martínez, J.; Saavedra, Á.; García-Nieto, P.J.; Piñeiro, J.I.; Iglesias, C.; Taboada, J.; Sanchoa, J.; Pastor, J. Air quality parameters outliers detection using functional data analysis in the Langreo urban area (Northern Spain). *Appl. Math. Comput.* **2014**, *241*, 1–10. [[CrossRef](#)]
24. Shaadan, N.; Jemain, A.A.; Latif, M.T. Anomaly detection and assessment of PM10, functional data at several locations in the Klang Valley, Malaysia. *Atmos. Pollut. Res.* **2015**, *6*, 365–375. [[CrossRef](#)]
25. Ignaccolo, R.; Mateu, J.; Giraldo, R. Kriging with external drift for functional data for air quality monitoring. *Stoch. Environ. Res. Risk Assess.* **2014**, *28*, 1171–1186. [[CrossRef](#)]
26. Wang, D.; Zhong, Z.; Kaixu, B.; Lingyun, H. Spatial and Temporal Variabilities of PM2.5 Concentrations in China Using Functional Data Analysis. *Sustainability* **2019**, *11*, 1620. [[CrossRef](#)]
27. Aneiros-Pérez, G.; Cardot, H.; Estévez-Pérez, G.; Vieu, P.H. Maximum ozone concentration forecasting by functional non-parametric approaches. *Environmetrics* **2004**, *15*, 675–685. [[CrossRef](#)]
28. Fernández de Castro, B.M.; González-Manteiga, W.; Guillas, S. Functional samples and bootstrap for predicting sulfur dioxide levels. *Technometrics* **2005**, *47*, 212–222. [[CrossRef](#)]
29. Quintela-del-Río, A.; Francisco-Fernández, M. Nonparametric functional data estimation applied to ozone data: Prediction and extreme value analysis. *Chemosphere* **2001**, *82*, 800–808. [[CrossRef](#)]
30. Besse, P.C.; Cardot, H.; Stephenson, D.B. Autoregressive forecasting of some functional climatic variations. *Scand. J. Stat.* **2000**, *27*, 673–687. [[CrossRef](#)]
31. Damon, J.; Guillas, S. The inclusion of exogenous variables in functional autoregressive ozone forecasting. *Environmetrics* **2002**, *13*, 759–774. [[CrossRef](#)]
32. Ruiz-Medina, M.D.; Espejo R.M. Spatial autoregressive functional plug-in prediction of ocean surface temperature. *Stoch. Environ. Res. Risk Assess.* **2012**, *26*, 335–344. [[CrossRef](#)]
33. Ruiz-Medina, M.D.; Espejo, R.M.; Ugarte, M.D.; Militino A.F. Functional time series analysis of spatio-temporal epidemiological data. *Stoch. Environ. Res. Risk Assess.* **2014**, *28*, 943–954. [[CrossRef](#)]
34. Alvarez-Liebana, J.; Ruiz Medina, M.D. Prediction of air pollutants PM10 by ARBX(1) processes. *Stoch. Environ. Res. Risk Assess.* **2019**, *33*, 1721–1736. [[CrossRef](#)]
35. Hsu K.J. Time series analysis of the interdependence among air pollutants. *Atm. Environ. Part B Urban Atmos.* **1992**, *26*, 491–503. [[CrossRef](#)]
36. Kadiyala, A.; Kumar, A. Vector time series models for prediction of air quality inside a public transportation bus using available software. *Environ. Prog. Sustain.* **2014**, *33*, 337–341. [[CrossRef](#)]
37. García-Nieto, P.J.; Sánchez-Lasheras, F.; García-Gonzalo, E.; de Cos Juez F.J. Estimation of PM10 concentration from air quality data in the vicinity of a major steelworks site in the metropolitan area of Avilés (Northern Spain) using machine learning techniques. *Stoch Environ. Res. Risk Assess.* **2018**, *32*, 3287–3298. [[CrossRef](#)]
38. Hedeker, D.; Mermelstein, R.J.; Demirtas, H. An Application of a Mixed-Effects Location Scale Model for Analysis of Ecological Momentary Assessment (EMA) Data. *Biometrics* **2008**, *64*, 627–634. [[CrossRef](#)]
39. Taylor, J.; Verbyla, A. Joint modelling of location and scale parameters of the t distribution. *Stat. Model.* **2004**, *4*, 91–112. [[CrossRef](#)]
40. Pugach, O.; Hedeker, D.; Mermelstein, R. A Bivariate Mixed-Effects Location-Scale Model with application to Ecological Momentary Assessment (EMA) data. *Health Serv. Outcomes Res. Methodol.* **2014**, *14*, 194–212. [[CrossRef](#)]
41. He, W.; Lawless, J.F. Bivariate location-scale models for regression analysis, with applications to lifetime data. *J. R. Statist. Soc. B* **2005**, *67 Pt 1*, 63–78. [[CrossRef](#)]
42. Jäntschi, L.; Bálint, D.; Bolboaca, S.D. Multiple Linear Regressions by Maximizing the Likelihood under Assumption of Generalized Gauss-Laplace Distribution of the Error. *Comput. Math. Methods Med.* **2016**, *2016*, 8578156. [[CrossRef](#)]

43. Rigby, R.A.; Stasinopoulos, D.M. Generalized additive models for location, scale and shape. *J. R. Stat. Soc. Ser. C* **2005**, *54*, 507–554. [[CrossRef](#)]
44. Karhunen, K. Zur Spektraltheorie Stochastischer Prozesse. *Annales Academiae Scientiarum Fennicae Series A1 Mathematica-Physica* **1946**, *54*, 1–7. Available online: <https://katalog.ub.uni-heidelberg.de/cgi-bin/titel.cgi?katkey=67295489> (accessed on 24 February 2020).
45. Febrero-Bande, M.; Oviedo de la Fuente, M. Statistical Computing in Functional Data Analysis: The R Package fda.usc. *J. Stat. Softw.* **2012**, *51*, 12. [[CrossRef](#)]
46. Dogruparmak, S.C.; Özbay, B. Investigating Correlations and Variations of Air Pollutant Concentrations under Conditions of Rapid Industrialization–Kocaeli (1987–2009). *Clean-Soil Air Water* **2011**, *39*, 597–604. [[CrossRef](#)]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).