Advances in Pure Mathematics, 2018, 8, 793-812
http://www.scirp.org/journal/apm
ISSN Online: 2160-0384
ISSN Print: 2160-0368

Comparison of Different Regularized and Shrinkage Regression Methods to Predict Daily Tropospheric Ozone Concentration in the Grand Casablanca Area

Halima Oufdou[1*], Lise Bellanger[2], Amal Bergam[1], Angélina El Ghaziri[3], Kenza Khomsi[4], El Mostafa Qannari[5]

[1]Laboratory MAE2D, University of Abdelmalek Essaadi, Larache, Morocco
[2]Laboratory of Mathematics Jean Leray UMR CNRS 6629, University of Nantes, Nantes, France
[3]Maison des Sciences de l'Homme Ange Guépin USR 3491, University of Nantes, Nantes, France
[4]National Meteorological Office, Casablanca, Morocco
[5]StatSC, ONIRIS, INRA, Nantes, France
Email: *oufdouhalima@gmail.com

How to cite this paper: Oufdou, H., Bellanger, L., Bergam, A., El Ghaziri, A., Khomsi, K. and Qannari, E.M. (2018) Comparison of Different Regularized and Shrinkage Regression Methods to Predict Daily Tropospheric Ozone Concentration in the Grand Casablanca Area. *Advances in Pure Mathematics*, **8**, 793-812.
https://doi.org/10.4236/apm.2018.810049

Received: July 21, 2018
Accepted: October 23, 2018
Published: October 26, 2018

Abstract

Tropospheric ozone (O3) is one of the pollutants that have a significant impact on human health. It can increase the rate of asthma crises, cause permanent lung infections and death. Predicting its concentration levels is therefore important for planning atmospheric protection strategies. The aim of this study is to predict the daily mean O3 concentration one day ahead in the Grand Casablanca area of Morocco using primary pollutants and meteorological variables. Since the available explanatory variables are multicollinear, multiple linear regressions are likely to lead to unstable models. To counteract the multicollinearity problem, we compared several alternative regression methods: 1) Continuum Regression; 2) Ridge & Lasso Regressions; 3) Principal component regression (PCR); 4) Partial least Square regression & sparse PLS and; 5) Biased Power Regression. The aim is to set up a good prediction model of the daily ozone in the Grand Casablanca area. These models are fitted on a training data set (from the years 2013 and 2014), tested on a data set (from 2015) and validated on yet another data set data (from 2015). The Lasso model showed a better performance for the prediction of ozone concentrations compared to multiple linear regression and its other alternative methods.

Keywords

Multiple Linear Regression, Multicollinearity, Penalized Regression, Statistical Forecasting, Tropospheric Ozone

1. Introduction

Tropospheric ozone (O3) is a dangerous air pollutant that threatens the human health [1]. Indeed, epidemiologic studies have shown that current ambient exposures are associated with reduced baseline lung function, exacerbation of asthma and premature mortality [2]. It is a secondary trace gas in the atmosphere, not directly emitted from any natural or anthropogenic source, but rather formed through a complex set of several chemical reactions in presence of sunlight [3].

As all the large cities in the world, Casablanca has a serious photochemical tropospheric ozone (O3) air pollution problem. The urban emission pattern of O3-forming pollutants is caused by meteorological factors: exposure to sunshine, temperature and wind speed and also by a series of atmospheric reactions involving precursor pollutants caused by car and industry emissions.

Various statistical methods are available to predict daily O3 [4] [5] [6] [7] Multiple linear regression (MLR) is frequently used by several environmental protection agencies involved in air quality monitoring (e.g. [8] [9] [10] etc.). The prediction ability of this type of models is generally satisfactory, notwithstanding the fact that, very often, the predictor variables are highly collinear. In the following, MLR will stand as the standard method to which alternative methods will be compared. In order to tackle the multicollinearity issue, various methods are proposed in the literature [11]. Ridge Regression [12] was certainly the first method proposed in this context. This method of analysis is based on a regularization strategy which aims at constraining the length (as measured by the L2 norm) of the vector of regression coefficients to be relatively small. Similarly, Lasso regression [13] follows the same principle as Ridge Regression, but, this time, the length of the regression coefficients is measured by L1 norm. Other alternative methods to MLR encompass Principal Component Regression [14] and Partial Least Squares regression [15] [16] [17]. These latter techniques were combined into a single approach, Continuum Regression (CR), proposed by Stone and Brooks [18]. Sundberg [19] shows that CR is also related to Ridge regression. Recently, a new biased regression strategy consisting in gradually shedding off the correlations among the independent variables was proposed by Qannari and El Ghaziri [20].

In this study, we compare different regression models to predict the daily mean O3 concentration in the Grand Casablanca area using O3 persistence and meteorological variables. We follow two successive stages. In the first stage, we fit statistical models using two years (2013-2014, calibration sets) of pollutants and observed meteorological data. In the second stage, in order to choose the best predictive model, we compare the prediction abilities using the observed dataset of 2015 (test set) and the meteorological forecasting dataset of 2015 (prediction dataset test). The aim of the study is to select the best model in terms of prediction ability.

2. Material and Methods

2.1. Data

The three years datasets used in this study are provided by the National Meteorological Office of Morocco (DMN). Their collection ranged from January 2013 to December 2015. A detailed description of the 25 available variables is given in **Appendix A**. The data consist of daily O3 pollutant concentrations, observed at "Jahid" monitoring site, located in the western center of the Casablanca city which is the most important industrial area (Figure 1). Following the DMN recommendation, we use in this study 23 meteorological variables such as temperature, humidity, duration of sunshine, wind direction, wind speed, precipitation, pressure, etc. also measured at the center of Casablanca. After a step of pre-processing of these data which involved in particular the imputation of missing values, we dispose of a dataset containing for each day the observed meteorological data, forecasted meteorological data acquired from the numeric model ALADIN-Maroc and the measured O3 concentrations. The period of the study is limited to the hot and sunny season (April-September) when ozone concentrations are at their maximum [21].

2.2. Exploratory Data Analysis

Inevitably, the collected data contain missing values and it is important to tackle

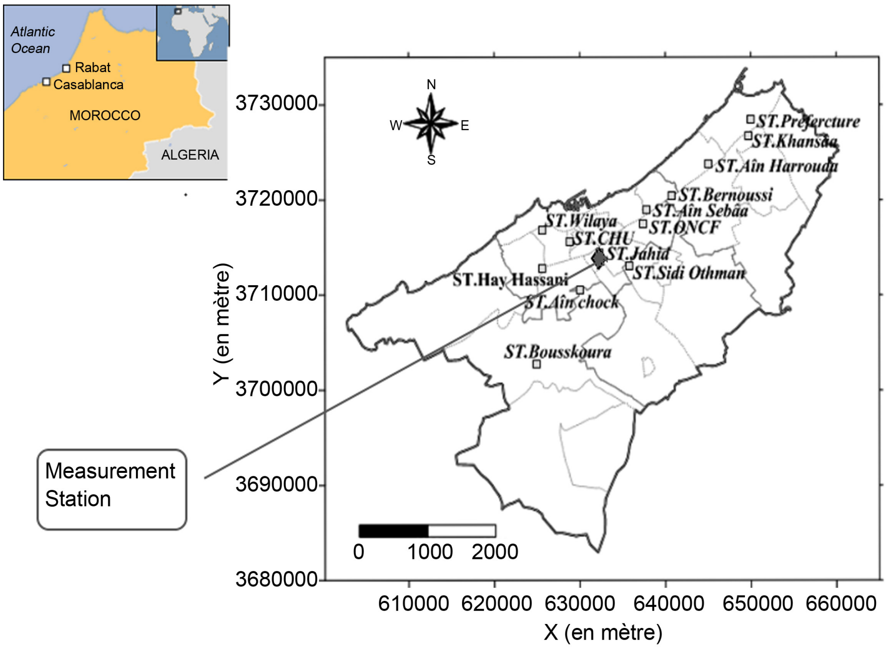

Figure 1. Map of Grand Casablanca area in the western center of Morocco. The measurement station El-Jahid is an urban one located in the center of Casablanca.

this problem before further analyses are performed. The reasons for which a value can be missing are numerous. For instance, in air quality applications, data can be missing due to a dysfunction of the equipment or an insufficient resolution of a sensor device. Therefore, it is necessary to identify missing values and choose an appropriate imputation technique in order to keep as much data as possible e.g. [22]. Various imputation procedures are used in practice. We can cite for instance regression imputation, nearest-neighbour imputation, random hot-deck imputation [23] [24] [25]. We choose the K-nearest neighbors (KNN) strategy because it is simple and efficient. With this technique, the imputation is based on the neighboring observations to each missing value [26] [27]. More precisely, missing values are replaced by values extracted from cases that are similar to the recipient with respect to the observed (*i.e.* non missing) characteristics.

Once the data were imputed, we applied a standardized Principal Components Analysis (PCA) to investigate the relationships between the variables and assess the degree of collinearity among the predictor variables [28] [29] [30].

2.3. Linear Modelling Approach

Several statistical models are available to predict tropospheric ozone concentration. Since the available explanatory variables are potentially highly correlated, we investigate alternative methods to the classical Multiple Linear Regression (MLR) that circumvent the problem of multicollinearity which is likely to lead to unstable models. The emphasis is put on: 1) classical regularized regression methods: Principal Components Regression [14], Partial Least Squares Regression [16], Sparse PLS [31] [32] and Continuum Regression introduced by Stone and Brooks in 1990; 2) Penalized regression methods: Ridge [12] and Lasso developed by Tibshirani [13] and finally; 3) Biased Power Regression recently introduced by Qannari and El Ghaziri [20].

2.3.1. Multiple Linear Regression (MLR)

We assume the MLR model using Equation (1):

$$O_{3_i} = \beta_0 + \beta_1 O_{3_{i-1}} + \sum\nolimits_{j=2}^{p} \beta_j varmeteo_i^j + e_i \tag{1}$$

where O_{3_i} : ozone concentration at day *i*; $O_{3_{i-1}}$: ozone concentration at day $i-1$ (*i.e.* the persistence); $varmeteo_i^j$: Meteorological variable *j* observed on day *i*.

Equation (1) can be written using usual matrix format after centring of the response variable:

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{e} \tag{2}$$

where $\boldsymbol{y}$ is an $(n \times 1)$ vector of centered dependant variable (O3 concentrations at day *i*), $\boldsymbol{X}$ is a $(n \times p)$ matrix of standardized predictors (Observed meteorological variables and O3 concentrations at day $i-1$), $\boldsymbol{\beta}$ is an $(p \times 1)$ vector of unknown regression coefficients and $\boldsymbol{e}$ is an $(n \times 1)$ vector of random errors. Classically, the distribution of $\boldsymbol{e}$ is assumed to be normal with mean equal to 0 and a variance covariance matrix equal to $\sigma^2 \boldsymbol{I}$; where $\boldsymbol{I}$ is the identity matrix.

The usual unbiased Ordinary Least Squares (OLS) estimator is expressed by (3) [33]:

$$\hat{\boldsymbol{\beta}}_{\text{OLS}} = \left(\boldsymbol{X}^{\text{T}}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^{\text{T}}\boldsymbol{y} \tag{3}$$

The prediction of $\boldsymbol{y}$ using OLS, $\hat{\boldsymbol{y}}_{\text{OLS}}$, is given by $\hat{\boldsymbol{y}}_{\text{OLS}} = \boldsymbol{X}\hat{\boldsymbol{\beta}}_{\text{OLS}}$. It is well known that this estimator is likely to lead to an unstable model and poor predictions in presence of quasi-collinearity among the predictors or in the case of the small sample and high dimensional setting.

2.3.2. Principal Component Regression (PCR)

The principal components regression (PCR) approach involves running PCA on the predictor variables and, thereafter, using the first m principal components (PC) with $1 \le m \le p$, $\boldsymbol{F}^1, \cdots, \boldsymbol{F}^m$, as the predictors in a linear regression model [14] [34].

The appropriate number, m, of first principal components to be introduced in the model can be determined in practice by a validation technique such as Leave One Out cross validation (LOO). The PCR model can be written as in Equation (4):

$$\boldsymbol{y} = \boldsymbol{F}_{(m)}\boldsymbol{\beta}_{\text{PCR}(m)} + \boldsymbol{e} \tag{4}$$

where $\boldsymbol{F}_{(m)}$ is a matrix with m columns containing the m first PCs. The Ordinary Least Squares (OLS) estimator of $\boldsymbol{\beta}_{\text{PCR}}$ is given by:
$\hat{\boldsymbol{\beta}}_{\text{PCR}} = \left(\boldsymbol{F}_{(m)}^{\text{T}}\boldsymbol{F}_{(m)}\right)^{-1}\boldsymbol{F}_{(m)}^{\text{T}}\boldsymbol{y}$.

It is easy to express this model in terms of the original variables by remarking that $\boldsymbol{F}_{(m)} = \boldsymbol{X}\boldsymbol{U}_{(m)}$, where $\boldsymbol{U}_{(m)}\left(\boldsymbol{U}^{\text{T}}\boldsymbol{U} = \boldsymbol{I}_p\right)$ is the matrix containing the m dominant normalised eigenvectors of $\boldsymbol{X}^{\text{T}}\boldsymbol{X}$. It follows that $\hat{\boldsymbol{\beta}}_{\text{PCR}} = \boldsymbol{U}_{(m)}^{\text{T}}\hat{\boldsymbol{\beta}}_{\text{OLS}}$.

PCR gives a biased estimate of the regression coefficients. If all the PCs are included in the model, we retrieve the usual MLR estimator, $\hat{\boldsymbol{\beta}}_{\text{OLS}}$.

2.3.3. Partial Least Squares Regression (PLS)

As with PCR, PLS regression, introduced by [35], consists in regressing y on components, also called latent variables, which are linear combinations of the p predictor variables.

The major difference between PCR and PLS regression is that, whereas PCR uses only $\boldsymbol{X}$ to construct the components to be used as regressors, PLS regression uses both $\boldsymbol{X}$ and $\boldsymbol{y}$ to determine such components. More precisely, the PLS components are determined sequentially and, at each step, we seek to determine a new component, constrained to be orthogonal to the components determined at the previous stages, so as to maximize the covariance between this component and the independent variable.

Suppose that m PLS components are determined. Again, the number m of latent variables to be introduced in the model can be selected by means of LOO cross validation technique in practice. These latent variables can be stacked into a matrix $\boldsymbol{T}_{(m)}$, $\boldsymbol{T}_{(m)} = \boldsymbol{X}\boldsymbol{W}_{(m)}$ where $\boldsymbol{W}_{(m)} = \left(\boldsymbol{w}_1, \cdots, \boldsymbol{w}_m\right)$ is the $\boldsymbol{X}$-weight matrix. Equation (5) gives the vector of fitted values obtained by regressing $\boldsymbol{y}$ on $\boldsymbol{T}_{(m)}$, namely:

$$\hat{\boldsymbol{y}}_{\text{PLS}}^{m} = \boldsymbol{T}_{(m)}\left(\boldsymbol{T}_{(m)}^{\mathrm{T}}\boldsymbol{T}_{(m)}\right)^{-1}\boldsymbol{T}_{(m)}^{\mathrm{T}}\boldsymbol{y} \tag{5}$$

We can show in Equation (6) the expression of PLS regression coefficients as:

$$\hat{\boldsymbol{\beta}}_{\text{PLS}} = \boldsymbol{W}_{(m)}\left(\boldsymbol{P}_{(m)}^{\mathrm{T}}\boldsymbol{W}_{(m)}\right)^{-1}\left(\boldsymbol{T}_{(m)}^{\mathrm{T}}\boldsymbol{T}_{(m)}\right)^{-1}\boldsymbol{T}_{(m)}^{\mathrm{T}}\boldsymbol{y} \tag{6}$$

where $\boldsymbol{P}_{(m)} = \boldsymbol{X}^{\mathrm{T}}\boldsymbol{T}_{(m)}\left(\boldsymbol{T}_{(m)}^{\mathrm{T}}\boldsymbol{T}_{(m)}\right)^{-1}$ and $\boldsymbol{W}_{(m)} = \boldsymbol{X}^{\mathrm{T}}\boldsymbol{y}/\left\|\boldsymbol{X}^{\mathrm{T}}\boldsymbol{y}\right\|$, where $\|\,\|$ is the L^2 norm.

PLS regression is often helpful to reduce the number of predictors to a small number of latent variables constructed by linear combinations of the columns of original predictors. It yields a biased estimate of the regression coefficients.

2.3.4. Sparse PLS Regression (SPLS)

The Sparse PLS method defined by [32] is a direct adaptation of the PLS regression method. It allows us to operate a dimensionality reduction using regression PLS.

In SPLS regression, $\boldsymbol{w}$ the first vector of loadings is sought as an optimal solution to:

$$\max_{w}\left(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{M}\boldsymbol{w}\right) \text{ subject to } \left\|\boldsymbol{w}^{\mathrm{T}}\boldsymbol{w}\right\| = 1,\ \|\boldsymbol{w}\|_1 \leq \eta,$$

where $\boldsymbol{M} = \boldsymbol{X}^{\mathrm{T}}\boldsymbol{y}\boldsymbol{y}^{\mathrm{T}}\boldsymbol{X}$, $\|\boldsymbol{w}\|_1$ is the L^1-norm of vector $\boldsymbol{w}$ and $\eta > 0$ is a scalar which controls the degree of sparsity.

The regression coefficients estimation of $\boldsymbol{y}$ on $\boldsymbol{X}$ is calculated in the following way: The coefficients of the non-selected variables are set to 0, and the coefficients of the selected variables are those obtained by means of the "standard" PLS regression. We can also give an expression of the SPLS regression coefficients defined by (7) [32]

$$\left(\hat{\boldsymbol{\beta}}_{\text{SPLS}}\right)_j = \begin{cases} \left(\hat{\boldsymbol{\beta}}_{\text{PLS}}\right)_j, & \text{if } \boldsymbol{w}_j \neq 0 \text{ and } j = 1, \cdots, m \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

The interest of the SPLS is two folds. On the one hand, thanks to the sparsity, it yields an easy to interpret model and, on the other hand, it prevents the problem of multicollinearity by using the PLS framework. SPLS estimator is biased comparing to OLS estimator. Moreover, SPLS is computationally efficient with a tunable sparsity parameter to select the important variables.

2.3.5. Continuum Regression (CR)

The CR prediction model is chosen from a continuum of candidates among which we find methods of analysis related to OLS estimation, PCR, PLSR. [19] gives a general overview of the continuum approach regression and shows how different methods relate to "least squares ridge regression". As with PCR and PLS regression, CR consists in a regression upon latent variables (*i.e.* optimal linear combinations of the independent variables). More precisely, these latent variables are determined in a sequential way where at each stage, a latent variable is defined so as to realize a balance between stability (as assessed by the variance of the latent variable) and prediction ability (as assessed by the correlation between the latent variable and the dependent variable $\boldsymbol{y}$). The prediction of $\boldsymbol{y}$ using CR, $\hat{\boldsymbol{y}}_{\text{CR}}$, is given by $\hat{\boldsymbol{y}}_{\text{CR}} = \boldsymbol{X}\hat{\boldsymbol{\beta}}_{\text{CR}}$. The CR achieves a reduction of the variance of estimator at the cost of introducing a small bias [18].

CR aims at transforming the explanatory variables into new latent predictors which are orthogonal to each other and constructed as linear combinations of the original predictors. It makes it possible to circumvent the problem of multicollinearity between predictors. However, the CR regression does not specifically aim at selecting a subset of variables [18].

2.3.6. Penalized Regression

Another general strategy to circumvent the problem of multicollinearity consists in imposing a constraint on the vector of regression coefficients. The two most popular methods in this context are Ridge and Lasso regressions.

- Ridge regression

Ridge regression is the first regularization procedure that was proposed to cope with the multicollinearity problem [12]. The Ridge estimator is given by (8):

$$\hat{\boldsymbol{\beta}}_{\text{R}} = \left(\boldsymbol{X}^{\text{T}}\boldsymbol{X} + k\boldsymbol{I}\right)^{-1}\boldsymbol{X}^{\text{T}}\boldsymbol{y} \tag{8}$$

where $k \geq 0$ is a constant to be selected. Note that if $k = 0$, the Ridge estimator amounts to the least-squares estimator.

Ridge estimator is obtained as a solution to the following least squares problem defined by (9):

$$\hat{\boldsymbol{\beta}}_{R} = \arg \min_{\boldsymbol{\beta}\in\mathbb{R}^{p}, \|\boldsymbol{\beta}\|^{2} \leq \delta} \left(\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|^{2}\right) \text{ where } \delta \geq 0 \tag{9}$$

There is a one to one correspondence between the Ridge parameter k and the upper bound, δ, imposed on the vector of regression coefficients, $\boldsymbol{\beta}$. From a practical point of view, these parameters can be selected by means of a cross-validation technique.

The Ridge regression shrinks the OLS estimators towards 0. It yields a biased estimator, but with a smaller variance than that of OLS estimator.

- Lasso regression

The Least Absolute Shrinkage and Selection Operator, or Lasso [13] is another penalized regression where L^2 penalty of ridge regression is replaced by an L^1 penalty: $\|\boldsymbol{\beta}\|_1 = \sum_{j=1}^{p}|\beta_j|$. This is a subtle change that has important consequences. Indeed, this constraint entails that some of the regression coefficients are shrunk exactly to zero. This means that this regression strategy operates de facto a selection of variables since the unimportant variables are discarded, their regression coefficients being equal to zero. Formally, the lasso estimator is given as a solution to the following optimization problem by (10):

$$\hat{\boldsymbol{\beta}}_{\text{Lasso}} = \arg \min_{\boldsymbol{\beta} \in \mathbb{R}^p, \|\boldsymbol{\beta}\|_1 \le \delta} \left(\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|^2 \right), \tag{10}$$

where $\delta \ge 0$.

The parameter δ controls the degree of sparsity and, in practice; it is determined by Leave One Out (LOO) cross validation procedure. The smaller this parameter is, the larger is the number of discarded variable. Contrariwise, if δ is larger than $\delta_0 = \sum_{j=1}^{p} \left| \hat{\boldsymbol{\beta}}_j \right|$ (where $\hat{\boldsymbol{\beta}}_j$ are the OLS estimators) then $\hat{\boldsymbol{\beta}}_{\text{Lasso}} = \hat{\boldsymbol{\beta}}_{\text{OLS}}$. Lasso regression has the double effect of shrinking the β coefficients, allowing to decrease the variance of the regression coefficients as with Ridge regression, and, more importantly, to operate an automatic selection of the variables, by cancelling out some β_j coefficients.

2.3.7. Biased Power Regression (BPR)

Recently, a new biased regression called Biased Power regression (BPR) strategy was proposed [20]. It consists in gradually shedding off the correlation among the independent variables by means of a tuning parameter α. More precisely, the BPR estimator of $\boldsymbol{\beta}$ is given by (11):

$$\hat{\boldsymbol{\beta}}_{\text{BP}} = \left(\boldsymbol{X}^{\text{T}} \boldsymbol{X} \right)^{\alpha - 1} \boldsymbol{X}^{\text{T}} \boldsymbol{y} \tag{11}$$

where α is a tuning parameter which ranges between 0 and 1.

In practice, α is selected using a cross validation procedure.

Clearly, when $\alpha = 0$, we retrieve the OLS estimator and as α increases, the variance-covariance matrix of the predictor variables is shrunk to the identity matrix. The prediction of y using BPR, $\hat{\boldsymbol{y}}_{\text{BP}}$ is given by $\hat{\boldsymbol{y}}_{\text{BP}} = \boldsymbol{X}\hat{\boldsymbol{\beta}}_{\text{BP}}$.

BP-regression shares the same properties as Ridge regression (see Section 2.3.4) and thus can highlight those variables whose coefficients become very small. However, it was not designed to select a subset of variables [20].

2.4. Evaluation of the Methods

To assess the prediction ability of the various models listed above on the Grand Casablanca O3 data, we performed a cross validation technique on a training set to determine the appropriate parameters (number of components, Ridge or lasso constant…) to be used in the prediction models. Using these parameters, the performance of the different models is assessed on the basis of a fresh data set. More precisely, we partitioned the available data into two complementary datasets: 1) summer period of 2013 and 2014 (called the training set) used to adjust the models; and 2) summer period of 2015 (called the validation set or testing set) used to "test" the models obtained in the training phase. The models are fitted on the training set used to predict the ozone responses for a) the observed meteorological data on 2015 of the validation set (obstest) and b) the forecasted meteorological data on 2015 for real validation (prevtest).

The performance of the models is measured with standard indicators defined by Equations (12)-(14) generally used to compare statistical models [36].

In a first stage, an internal validation (2013 and 2014 datasets) is performed on the basis of the following criteria in order to assess the quality of the model adjustment:

The multiple correlation coefficient R^2 allows us to assess the quality of the adjustment based on the training set: $R^2 = \frac{\sum_{i=1}^{n_{\text{train}}} \left(y_i - \hat{y}_i \right)^2}{\sum_{i=1}^{n_{\text{train}}} \left(y_i - \bar{y} \right)^2}$, where n_{train} is the size of training sample.

The Root Mean Squared Errors (RMSE): This is computed according to the following expression:

$$\text{RMSE} = \sqrt{\frac{1}{n_{\text{train}}} \sum_{i=1}^{n_{\text{train}}} \left(y_i - \hat{y}_i \right)^2} \qquad (12)$$

The smallest value of this criterion corresponds to the best adjustment of the model.

For the external validation (on summer 2015 observed dataset), the following criterion is used to assess the prediction ability of the models [37]:

The Root Mean Squared Errors of Prediction (RMSEP). This criterion is similar to RMSE but, this time, the validation data set is used instead of the training data set.

$$\text{RMSEP}_{\text{obs}} = \sqrt{\frac{1}{n_{\text{obstest}}} \sum_{i=1}^{n_{\text{obstest}}} \left(y_i - \hat{y}_i \right)^2} \qquad (13)$$

where n_{obstest} is the size of the observed the validation set (obstest).

Obviously, the best predictive model corresponds to the smallest RMSEP.

The following criterion is used to assess the performance of the models with observed meteorological data (obstest) and real meteorological forecast data (prevtest) for summer period of 2015.

In the same way, we define the RMSEP of prevision based on the forecasted dataset as:

$$\text{RMSEP}_{\text{prev}} = \sqrt{\frac{1}{n_{\text{prevtest}}} \sum_{i=1}^{n_{\text{prevtest}}} \left(y_i - \hat{y}_i \right)^2} \qquad (14)$$

where n_{prevtest} is the size of the sample size of the forecasted data (prevtest).

3. Results and Discussion

Experiments were run on an Intel(R) Core(TM) i7-6600U CPU computer with 2.60 GHz, 8 Go in RAM, Windows 10 Professional 64 bits.

All the statistical analyses were performed using the free software R. (http://www.rproject.org/).

3.1. Data Description

In this study, the dataset is composed of 25 explanatory variables. **Appendix A** gives the abbreviation of these variables.

Table 1 provides descriptive statistics of the meteorological variables and tropospheric ozone concentrations calculated on the data with missing values.

Table 1. Statistics of measured variables at Grand Casablanca area from 01 April 2013 to 30 September 2014.

Variable	Min	Max	Mean	St.Dev	NA
TMPMAX	16.2	37.5	24.5	3.09	0
TMPMIN	8.20	23.50	18.35	3.02	0
TMPMOY	12.40	29.90	21.45	2.88	0
RRQUOT	0.00	19.30	0.39	1.98	0
DRINSQ	0.00	13.30	9.72	2.79	0
HUMREL06h	50.00	100.0	87.42	8.00	0
HUMREL12h	34.00	95.00	68.32	8.78	0
HUMREL18h	28.00	97.00	75.66	9.66	0
PRESTN06h	9997.7	1017.3	1008.2	2.97	0
PRESTN12h	997.7	1016.5	1008.9	2.91	0
PRESTN18h	999	1016	1008	2.88	0
FFVM06h	0.00	4.00	1.55	0.80	3
FFVM12h	0.00	6.00	3.58	0.98	4
FFVM18h	0.00	7.00	3.46	1.04	4
DDVM06degre	0.00	360.0	176.4	117.87	3
DDVM12hDEG	0.00	360.0	227.3	141.63	4
DDVM18hDEG	0.00	360.0	189.2	152.21	4
Vx06	−2.95	3.46	−0.05	1.06	3
Vx12	−5.91	3.94	−0.59	1.98	4
Vx18	−5.91	4.50	−0.10	1.84	4
Vy06	−4.00	4.00	0.08	1.38	3
Vy12	−3.06	6.00	2.75	1.39	4
Vy18	−5.36	6.00	2.79	1.36	4
O3veilleJahid	10.00	130.0	52.83	25.66	23
O3Jahid	10.00	130.0	52.84	25.62	23

Minimum, maximum, mean and standard deviation statistics are provided to describe the characteristics of the data set.

The 2013 and 2014 studied periods are characterized by high temperatures. We can notice that, in the Grand Casablanca Area, the maximal temperature (TMPMAX) can go up to 37.5˚C and the minimal temperature is 16.2˚C. The maximal daily total sunshine duration is of 13.3 hours. We can also notice that there is almost no rain is these periods (RRQUOT). The Wind strength average is relatively high at 18 hours (FFVM18h = 7 m/s). The O3 concentrations are between 10 and 130 μg/m^3.

There are in total 90 missing values for the 366 recording days, distributed on 14 variables. This represents around 2% of missing values to be imputed before the prediction models are performed.

3.1.1. Missing Values Imputation

As mentioned above, a strategy based on the K-nearest neighbors was performed. Different K values were used in the literature and the choice of K = 10 led to the best results [38].

3.1.2. Multivariate Analysis

Figure 2 shows the scatter plots associated with pairs of the available variables. It highlights the pairwise correlations between these variables. We also indicate in Figure 2, the histograms associated with each variable and the correlation coefficients between each pair of variables. For instance, we can see that there is a high correlation between O3veilleJahid and O3 Jahid (the two last columns) with a correlation coefficient equal to 0.92. We can also observe large correlations (around 0.94) between the first three explanatory variables (TMPMAX and TMPMOY, TMPMIN and TMPMOY).

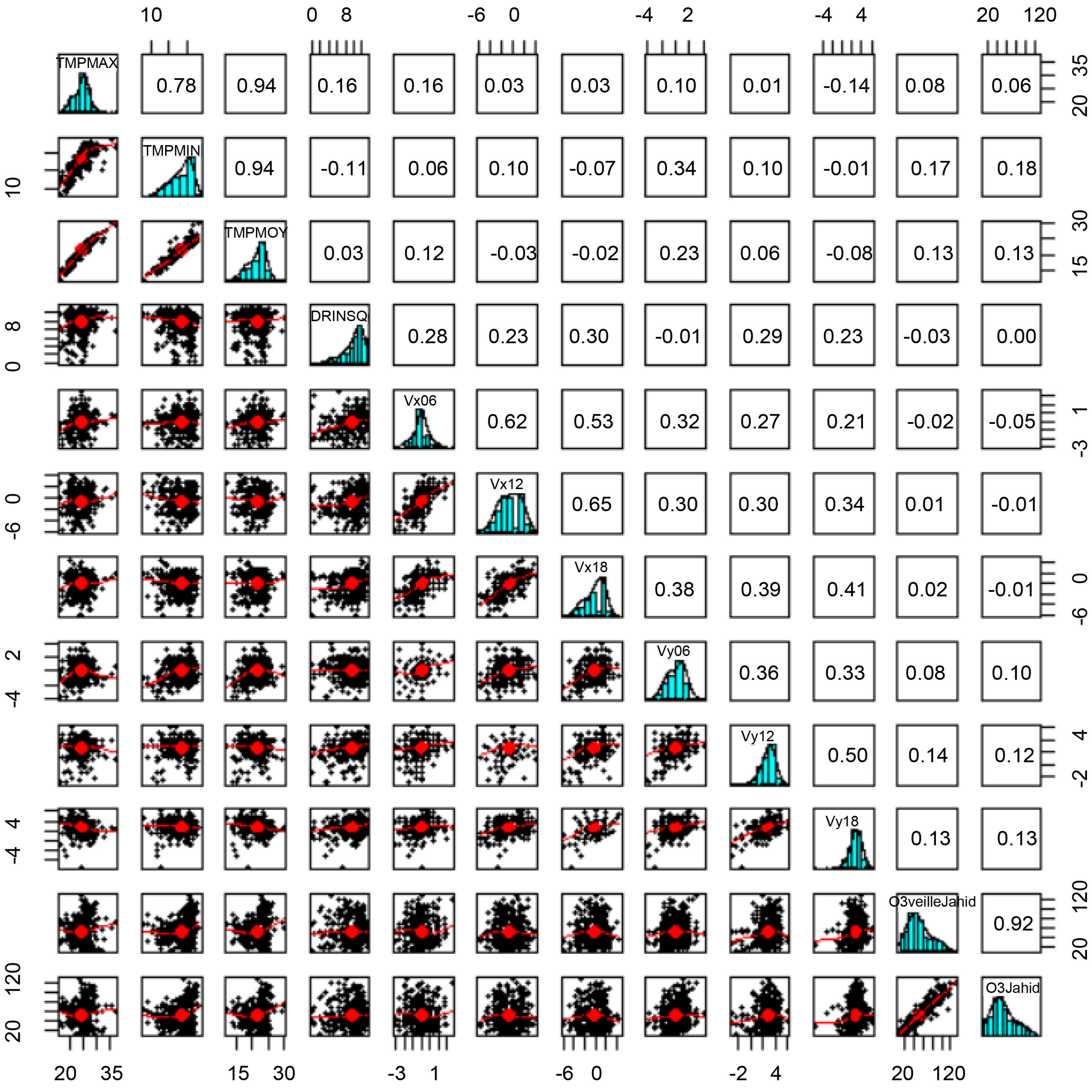

Figure 2. Scatter plots highlighting the correlations between pairs of variables.

The diagonal entries show the histograms associated with the various variables and the upper entries indicate the coefficients of correlations between pairs of variables.

We also performed a PCA on the imputed dataset. PCA is run on the complete data (2013 and 2014) after imputation and standardization of the variables. The data is composed of 366 days (from April to September of 2013 and 2014) and 24 variables. The first five principal components recover up to 65% of the total variance (Table 2). In the following, only the results related to the first two principal components which recover around 40% of the total inertia are shown.

Figure 3 shows the correlations of the explanatory variables with the first two principal components. The variable O3 Jahid is superimposed as an illustrative variable with a blue arrow to depict its relationships with the explanatory variables. This figure highlights the strong correlations among the variables, which may be harmful for the prediction models.

The first PC is linked to wind direction (Vx06, Vx12, Vy12 and Vx18) and pressure variables (PRESTN at 06 h, 12 h and 18 h). We can notice, for example, that variables TMPMAX, TMPMIN and TMPMOY as well as PRESTN06h, PRESTN12h and PRESTN18h are strongly correlated. A strong correlation also exists between variables Vx06, Vy06, Vx18 and Vx12. O3veilleJahid variable is very correlated to O3 Jahid but it is not very well represented in the plan (PC1-PC2).

3.2. Prediction Models

In this section, we compare the results obtained from the different regression models described in section 2.3, namely: 1) the Multiple Linear Regression (MLR) model applied to all the variables of the dataset (24 variables), 2) the Reduced MLR with seven variables selected by means of Akaike Information Criterion (AIC) [39] [40], 3) The Principal Component Regression (PCR) model, 4) PLS and sparse PLS models, 5) Continuum Regression (CR), 6) Ridge and Lasso

Table 2. Percentage of total variance recovered by the principal components.

Component	Eigenvalue	Percentage of variance	Cumulative percentage of variance
comp 1	4.83	23.02	23.02
comp 2	3.67	17.47	40.49
comp 3	1.91	9.11	49.60
comp 4	1.77	8.43	58.04
comp 5	1.52	7.27	65.31
comp 6	1.30	6.19	71.51
comp 7	1.01	4.82	76.33
comp 8	0.99	4.73	81.06
comp 9	0.83	3.93	84.99
comp 10	0.64	3.05	88.05

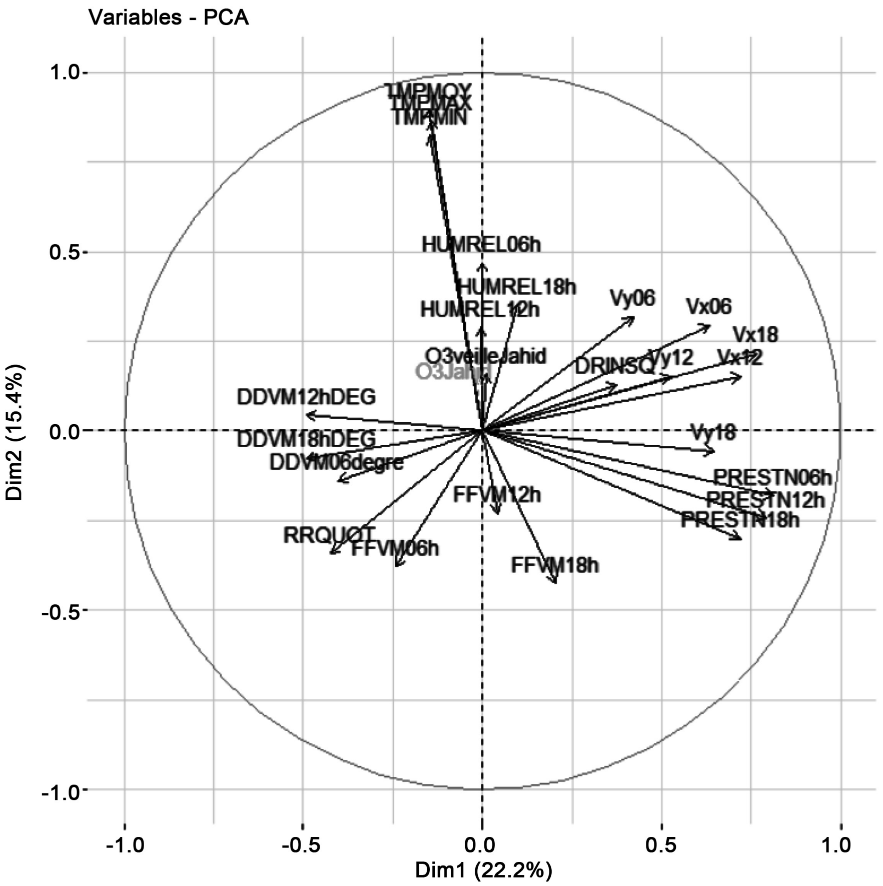

Figure 3. PCA correlation circle for Grand Casablanca data. Illustrative variable (O3) is represented with blue dashed arrow.

regressions, 7) Biased Power regression (BP-regression).

A cross validation procedure (LOO) is applied on the data collected during the period extending from April 1st to September 30th in 2013 and 2014 (training data) to determine for each model the parameters (number of components, Ridge and Lasso parameters…) leading to the minimum of the Root Mean Squared Error (RMSE). Then, the prediction ability according to the Root Mean Squared Error Predicted (RMSEP) of the various models is assessed on the basis of: 1) observed data (test data), $\text{RMSEP}_{\text{obs}}$, and 2) forecasted data, $\text{RMSEP}_{\text{prev}}$ from the summer period of 2015.

Table 3 shows the results of the various methods according to several criteria (RMSE, R^2, RMSEPobs and RMSEPprev).

Concerning the adjustment of the model on training data (internal validation), not surprisingly, the MLR model leads to the lowest RMSE (9.503), but the other models lead to close values and take into account multicollinearity problem. However, if the goal is to get the best predictive model, the RMSE alone is unsufficient and we need to analyze the RMSEP to assess the predictive quality of each model.

As for the criterion $\text{RMSEP}_{\text{obs}}$ in external validation, Lasso, Ridge, PLS, BP

Table 3. Comparison of different models according RMSE, R^2 and RMSEP criteria.

Model	MLR	Reduced MLR	PCR	PLS	Sparse PLS	CR	Ridge	Lasso	BP Reg
Nb var	24	7	24	24	14	24	24	11	24
Parameter			ncp = 10	ncp = 5	ncp = 3 $\eta = 0.56$	ncp = 1 $\alpha = 0.1$	λopt = 9.322	Fract = 0.2	$\alpha = 0.01$
RMSE	9.503	9.587	10.521	9.59	9.703	10.11	9.537	9.676	9.535
R^2	0.862	0.859	0.831	0.859	0.858	0.872	0.818	0.829	0.834
RMSEPobs	11.84	11.68	13.45	11.74	12.24	11.83	11.73	11.58	11.74
RMSEPprev	15.40	14.49	15.80	13.49	14.92	15.21	14.98	12.74	14.39

regressions and CR outperform the other methods. Lasso shows the best predictive ability since it has the smallest RMSEPobs. Moreover, this method of analysis has yet another advantage since the model is based on fewer predictive variables (11 variables such as TMPMAX, TMPMIN, DRINSQ, HUMREL12h, PRESTN06h, FFVM06h, FFVM18h, DDVM12h, Vx06h, Vy06h and O3veilleJahid) than the other models with the exception of the reduced model (7 predictive variables such as TMPMIN, TMPMOY, DRINSQ, PRESTN06, Vx06, Vx12, O3veilleJahid). Among these significant variables, TMPMIN and TMPMOY are strongly correlated so the reduced model does not solve the multicollinearity problem by comparing it to the Lasso model.

Most important are the results of the RMSEP$_{prev}$ based on the forecasted meteorological data for 2015. We recall that the forecasted meteorological data will be the data used on daily basis to predict the O3 concentration as obtained by Aladin-Maroc numerical forecasted model. It turns out that Lasso has by far the best RMSEP_prev (equal to 12.74), a value close to its RMSEP_obs (11.58) followed by the PLS and BP regression model. However, these last two models keep all the predictive variables, unlike the Lasso model, which keeps fewer variables, thus obtaining a model that is simple and easy to interpret.

Figure 4 shows a good correlation (around 0.723) between observed O3 and forecasted O3 data one day ahead obtained with the Lasso regression model only in 2015.

The most important finding (Table 3, Figure 4) is that the Lasso regression model has the best performance in predicting O3 concentrations in Jahid compared to the other models. Moreover, it clearly gives stable regression coefficients compared to Reduced MLR model. Table A1 of explanatory variables and Table B1 of the coefficients estimated by the models used in this study shows that the explanatory variables most retained by the models are: TMPMAX, TMPMIN, DRINSQ, PRESTNO6h, Vx06 and O3veilleJahid. Indeed, the formation of ozone in the Grand Casablanca area is related more particularly to: 1) the maximum and minimum daily temperature; 2) the period of intense sunshine; 3) the weak wind that accumulates the massive concentration of ozone and; 4) the previous day's concentration, which, to a large extent, determines the next day's

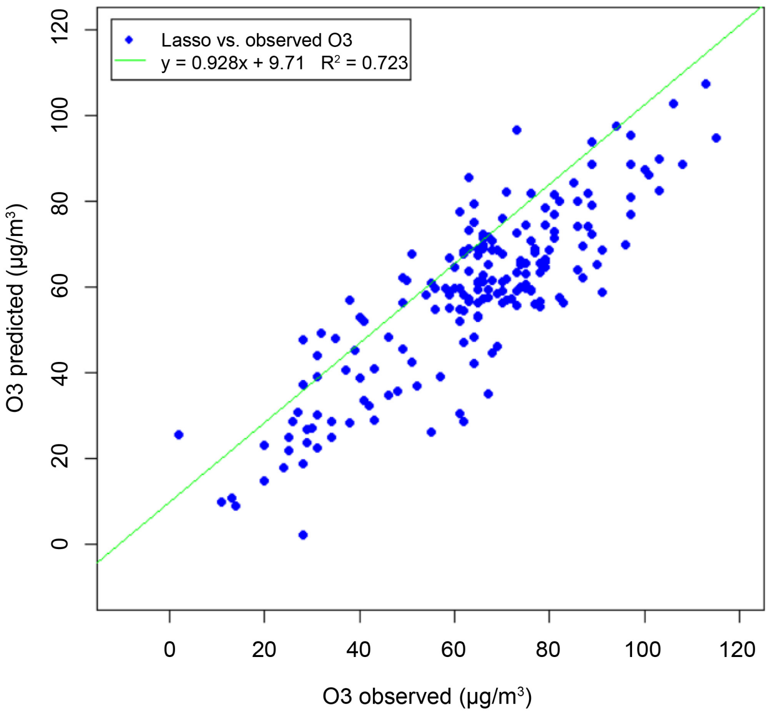

Figure 4. Predicted O3 using the Lasso regression model versus observed O3.

ozone concentration.

4. Conclusions

Starting with a multiple linear regression model, which is plagued by multicollinearity among the predictor variables, we have considered nine more or less recent alternative methods to relate meteorological and pollution variables. The emphasis was put on the prediction ability of the daily tropospheric ozone of these models in the Grand Casablanca area as the first comparative study of its type in such region.

We proposed the selected Lasso model based on a comparison of several linear forecasting methods to reduce the multicollinearity problem. The results obtained over two years of training data (2013 and 2014), verified on observed data (2015) and validated on forecast data (2015) show that the Lasso model has the best predictive capacity O3 for the Jahid station located in Grand Casablanca area. Moreover, using the dataset of 2015, Lasso model still gives the best predictive ability for O3 in Jahid station. The Lasso model presents the interest of being relatively simple and easily interpretable. The choice of this model is explained by the fact that it yields the best criteria in comparison to the alternative models discussed in this paper. These criteria include R^2, RMSE, RMSEPobs and RMSEPprev. Furthermore, besides yielding a more stable model than multiple linear regression, Lasso is based on a relatively small number of explanatory variables. This feature presents a significant advantage for the daily prediction of the ozone concentration in the Grand Casablanca.

This contribution proposes the first linear model of daily O3 concentration forecast in Morocco and more particularly in the Grand Casablanca area.

In perspective, we plan to widen our study by comparing the performances of the Lasso model with those of other non-parametric models and we will add more data (2017-2018) to ensure model validation. The most appropriate forecast model will be routinely implemented by the National Meteorological Office of Morocco (DMN).

Acknowledgements

The National Meteorological Office of Morocco (Direction de la Météorologie Nationale DMN) is gratefully acknowledged for providing the necessary data for undertaking the present study.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Anenberg, S.C., Horowitz, L.W., Tong, D.Q. and West, J.J. (2010) An Estimate of the Global Burden of Anthropogenic Ozone and Fine Particulate Matter on Premature Human Mortality Using Atmospheric Modeling. *Environmental Health Perspectives*, **118**, 1189-1195. https://doi.org/10.1289/ehp.0901220

[2] Malig, B.J., Pearson, D.L., Chang, Y.B., Broadwin, R., Basu, R., Green, R.S. and Ostro, B. (2016) A Time-Stratified Case-Crossover Study of Ambient Ozone Exposure and Emergency Department Visits for Specific Respiratory Diagnoses in California (2005-2008). *Environmental Health Perspectives*, **124**, 745-753. https://doi.org/10.1289/ehp.1409495

[3] Russell, B., Cooley, D., Porter, W. and Heald, C. (2016) Modeling the Spatial Behavior of the Meteorological Driver's Effects on Extreme Ozone. *Environmetrics*, **27**, 334-344. https://doi.org/10.1002/env.2406

[4] Bel, L., Bellanger, L., Bobbia, M., Ciuperca, G., Dacunha-Castelle, D., Gilibert, E., Jackubowicz, P., Oppenheim, G. and Tomassone, R. (1998) On Forecasting Ozone Episodes in the Paris Area. *Listy Biometryczne-Biometrical Letters*, **35**, 37-66.

[5] Bel, L., Bellanger, L., Bonneau, V., Ciuperca, G., Dacunha-Castelle, D., Deniau, C., Ghattas, B., Misiti, M., Misiti, Y., Oppenheim, G., Poggi, J.M. and Tomassone, R. (1999) Eléments de comparaison de prévisions statistiques des pics d'ozone. *Revue de Statistique Appliquée*, **XLVII**, 7-25.

[6] Besse, P., Milhem, H., Mestre, O., Dufour, A. and Peuch, V.H. (2007) Comparaison de techniques de <Data Mining> pour l'adaptation statistique des prévisions d'ozone du modèle de chimie-transport MOCAGE. *Pollution atmosphérique*, **195**, 285-292. https://doi.org/10.4267/pollution-atmospherique.1442

[7] Tamas, W. (2015) Prévision statistique de la qualité de l'air et d'épisodes de pollution atmosphérique en Corse: Génie des procédés. Ph.D. Thesis, Université de Corse Pascale Paoli, Français, 254 p.

[8] Lethrosne, M. (2008) Adaptation statistique des prévisions d'ozone issues du système Esméralda. Rapport de stage Master 2 Ingénierie Statistique à Airparif. Université de Rennes.

[9] Oufdou, H. (2010) Réalisation de modèles de prévision par adaptation statistique. Rapport de stage Master 2 Ingénierie Statistique à AIRAQ. Université Bordeaux2.

[10] Eljohra, B. (2005) Prévisibilité à 24 heures des concentrations en ozone troposphérique à Casablanca. Centre National de Recherches Météorologiques, Service Météorologie Sectorielle. DMN Maroc.

[11] Hastie, T., Tibshirani, R. and Friedman, J. (2009) The Element of Statistical Learning: Data Mining, Inference, and Prediction. Springer, Berlin. https://doi.org/10.1007/978-0-387-84858-7

[12] Hoerl, A. and Kennard, R. (1970) Ridge Regression: Biased Estimation for Non Orthogonal Problems. *Technometrics*, **12**, 55-67. https://doi.org/10.1080/00401706.1970.10488634

[13] Tibshirani, R. (1996) Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society*: *Series B*, **58**, 267-288.

[14] Joliffe, I.T. (1982) A Note on the Use of Principal Components in Regression. *Journal of the Royal Statistical Society*: *Series C*, **31**, 300-303. https://doi.org/10.2307/2348005

[15] Höskuldsson, A. (1988) PLS Regression Methods. *Journal of Chemometrics*, **2**, 211-228. https://doi.org/10.1002/cem.1180020306

[16] Tenenhaus, M., Gauchi, J.P. and Ménardo, C. (1995) Régression PLS et applications. *Revue de Statistique Appliquée*, **43**, 7-63.

[17] Tenenhaus, M. (1998) La régression PLS, théorie et pratique. Technip, Paris.

[18] Stone, M. and Brooks, R.J. (1990) Continuum Regression: Cross-Validated Sequentially Constructed Prediction Embracing Ordinary Least Squares, Partial Least Squares and Principal Components Regression. *Journal of the Royal Statistical Society*: *Series B*, **52**, 237-269.

[19] Sundberg, R. (1993) Continuum Regression and Ridge Regression. *Journal of the Royal Statistical Society*: *Series B*: *Methodological*, **55**, 653-659. http://www.jstor.org/stable/2345877

[20] Qannari, E.M., El Ghaziri, A. and Hanafi, M. (2017) Biased Power Regression: A New Biased Estimation Procedure in Linear Regression. *Electronic Journal of Applied Statistical Analysis*, **10**, 160-179.

[21] Houze, M.L. (1999) Concentrations en ozone dans les agglomérations dijonnaise et chalonnaise et conditions météorologiques (avril-août 1998), Mémoire de maîtrise de géographie. Université de Bourgogne, 67 p.

[22] Little, R.J.A. and Rubin, D.B. (2002, 2014) Statistical Analysis with Missing Data. John Wiley & Sons, Hoboken.

[23] Sande, I.G. (1983) Hot-Deck Imputation Procedures. *Incomplete Data in Sample Surveys*, **3**, 334-350.

[24] Ford, B.L. (1983) An Overview of Hot-Deck Procedures. *Incomplete Data in Sample Surveys*, **2**, 185-207.

[25] Fuller, W.A. and Kim, J.K. (2005) Hot Deck Imputation for the Response Model. *Survey Methodology*, **31**, 139.

[26] Rubin, D.B. (1987) Multiple Imputation for Non Response in Surveys. Wiley, New York. https://doi.org/10.1002/9780470316696

[27] Beretta, L. and Santaniello, A. (2016) Nearest Neighbor Imputation Algorithms: A Critical Evaluation. *BMC Medical Informatics and Decision Making*, **16**, 74. https://doi.org/10.1186/s12911-016-0318-z

[28] Jolliffe, I.T. (2002) Principal Component Analysis. Second Edition, Chapter 7.

[29] Husson, F., Le, S. and Pages, J. (2010) Exploratory Multivariate Analysis by Example Using R. Chapman and Hall, London. https://doi.org/10.1201/b10345

[30] Bellanger, L. and Tomassone, R. (2014) Exploration de données et méthodes statistiques: Data analysis & Data mining avec R. Collection Références Sci, Editions Ellipses, Paris, 480 p.

[31] Le Cao, K.A., Rossouw, D., Robert-Granié, C. and Besse, P. (2008) A Sparse PLS for Variable Selection When Integrating Omics Data. *Statistical Applications in Genetics and Molecular Biology*, **7**, 32. https://doi.org/10.2202/1544-6115.1390

[32] Chun, H. and Keleş, S. (2010) Sparse Partial Least Squares Regression for Simultaneous Dimension Reduction and Variable Selection. *Journal of the Royal Statistical Society*: *Series B*, *Statistical Methodology*, **72**, 3-25. https://doi.org/10.1111/j.1467-9868.2009.00723.x

[33] Draper, N.R. and Smith, H. (1998) Selecting the "Best" Regression Equation. In: *Applied Regression Analysis*, 3rd Edition, John Wiley & Sons, Inc., Hoboken, 473-504.

[34] James, G., Witten, D., Hastie, T. and Tibshirani, R. (2013) An Introduction to Statistical Learning with Applications in R. Springer Science + Business Media, New York.

[35] Wold, H. (1966) Estimation of Principal Components and Related Models by Iterative Least Squares. In: Krishnaiah, P.R., Ed., *Multivariate Analysis*, Academic Press, New York, 391-420.

[36] Abudu, S., Cui, C.L., King, J.P. and Abudukadeer, K. (2010) Comparison of Performance of Statistical Models in Forecasting Monthly Streamflow of Kizil River, China. *Water Science and Engineering*, **3**, 269-281.

[37] Sayegh, A., Munir, S. and Habeebullah, T.M. (2014) Comparing the Performance of Statistical Models for Predicting PM10 Concentrations. *Aerosol and Air Quality Research*, **14**, 653-665. https://doi.org/10.4209/aaqr.2013.07.0259

[38] Batista, G.E.A.P.A. and Monard, M.C. (2002) K-Nearest Neighbour as Imputation Method: Experimental Results. Technical Report, ICMC-USP.

[39] Akaike, H. (1974) A New Look at the Statistical Model Identification. *IEEE Transactions on Automatic Control*, **19**, 716-723. https://doi.org/10.1109/TAC.1974.1100705

[40] Zhang, Z. (2016) Variable Selection with Stepwise and Best Subset Approaches. *Annals of Translational Medicine*, **4**, 136. https://doi.org/10.21037/atm.2016.03.35

Appendix A

Table A1. Variables abbreviation and units of measurement.

Abbreviation	Variable	Unit
TMPMAX	Maximal temperature	˚C
TMPMIN	Minimal temperature	˚C
TMPMOY	Average temperature	˚C
RRQUOT	Total precipitation	mm
DRINSQ	Sunshine duration	heure
HUMREL06h	Relative humidity at 06 h	%
HUMREL12h	Relative humidity at 12 h	%
HUMREL18h	Relative humidity at 18 h	%
PRESTN06h	Pressure at the station level at 06 h	hpa
PRESTN12h	Pressure at the station level at 12 h	hpa
PRESTN18h	Pressure at the station level at 18 h	hpa
FFVM06h	Wind force at 06 h	m/s
FFVM12h	Wind force at 12 h	m/s
FFVM18h	Wind force at 18 h	m/s
DDVM06h	Wind direction at 06 h	degree
DDVM12h	Wind direction at 12 h	degree
DDVM18h	Wind direction at 18 h	degree
Vx06	Horizontal wind at 06 h	m/s
Vx12	Horizontal wind at 12 h	m/s
Vx18	Horizontal wind at 18 h	m/s
Vy06	Vertical wind at 06 h	m/s
Vy12	Vertical wind at 12 h	m/s
Vy18	Vertical wind at 18 h	m/s
O3veilleJahid	Ozone concentrations of the day before	μg/m^3
O3veille	Ozone concentrations	μg/m^3

Appendix B

Table B1. Comparison of regression coefficients estimated by the different models.

Variables	Complete Reg	Reduced Reg	PCR	PLS	SPLS	CR	Ridge	Lasso	BP Reg
TMPMAX	24.55	0.00	0.03	−1.06	−1.41	−1.74	−2.52	−0.55	21.62
TMPMIN	30.45	6.99	1.13	1.09	1.60	2.16	2.85	0.79	27.38
TMPMOY	−51.62	−6.55	0.61	−0.001	0.08	0.17	−0.003	0.00	−45.91
RRQUOT	−0.08	0.00	0.84	0.27	0.00	0.00	−0.03	0.00	−0.08
DRINSQ	2.15	2.11	−1.47	0.66	0.92	1.66	1.97	1.02	2.08
HUMREL06h	0.51	0.00	−0.40	−0.43	0.00	0.10	0.36	0.00	0.48
HUMREL12h	0.27	0.00	1.37	0.84	0.00	0.53	0.33	0.14	0.28
HUMREL18h	−0.62	0.00	−1.33	−0.36	0.00	−0.34	−0.48	0.00	−0.59
PRESTN06h	−1.46	−1.64	−0.51	−1.04	−0.65	−0.97	−1.15	−0.87	−1.42
PRESTN12h	−0.02	0.00	−0.46	−0.89	−0.41	−0.43	−0.28	0.00	−0.05
PRESTN18h	0.08	0.00	−0.54	−0.85	−0.39	−0.15	0.06	0.00	0.06
FFVM06h	0.27	0.00	−0.19	0.64	0.45	0.41	0.31	0.26	0.28
FFVM12h	0.54	0.00	0.37	−0.03	0.69	0.31	0.42	0.00	0.52
FFVM18h	0.41	0.00	0.35	0.47	1.18	0.43	0.41	0.52	0.41
DDVM06deg	0.05	0.00	0.66	0.45	0.56	0.22	0.11	0.00	0.07
DDVM12hDEG	−0.11	0.00	−0.87	−0.60	0.00	−0.36	−0.19	−0.11	−0.11
DDVM18hDEG	−0.58	0.00	−0.18	−0.09	0.00	−0.39	−0.54	0.00	−0.56
Vx06	−1.33	−1.35	−1.16	−1.23	−0.94	−1.34	−1.31	−0.82	−1.31
Vx12	1.28	1.10	1.03	0.39	0.00	0.62	0.98	0.00	1.22
Vx18	−0.70	0.00	0.38	−0.24	0.51	−0.61	−0.71	0.00	−0.68
Vy06	0.69	0.00	−2.20	0.26	0.00	0.79	0.73	0.53	0.69
Vy12	−0.97	0.00	1.86	0.41	0.00	−0.19	−0.69	0.00	0.91
Vy18	0.24	0.00	1.75	1.18	0.00	0.67	0.36	0.00	0.27
O3veilleJahid	23.36	23.26	22.34	23.03	23.31	23.21	22.75	23.14	22.97