POLITECNICO DI MILANO
Master Degree in Mathematical Engineering
Industrial and Information Engineering
Department of Mathematics

Air Quality forecasting

Bayesian Statistics Project Report

Group members:
Datena Maria Alfredo
Dell'Atti Martina
Palummo Alessandro

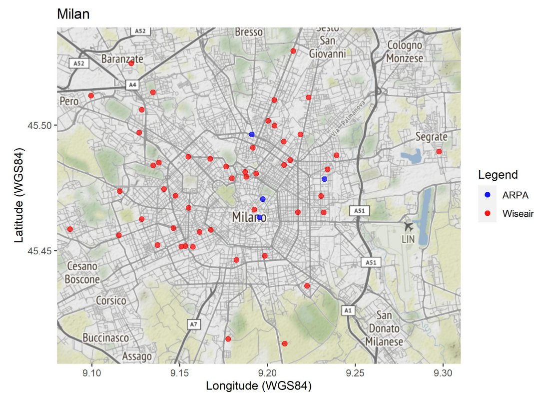
Academic Year 2020-2021

Contents

1 Introduction

1.1 Wiseair

In Lombardia, ARPA is the institutional agency in charge of measuring and publishing certified air quality data. ARPA's measures are based on accurate sensing stations. The main issue of these stations is the fact they are costly and bulky, so few of them can be deployed (2 in Milan for PM2.5). This is a problem because the concentration of Particulate Matters varies at a hyperlocal scale, so 2 stations are not enough to capture the phenomenon.



Other countries have similar problems: therefore, institutions and private companies are developing technological solutions to obtain data complementary to traditional sensing stations. In particular, traditional sensing stations are based on the gravimetric technique, while newer sensors are based on laser-scattering. The former technique is more accurate, while the latter requires cheaper and smaller sensors (5cm instead of 2m) and can provide data in real-time. Wiseair is a startup company that has designed new sensors that are low-cost, user-friendly and easy to install: Arianna. In July 2020, after the lockdown, Wiseair distributed the first 50 sensors. The data measured by the sensors are available for citizens to access on our iOs and Android applications and are accessible via API.

1.2 Objective

Wiseair's goal is to provide citizens useful information about air quality, for example for planning outdoor activities. Up to now they can only inform about the current levels of pollutants in the air. We aim at forecasting future values of Particulate Matters so that prevision about the quality of the air can be maden available to the citizens.

1.3 The dataset

Our data as collected by Wiseair's stations go from September 2020 to November 2020. Each observation in the dataset is a measurement from a single Arianna station. Information which are available without any further transformation are:

- the time coordinate for the measurement, i.e. day and hour when the measurement has taken place.
- an identifier of the sensor which has taken the measurement.
- one variable for each type of particular matter registered by the sensor. We have access to PM2.5, PM4, PM10 values. The different PM differs for the dimension of the particulate measured, respectively, less or equal than $1\mu m$, $2.5\mu m$, $10\mu m$. They are measured in $\mu g/m^3$.
- The temperature and the humidity measured by the Arianna which has been detected at the moment of measuring the levels of PM.
- latitude and longitude of the sensor which takes the measurement.

2 Data Preprocessing

From the available data we extract other usefull informations

- a dummy variable which indicates if the measurement has been performed during a weekday or during the weekend.
- a dummy variable which indicates if the measurement has been taken during the night or during the day

Morover to enrich the data with other usefull atmospheric confitions we have integrated the available data with data coming from ARPA sensor

- the wind speed registered by ARPA (measured in m/s)
- the wind angle registered by ARPA with respect to the north (measured in degrees)
- the amount of rain fallen (measured in mm)
- the global solar radiation (measured in W/m^2

Originally we had data with a non regular frequency, i.e. the frequency of measurements of a sensor is different from others. For example, some sensors make two measurements per hour, others make five measurements per hour. To have equally spaced points we decided to perform a first hourly average of the measured quantities. We choose to aggregate hourly because of two reasons:

1. This choice does not remove eventual patterns during the day.
2. Weather data (rain, wind speed, wind direction, global solar radiation) from ARPA are available with hourly frequency. This require to have hourly aggregated data in order to perform the integreation of available data with data coming from ARPA.

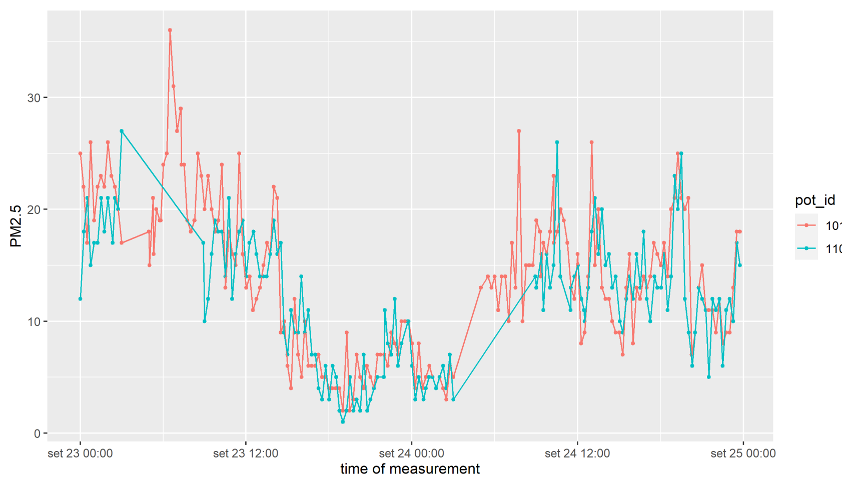
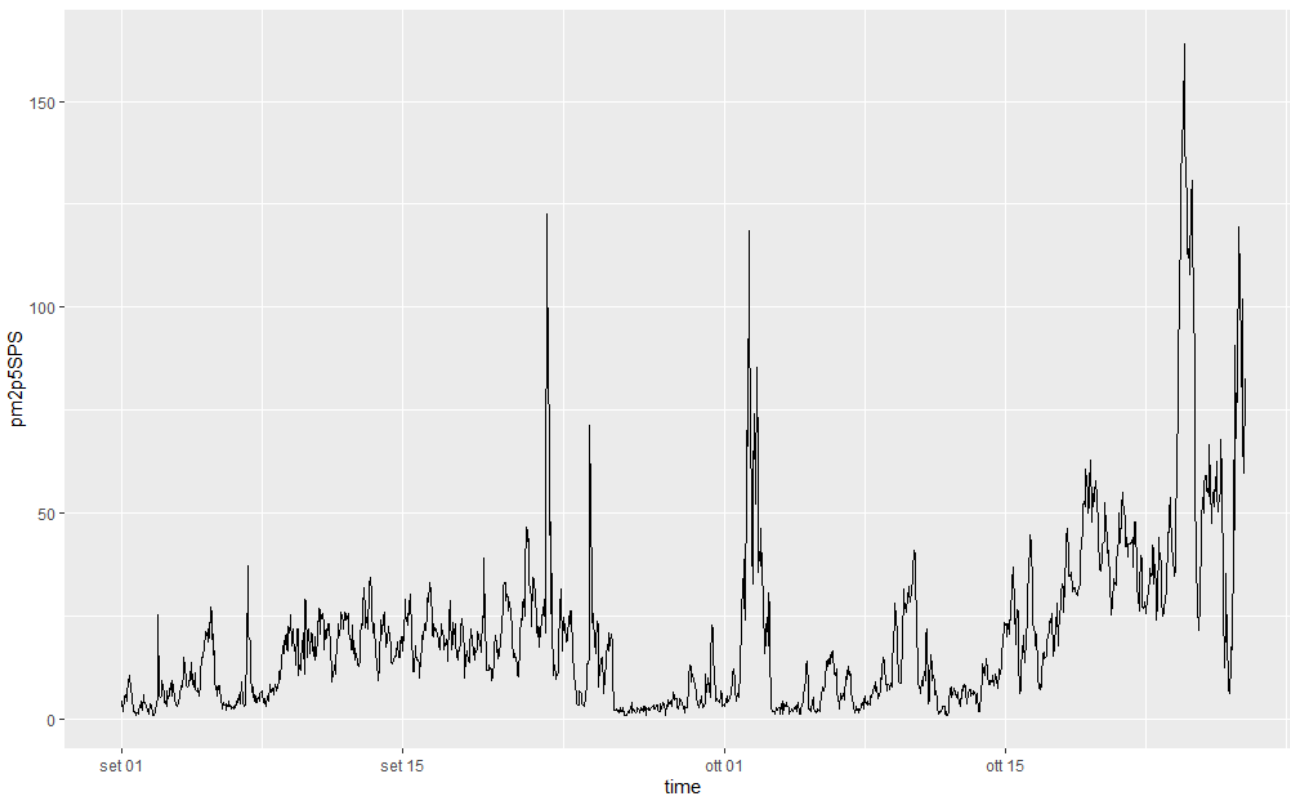


Figure 1: Plot of two Arianna stations: each Arianna station has its own frequency of measurement depending on the charge level of the station

In our dataset sensors are identified by an unique identifier. From now on, we'll focus on an **univariate analysis**.

Sometimes can happen that sensor does not send any data; from our perspective this translates in the presence of missing values in the time series. The selection of what sensor consider has been based on the percentage of missing values and their distribution over time. Indeed there are sensors which have more than 40% of their point missing. Such large amount of missing information makes the recordered quantites not suitable for our study. We instead focus on sensors which operate continously along the considered period, or at least have a small fraction of their value missing. Moreover if missing data are distributed over all the considered period (we have missing values in correspondence of just few hours) and do not span a large interval of time they can be resolved by interpolation. The following plot shows the selected time series for our analysis.



2.1 Stationarity

A central characteristics of a time series is stationarity: if a series is stationary this means that its mean, variance and autocorrelation at any lag does not depend on the specific time instant t where they are considered. A stationary time series then presents a constant mean and points are fluctuations more or less large around such mean. The plot above shows that our target series cannot be regarded as stationary. The presence of large peaks and of a non constant trend over time are clearly property of non stationary processes. Models which aim at predicting this phenomenon must hence work well without stationarity assumptions.

3 Static Models

Static models are models where the parameters coefficients remain unchanged after the fitting process. As we will see this will consitute a problem to forecast the data under analysis. They indeed make the specific assumption of stationary data, which is not the case for our study. Anyway we start focusing on this class of simple models also to start inspecting the main characteristics of the series and the most usefull regressors.

3.1 AR(2) Model

As a first proposal, we choose an Autoregressive Model of order 2 for the univariate case. We make this choice focusing on the analysis of the autocorrelation (ACF) and partial autocorrelation function (PACF) of the data in analysis: since ACF tails off and PACF cuts off after two lags, then the simplest choice is to assume an AR(2) model (see figure 8 and 9).

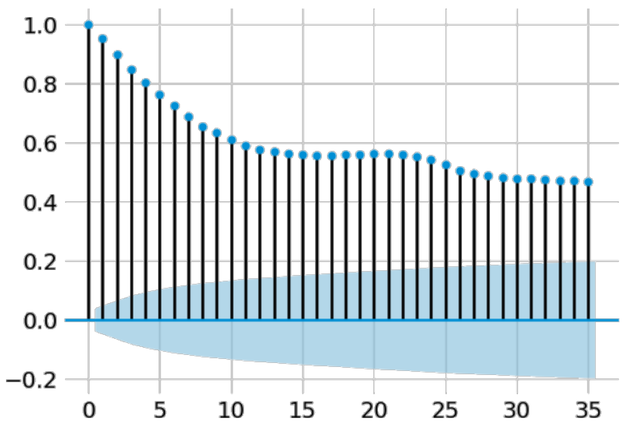


Figure 2: Autocorrelation plot for the 1091 pot series

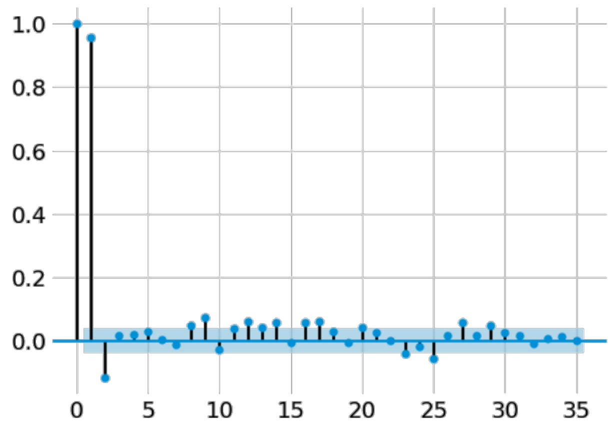


Figure 3: Partial autocorrelation plot for the 1091 pot series

Formally the model can be represented as follow:

$$y_t = \phi_0 + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \epsilon_t \qquad\qquad \epsilon_t \sim \mathcal{N}(0, \sigma^2)$$

where ϵ_t is a sequence of uncorrelated error terms and the ϕ_i are constant parameters.

We assume a conjugate model to explain the AR(2) structure of the time series. If the first p = 2 values of the series are known and viewed as fixed constants, we do not have to put a prior for the initial condition, morover the total number of points can be defined as $T = n + p$. Given

- $Y = [y_{n+2}, \dots, y_{n+p}]$ the response variable PM2.5
- $\Phi = [\phi_0, \phi_1, \phi_2]^T$ a 3×1 vector of parameters relative to the autoregressive part
- $y_t | y_{t-2}, y_{t-1}, \Phi, \sigma^2 \sim \mathcal{N}({f_t}^T \Phi, \sigma^2)$ where ${f_t}^T = (1, y_{t-1}, y_{t-2})$
- $F = [f_{n+2}, ..., f_{n+p}]$ is a $3 \times n$ design matrix

The conditional density of Y given the first 2 values is

$$p(Y|y_1, y_2, \Phi, \sigma^2) = \prod_{t=3}^{T} \mathcal{N}(y_t|{f_t}^T \Phi, \sigma^2) = \mathcal{N}_n(Y|F^T \Phi, \sigma^2 \mathcal{I}_n)$$

Overall, from a bayesian perspective, the model is so formulated:

LIKELIHOOD

$$Y|y_1, y_2, \Phi, \sigma^2 \sim \mathcal{N}_n(F^T\Phi, \sigma^2\mathcal{I}_n)$$

PRIORS

$$\pi(\Phi, \sigma^2) = \pi(\Phi|\sigma^2)\pi(\sigma^2)$$
$$\Phi|\sigma^2 \sim \mathcal{N}_3(\mu_0, \sigma^2 B_0)$$
$$\sigma^2 \sim inv - Gamma\Big(\frac{\nu_0}{2}, \frac{\nu_0\sigma_0^2}{2}\Big)$$

$\mu_0 \in \mathcal{R}^2$, B_0 is a 2×2 matrix, $\nu_0, \sigma_0 > 0$

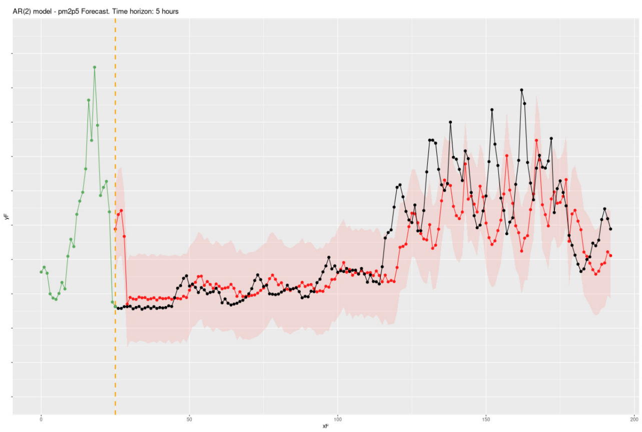


Figure 4: Forecasting plot of PM2.5 values with a time horizon of 5 hours

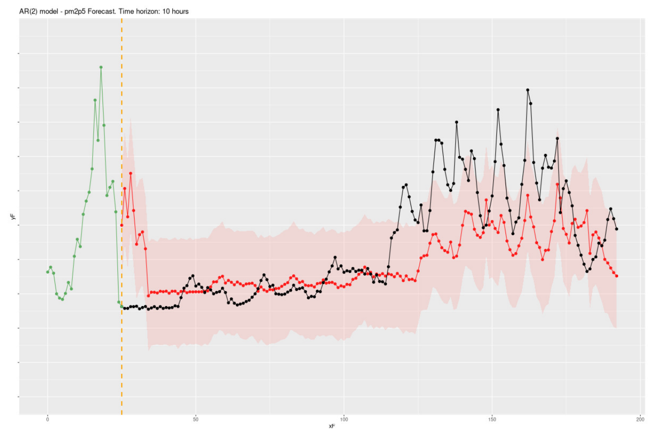


Figure 5: Forecasting plot of PM2.5 values with a time horizon of 10 hours

In the graphs above we plot the prediction of the AR(2) model with a time horizon of 5 hours and 10 hours: in the first plot the red points are the prediction using the series' values until 5 hour before, while in the second plot the red points represent the prediction using the series' values until 10 hour before. Therefore, knowing the values of particulate matter of the last two hours until now, we make prediction for the next 5 hours and for the next 10 hours. At the end we plot just the prediction at time t+5 and t+10. As we can see, the prediction becomes very bad as the time horizon increases and this is because this model is too simple and does not consider important factors that have influence on the time series.

3.2 ARX(7) Model

In order to enrich the AR(2) model and make a more accurate prediction we propose a model that has not only the autoregressive part but also a regressive part, so including covariates temperature, humidity, rain and wind. For the moment, we consider the same order for all the covariates and we choose order 7 because the number of parameters grows with the order p. So, due to our limited computational capacity we have adopted p equal to 7. The conditional density of Y given the first p values is

$$y_t|y_{t-1}, .., y_{t-p}, X, \beta, \sigma^2 \sim \mathcal{N}(y_t|{f_t}^T\beta, \sigma^2) \qquad \epsilon_t|\sigma^2 \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma^2)$$
$$\Rightarrow p(Y|y_{t-1}, .., y_{t-p}, X, \beta, \sigma^2) = \prod_{t=p_{max}}^{n+p_{max}} \mathcal{N}(y_t|{f_t}^T\beta, \sigma^2) = \mathcal{N}_n(Y|F^T\beta, \sigma^2)$$

LIKELIHOOD

$$Y|y_{t-1}, .., y_{t-p}, X, \beta, \sigma^2 \sim \mathcal{N}_n(Y|F^T\beta, \sigma^2)$$

PRIORS

$$\beta|\sigma^2 \sim \mathcal{N}_k(\mu_0, \sigma^2 B_0)$$
$$\sigma^2 \sim inverse - gamma\Big(\frac{\nu_0}{2}, \frac{\nu_0\sigma_0^2}{2}\Big)$$

- $Y = [y_{pmax}, .., y_{p_{max}+n}]$, where $p_{max} = max(p, p_1, p_2, p_3, p_4)$
- $X = [x_1, x_2, x_3, x_4]$ is the vector of covariates: temperature, humidity, rain, wind, where $x_i = [x^i_{t-1}, .., x^i_{t-p_i}]$

- $f_t^T = [1, y_{t-1}, .., y_{t-p}, X]$ is the autoregressive part together with the vector of covariates
- $F = [f_{p_{max}}, ..., f_{p_{max+n}}]$ is the design matrix of known regressors.

3.3 Ridge Regression

The model with p=7 is

$$y_t = \beta_0 + \beta_1 y_{t-1} + ... + \beta_7 y_{t-7} + \beta_8 x^1_{t-1} + ... + \beta_{14} x^1_{t-7} + ... + \beta_{29} x^4_{t-1} + ... + \beta_{35} x^4_{t-7} + \epsilon_t$$

Therefore, we have 7 lags for each of the 5 variables with a total of 36 covariates. For this reason we perform a covariate selection using ridge regression in order to choose an appropriate order p for each variable and discard lags which a posteriori does not show to be significant.

LIKELIHOOD

$$Y|y_{t-1}, .., y_{t-p}, X, \beta, \sigma^2 \sim \mathcal{N}_n(Y|F^T\beta, \sigma^2)$$

PRIORS

$$\beta_j|\lambda \sim \mathcal{N}\Big(0, \frac{1}{\lambda}\Big) \qquad j = 1, \ldots, k \qquad k = (p+1) + p_1 + p_2 + p_3 + p_4$$

$$\lambda \sim gamma(a_\lambda, b_\lambda)$$

$$\sigma^2 \sim inverse - gamma(a_{\sigma^2}, b_{\sigma^2})$$

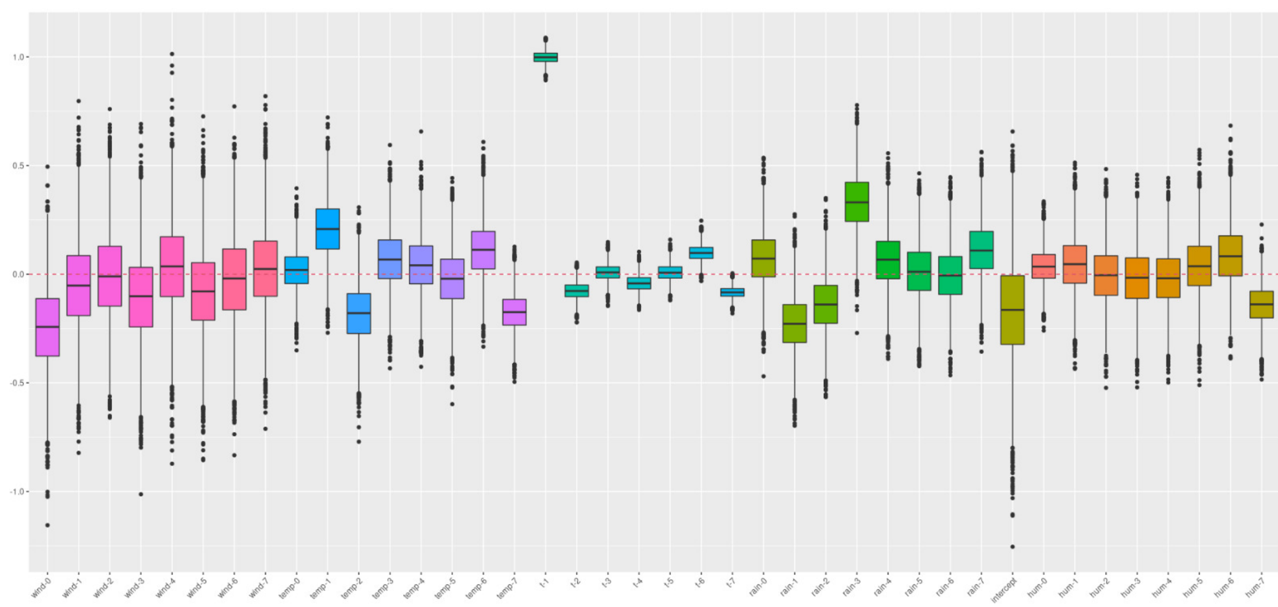


Figure 6: Boxplots from ridge regression

Boxplots above show the posterior distribution of parameters after the fit. Covariates can be considered significant for the target if their 90% credible interval does not contain the 0. Posterior inferences shows that the meaningfull covariates selected by ridge are:

- lag 1-2 for the autoregressive part $\Rightarrow$ y_{t-1}, y_{t-2}
- lag 1 for humidity $\Rightarrow$ x^1_{t-1}
- lag 1-2-3 for rain $\Rightarrow$ $x^2_{t-1}, x^2_{t-2}, x^2_{t-3}$
- t lag 1-2 for temperature $\Rightarrow$ x^3_{t-1}, x^3_{t-2}
- the covariate wind seems to be not significant

Surprisingly the covariate wind seems not to be significant, or at least weakly significant for its lag 0. As we will see discarding the effect of the wind from the model is a bad choice, because we have strong prior domain knowledge about the effect of the wind on the PM concentration.

At the end, we have 9 covariates from the oringinal 36.

3.4 Hourly SARX: ARX with seasonal effect

The next natural extension of the ARX model (using the result coming from Ridge in terms of covariate selection) is a Seasonal ARX model, in order to include also hypothetical periodical components.

Seasonality can be modeled as random effects which contribute to the mean of the response variable y_t when there are some specific periodical conditions, i.e. a specific hour during the day or a specific day during the week. Those specific periodical conditions are modeled by the indicator δ_t^i. In this sense the seasonal ARX can be considered as a mixed effect model, where there is a common mean between all the observations (represented by the ARX part) and the seasonal effects are variations from this common mean.

We first focus on a hourly SARX model, namely we assume a periodical component of period one day. In practice we introduce 24 different parameters γ_i, one for each hour of the day. The effect of each γ_i is active when hour i is met during the day.

Both regressive and autoregressive part are kept fixed as in a standard ARX model, and hence SARX can be viewed as an extension of it.

The model is written as follow

$$
\begin{aligned}
&y_t = \mu_t + \epsilon_t, \qquad \epsilon_t \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma^2) \qquad t = p_{max} : n, \quad p_{max} = max(p, p_1, p_2, p_3) \\
&\mu_t = {f_t}^T \alpha + {x_t}^T \beta + S_t \\
&S_t = \sum_{i=0}^{T-1} \gamma_i \delta_t^i \qquad T = 24 \\
&\delta_t^i = \begin{cases} 1 & \text{if at time t is hour i} \\ 0 & \text{otherwise} \end{cases}
\end{aligned}
$$

- ${f_t}^T = (1, y_{t-1}, ..., y_{t-p})$ is the autoregressive term
- ${x_t}^T = (X_{t-1}^1, ..., X_{t-p_1}^1, ..., X_{t-1}^3, ..., X_{t-p_3}^3)$ is the regressive term

The mean of the response variable for all times $t = p_{max}, \ldots, n$ can be written in vectorial notation as:

$$\mu = F^T \alpha + X^T \beta + \Delta \gamma$$

Δ is a $24 \times n$ matrix having for each row t, 1 in position j if $\delta_t^j = 1$, 0 otherwise

The model is then formulated as follow:

LIKELIHOOD	PRIORS	
$Y \mid F, \alpha, X, \beta, \Delta, \gamma, \sigma^2 \sim \mathcal{N}_n(Y \mid \mu, \sigma^2)$	$\sigma^2 \sim inverse-gamma(a_\sigma^2, b_\sigma^2)$	
	$\alpha_i \overset{i.i.d.}{\sim} \mathcal{N}(a_0, \sigma_\alpha^2)$	$i = 0, .., p$
	$\beta_j \overset{i.i.d.}{\sim} \mathcal{N}(b_0, \sigma_\beta^2)$	$j = 1, .., k, \quad k = p_1 + p_2 + p_3$
	$\gamma_h \mid \sigma_\gamma \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma_\gamma)$	$h = 0, .., 23$
	$\sigma_\gamma \sim inverse-gamma(a_\gamma, b_\gamma)$	

The boxplots show the a posteriori distribution of gamma variables, namely the seasonal effects over a day. As is clear there is a peak during the morning, at about 6 or 7 am, meaning that there is an increase of the level of Particulate Matter, and such increment happen periodically each day. During the evening and night there is instead a decrease of pollution. We would expected a second peak during the afternoon, at about 5 or 6 pm, possibly due to people going home after work but this is not visible from this sensor. A possible explanation could be found in the location of the sensor, which is placed not in the center of the city but in periphery, or in the restrictions due to the pandemic period since they could have modified habits of people.

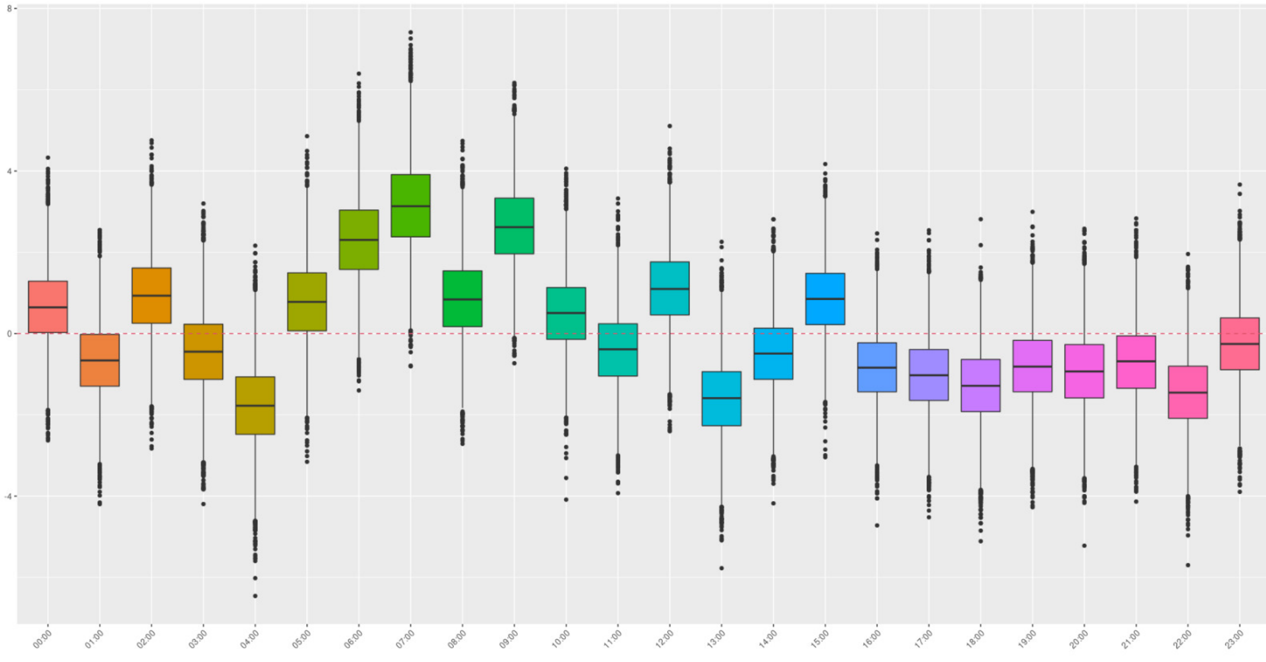
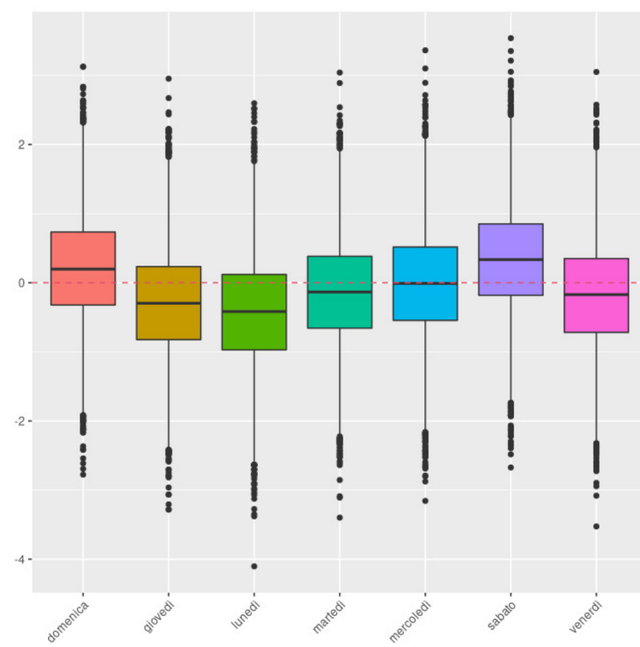


Figure 7: Boxplots of gamma

3.5 Daily SARX: ARX with daily seasonal effect

We have also tried to inspect periodical components with weekly period. The model is the same as the SARX with hourly effect, what changes is the period of the seasonal component (here 7 days) and the definition of the indicators γ_t^i. Anyway, as the posterior distribution show, there is no relevant periodical component over the week, since posterior distributions of γ are centered in zero.



$$y_t = \mu_t + \epsilon_t, \qquad \epsilon_t \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma^2)$$
$$t = p_{max} : n, \quad p_{max} = max(p, p_1, p_2, p_3)$$
$$\mu_t = {f_t}^T \alpha + {x_t}^T \beta + S_t$$
$$S_t = \sum_{i=0}^{T-1} \gamma_i \delta_t^i \qquad T = 7$$
$$\delta_t^i = \begin{cases} 1 & \text{if at time t is hour i} \\ 0 & \text{otherwise} \end{cases}$$

As before, let $\mu = F^T \alpha + X^T \beta + \Delta \gamma$, where Δ is a $7 \times n$ matrix having for each row t, 1 in position j if $\delta_t^j = 1$, 0 otherwise, then:

LIKELIHOOD

$$Y|F,\alpha,X,\beta,\Delta,\gamma,\sigma^2 \sim \mathcal{N}_n(Y|\mu,\sigma^2)$$

PRIORS

$$\begin{aligned}
&\sigma^2 \sim inverse-gamma(a_\sigma^2, b_\sigma^2) && \\
&\alpha_i \overset{i.i.d.}{\sim} \mathcal{N}(a_0,\sigma_\alpha^2) && i = 0,..,p \\
&\beta_j \overset{i.i.d.}{\sim} \mathcal{N}(b_0,\sigma_\beta^2) && j = 1,..,k, \quad k = p_1+p_2+p_3 \\
&\gamma_h|\sigma_\gamma \overset{i.i.d.}{\sim} \mathcal{N}(0,\sigma_\gamma) && h = 0,..,6 \\
&\sigma_\gamma \sim inverse-gamma(a_\gamma, b_\gamma) &&
\end{aligned}$$

3.6 Limitation of static models

We perform model selection using both WAIC and BIC: as a result SARX seems to be the best model according to both indexes, since it reaches the minimum WAIC and maximum BIC.

model	**WAIC**	**BIC**
AR(2)	8948.05	-18405.63
ARX(7)	8830.286	-18175.14
regularized ARX(7)	20150.2	-23111.88
SARX(2,2,1,3)	8820.29	-18122.96

Hence adding more information to the model helps. Anyway results returned by these indexes are misleading. Indeed we need to be critical: if we simulate the SARX prediction as function of the time horizon we see that it is not so better than the simplest AR(2) model seen so far. We can see that the SARX model performs better when there is some stable and periodical behaviour of the series, but both models completely fails in the forecast.

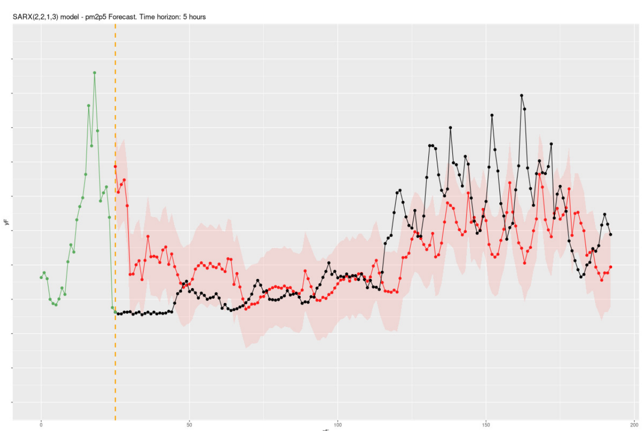


Figure 8: Forecasting plot of PM2.5 values with time horizon of 5 hours

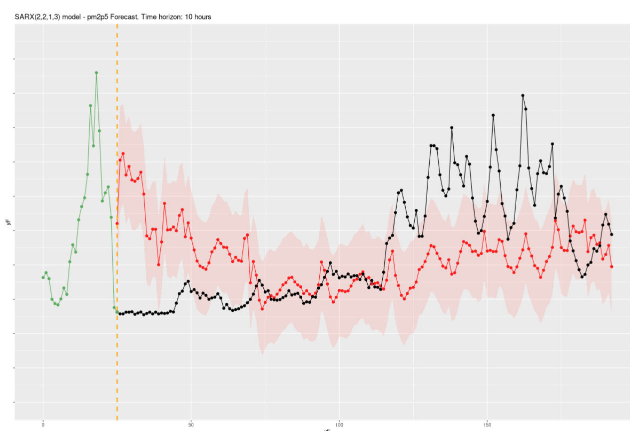


Figure 9: Forecasting plot of PM2.5 values with time horizon of 10 hours

The reason for such poor performances must be found in the stationarity assumptions all model presented so far are doing. Static coefficients models are not capable to react to structural changes in the series dynamics, they expect the target series to mantain a constant mean and variance. The presence of peaks in the target PM series is clearly an evidence of non stationarity and is not surprising to see such models fail.

4 Dynamic Linear Models

All the models presented so far are static: this means that we take our data and fit them to some parametric model. Once the model has been fitted, posterior distributions of parameters are never changed. This actually can't allow the model to adapt itself to big changes in the series dynamics. So we now turn our attention on Dynamic models, i.e. fully Bayesian's models where parameters are a function of time. They allow to update our prior belief as new data are collected in real time, and use our posterior distribution as prior belief for the next data arrival, in a sort of feedback loop system. This allow the model to keep track of the ongoing development of the series and better resolve the non stationary may arise.

4.1 Exponential Smoothing and Holt-Winters' method

Simple exponential Smoothing is a forecasting method which uses the last forecast and adjusts it using the forecast error and a parameter, α, set between 0 and 1.

$$\hat{y}_{t+1} = \hat{y}_t + \alpha(y_t - \hat{y}_t)$$

another way of writing this equation is:

$$\hat{y}_{t+1} = \alpha y_t + (1 - \alpha)\hat{y}_t$$

This shows that the forecast is produced by a weighted sum of the last observation and the last forecast. α is the weight which is assigned to the last observation. By iterating the formula, this expression can be obtained:

$$\hat{y}_{t+1} = \sum_{j=1}^{t} \alpha(1 - \alpha)^{t-j} y_j + (1 - \alpha)^t \hat{y}_1$$

This shows that the forecast consists of a weighted average of the observations, with the weight exponentially decreasing for observations farther in time, hence the name "exponential smoothing".

Holt (1957) extended simple exponential smoothing to linear exponential smoothing to allow forecasting data with trends and seasonality, adding two parameters and two equations:

$$\begin{aligned}
l_t &= \alpha(y_t - c_{t-p}) + (1 - \alpha)(l_{t-1} + b_{t-1}) \\
b_t &= \beta(a_t - a_{t-1}) + (1 - \beta)b_{t-1} \\
c_t &= \gamma(y_t - a_{t-1} - b_{t-1}) + (1 - \gamma)c_{t-p} \\
\hat{y}_{t+h|t} &= a_t + b_t h + c_{t+h-p}
\end{aligned}$$

where a_t is called "level",and is an estimate of the value of the time series at time t, and b_t is an estimation of the "slope" or "growth", of the time series at time t. This value is itself estimated at each step starting from its previous values and the difference between the two last "level" values. c_t is the seasonal term, with periodicity p

What could make this model more suitable for the analysis of long time series (¿1000 observations) is the flexibility of the level, trend and seasonality terms, which change with the behaviour of the time series according to the smoothing parameters α, β and γ. These parameters ultimately control the rate at which the model adapts to new behaviours of the time series.

4.1.1 State Space model with additive error

A model for an observed phenomenon can be obtained using the exponential smoothing framework by complementing the above system with the equation:

$$y_t = a_{t-1} + b_{t-1} + c_{t-p} + \varepsilon_t$$

Where $\{\varepsilon_t\}$ are uncorrelated, homoscedastic normally distributed errors, with standard deviation σ. This formulation has an equivalent linear formulation by constructing opportune matrices:

$$Y = M\psi + L\varepsilon$$

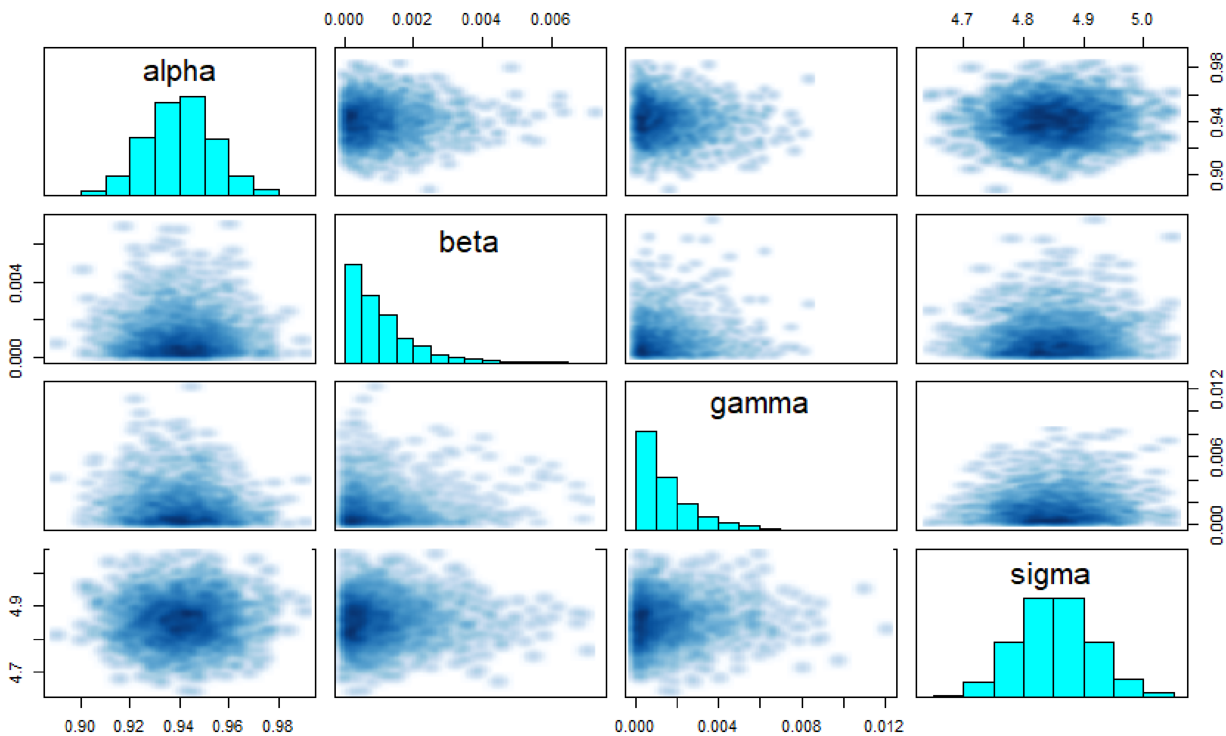
Where M is a known $n \times (p+1)$ full-rank matrix, L is a lower triangular matrix constructed with the smoothing parameters α, β, γ, while ψ is a vector of unknown initial conditions for level, growth and seasonal terms. The likelihood function for this model can be computed:

$$\sigma^{-n} exp - \frac{1}{2\sigma^2}(\tilde{\psi} - \psi)' X' X (\tilde{\psi} - \psi) \times exp - \frac{1}{2\sigma^2}(L^{-1}Y)'(I - P_X) L^{-1} Y$$

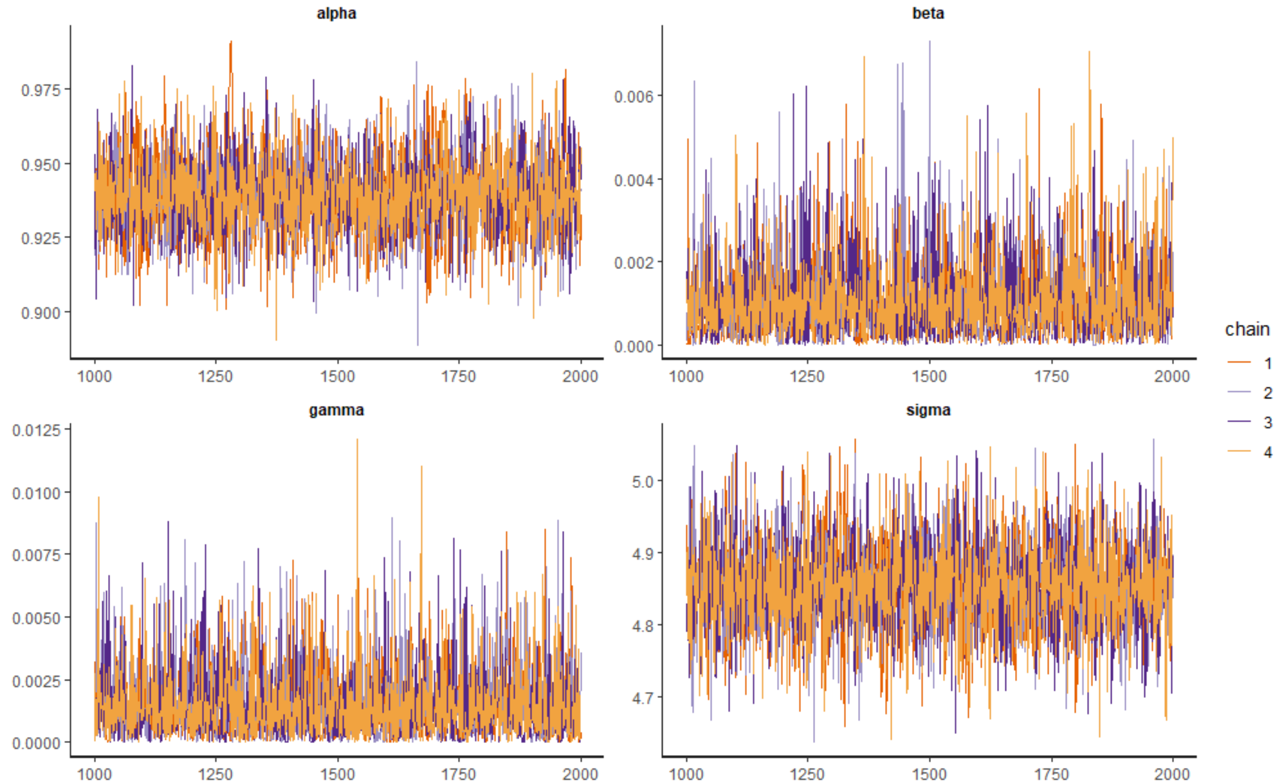
where X is the matrix $L^{-1}M$, P_X is the orthogonal projection matrix on the column space of X: $(P_X = X(X'X)^{-1}X'$. While $\tilde{\psi} = (X'X)^{-1}X'L^{-1}Y$

4.1.2 Model Performance

We have performed MCMC sampling of this model on our data regarding the values of PM 2.5 registered by an Arianna station. The sampling has been performed by compiling a STAN program. Below is a plot of the sampling of the model parameters (excluding the vector of initial conditions):



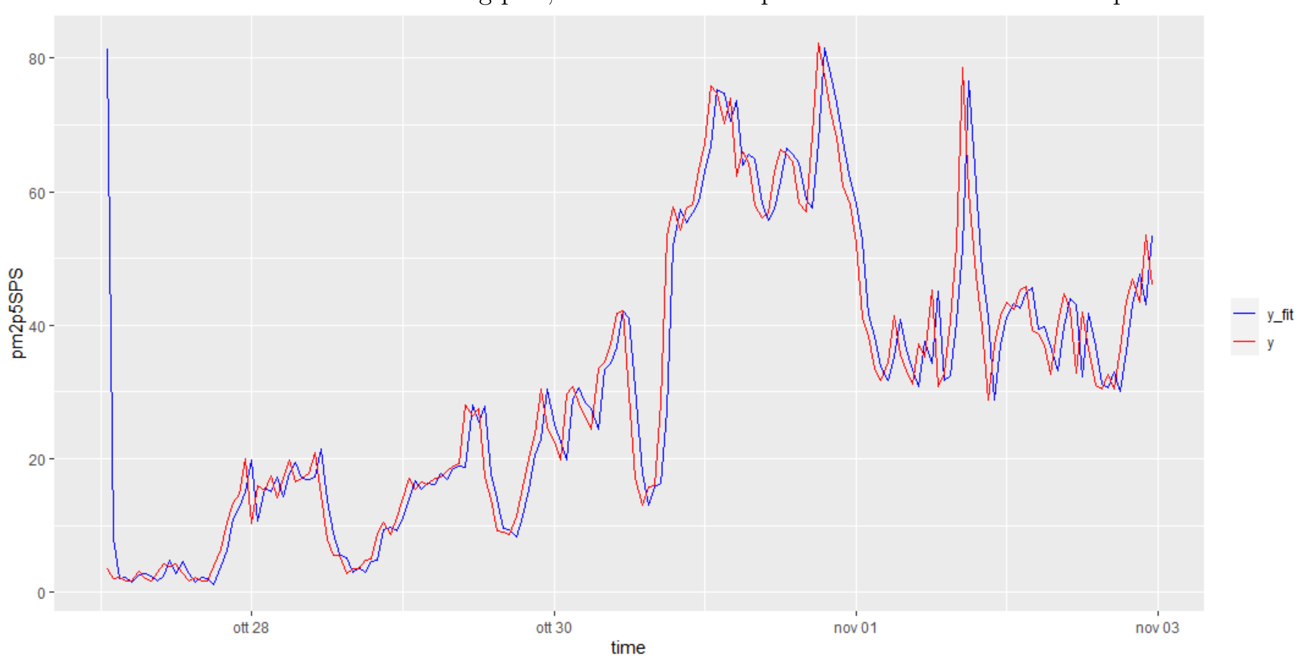
The traceplots of the main parameters (α, β, γ) of the model show a good behavior of the MCMC algorithm.



The posterior distributions of γ and β have most of their mass very close to zero, which means that the model does not adapt the seasonal and trend components to the evolution of the time series, or that these components do not have a great impact on the fitted model.

The posterior distribution of α has its mass concentrated on a point which is close to 1 (posterior mean: 0.9402484). This means that the behaviour of the model will be similar to an $AR(1)$ process.

This can also be observed on the following plot, which shows the predicted values on a test sample:



Overall we can conclude that the model is not able to represent a possible seasonal behaviour of the time series. This model can account for the seasonal shape to change over time, but does not account for a change in the seasonal period. The input data does not show a distinct seasonal pattern over the whole considered time period. The data does seem to show some seasonality in in some specific time segments, but this possible seasonal behaviour changes both in shape and in period during the time considered.

A solution for this model to work would be to segment the time series in appropriate subsets and fit this same model to each of these segments. The model fitted to the latest segment would then be used to perform a prediction.

4.2 The general univariate DLM

Dynamic linear models are state space models where we assume that there is an hidden dynamic of the system generating the observation Y_t. The state is represented by a set of latent, or unobserved, variables, each one evolving according to a specific dynamic.

The structural property enabling dynamic modeling in DLM is conditional independence: given the present, the future is independent of the past. DLM are markovian models, where all the information required to forecast the future is contained in the current parameter vector $\boldsymbol{\theta_t}$ and all the information concerning the future is contained in the posterior $\boldsymbol{\theta_t}|\mathcal{D}_t$, where $\mathcal{D}_t$ is the information set at time t, i.e. the set of all the known informations up to time t.

Let Y_t be a vector of observations, the general univariate normal DLM is defined by the quadruple $\{F_t, G_t, \sigma_t^2, W_t\}$ for each time t. This quadruple defines the model between Y_t and the parameter vector $\boldsymbol{\theta_t}$ by means of the following distributions

$$Y_t|\boldsymbol{\theta_t}, \mathcal{D}_{t-1} \sim \mathcal{N}(F_t^T\boldsymbol{\theta_t}, \sigma_t^2)$$
$$\boldsymbol{\theta_t}|\boldsymbol{\theta_{t-1}}, \mathcal{D}_{t-1} \sim \mathcal{N}(G_t\boldsymbol{\theta_{t-1}}, W_t)$$

Equivalently we can write

$$\begin{array}{lll} Y_t = F_t^T\boldsymbol{\theta_t} + v_t & v_t \sim \mathcal{N}(0, \sigma_t^2) & \text{Observation equation} \\ \boldsymbol{\theta_t} = G_t\boldsymbol{\theta_{t-1}} + \boldsymbol{\omega_t} & \boldsymbol{\omega_t} \sim \mathcal{N}(\mathbf{0}, W_t) & \text{System equation} \end{array}$$

We assume the error sequences v_t and $\boldsymbol{\omega_t}$ to be internally and mutually independent. This equations relates the response Y_t to $\boldsymbol{\theta_t}$ via a dynamic linear regression with a normal error structure having known, though possibly time varying, observational variance σ_t^2. Given time t:

- F_t is an $n \times 1$ design matrix of known values of independent variables
- θ_t is a $n \times 1$ state vector
- G_t is an $n \times n$ known matrix, describing the state dynamic, named evolution matrix
- W_t is an $n \times n$ known variance matrix

Note that, given θ_t, the distribution of the response Y_t is independent of any past observation $Y_{t-1}, Y_{t-2}, \ldots, Y_0$ and $\theta_{t-1}, \theta_{t-2}, \ldots, \theta_0$. In the same way, given θ_{t-1} the distribution of the state vector is independent of any past observation $Y_{t-1}, Y_{t-2}, \ldots, Y_0$ and $\theta_{t-2}, \theta_{t-3}, \ldots, \theta_0$.

We call $\mu_t = F_t^t\boldsymbol{\theta_t}$ mean response, or level, of the series. Morover we assume that the information set at time t, namely $\mathcal{D}_t$, contains the values of W_t, F_t, G_t for any time $t > 0$ as well as all past response observations $Y_{t-1}, Y_{t-2}, \ldots, Y_0$.

The requirement that W_t, F_t, G_t must be known for any time t implies that when forecasting k step ahead, the value of possible regressive terms must be known up to time t+k. We assume for our application that values for the covariates are known and supplied by an external font (for example, since all our regressors concern wheater conditions, we can use wheater forecasts as input for our model)

If the couple $\{F_t, G_t\}$ is constant for all t, then the DLM is referred to as **time series DLM**, or **TSDLM**. If morover the observational variance σ_t^2 and the state evolution variance W_t are constant for all t as well, the TSDLM is referred to as **constant DLM**. A constant DLM is characterized by the single quadruple $\{F, G, \sigma^2, W\}$.

The main result about normal DLM theory follows: intuitively the central characteristic of the normal model is that at any time, existing information about the system is represented and sufficiently summarised by the posterior distribution for the current state vector.

In the following we assume that $\mathcal{D}_0$ contains our prior knowledge about the state vector as well as the values of $\{F_t, G_t, \sigma_t^2, W_t\}$ for any t, so that at any time instant $t > 0$, the information set $\mathcal{D}_t$ is given by $\mathcal{D}_{t-1} \cup \{Y_t\}$.

The general conjugate DLM model is defined by the following equations:

$$\begin{cases} Y_t = F_t^T \boldsymbol{\theta_t} + v_t \quad v_t \sim \mathcal{N}(0, \sigma^2) & \text{Observation equation} \\ \boldsymbol{\theta_t} = G_t \boldsymbol{\theta_{t-1}} + \boldsymbol{\omega_t} \quad \boldsymbol{\omega_t} \sim \mathcal{N}(\mathbf{0}, \sigma^2 W_t) & \text{System equation} \\ \boldsymbol{\theta_0} | \mathcal{D}_0, \sigma^2 \sim \mathcal{N}(\boldsymbol{m_0}, \sigma^2 C_0) & \text{Initial information} \\ \sigma^2 | \mathcal{D}_0 \sim inv\Gamma\left(\frac{n_0}{2}, \frac{n_0 S_0}{2}\right) & \end{cases}$$

It can be prooven (see appendix) that for any time $t > 0$ the one step forecast and posterior distributions can be obtained sequentially as:

$$\begin{aligned}
&\sigma^2 | \mathcal{D}_{t-1} \sim inv\Gamma\left(\frac{n_{t-1}}{2}, \frac{n_{t-1} S_{t-1}}{2}\right) & & & \text{Prior for } \sigma^2 \\
&\boldsymbol{\theta_t} | \mathcal{D}_{t-1}, \sigma^2 \sim \mathcal{N}(\boldsymbol{a_t}, \sigma^2 R_t) & R_t &= G_t C_{t-1} G_t^T + W_t & \text{Prior for } \boldsymbol{\theta_t} \\
& & \boldsymbol{a_t} &= G_t \boldsymbol{m_{t-1}} & \\
&Y_t | \mathcal{D}_{t-1}, \sigma^2 \sim \mathcal{N}(f_t, \sigma^2 Q_t) & Q_t &= F_t^T R_t F_t + 1 & \text{One step forecast} \\
& & f_t &= F_t^T \boldsymbol{a_t} & \\
&\sigma^2 | \mathcal{D}_t \sim inv\Gamma\left(\frac{n_t}{2}, \frac{n_t S_t}{2}\right) & n_t &= n_{t-1} + 1 & \text{Posterior for } \sigma^2 \\
& & n_t S_t &= n_{t-1} S_{t-1} + e_t^2 Q_t^{-1} & \\
&\boldsymbol{\theta_t} | \mathcal{D}_t, \sigma^2 \sim \mathcal{N}(\boldsymbol{m_t}, \sigma^2 C_t) & m_t &= \boldsymbol{m_{t-1}} + A_t e_t \quad C_t = R_t - A_t Q_t A_t^T & \text{Posterior for } \boldsymbol{\theta_t} \\
& & A_t &= R_t F_t Q_t^{-1} \quad e_t = Y_t - f_t &
\end{aligned}$$

In absence of a conjugate structure MCMC algorithm can be adopted to sample from the posterior distribution of parameters. Anyway the recursive update structure still applies.

4.2.1 Principle of superposition

We pose the problem of defining system matrices F_t and G_t. In applications, models are usually constructed by combining two or more component DLMs, each of which captures an individual feature of the real series under study. The construction of complex DLMs from component DLMs is referred to as superposition. The following holds:

Consider h time series $Y_{it} \quad i = 1, \dots, h$, each one generated by a DLM $\mathcal{M}_i$ described by the quadruple $\{F_{it}, G_{it}, \sigma_{it}^2, W_{it}\}$. Let assume the state vector $\boldsymbol{\theta_{it}}$ of model $\mathcal{M}_i$ to be of dimension $n_i \times 1$. Denote with v_{it} and $\boldsymbol{\omega_{it}}$ the observational and evolution error od model $\mathcal{M}_i$. The state vectors are distinct, and for all distinct $i \neq j$, the series v_{it} and ω_{it} are mutually independent of the series v_{jt} and ω_{jt}. Then the series

$$Y_t = \sum_{i=1}^{h} Y_{it}$$

follows the n-dimensional DLM $\{F_t, G_t, \sigma_t^2, W_t\}$ where $n = \sum_{i=1}^{h} n_i$ and

$$F_t = \begin{bmatrix} F_{1t} \\ F_{2t} \\ \vdots \\ F_{ht} \end{bmatrix} \qquad G_t = \begin{bmatrix} G_{1t} & 0 & \dots & 0 \\ 0 & G_{2t} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & G_{ht} \end{bmatrix} \qquad \sigma_t^2 = \sum_{i=1}^{h} \sigma_{it}^2 \qquad W_t = \begin{bmatrix} W_{1t} & 0 & \dots & 0 \\ 0 & W_{2t} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & W_{ht} \end{bmatrix}$$

and the state vector $\boldsymbol{\theta_t}$ is obtained by concatenating the state vectors of each single model $\mathcal{M}_i$. The final message to take home is that **a linear combination of DLMs is a DLM**. In practice, any complex target series Y_t we want to forecast can be seen as a linear combination of simple components: a **trend** describing the slow variation dynamic of the series, a **seasonal** component, capturing periodical patterns inside the data, a **regressive** and **autoregressive** component which incorporate external information helpfull in forecasting the target series Y_t inside the model.

Our focus is now in presenting the main component DLM which will be used in our case study.

Unless otherwise specified, the state variance matrix W is assumed to be diagonal, i.e. $W = [w_1, \ldots, w_p]$ if W is a $p \times p$ matrix.

4.3 Polynomial trend models

In time series polynomial trend models prove useful in describing trends that are generally viewed as smooth developments over time. Relative to the sampling interval of the series and the required forecast horizons, such trends are usually well approximated by low-order polynomial functions of time. Indeed, a first or second order polynomial component DLM is often quite adequate for short term forecasting. Polynomial DLMs are a subset of TSDLM, hence both F_t and G_t are constant over time and equal to F and G respectively.

4.3.1 Second order Polynomial trend model

System matrices in this case are as follow:

$$F = \begin{bmatrix} 1 & 0 \end{bmatrix} \qquad G = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \qquad W_t = \begin{bmatrix} w_{t,1} & 0 \\ 0 & w_{t,2} \end{bmatrix}$$

Let consider the following parametrization

$$\boldsymbol{\theta_t} = \begin{bmatrix} \theta_{t,1} \\ \theta_{t,2} \end{bmatrix} = \begin{bmatrix} \mu_t \\ \beta_t \end{bmatrix}$$

where μ_t takes the meaning of level of the series while β_t of increment of the level μ_t between time $t-1$ and t (indeed, the first derivative of the level). Overall the second order polynomial DLM is defined by equations:

$$\begin{cases} Y_t = \mu_t + v_t \quad v_t \sim \mathcal{N}(0, \sigma^2) & \text{Observation equation} \\ \mu_t = \mu_{t-1} + \beta_{t-1} + \omega_{t,1} \quad \omega_{t,1} \sim \mathcal{N}(0, w_1) & \text{System equation} \\ \beta_t = \beta_{t-1} + \omega_{t,2} \quad \omega_{t,2} \sim \mathcal{N}(0, w_2) & \\ \boldsymbol{\theta_0} | \mathcal{D}_0 \sim \mathcal{N}(m_0, C_0) & \text{Initial information} \\ w_i \sim \text{LogNormal}(a, b) \quad i = 1, 2 & \text{Parameter priors} \\ \sigma^2 \sim \text{inverse-gamma}(c, d) & \end{cases}$$

where the error sequences v_t and $\boldsymbol{\omega_t}$ are internally independent and mutually independent, and independent of initial information $\boldsymbol{\theta_0}|\mathcal{D}_0$. We can see why β_t takes the meaning of first derivative of the level μ_t. Indeed, let us not consider the stochastic variation affecting μ_t, thus from the system equation we can write

$$\mu_t - \mu_{t-1} = \frac{\Delta \mu_t}{\Delta t} \overset{\Delta t = 1}{=} \beta_{t-1}$$

hence β_{t-1} indicates the variation of the series level μ_t between the time instant $t-1$ and t, excluding a small fluctuation. Notice that β_t itself exhibit a slow variation dynamic, (random walk dynamic). Thus in such model, β_t and as a consequence the increment of μ_t, changes only by the effect of random fluctuations. Inference is performed using general DLM theory seen in previous sections.

4.3.2 Semilocal Linear trend model

The local linear trend is good only for short term forecasting. For this reason is common to consider a slight variation of the second order polynomial model. A semilocal linear trend model assumes the level component μ_t to evolve according to a random walk, but the slope component β_t moves according to an AR(1) process centered on a potentially nonzero value D. Generally a semilocal linear trend allow to achieve better results in terms of long term forecast. The DLM is described by the following set of equations:

$$
\begin{cases}
y_t = \mu_t + \epsilon_t \quad \epsilon_t \sim \mathcal{N}(0, \sigma^2) & \text{Observation equation} \\
\\
\mu_t = \mu_{t-1} + \beta_{t-1} + \omega_t^{(1)} \quad \omega_t^{(1)} \sim \mathcal{N}(0, w_1) & \text{system equation} \\
\beta_t = D + \alpha(\beta_{t-1} - D) + \omega_2 \quad \omega_t^{(2)} \sim \mathcal{N}(0, w_2) & \\
\boldsymbol{\theta_0} | \mathcal{D}_0 \sim \mathcal{N}(m_0, C_0) & \text{Initial information} \\
w_i \sim \text{LogNormal}(a, b) \quad i = 1, 2 & \text{Parameter priors} \\
\sigma^2 \sim \text{inverse-gamma}(c, d) & \\
D \sim \mathcal{N}(\mu, \Sigma) & \\
\alpha \sim \mathcal{N}(0, 1) \text{ truncated on (-1,1)} &
\end{cases}
$$

Note that $\beta_t = D + \alpha(\beta_{t-1} - D) +_t^{(2)}$ for $|\alpha| < 1$ is an Autoregressive model of order 1 centered in D. A stationary AR(1) process is less variable than a random walk when making projections far into the future, so this model often gives more reasonable uncertainty estimates when making long term forecast.

$$
\beta_t = D + \alpha(\beta_{t-1} - D) + w_t^{(2)} \quad w_t^{(2)} \sim \mathcal{N}(0, w_t^{(2)}) \Rightarrow \beta_t = \alpha(\beta_{t-1} - D) + w_t \quad w_t^{(2)} \sim \mathcal{N}(D - \alpha D, w_t^{(2)})
$$

4.4 Form free Seasonal models

We use seasonal models to capture patterns in the data which happen with a fixed period. In the most general form, seasonal models do not take any specific functional form. When the seasonal factors relating to a period may take any arbitrary real value, the seasonal pattern is termed form free.

Let p the period of the seasonal pattern, then we define

$$
\boldsymbol{E_p} = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix} \qquad P = \begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \\ 1 & 0 & 0 & \dots & 0 \end{bmatrix}
$$

It can be prooven that matrix P is p-cyclic, namely $P^{k+np} = P^k$. In practice this allow to rotate the state vector every p steps, in a way that the same factor ψ_p contributes to the model output Y_t every p time stesp. Using $\boldsymbol{E_p}$ and P so defined, the form free seasonal factor DLM is written as:

$$
\begin{cases}
Y_t = \boldsymbol{E_p^T} \boldsymbol{\psi_t} + v_t \quad v_t \sim \mathcal{N}(0, \sigma^2) & \text{Observation equation} \\
\boldsymbol{\psi_t} = P \boldsymbol{\psi_{t-1}} + \boldsymbol{\omega_t} \quad \boldsymbol{\omega_t} \sim \mathcal{N}(\mathbf{0}, W) & \text{System equation} \\
\boldsymbol{\phi_0} | \mathcal{D}_0 \sim \mathcal{N}(m_0, C_0) & \text{Initial information} \\
w_i \sim \text{LogNormal}(a, b) \quad i = 1, \dots, p & \text{Parameter priors} \\
\sigma^2 \sim \text{inverse-gamma}(c, d) &
\end{cases}
$$

The previous definition is usefull in describing a purely seasonal model. Generally we decompose a set of p seasonal factors in two components: a deseasonalised level of the series and p seasonal deviations from this level. The seasonal deviations from the series level are called seasonal effects.

Due to identifiability problems, we need to constraint the $\boldsymbol{\psi_t}$ vector. The commonest constraint is the zero-sum constraint. Hence a form free seasonal effect DLM is defined exactly as before but where the vector $\boldsymbol{\psi_t}$ satisfies $\sum_{i=0}^{p-1} \psi_{t,i} = 0$. The superposition of a seasonal effect model with a polynomial trend model can be used to describe a series Y_t with a smooth development over time affected by seasonal fluctuation around its trend.

For example, the following describes a second-order polynomial seasonal effects model:

$$
\begin{cases}
Y_t = \mu_t + \phi_{t,0} + v_t \quad v_t \sim \mathcal{N}(0,\sigma^2) & \text{Observation equation} \\
\mu_t = \mu_{t-1} + \beta_{t-1} + \omega_t^1 \quad \omega_t^1 \sim \mathcal{N}(0,w_1) & \text{Trend dynamic} \\
\beta_t = \beta_{t-1} + \omega_t^2 \quad \omega_t^2 \sim \mathcal{N}(0,w_2) & \\
\phi_{t,r} = \phi_{t-1,r+1} + \omega_{t,r} \quad \omega_{t,r} \sim \mathcal{N}(0,w_r) \quad r = 1,\dots,p-2 & \text{Seasonal dynamic} \\
\phi_{t,p-1} = \phi_{t-1,0} + \omega_{t,p-1} \quad \omega_{t,p-1} \sim \mathcal{N}(0,w_{p-1}) &
\end{cases}
$$

which poses a set of p+2 latent variables (two variables for the trend part and p seasonal effects).

4.5 Regressive models

In forecasting a time series Y_t is generally not enought to use information coming only from the series itself or on some periodical patterns. For this reason is fundamental to add external informations which help the model in predicting the occurence of an event before the event happen. Let $X_1, \dots, X_n$ be n independent time series. The value of the i th variable X_i at each time t is assumed known. For $t = 1, \dots,$ let the regression vector F_t be given by

$$F_t = \begin{bmatrix} X_{t1}, \dots, X_{tn} \end{bmatrix}$$

. For each t then, the model equations for a regressive model are:

$$
\begin{cases}
Y_t = F_t \boldsymbol{\theta_t} + v_t \quad v_t \sim \mathcal{N}(0,\sigma^2) & \text{Observation equation} \\
\boldsymbol{\theta_t} = \boldsymbol{\theta_{t-1}} + \boldsymbol{\omega_t} \quad \boldsymbol{\omega_t} \sim \mathcal{N}(\mathbf{0},W) & \text{System equation} \\
\boldsymbol{\theta_0} | \mathcal{D}_0 \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma) & \text{Initial Information} \\
w_i \sim \text{LogNormal}(a,b) \quad i = 1,\dots,n & \text{Parameter priors} \\
\sigma^2 \sim \text{inverse-gamma}(c,d) &
\end{cases}
$$

the latent vector θ_t is assumed to follow a random walk, or slow variation, dynamic.

In case of autoregression of order p, (AR(p) models) we consider the general DLM formulation with system matrices defined as follow:

$$
\boldsymbol{F_t} = \boldsymbol{E_p} = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix} \qquad G_t = \begin{bmatrix} \nu_1 & \nu_2 & \nu_3 & \dots & \nu_p \\ 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{bmatrix}
$$

4.6 Proposed Dynamic Linear Model

Because the target series presents big changes from one time step to the next one, a trend model is not suitable to describe the series because we can't imagine an enought smooth dynamic of the series over time. As a result the proposed model does not contain any explicit trend dynamic.

We propose a dynamic linear model composed by the superposition of the following component DLM:

- **Seasonal component**: since we use daily aggregated data we set a day of the week effect seasonality. As a result we introduce 7 latent variables, one for each day of the week, whose dynamics is described by a form free seasonal effect DLM.
- **Autoregressive component**: we suppose that the value of PM at time t is not independent from the value of PM registered during the past days. For this reason we introduce an autoregressive dynamic of order 2.
- **Regressive component**: adding external information to the model in order to give to it information to forecast the increment or decrement of the target series before such change actually happen is fundamental. The following regressors are supplied to the model
  - temperature
  - humidity
  - wind
  - wind direction

– rainfall intensity
– global solar radiation

The overall model is so formulated:

$$
\begin{cases}
Y_t = \boldsymbol{E_7}\boldsymbol{\phi_t} + F_t\boldsymbol{\theta_t} + \boldsymbol{E_2}\alpha_t + v_t \quad v_t \sim \mathcal{N}(0,\sigma^2) & \text{Observation equation} \\
 & \text{Seasonal dynamic} \\
\phi_{t,r} = \phi_{t-1,r+1} + \omega_{t,r} & \omega_{t,r} \sim \mathcal{N}(0, w_r^{\phi}) \quad r = 1,\dots,5 \\
\phi_{t,6} = \phi_{t-1,0} + \omega_{t,p-1} & \omega_{t,6} \sim \mathcal{N}(0, w_6^{\phi}) \\
 & \text{Regressive dynamic} \\
\boldsymbol{\theta_t} = \boldsymbol{\theta_{t-1}} + \boldsymbol{\omega_t} & \boldsymbol{\omega_t} \sim \mathcal{N}(\boldsymbol{0}, W^{\theta}) \\
 & \text{Autoregressive dynamic} \\
\alpha_t = \nu_1\alpha_{t-1} + \nu_2\alpha_{t-2} + \omega_t & \omega_t \sim \mathcal{N}(0, w_1^{\alpha}) \\
\alpha_{t-1} = \alpha_{t-2} + \omega_t & \omega_t \sim \mathcal{N}(0, w_2^{\alpha}) \\
\boldsymbol{\phi_0}|\mathcal{D}_0,\sigma^2 \sim \mathcal{N}(\boldsymbol{m_0}^{\phi}, \sigma^2 C_0^{\phi}) & \text{Initial information} \\
\boldsymbol{\theta_0}|\mathcal{D}_0,\sigma^2 \sim \mathcal{N}(\boldsymbol{m_0}^{\theta}, \sigma^2 C_0^{\theta}) & \\
\boldsymbol{\alpha_0}|\mathcal{D}_0,\sigma^2 \sim \mathcal{N}(\boldsymbol{m_0}^{\alpha}, \sigma^2 C_0^{\alpha}) & \\
\sigma^2|\mathcal{D}_0 \sim \text{inv}\Gamma(a,b) & \text{Parameter priors} \\
w_r^{\phi} \sim \text{LogNormal}(c_r^{\phi}, d_r^{\phi}) \qquad r = 1,\dots,6 & \\
w_i^{\theta} \sim \text{LogNormal}(c_j^{\theta}, d_j^{\theta}) \qquad j = 1,\dots,n & \\
w_i^{\alpha} \sim \text{LogNormal}(c_k^{\alpha}, d_k^{\alpha}) \qquad k = 1,2 & \\
\nu_i \sim \mathcal{N}(0,1) \text{ truncated on (-1,1)} \qquad i = 1,2 & \\
\boldsymbol{m_0^{\phi}}, \boldsymbol{m_0^{\theta}}, \boldsymbol{m_0^{\alpha}}, C_0^{\phi}, C_0^{\theta}, C_0^{\alpha}, c_r^{\phi}, c_j^{\theta}, c_k^{\alpha}, d_r^{\phi}, d_j^{\theta}, d_k^{\alpha} \quad \forall r,j,k & \text{constants}
\end{cases}
$$

4.7 Regressive components

4.7.1 Rainfall Intensity

A major factor in the reduction of the levels of PM are the precipitations. Actually the values of PM are not strictly related to the amount of rain fallen during the day, but there is a significant correlation between decrement of pollutants and intensity of precipitation: the stronger the precipitation event, the higher the reduction of PM10, which indicates that precipitation has a wet scavenging effect on PM10. The change of PM10 before and after precipitation is related to the initial concentration of PM10 before precipitation. The higher the initial concentration of PM10 is, the greater the removal of PM10 by precipitation will be; that is, when the air quality is relatively poor, the precipitation will play a role in scavenging pollutants.

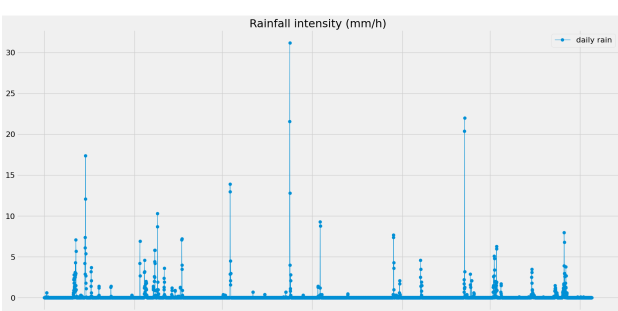


Figure 10: Rainfall intensity

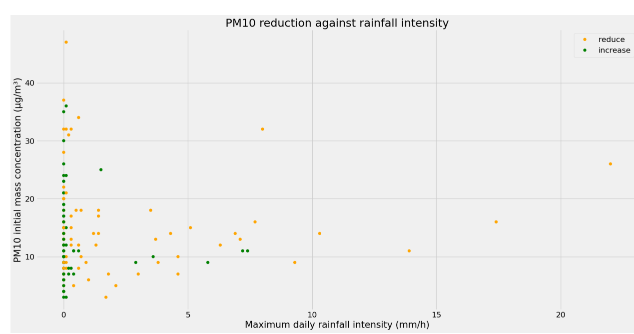


Figure 11: Reduction of PM as function of rainfall intensity

As we can see from the plot, precipitation events are often related to a decrement of the level of pollutants before and after the event. However, we can note that even if there is no rain the air quality can be good. Note that stronger rain events are always associated to a reduction of PM10 levels.

To capture the intensity of the precitpitation, we define the rainfall intensity as the area under the curve of the rain time series between time t and t-1, that indicate how much millimeter of water are fallen during one hour:

$$\begin{aligned}\text{rain}_{min} &= \min\left(\text{rain}_t, \text{rain}_{t-1}\right)\\ \text{rain}_{max} &= \max\left(\text{rain}_t, \text{rain}_{t-1}\right)\\ \text{Rainfall Intensity} &= \text{rain}_{min} + \frac{1}{2}\Big(\text{rain}_{max} - \text{rain}_{min}\Big)\end{aligned}$$

Because we have daily aggregated data, we take the maximum value of rainfall intensity registered during the day, for each day. The scatterplot of figure 11 has on the x-axis the rainfall intensity at time t and on the y-axis the PM10 values before the precipitation at time t (i.e. at time t-1). As we can see, if there is an heavy rain then the level of pm decrease, if there is no rain or a light rain then the pollutant in the air tends to increase. Therefore, the rainfall intensity is a good predictor for the PM values, especially to predict a reduction in the level of PM.

4.7.2 Wind speed and wind direction

Wind is another major factor for the reduction of pollutant particles in the air. Differently from other variables, wind encode a vectorial information: the intensity of the wind (expressed in m/s) and its direction (expressed in degrees with respect to the north). Stronger winds have a stronger effect in reducing the level of PM. Moreover, due to the position of the city of Milan, if the wind comes from north going to south, these winds have a much more stronger effect in the reduction of PM levels.

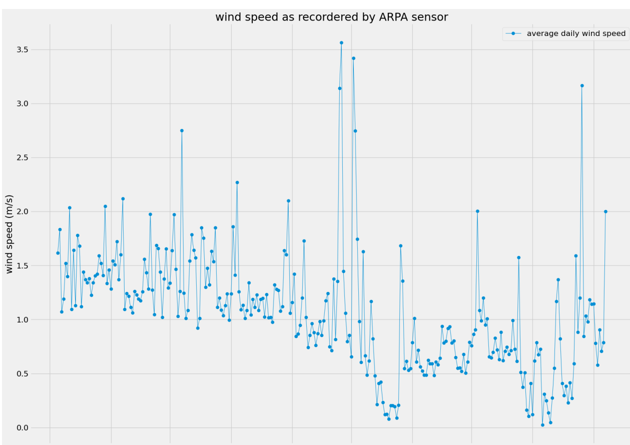


Figure 12: Average wind speed

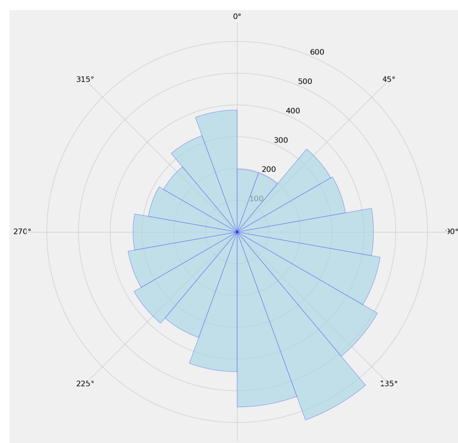


Figure 13: Distribution of wind direction

From the wind time series we hence extract the following variables:

- wind speed, as recordered by the sensor
- the sine and cosine of the angle with respect to the north
- 4 dummy variables, one for each main direction of the wind, namely: north-east, north-west, south-west, south-east. For example the dummy for direction north-east is 1 if the average angle of the wind observed during the day is between 0 and 89 degrees.

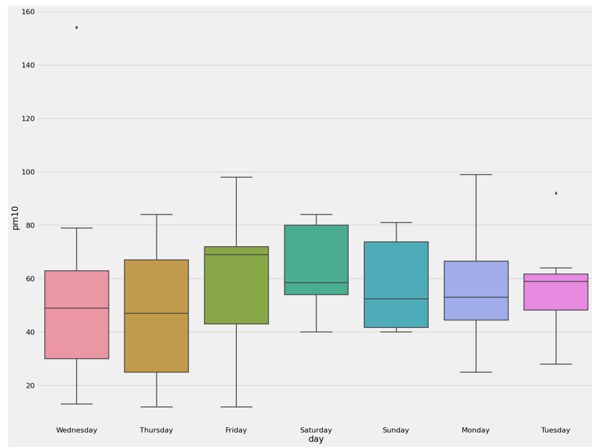
Note that having introduced 4 dummy variables explicitly allow us to incorporate our prior knowledge about the clean effect of winds going to south: we will put a strong negative value on the parameters describing the prior distribution of these factors to indicate that winds going to south contribute in a strong reduction in the level of

PM.

$$\begin{bmatrix}\theta_{\text{wind speed}}\\ \theta_{\text{sin}}\\ \theta_{\text{cos}}\\ \theta_{\text{NE}}\\ \theta_{\text{SE}}\\ \theta_{\text{SW}}\\ \theta_{\text{NW}}\end{bmatrix}\Bigg|\mathcal{D}_0,\sigma^2 \sim \mathcal{N}(\boldsymbol{m}_0^{\text{wind}}, \sigma^2 C_0^{\text{wind}}) \qquad \boldsymbol{m}_0^{\text{wind}} = \begin{bmatrix}\hat{\alpha_{\text{wind speed}}}\\ \hat{\alpha_{\text{sin}}}\\ \hat{\alpha_{\text{cos}}}\\ 0\\ \mathbf{-10}\\ \mathbf{-10}\\ 0\end{bmatrix}$$

4.7.3 Seasonal block prior

About the seasonal block, we need to incorporate our prior knowledge about the mean levels of PM registered during the week. In order to accomplish this, we have taken past recorded data from ARPA stations for the same period used to test our model (namely 1 january 2020 - 1 march 2020) and computed the mean PM concentration grouped by day. The resulting values are used as mean for the normal distribution defining the seasonal block state vector prior. Boxplots show the empirical distribution of particulate matter computed using old data.

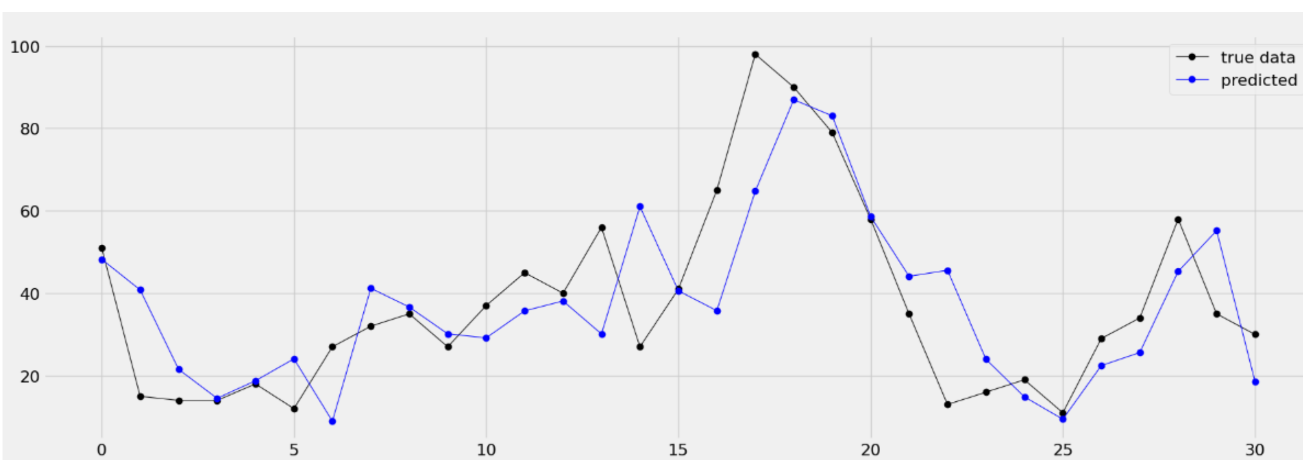


4.7.4 Other regressors

Other regressors such temperature, humidity and global solar radiation are given as input to the model without any other transformation.

4.8 Model performance and criticism

We report the one step predictive distribution of the model. The black line reports the true PM recordered by the sensor while the blue line corresponds to the mean of the posterior distribution of Y_t one step ahead.



As can be seen, the model commits big errors when it has to forecast an increase of the level of PM, and such increments can't be explained by the seasonal component nor the regressive part. The main reason for this lack of performance in anticipating an increment of the values of PM may be found in the supplied regressors. From domain literature is known that weather factors are only responsable for a decrease of the pollutant levels in the air. Temperature may be indirectly correlated with an increment of the level of PM, due to home heatings. But it can't explain the huge increment in the level of PM which may happen in a relatively short period of time. The model is on the contrary good in predicting a decrease of the level of PM, meaning that the supplied regressors are indeed usefull to forecast a decrement of the level of particulate matter.

We think that the model could be improved by supplying to it further information which may be helpfull in forecasting an increment of the level of PM.