Air Quality Forecasting

Datena Amedeo, Dell'Atti Martina, Palummo Alessandro

18 February 2021

POLITECNICO
MILANO 1863

About the project

- Our goal is to model PM (Particulate Matter) air concentration in Milan.
- Our dataset consists of PM values and other atmospheric data collected by a number of sensors in Milan, plus PM concentration detected by ARPA stations
- our response variable is univariate, but we are including a number of regressors in our models, such as:
  - temperature
  - humidity
  - wind speed and wind direction
  - rain amount and rain intensity

When we met before

The models presented last time were not performing well:

- All the models were static, in the sense that parameters are kept fixed once the model is fitted
- Static models are not able to adapt to big changes in the time series dynamics

When we met before

The models presented last time were not performing well:

- All the models were static, in the sense that parameters are kept fixed once the model is fitted
- Static models are not able to adapt to big changes in the time series dynamics

$\Rightarrow$ **Solution:Dynamic Models**

- ▷ Allow the model to keep track of the ongoing development of the series
- ▷ Can handle non-stationary processes, missing values as well as time varying variance.
- ▷ Allow for regression coefficients to change over time
- ▷ Allow for local trend and seasonality adjustments

Exponential Smoothing

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:

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

And, recursively:

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

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

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

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, homoschedastic normally distributed errors, with standard deviation σ. This formulation has an equivalent linear formulation by constructing opportune matrices:

$$\boldsymbol{Y} = M\boldsymbol{\psi} + L\boldsymbol{\varepsilon}$$

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 $\boldsymbol{\psi}$ is a vector of unknown initial conditions for level, growth and seasonal terms.

The **likelihood** of this model is:

$$f(\boldsymbol{Y}|\boldsymbol{\psi},\boldsymbol{\theta},\sigma) = \sigma^{-n} exp\left\{-\frac{1}{2\sigma^2}(\tilde{\boldsymbol{\psi}}-\boldsymbol{\psi})' X' X(\tilde{\boldsymbol{\psi}}-\boldsymbol{\psi})\right\} \times exp\left\{-\frac{1}{2\sigma^2}(L^{-1}\boldsymbol{Y})'(I-P_X)L^{-1}\boldsymbol{Y}\right\}$$

where

- X is the matrix $L^{-1}M$.
- P_X is the orthogonal projection matrix on the column space of X: $(P_X = X(X'X)^{-1}X')$.
- $\boldsymbol{\psi} = (b_0, c_{1-p}, ..., c_0)'$ is the vector of initial conditions with the restriction $a_0 + b_0 = 0$
- $\tilde{\boldsymbol{\psi}} = (X'X)^{-1}X'L^{-1}\boldsymbol{Y}$
- $\boldsymbol{\theta} = (\alpha, \beta, \gamma)'$

Priors:

$$f(\sigma,\boldsymbol{\psi}) \propto \sigma^{-1}\chi_{(0,+\infty)}$$
$$\alpha \sim \mathcal{U}[0,1]$$
$$\beta \sim \mathcal{U}[0,1]$$
$$\gamma \sim \mathcal{U}[0,1]$$

Dynamic Linear Models - general definition

Let

- $Y_t \in \mathcal{R}$ be the observation at time t
- $\theta_t \in \mathcal{R}^n$ be the state vector at time t
- F_t be an $n \times 1$ known design matrix
- G_t be an $n \times n$ describing the state dynamic, named evolution matrix
- W_t be an $n \times n$ known covariance matrix Let assume that the set $\mathcal{D}_0$ contains prior knowledge about the state vector as well as the values of $\{F_t, G_t, \sigma_t^2, W_t\}$ for any t.

  The general univariate Dynamic Linear Model is defined by the system of equations

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

Dynamic Linear Models - general definition

Let

- $Y_t \in \mathcal{R}$ be the observation at time t
- $\theta_t \in \mathcal{R}^n$ be the state vector at time t
- F_t be an $n \times 1$ known design matrix
- G_t be an $n \times n$ describing the state dynamic, named evolution matrix
- W_t be an $n \times n$ known covariance matrix Let assume that the set $\mathcal{D}_0$ contains prior knowledge about the state vector as well as the values of $\{F_t, G_t, \sigma_t^2, W_t\}$ for any t.

  The general univariate Dynamic Linear Model is defined by the system of equations

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

  - When forecasting more than one step ahead, the corresponding quantities F_{t+k} , G_{t+k} , and W_{t+k} must belong to the current information set $\mathcal{D}_t$.
  - Future values of regressors must be known up to time $t + k$ given we are at time t. We assume to have forecasts about future values of the regressors provided by external sources.

Prior for σ^2

$$\sigma^2|\mathcal{D}_{t-1} \sim inv-gamma\left(\frac{n_{t-1}}{2}, \frac{n_{t-1}S_{t-1}}{2}\right)$$

Prior for $\boldsymbol{\theta_t}$

$$\boldsymbol{\theta_t}|\mathcal{D}_{t-1}, \sigma^2 \sim \mathcal{N}(\boldsymbol{a_t}, \sigma^2 R_t)$$

$$R_t = G_t C_{t-1} G_t^T + W_t$$
$$\boldsymbol{a_t} = G_t \boldsymbol{m_{t-1}}$$

One step forecast

$$Y_t|\mathcal{D}_{t-1}, \sigma^2 \sim \mathcal{N}(f_t, \sigma^2 Q_t)$$

$$Q_t = F_t^T R_t F_t + 1$$
$$f_t = F_t^T \boldsymbol{a_t}$$

Posterior for σ^2

$$\sigma^2|\mathcal{D}_t \sim inv-gamma\left(\frac{n_t}{2}, \frac{n_t S_t}{2}\right)$$

$$n_t = n_{t-1} + 1$$

$$n_t S_t = n_{t-1} S_{t-1} + e_t^2 Q_t^{-1}$$

Posterior for $\boldsymbol{\theta_t}$

$$\boldsymbol{\theta_t}|\mathcal{D}_t, \sigma^2 \sim \mathcal{N}(\boldsymbol{m_t}, \sigma^2 C_t)$$

$$m_t = \boldsymbol{m_{t-1}} + A_t e_t \quad C_t = R_t - A_t Q_t A_t^T$$
$$A_t = R_t F_t Q_t^{-1} \quad e_t = Y_t - f_t$$

The superposition principle

How can we define G_t and F_t?

Theorem

Given h time series $Y_{it} \quad i = 1, \ldots, h$, each one generated by a DLM $\mathcal{M}_i$ described by the quadruple $\{F_{it}, G_{it}, \sigma^2_{it}, W_{it}\}$. Let $\boldsymbol{\theta_{it}} \in \mathcal{R}^{n-1}$ and let v_{it} and $\boldsymbol{\omega_{it}}$ the observational and evolution error of model $\mathcal{M}_i$. Morover assume the state vectors distinct, and for all distinct $i \neq j$, the series v_{it} and ω_{it} be 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 $\{F_t, G_t, \sigma^2_t, W_t\}$ where $n = \sum_{i=1}^{h} n_i$ and

$$F_t = \begin{bmatrix} F_{1t} \\ F_{2t} \\ \vdots \\ F_{ht} \end{bmatrix} \quad G_t = \begin{bmatrix} G_{1t} & 0 & \ldots & 0 \\ 0 & G_{2t} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & G_{ht} \end{bmatrix} \quad \sigma^2_t = \sum_{i=1}^{h} \sigma^2_{it} \quad W_t = \begin{bmatrix} W_{1t} & 0 & \ldots & 0 \\ 0 & W_{2t} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & W_{ht} \end{bmatrix}$$

The state vector $\boldsymbol{\theta_t}$ is obtained by concatenating the state vectors of each single model $\mathcal{M}_i$.

Let the level μ_t and the slope β_t two latent variables:

Semilocal linear trend

$$\begin{cases} Y_t = \mu_t + v_t & v_t \sim \mathcal{N}(0, \sigma^2) \\ \mu_t = \mu_{t-1} + \beta_{t-1} + \omega_{t,1} & \omega_{t,1} \sim \mathcal{N}(0, w_1) \\ \beta_t = D + \alpha(\beta_{t-1} - D) + \omega_{t,1} & \omega_{t,2} \sim \mathcal{N}(0, w_2) \\ \boldsymbol{\theta_0} | \mathcal{D} \sim \mathcal{N}(\boldsymbol{m_0}, C_0) \end{cases}$$

Priors

$$\begin{cases} \sigma^2 | \mathcal{D}_0 \sim \text{inv-gamma}(c, d) \\ w_1 \sim LogNormal(a^1, b^1) \\ w_2 \sim LogNormal(a^2, b^2) \\ \alpha \sim \mathcal{N}(0, 1) \text{ truncated on (-1,1)} \\ D \sim \mathcal{N}(d, \sigma_d^2) \end{cases}$$

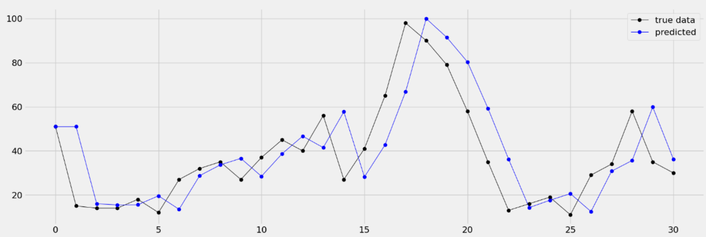


Figure: Semilocal linear trend model

Form Free Seasonal models

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, define $\boldsymbol{E_p}$ and the $p \times p$ permutation matrix P as

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

A form free seasonal effect model is described by the following system:

$$\begin{cases} Y_t = \boldsymbol{E_p^T} \boldsymbol{\phi_t} + v_t \quad v_t \sim \mathcal{N}(0, \sigma^2) \\ \boldsymbol{\phi_t} = P \boldsymbol{\phi_{t-1}} + \boldsymbol{\omega_t} \quad \boldsymbol{\omega_t} \sim \mathcal{N}(\mathbf{0}, W) \\ \\ \boldsymbol{\phi_0} | \mathcal{D}_0 \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma) \\ w_i \sim \text{LogNormal}(a, b) \quad i = 1, \dots, p \\ \sigma^2 \sim \text{inv-gamma}(c, d) \end{cases}$$

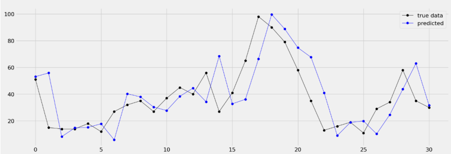


Figure: Superposition of a semilocal linear trend model and a week seasonality

Let $X_1, \ldots, X_n$ be n independent time series. The value of the i th variable X_i at each time t is assumed known. For $t = 1, \ldots,$ let the regression vector F_t be given by

$$F_t = \left[X_{t1}, \ldots, X_{tn}\right]$$

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

In case of autoregression of order p:

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

where $\nu_1, \ldots, \nu_p \sim \mathcal{N}_p(0, I_p)$ truncated on (-1,1).

Overview of the proposed DLM

- The series to forecast presents big changes from one time step to the next one. Y_t is not enought smooth to be described by a linear trend component. **The model does not include any explicit trend dynamic**
- **Seasonal component**: 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**: The following regressors are supplied to the model
  - temperature
  - humidity
  - wind
  - wind direction
  - rainfall intensity
  - global solar radiation

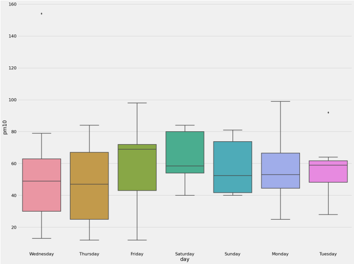
We set a day of the week effect. We have 7 latent variables each one describing the effect of the day over the mean level of PM10. Let P a 7×7 permutation matrix

$$\begin{cases} Y_t = \boldsymbol{E_7^T} \boldsymbol{\phi_t} + v_t \quad v_t \sim \mathcal{N}(0, \sigma^2) & \text{Observation equation} \\ \boldsymbol{\phi_t} = P\boldsymbol{\phi_{t-1}} + \boldsymbol{\omega_t} \quad \boldsymbol{\omega_t} \sim \mathcal{N}(\mathbf{0}, W) & \text{System equation} \\ \\ \boldsymbol{\phi_0} | \mathcal{D}_0, \sigma^2 \sim \mathcal{N}(\boldsymbol{m_0}, \sigma^2 C_0) & \text{Initial information} \\ w_i \sim \text{LogNormal}(a, b) \quad i = 1, \ldots, p \\ \sigma^2 \sim \text{inv-gamma}(c, d) \end{cases}$$

How set initial information $\boldsymbol{m_0}$ and C_0?

- We take PM10 values in the period from 01-01-2020 to 01-03-2020 and compute the distribution of pollutants grouped by days
- Set

$$\boldsymbol{m_0} = \begin{bmatrix} \bar{Y}_{mon} \\ \bar{Y}_{Thu} \\ \vdots \\ \bar{Y}_{sun} \end{bmatrix} \qquad C_0 = Diag \begin{bmatrix} \bar{\sigma}^2{}_{mon} & \ldots & \bar{\sigma}^2{}_{sun} \end{bmatrix}$$

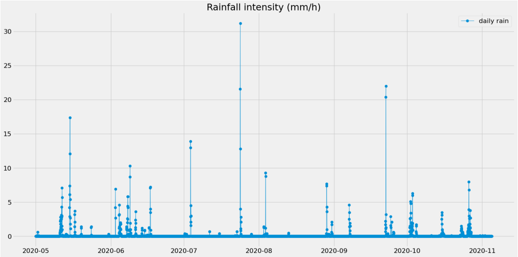


Rainfall intensity

Instead of considering the mean precipitation levels during the day is better to consider another measure which takes into account the intensity of the metereological event:

$$\begin{aligned} \text{rain}_{min} &= \min\left(\text{rain}_t, \text{rain}_{t-1}\right) \\ \text{rain}_{max} &= \max\left(\text{rain}_t, \text{rain}_{t-1}\right) \end{aligned}$$

$$\text{Rainfall Intensity} = \text{rain}_{min} + \frac{1}{2}\left(\text{rain}_{max} - \text{rain}_{min}\right)$$

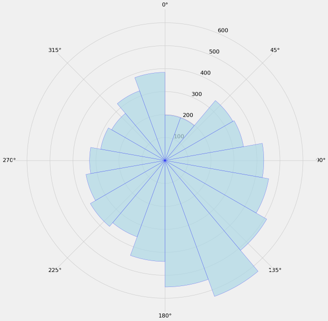


The intensity of the rain determines the reduction of PM: stronger rainfall intensity causes a reduction of PM, light rain has not a strong impact on the levels of PM

Not only the wind intensity but also its direction determine a change in the level of pollutants: **winds going from north to south clean the air**

We introduce the following regressors:

- wind speed (measured in m/s)
- sine and cosine of the wind angle with respect to north direction
- 4 dummy variables, one for each main direction of the wind: NE, SE, SW, NW.
  To incorporate our prior knowledge we set strong negative effects on the dummies SE and SW in the prior distribution of the latent state



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

Overall proposed model

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

Implementation

The model has been implemented using python and STS module from Tensorflow probability API

- Model has been defined using the DLM components blocks presented before, once proper data and prior distributions have been defined for each block.
- The overall model is given by superposition (i.e. linear combination) of the single building blocks.
- An MCMC sample is obtained from the posterior distribution of model parameters.
- Given the sample obtained at previous step, for each data point received by the model a forward kalman update pass is performed, producing the one step forecast distribution and the posterior distribution for the latent state vector.
- The obtained posterior is used as prior for the next iteration (i.e. when a new data point is received).

Posterior Inferences and criticism

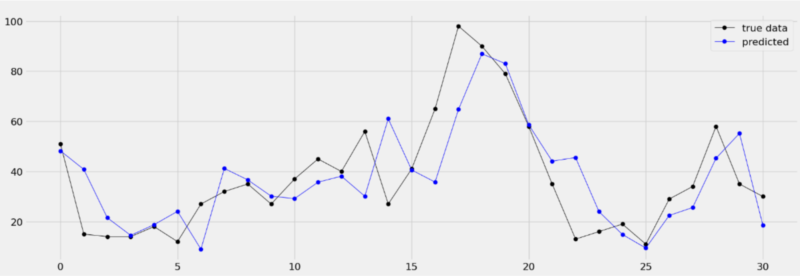


Figure: one step prediction

- As is clear the model is not able to anticipate an increment of the level of pollutants, but **it waits to see the increment before correcting its prediction**.
- The only weather conditions are not enough to predict an increment of PM: domain literature always correlate meteorological factors to a decrease of the level of pollutant.

Bibliography

- Mike West, Prado Raquel, "Time Series Modeling, Computation,and Inference"
- J D Bermúdez, J V Segura, E Vercher, "Bayesian forecasting with the Holt–Winters model", Journal of the Operational Research Society,2010
  https://doi.org/10.1057/jors.2008.152
- Mike West, Jeff Harrison, "Bayesian Forecasting and Dynamic Models"
- https://www.tensorflow.org/probability/api_docs/python/tfp/sts
- Zhen Liu, Luming Shen, Chengyu Yan, Jianshuang Du, Yang Li, Hui Zhao ,"Analysis of the Influence of Precipitation and Wind on PM2.5 and PM10 in the Atmosphere", 2020
  https://doi.org/10.1155/2020/5039613