Forecasting particulate matter concentrations in an indoor environment

R. Ouaret[1,2], A. Ionescu[1], V. Petrehus[3], Y. Candau[1] and O. Ramalho[2]

[1] Université Paris-Est, CERTES (EA 3481), UPEC, F-94010, Créteil, France
[2] Université Paris-Est, CSTB, Champs-sur-Marne, F-77447, France
[3]Department of Mathematics and Informatics, Technical University of Civil Engineering, Bucharest, Romania

Keywords: open-plan office, particulate matter, seasonal trend decomposition, autoregressive model
Presenting author email: rachid.ouaret@u-pec.fr, r.ouaret@gmail.com

This study focuses on the statistical forecasting of particulate matter concentrations in a real indoor environment.

The monitoring campaign was performed in an open-plan office, during 6 months (from February to June 2012). The indoor air particulate matter was sampled hourly using an optical particle counter (Dust Monitor 1.108, Grimm), which provided the number of particles per liter, for fifteen size bins within a range of 0.3-20 µm. Due to some measurement problems and given that time series corresponding to some size bins were strongly correlated, only five size bins were selected in this study: 0.3-0.4 µm, 0.8-1 µm, 1.6-2 µm, 4-5 µm, 7.5-10 µm.

The five time series corresponding to the selected size bins were not second order stationary. Consequently, data was transformed by ln(1+N), with N number of particles for each size bin. A hybrid STL-ARIMA model was applied to the transformed data for forecasting purposes.

The STL-ARIMA model is based on a 2-step procedure. Firstly, each time series is decomposed into trend, seasonal and remainder using a Seasonal-Trend Decomposition procedure based on Loess smoothing (Cleveland *et al*., 1990)-STL. After filtering the seasonal component, an autoregressive ARIMA model is fitted to the (trend+remainder) component and used for forecasting. Assuming that the seasonal component doesn't change (or changes extremely slowly), its forecast is obtained by taking the highest frequency of the time series spectrum. The final forecast is obtained by adding to the ARIMA forecast the seasonal component forecast.

Data was splitted in two subsets, for training and test. The forecasting performance of the model was assessed for a 72-hour horizon and compared to the test set. The resulting forecasts of the log-transformed data with the prediction intervals are shown in Figure 1.

In order to calculate the forecast accuracy, the results were transformed in original units by the exponential function and then compared to the raw test sample. The Mean Absolute Percentage Error (MAPE) was calculated for the whole forecast horizon. The MAPE obtained for a 72-hour horizon for $PM_{0.3\text{-}0.4}$ was 32.5% and for $PM_{7.5\text{-}10}$, 36.4%.

Graphical representations and MAPE indices show that the STL-ARIMA model performs well in forecasting throughout the entire test period, reproducing the overall mean fluctuation and capturing the most important feature of the time variability, with a MAPE index about 35%.

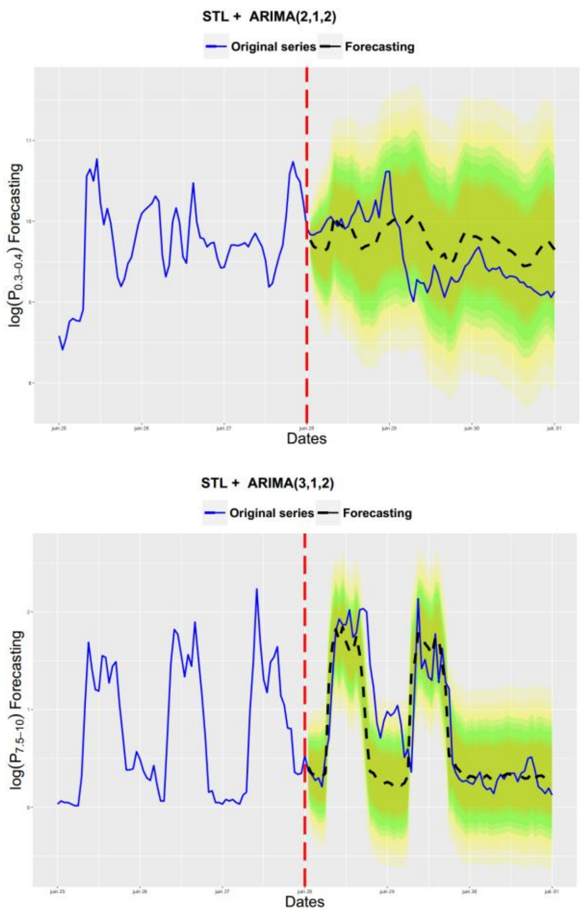

Figure 1. Forecasting PM time series using STL-ARIMA model. Example for $PM_{0.3\text{-}0.4}$ and $PM_{7.5\text{-}10}$ concentrations in number by liter, represented on a logarithmic scale. Forecasting (dashed black curve) follows after the dashed red vertical line.

This work was supported by ADEME and MEDDE within the frame of the PRIMEQUAL-PREDIT program, TRIBU project, under grant 12-MRES-PRIMEQUAL-4-CVS-09).

Box, G.; Jenkins, G. (1970). *Time Series Analysis: Forecasting and Control.* San Francisco: Holden-Day.

Cleveland, R.B., Cleveland, W.S., McRae J.E. and Terpenning, I. (1990) *Journal of Official Statistics* **6**, 3-73.