Early warning system for PM_{10} in Sofia

H. Branzov, V. Nikolov, D. Syrakov, N. Neykov, P. Neytchev, H. Galabova, N. Neykova, M. Prodanova

National Institute of Meteorology and Hydrology

INNOAIR webinar, Sofia, Bulgaria, 24 September 2021

Outline

Early warning system for forecasting the potential of particulate matter (PM_{10}) concentrations at 5 different monitoring stations Druzhba, Hipodruma, Nadezhda, Pavlovo and Kopitoto in Sofia area

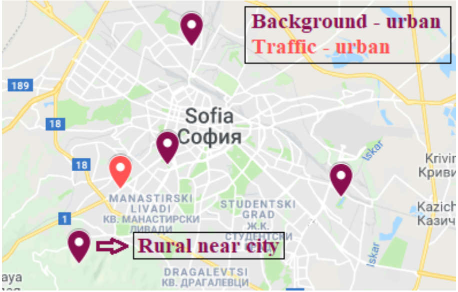
The Early Warning System for PM_{10} in Sofia

The early warning system for forecasting the next 72 hours of PM_{10} concentrations arises from the collaboration between Sofia Municipality and the National Institute of Meteorology and Hydrology.

NIMH research team:

prof. H. Branzov, PhD - project leader;
prof. D. Syrakov, Dr.Sci., V. Nikolov,
prof. N. Neykov, Dr.Sci., assoc.prof. P. Neytchev, PhD,
M. Prodanova, H. Galabova, N. Neykova

PM_{10} data



- The Early Warning System is based on PM_{10} data measured at 5 stations of the ExEA located at Druzhba, Pavlovo, Hipodruma, Kopitoto and Nadezhda.
- The data series consists of 67728 hourly averaged PM_{10} $\mu g/m^3$ concentrations for the period 01.01.2014-20.09.2021 for each of these 5 stations.

PM_{10} distributions at Pavlovo in different time scales

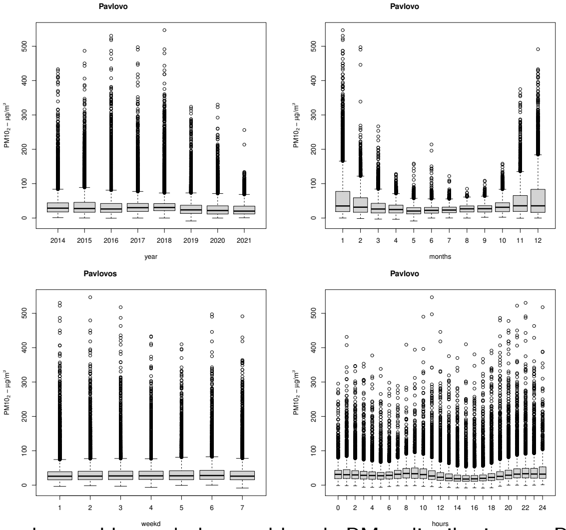


Figure: Annual, monthly, week days and hourly PM_{10} distributions at Pavlovo. PM_{10} are higher during the winter season and exhibit strong seasonal and hourly variation.

Pavlovo: PM_{10} distribution

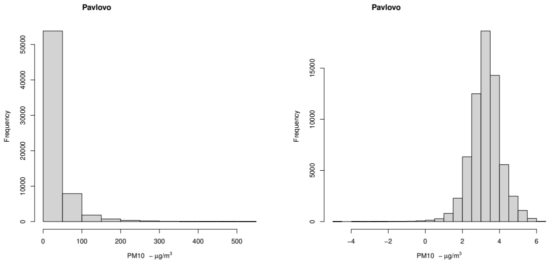


Figure: PM_{10} (left) and $\log(PM_{10})$ (right) distributions at Pavlovo for the period 01.01.2014-20.09.2021 yrs.

Pavlovo: missing PM_{10} data imputation

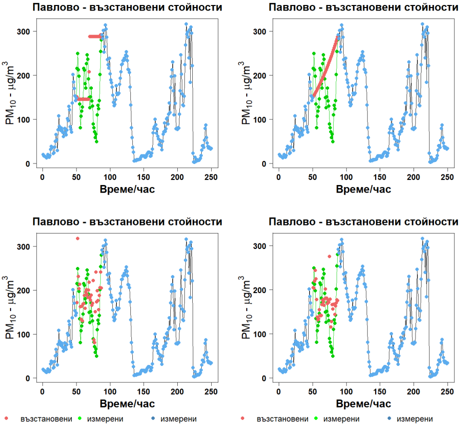


Figure: Missing data imputation based on: replacement with last-first (upper left); linear interpolation (upper right); gamma and log-normal time series autoregression models and Hipodruma PM_{10} data as predictors (lower panel plots). Observed PM_{10} data are in blue and green, recovered data are in red.

The Early Warning System Models

- The System is based on a combination of two types of models - numerical weather prediction model and stochastic model.
- The Weather Research and Forecasting (WRF) Model is a mesoscale numerical weather prediction system designed to serve both atmospheric research and operational forecasting needs.
- The WRF output delivers 72 hours forecasts which characterize the atmospheric conditions in Sofia area for the next 3 days from ground surface up to 5500 m height.
- The WRF output derivatives, the lags of these derivatives and the PM_{10} lags from 72 hours to 144 hours serve as input predictors into the stochastic model.

The WRF model derivatives

t2	[C]	surface temperature at 2m
td2	[C]	dewpoint temperature
ts	[C]	soil temperature
ts/t2	[C]	surface temperature ratio
wd10	[deg]	wind direction at 10m
ws10	[m/s]	wind speed at 10m
Q2	[g/kg]	specific humidity
U*	[m/s]	roughness velocity
Moi	[1/m]	inverse Monin-Obukhov length
PBLH	[m]	Planetary Boundary Layer (PBL) height
PBLT	[m^2/s^2]	PBL kinetic turbulent energy
PBLL	[m]	PBL horizontal scale
CL		low cloudiness level (ceiling level)
Pasq		Pasquil stability parameter
W	[m/s]	mean vertical velocity in the lowest 50m layer
G20	[C/100m]	temperature gradient of layer 0-20m
G50	[C/100m]	temperature gradient of layer 20-50m
G120	[C/100m]	temperature gradient of layer 50-120m
G200	[C/100m]	temperature gradient of layer 120-200m
G300	[C/100m]	temperature gradient of layer 200-300m
G500	[C/100m]	temperature gradient of layer 300-500m
G800	[C/100m]	temperature gradient of layer 500-800m
RAINC	[mm]	convective precipitation
RAINNC	[mm]	nonconvective precipitation

These derivatives are used as input predictors in the stochastic model.

Stochastic PM_{10} Forecasting Model

- Let $y_t = \log(\mathrm{PM10}_t)$ and $x_t = (y_{t-r}, \ldots, y_{t-d}, x_{t1}, \ldots, x_{tm})^T$ is the vector of $\mathrm{PM10}_t$ lags and WRF derivatives for $t = 1, \ldots, T$
- The distribution of y_t is highly complex; Thus it is approximated by mixture of distributions, Markov switching time series regression models

$$\varphi(y_t; x_t, \Psi) = \sum_{j=1}^{J} \pi_j(x_t, \beta_j)\psi_j(y_t; x_t, \theta_j)$$

$$\varphi(y_t; x_t, \Psi) = \begin{cases} \psi_1(y_t; x_t, \theta_1) & \textit{with probability} \quad \pi_1(x_t, \beta_1) \\ \psi_2(y_t; x_t, \theta_2) & \textit{with probability} \quad \pi_2(x_t, \beta_2) \\ \ldots & \ldots \\ \psi_J(y_t; x_t, \theta_J) & \textit{with probability} \quad \pi_J(x_t, \beta_J) \end{cases}$$

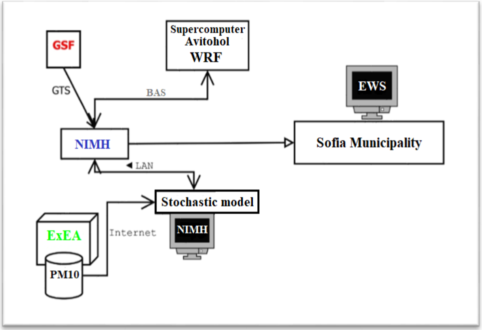
- $\Psi = (\beta_1, \ldots, \beta_k, \theta_1, \ldots, \theta_k)^T$ - vector of unknown coefficients (parameters);
- $\psi_j(y_t; x_t, \theta_j)$ - jth autoregressive time series model with Gaussian error;
- $\pi_j(x_t, \beta_j) = \frac{\exp(x_t^T \beta_j)}{\sum_{j=1}^{J} \exp(x_t^T \beta_j)}$ - jth multinomial probability; $\sum_{j=1}^{J} \pi_j(x_t, \beta_j) = 1$

- The likelihood function (the joint distribution of y_t) is defined as

$$L(\Psi) = \prod_{t=r+1}^{T} \varphi(y_t; x_t, \Psi) = \prod_{t=r+1}^{T} \sum_{j=1}^{J} \pi_j(x_t, \beta_j)\psi_j(y_t; x_t, \theta_j)$$

- The parameter Ψ is obtained as maximum of $L(\Psi)$ using LASSO regularization.

The Early Warning System Architecture



1. The PM_{10} data from ExEA are delivered at the NIMH server.
2. The WRF 72 hours forecast derivatives are delivered and updated twice daily at 12:00 am and 12:00 pm and serve as input to the stochastic models for each of the 5 locations;
3. The next 72 hours forecasts of PM_{10} for each of the 5 locations are delivered at the Sofia municipality twice daily at 12:00 am and 12:00 pm.
4. The output WRF model derivatives, the PM_{10} measured concentrations and predicted values are stored for further analyses.

Pavlovo: observed versus forecast PM10

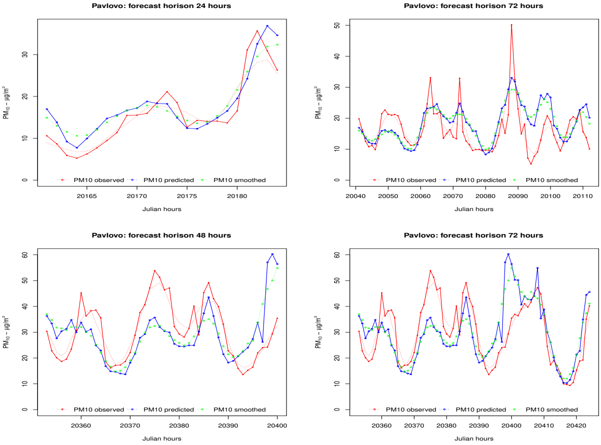


Figure: Pavlovo observed (red) data from tested sample and forecast (blue) PM_{10} for different horizons. Smoothed line of the forecast are colored in green.

Airthings, ExEA Hipodruma and ExEA Pavlovo data quality

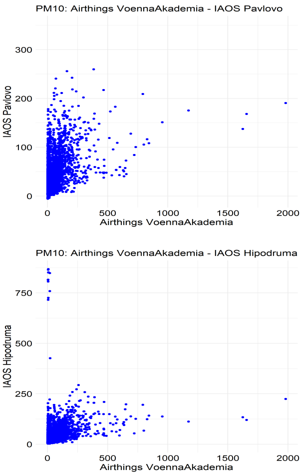
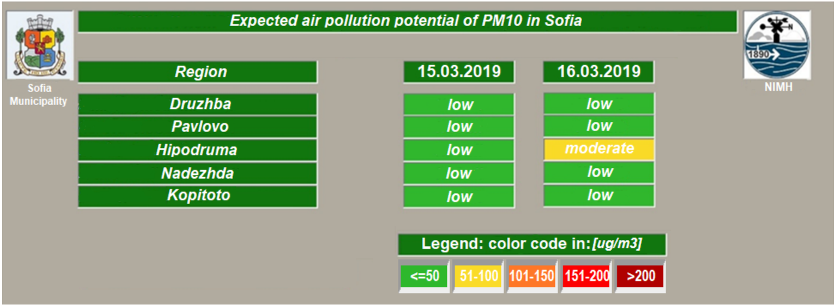
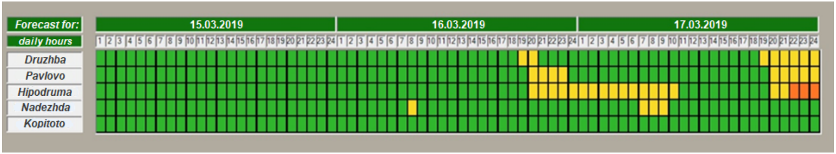


Figure: Now the question is to use or not to use the Airthings data about better model forecast at the ExEA stations.

The EWS output control panel



The upper panel is public. For each one of the 5 locations in Sofia, the next 2 days forecast is given in categories - low, moderate, high and very high.



The lower panel is available for municipality experts. The numerical forecast for the next 72 hours is given in the same categories.

Pavlovo 1st and 2nd days PM_{10} Forecast Assessment

Table: Pavlovo PM_{10} model: 1st day forecast misclassification table

	Observed			
		00-49	50-$\gg$	error
Forecast	00-49	84.3 %	3.1 %	3.60 %
	50-$\gg$	1.5 %	11.1 %	11.92 %
total				4.64 %

Table: Pavlovo PM_{10} model: 2nd day forecast misclassification table

	Observed			
		00-49	50-$\gg$	error
Forecast	00-49	83.8 %	4.7 %	5.61 %
	50-$\gg$	3.2 %	8.3 %	38.55 %
total				7.9 %

Each table shows the observed versus forecasted frequencies in percentage.

Druzhba and Nadezhda PM_{10} Forecast Assessment

Table: Druzhba PM_{10} model: 1st day forecast misclassification table

		Observed		
		00-49	50-≫	error
Forecast	00-49	77.9 %	4.4 %	5.36 %
	50-≫	6.2 %	11.4 %	35.28 %
total				10.64 %

Table: Nadezhda PM_{10} model: 1st day forecast misclassification table

		Observed		
		00-49	50-≫	error
Forecast	00-49	90.1 %	4.9 %	5.19 %
	50-≫	1.1 %	3.9 %	21.11 %
total				5.98 %

Each table shows the observed versus forecasted frequencies in percentage.

Thank you for your attention!