21st International Conference on
Harmonisation within Atmospheric Dispersion Modelling for Regulatory Purposes
27-30 September 2022, Aveiro, Portugal

$PM_{2.5}$ PREDICTIONS FOR URBAN MONITORING SITES IN BUDAPEST USING STATISTICAL FUSION OF CAMS AIR QUALITY MODELS

Adrienn Varga-Balogh [1], Ádám Leelőssy [1] and Róbert Mészáros [1]

[1] Eötvös Loránd University, Institute of Geography and Earth Sciences, Department of Meteorology, Pázmány Péter sétány 1/A, 1117 Budapest, Hungary

Abstract: For urban monitoring sites, air quality forecasting with atmospheric chemistry transport models have limitations due to the complexity of air pollution sources in cities. In the winter, communal heating is the main source of air pollutants in Hungary. Winter stagnation events with low wind speed can increase concentrations and cause the deviation of air quality guidelines within the city.

For Budapest, the capital of Hungary, a time-dependent downscaling method was used to predict the daily mean of $PM_{2.5}$ concentration for the heating seasons in 2018–2021. Nine individual models of the Copernicus Atmosphere Monitoring Service (CAMS) were compared to six urban measurement points' $PM_{2.5}$ data in Budapest. Downscaled predictions were produced by the linear combination of the CAMS models using spatially constant and time-dependent weights fit on the previous 10-days long period. The last 10-day bias was also corrected in the models.

Downscaling generally reduced the root mean square error (RMSE) for the heating season, especially for the smog episodes, as the method reduced the high underestimation of $PM_{2.5}$ in contaminated periods. Predictions from the model fusion were more efficient in smog episodes and had similar overall efficiency to the bias-corrected ensemble. The fusion of the CAMS models leads to a more accurate forecast of wintertime $PM_{2.5}$ peaks in urban monitoring sites of Budapest than using any of the individual models.

Key words: *PM2.5; CAMS; Budapest; air quality; data fusion*

INTRODUCTION

Most cities have been facing polluted air conditions. In the case of Budapest, the capital city of Hungary, especially winter stagnation events can cause heavily polluted episodes due to residential heating.

The $PM_{2.5}$ forecasts of the Copernicus Atmosphere Monitoring Service (CAMS) models were used. CAMS includes the air quality forecast of several models developed independently: CHIMERE, EMEP, LOTOS-EUROS, MATCH, MOCAGE, SILAM, EURAD-IM, DEHM (from 2019), GEM-AQ (from 2019), and the multi-model ENSEMBLE forecast as the median of the individual values.
Air quality data is available from the Hungarian Air Quality Network, which has six stations measuring $PM_{2.5}$ in Budapest and providing hourly $PM_{2.5}$ measurements.

METHODS

A time-dependent linear combination for CAMS models were presented by Sofiev et al. (2017). The data of $PM_{2.5}$ monitoring sites of Budapest were used to fit the linear combination on a 10-day training period. The forecasted concentrations by the fusion model $\left(c_{fusion}(x,t)\right)$ at location x and time t:

$$c_{fusion}(x,t) = w_{0,t} + \sum_{i=1}^{M} w_{i,t} c_{i,x,t} \,, \tag{1}$$

where the prediction $(c_{i,x,t})$ was the uncorrected/bias-corrected $PM_{2.5}$ forecast of the i^{th} model for a measurement station from the nearest gridpoint. M is the number of models used, seven in the winters of 2018–2019, nine afterwards (in the winters of 2019–2020, 2020–2021 and 2021–2022). In June 2022, two extra models were added. The weights (w_t) were fit for each day on the previous 10-day training period by minimizing the J_t cost function for all available sites.

$$J_t = \sqrt{\frac{1}{X \cdot T} \sum_{x=1}^{X} \sum_{\tau=T-d}^{t-d} \left(c_{fusion}(x,t) - c_{obs}(x,t)\right)^2} + R. \quad (2)$$

The weights are variant in time for each model, and spatially consistent. The regularization term R minimizes the differences in weights between the consecutive timesteps and among models:

$$R = \alpha \sum_{i=1}^{M} \left(w(i,t) - \frac{1}{M}\right)^2 + \beta \sum_{i=1}^{M} \left(w(i,t) - w(i,t-1)\right)^2. \quad (3)$$

The model bias was calculated on a rolling 10-day period, then added to the next days' forecast. This way, a bias-corrected dataset was created. For the heating seasons, the optimized fusion forecasts were produced from both the CAMS model forecasts and the bias-corrected model-forecasts.

The performance of the models was measured in terms of mean absolute error (bias), root mean square error (RMSE), Pearson correlation (r) and accuracy of EAQI (European Air Quality Index) categories.

RESULTS

For the heating seasons (defined between 15 October – 15 April), in the winters of 2018, 2019, 2020 and 2021 the $PM_{2.5}$ levels and forecasts were investigated. The analysis of the 2018–2019 period was described in detail in Varga-Balogh et. al. (2020).

In Fig. 1., the time series of the forecasts is presented for the last three winter periods. The CAMS ENSEMBLE and fusion model is shown with the $PM_{2.5}$ values measured at Kőrakás park measurement station. Results from the bias-corrected models are added with dotted lines. In Hungary, stagnation events can occur often in the winters, when persistent anticyclonic conditions with low windspeeds and weak mixing in the lower troposphere lead to poor air quality situations. Each winter showed different weather conditions. In January 2020, a longer period appeared with high concentrations of $PM_{2.5}$, while the last winter was different: only short periods were above the WHO guideline (25 μg m^{-1}) concentration. A general overestimation was observable by the CAMS ENSEMBLE.

The errors of the bias-corrected CAMS ENSEMBLE is a consequence of the 10-day bias-correction period. When the uncorrected model had given relatively strong underestimation for the previous days, the correction shifted the forecast to overestimation especially in cases of longer periods with high $PM_{2.5}$ concentrations followed by a relatively clear period. (E.g.: the end of the heating season in 2020, end of December 2019.)

The errors of the fusion model can also be attributed to the 10-day training period. In winter stagnation episodes (anticyclonic conditions), high concentrations of $PM_{2.5}$ is followed by a cold front with high windspeed and rapid cleansing in air pollution. The rapid improvement of air pollution is not caught by the fusion model. (End of January 2020.)

The forecast for the heating season of 2021–2022 led to errors both in cases of ENSEMBLE and fusion models due to the high variability of air pollution.

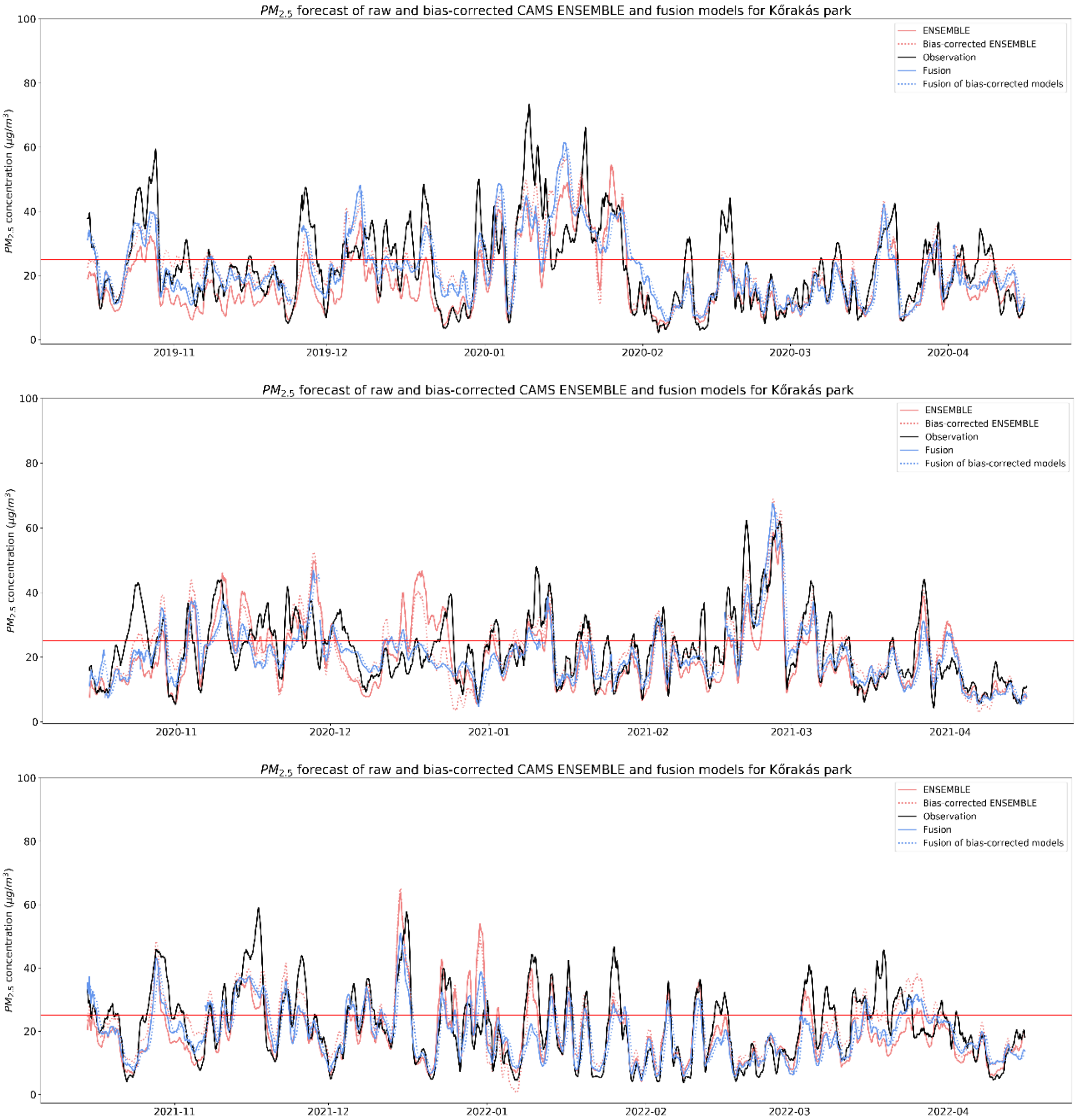

Figure 1. Time series of $PM_{2.5}$ measurements at Kőrakás park measurement station with black line (24h moving average), uncorrected (red) and bias-corrected (dotted red) CAMS ENSEMBLE, and the optimized fusion model of uncorrected (blue) and bias-corrected models (dotted blue) for the heating seasons of 2019–2020, 2020–2021 and 2021–2022

In terms of bias, RMSE, Pearson correlation (r), the CAMS ENSEMBLE performed nearly as the fusion models. The bias-correction improved the forecasts of the individual models as well as the performance of CAMS ENSEMBLE, however in some cases, the fusion of uncorrected models performed better than the fusion of bias-corrected models. The validation statistics for the heating season of 2021–2022 is presented in Figure 2.

The time series of the model-weight was also investigated (Figure 3). Although in the winter of 2018–2019, the high air pollution levels well-correlated with the high weights of the SILAM model (Varga-Balogh et. al., 2020), in the rapidly changing air quality of the 2021–2022 heating season (no stagnation event occurred), the SILAM model was overperformed by the others.

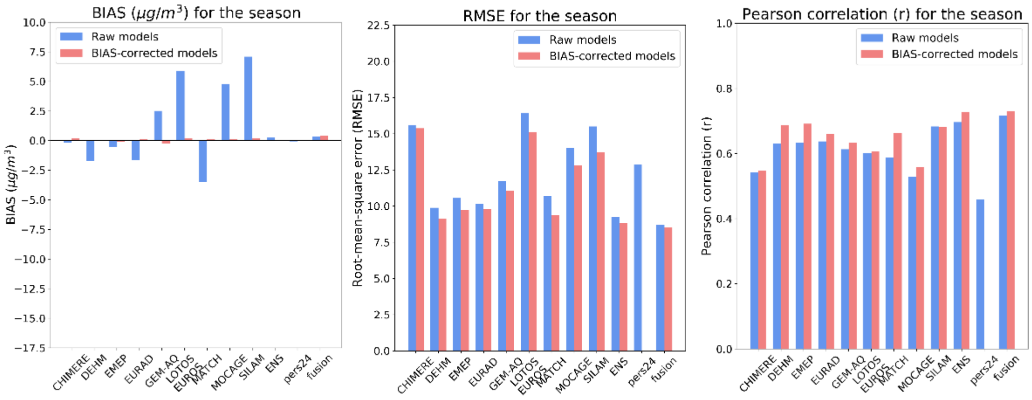

Figure 2. Validation for the heating season of 2021–2022. Bias, RMSE and Pearson correlation is presented for the uncorrected (blue) and bias-corrected (red columns) CAMS, CAMS ENSEMBLE, 24-h persistence, and fusion models.

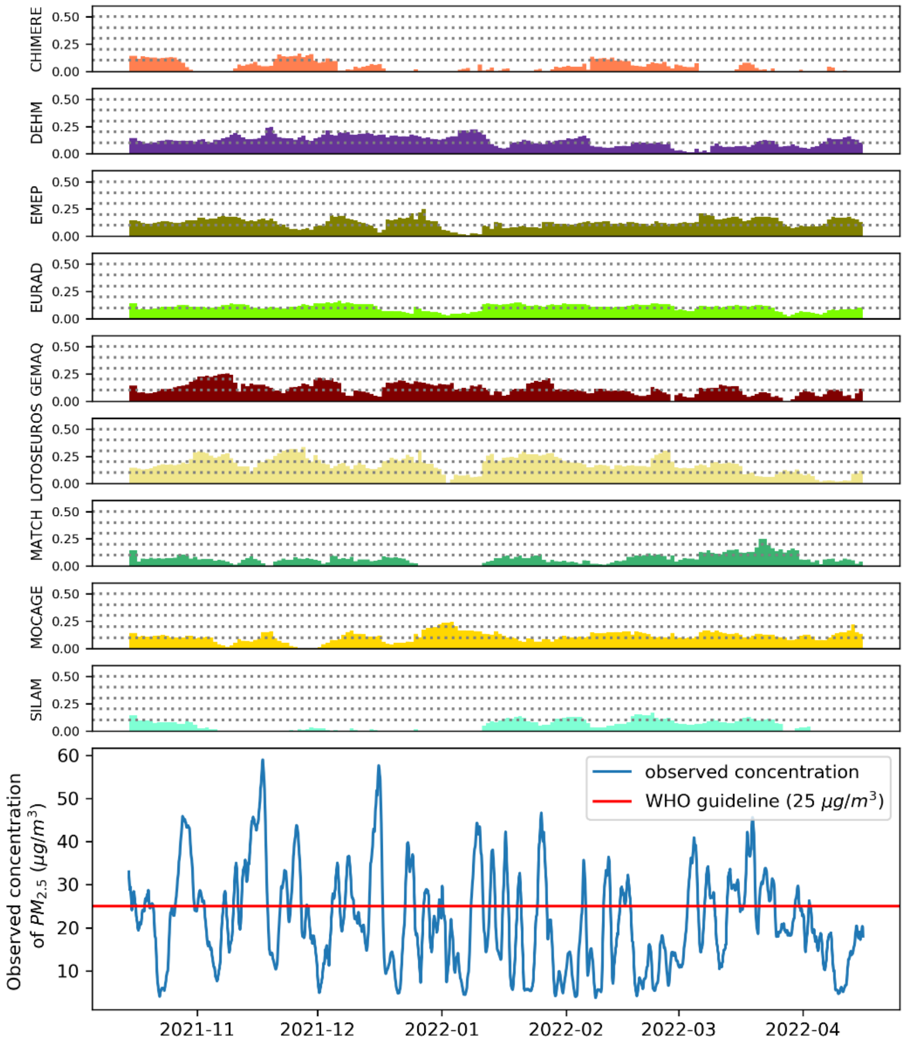

Figure 3. Time series of the applied model-weights and observed concentrations of $PM_{2.5}$ at Kőrakás park measurement station in the heating season of 2021–2022.

CONCLUSIONS

The CAMS $PM_{2.5}$ forecasts were compared to air quality measurement stations of Budapest for four heating seasons. Furthermore, bias was corrected on the previous 10-day data, and also a 10-day training period was applied for an optimization to produce a linear combination of the model forecasts. The CAMS ENSEMBLE was better than individual models in terms of bias, RMSE and Pearson correlation (r). Bias-corrected models mostly performed better than the uncorrected models in $PM_{2.5}$ forecasts, especially ENSEMBLE forecast improved for all winters with bias-correction.

Fusion model performs nearly as ENSEMBLE forecast, however in winter stagnation events, it performs better than CAMS and CAMS ENSEMBLE models.

The time series of the weights were examined to see which model performed best. Model weights were found to be strongly weather-dependent and variable among winters with many and no stagnation events.

FUNDING

This work was supported by the National Research, Development and Innovation Office of Hungary (grant No. 128805 and 128818). The research was funded by the National Multidisciplinary Laboratory for Climate Change, RRF-2.3.1-21-2021 project.

REFERENCES

Sofiev, M., Ritenberga, O., Albertini, R., Arteta, J., Belmonte, J., Bernstein, C.G., Bonini, M., Celenk, S., Damialis, A., Douros, J., et al., 2017: Multi-model ensemble simulations of olive pollen distribution in Europe in 2014: Current status and outlook. *Atmos. Chem. Phys.*, **17**, 12341–12360.

Varga-Balogh, A., Leelőssy, Á., Lagzi, I., Mészáros, R., 2020: Time-Dependent Downscaling of PM2.5 Predictions from CAMS Air Quality Models to Urban Monitoring Sites in Budapest. *Atmosphere*, **11**, 669.