Supplement of Atmos. Chem. Phys., 24, 1673–1689, 2024
https://doi.org/10.5194/acp-24-1673-2024-supplement
© Author(s) 2024. CC BY 4.0 License.

Atmospheric Chemistry and Physics
Open Access
EGU

Supplement of

Technical note: Accurate, reliable, and high-resolution air quality predictions by improving the Copernicus Atmosphere Monitoring Service using a novel statistical post-processing method

Angelo Riccio and Elena Chianese

Correspondence to: Angelo Riccio (angelo.riccio@uniparthenope.it)

The copyright of individual parts of the supplement might differ from the article licence.

S1 Detailed description of spatial and spatio-temporal predictors

Predictor	Description, source, and spatial and temporal resolution
Resident population	Resident population surveyed by the Italian Institute of Statistics. Source: ISTAT, https://www.istat.it. Available as vector data for each of the 366,863 population areas related to the national census of the year 2011. Data are remapped to each buffer based on cell block intersections
Imperviousness density	Soil sealing at the pixel level. Source: ISPRA download centre, https://www.isprambiente.gov.it/it/attivita/suolo-e-territorio/suolo/copertura-del-suolo/high-resolution-layer. Average over the three years 2017-2019, available as raster data at 10m spatial resolution. Data are remapped as the percentage of soil sealing within the buffer distance
Built-up density	Percentage of building and no-building class within the sealing arrangement. Source: Copernicus Land Monitoring Service, https://land.copernicus.eu/en/products/high-resolution-layer-impervious-built-up. Average over the three years 2017-2019, available as raster data at 10m spatial resolution. Data are remapped as the percentage of soil sealing within the buffer distance
Land cover	Corine Land Cover (CLC). Source: Copernicus Land Monitoring Service, https://land.copernicus.eu/en/products/high-resolution-layer-impervious-built-up. Average over the three years 2017-2019, available as raster data at 10m spatial resolution. Data are remapped as percentage covered by four classes (high urban development, low urban development/industrial/other artificial areas, agricultural areas, forest and semi-natural areas) within the buffer distance
Road density	Road segments. Source: Open Street Map database, https://download.geofabrik.de. Available as vector data as of 2022-11-12. The data are remapped as the sum of the length of all road segments within the buffer distance
Main roads	Road segments. Source: Open Street Map database, https://download.geofabrik.de. Available as vector data as of 2022-11-12. The data are remapped as the sum of the lengths of all road segments within the buffer distance. Data are remapped as the sum of the main road segments (highways and trunks) within the buffer distance
Precipitation	Total daily precipitation (m). Source: ECMWF ERA5 database, http://https://www.ecmwf.int/. Daily time resolution, available as raster data with spatial resolution $0.1^\circ \times 0.1^\circ$. Data are remapped to each buffer on the basis of cell block intersections
Wind speed and direction	Wind speed and direction (m). Source: ECMWF ERA5 database, http://https://www.ecmwf.int/. Hourly time resolution, available as raster data with $0.1^\circ \times 0.1^\circ$ spatial resolution, retrieved at 12 UTC for each day. Data are remapped to each buffer on the basis of cell block intersections
PBL	Planetary Boundary Layer height (m). Source: ECMWF ERA5 database, http://https://www.ecmwf.int/. Hourly time resolution, available as raster data with spatial resolution $0.1^\circ \times 0.1^\circ$, retrieved at 00 and 12 UTC for each day. Data are remapped to each buffer on the basis of cell block intersections

Table S1. Purely spatial and spatio-temporal predictors used during the post-processing stage.

S2 Skill score of ensemble models

To assess the value of the raw CAMS air quality forecasts, we here introduce the same approach described in Murphy (1988). To be precise, we measure the added value by means of the skill score SS, defined as:

$$SS = 1 - \frac{RMSE_f}{RMSE_r} \tag{1}$$

where $RMSE_f$ is the root mean square error of forecasts, and $RMSE_r$ is the root mean square error of the reference used as no-skill baseline. The observations of the previous day are used as a reference baseline; in this case, the skill score measures the accuracy of the CAMS forecast in predicting the next-day value compared to the hypothesis of persistence, i.e., that the concentration does not change from the previous day. Note that SS is positive when the forecast accuracy is greater than the reference baseline accuracy, and the added value becomes more and more important as the skill score approaches one. Furthermore, negative values of the skill score mean that, on average, the performance of the persistence hypothesis exceeds that of the raw CAMS forecast.

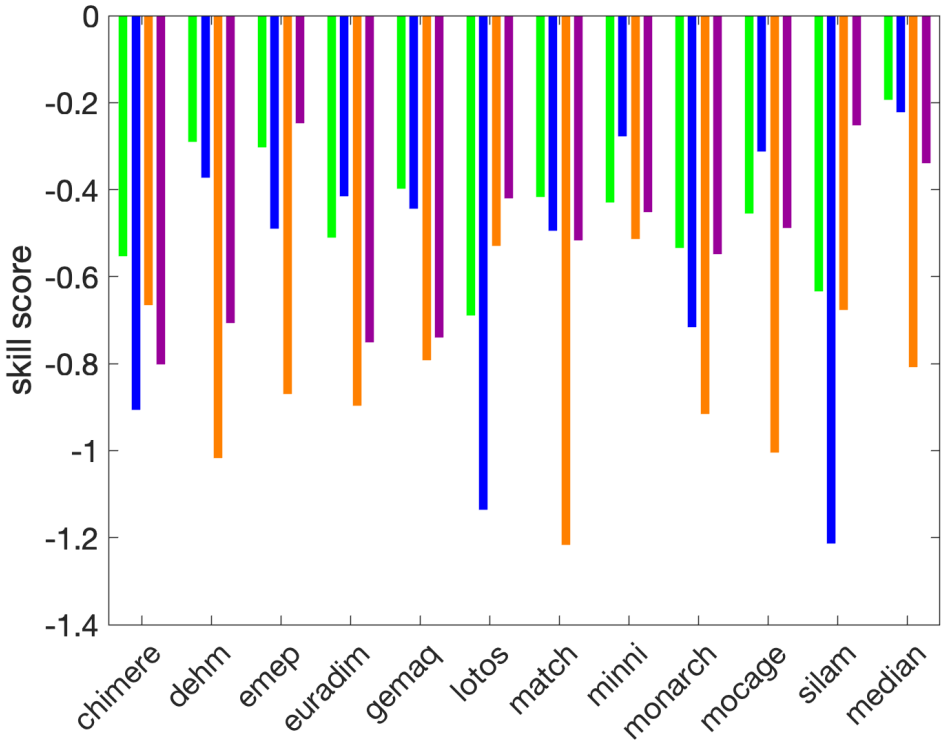

Figure S1. Skill score for the CAMS models. For each model the skill score is reported for the 24-hour look-ahead forecast during the year 2022 compared to the prediction based on the persistence of the previous day concentration for PM_{10} (green), $PM_{2.5}$ (blue), NO_2 (orange) and O_3 (purple bars)

The results are reported in Figure S1 where the CAMS results for the next day prediction against persistence are evaluated in terms of the skill score defined in (1). The persistence-based forecast (from the observed previous day values) performs consistently better than the model-derived values, so that the skill score is systematically negative for all models and pollutants. In particular, for the 1-hour NO_2 daily maximum, persistence-based prediction allows almost halving the error in the next day

prediction for almost all models, indicating the problems they have in predicting the concentration peaks on a small time scale, probably due to the low spatial resolution. Also note that in some cases the skill score is even lower than -1, meaning that the root mean square error of the raw CAMS predictions is more than double that obtained by exploiting the persistence assumption. The median model is only partially able to remedy this condition, usually showing an improvement over the prediction made by the individual models but with a still negative skill score. Even if we disentangle results among the different area type monitoring stations (data not shown), the same general conclusions about the skill of the raw CAMS predictions still continue to be valid.

S3 The ensemble model output statistical approach

A number of different pdfs have been proposed for f in (1): normal, truncated normal, logistic, gamma, and other distributions; the reader is referred to Wilks (2018) for a detailed discussion and comparison. Among all possible choices, after an exploratory phase, we found that an effective approach consists of selecting the gamma distribution, $\mathcal{G}(\alpha, \beta)$.

The gamma probability distribution function (pdf) is:

$$\mathcal{G}_{\alpha,\beta}(x) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} \quad \text{for } x > 0 \quad \alpha, \beta > 0 \tag{2}$$

characterised by the 'shape' parameter α and the 'inverse scale' parameter β. $\Gamma(\cdot)$ is the gamma function. The shape and inverse scale parameters are related to the predicted mean, $\mu = \alpha/\beta$ and the variance, $\sigma^2 = \alpha/\beta^2$, which in turn are related to the ensemble forecasts, $x_1, \ldots, x_m$, by the relations shown in (2).

Gneiting et al. (2007) proposed to evaluate the coefficients in (2) using a diagnostic approach based on minimisation of the *continuous-ranked probability score* (*crps*). The *crps* is the integral of the Brier scores at all possible threshold values t for the continuous predictand (Toth et al., 2003). In simple terms, the *crps* is defined as:

$$crps(F_{\alpha,\beta}, y) = \int_{-\infty}^{\infty} \left[F_{\alpha,\beta}(t) - H(t-y)\right]^2 dt \tag{3}$$

where $H(t-y)$ is the Heaviside function and takes the value 0 when $t < y$ and 1 otherwise, and $F_{\alpha,\beta}$ is the cumulative distribution function (cdf) corresponding to the pdf in (2). The closed form of *crps* for the gamma pdf has been obtained by Scheuerer and Möller (2015), making the minimisation procedure easy and fast. For an observation-forecast pair $(y, \mathbf{x})$, the *crps* closed form reads:

$$crps(y, \mathbf{x}) = y\left(2F_{\alpha,\beta}(y) - 1\right) - \frac{\alpha}{\beta}\left(2F_{\alpha+1,\beta}(y) - 1\right) - \frac{1}{\beta \mathcal{B}\left(\frac{1}{2}, \alpha\right)} \tag{4}$$

with y being the observation, and $\mathcal{B}$ the beta function. The forecast vector $\mathbf{x} = (x_1, \ldots, x_m)$ enters (4) through the shape and inverse scale parameters. Their expressions in terms of expected mean and variance read $\alpha = \mu^2/\sigma^2$ and $\beta = \mu/\sigma^2$. The mean and variance, in turn, depend on the coefficients used in (2). In case of a training set of observations and forecasts, the quantity to be minimised is

$$crps = \frac{1}{N} \sum_{i=1}^{N} crps(y_i, \mathbf{x}_i) \tag{5}$$

with i denoting the ith observation-forecast pair and N is the total number of pairs in the training set.

As also implemented in Gneiting et al. (2005), we strengthened the estimate of the coefficients in (2a), in order to avoid negative values, which can be caused by collinearities among the members of the ensemble.

To estimate the expected mean and variance from the *crps* minimisation in (5), we are left with the selection of the length of the training period. This aspect was already faced in Bertrand et al. (2022) and Gneiting et al. (2005), where different sliding

windows, 3 to 62 days, were considered. Of course, there is a trade-off in selecting a specific training length: a longer training period reduces the statistical variability in the estimation of coefficients; a shorter training period is able to adapt more rapidly to different conditions, for example, meteorological perturbations or changes in the emission scenarios. In our case, we found that even a very short training period is able to achieve good performance. The results shown in this work refer to a sliding training period of three days; that is, all predictions for the next day were obtained using air quality and meteorological data from the previous three days. For each day, this process was applied repeatedly, mimicking an operational forecasting system.

S4 The spatio-temporal statistical model

Results from the first stage are used to feed a second stage, in which we introduce additional spatio-temporal predictors. For a given calibrated ensemble prediction, $y(t, s_i)$, at time t and spatial location s_i, we assumed the model shown in (3). In this case, we model the residual as a first-order autoregressive model with spatially correlated innovations $\omega(t, s_i)$:

$$\xi(t, s_i) = a\xi(t-1, s_i) + \omega(t, s_i) \tag{6}$$

for $t = 2, \ldots, T$ and $|a| < 1$, $\xi(t, s_i)$ derives from the stationary distribution $\xi \sim \mathcal{N}(0, \sigma_\omega^2/(1-a^2))$, where $\mathcal{N}(\eta, \varepsilon^2)$ denotes the Gaussian distribution with mean η and variance ε^2. Moreover, $\omega(t, s_i)$ is assumed to be temporally independent and characterised by the spatio-temporal covariance function

$$\mathrm{Cov}\left(\omega(t, s_i), \omega(t', s_j)\right) = \begin{cases} 0 & \text{if } t \neq t' \\ \sigma^2 \mathcal{C}(h) & \text{if } t = t' \end{cases} \tag{7}$$

for $i \neq j$. The purely spatial correlation function $\mathcal{C}(h)$ depends on spatial location s_i and s_j only through the Euclidean distance $h = \| s_i - s_j \| \in \mathcal{R}$; therefore, the process is, according to the nomenclature used in the geostatistical literature, a second-order stationary and isotropic process (Cressie and Wikle, 2015). For the specification for the purely spatial covariance function, $\mathcal{C}(h)$, we follow the common choice of the Mat'ern function.

$$\mathcal{C}(h) = \frac{1}{\Gamma(\nu) 2^{\nu-1}} (kh)^\nu K_\nu(kh) \tag{8}$$

with K_ν denoting the modified Bessel function of the second kind and order $\nu = 1$. The parameter ν measures the degree of smoothness of the process; instead, $k > 0$ is a scaling parameter related to the range ρ, that is, the distance at which the spatial correlation becomes small. In particular, we use the empirically derived definition $\rho = \sqrt{8\nu/k}$, with ρ corresponding to the distance where the spatial correlation is close to 0.1 (Lindgren et al., 2011). This kind of model is well discussed and widely used in the literature on air quality, thanks to its flexibility in modelling the effect of relevant predictors, as well as space and time dependence (Blangiardo et al., 2013; Cameletti et al., 2013; Fioravanti et al., 2021; Konstantinoudis et al., 2022).

For the second stage, a three-day training period was also chosen and the estimation process was repeated for each day.

S5 Sensitivity, specificity, threat score and reliability for the prediction dataset

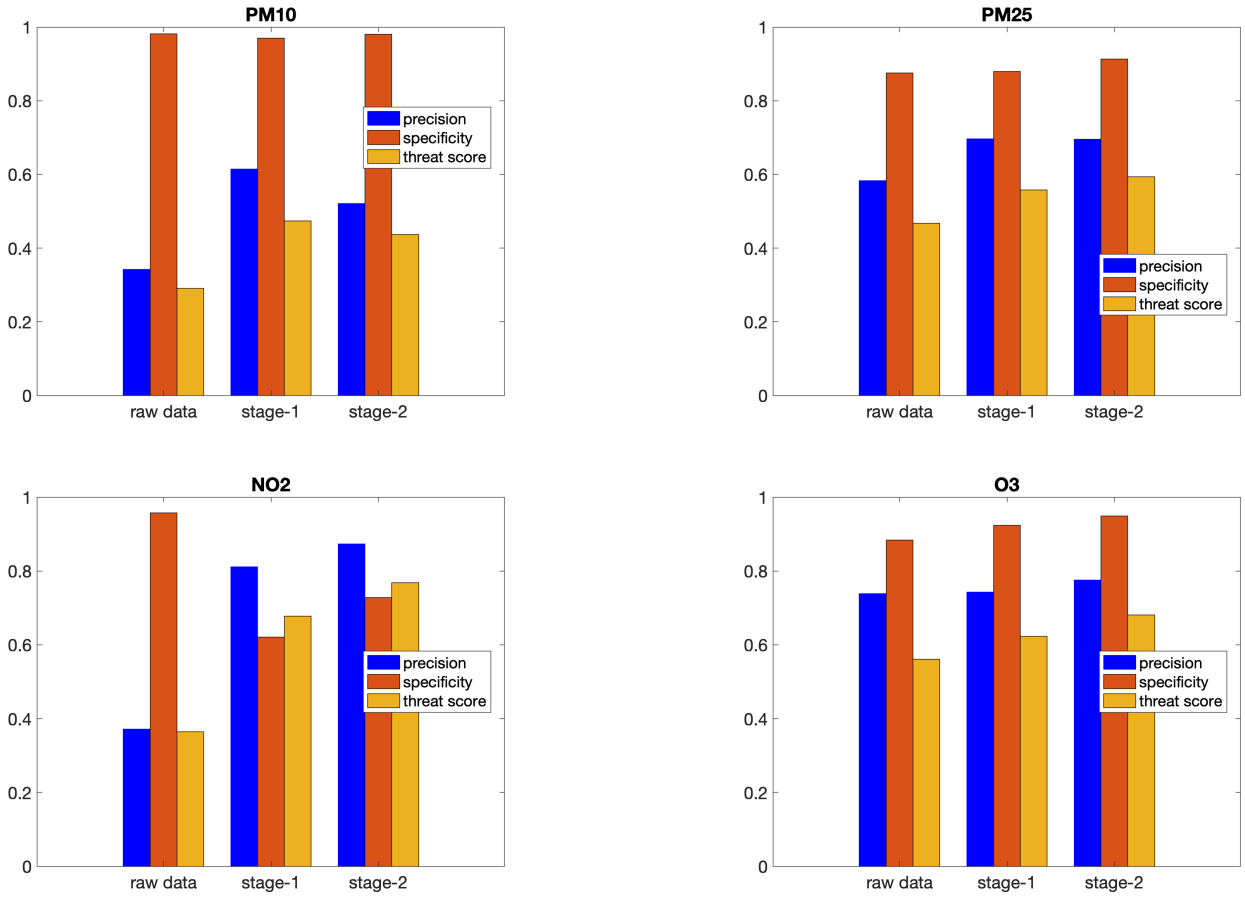

Figure S2. Scores (sensitivity, specificity and threat score) for the prediction dataset for PM_{10} (upper-left panel), $PM_{2.5}$ (upper-right panel), NO_2 (lower-left panel) and O_3 (lower-right panel). The blue bars correspond to the raw CAMS results, whereas the results after the application of the first and second stage are reported as orange and yellow bars, respectively. The number of exceedances (both for observations and predictions) is defined according to the new WHO guidelines: 45 μg/m^3 for daily PM_{10}, 15 μg/m^3 for daily $PM_{2.5}$, 100 μg/m^3 for the maximum 8-hour daily value for O_3, and 25 μg/m^3 for daily NO_2.

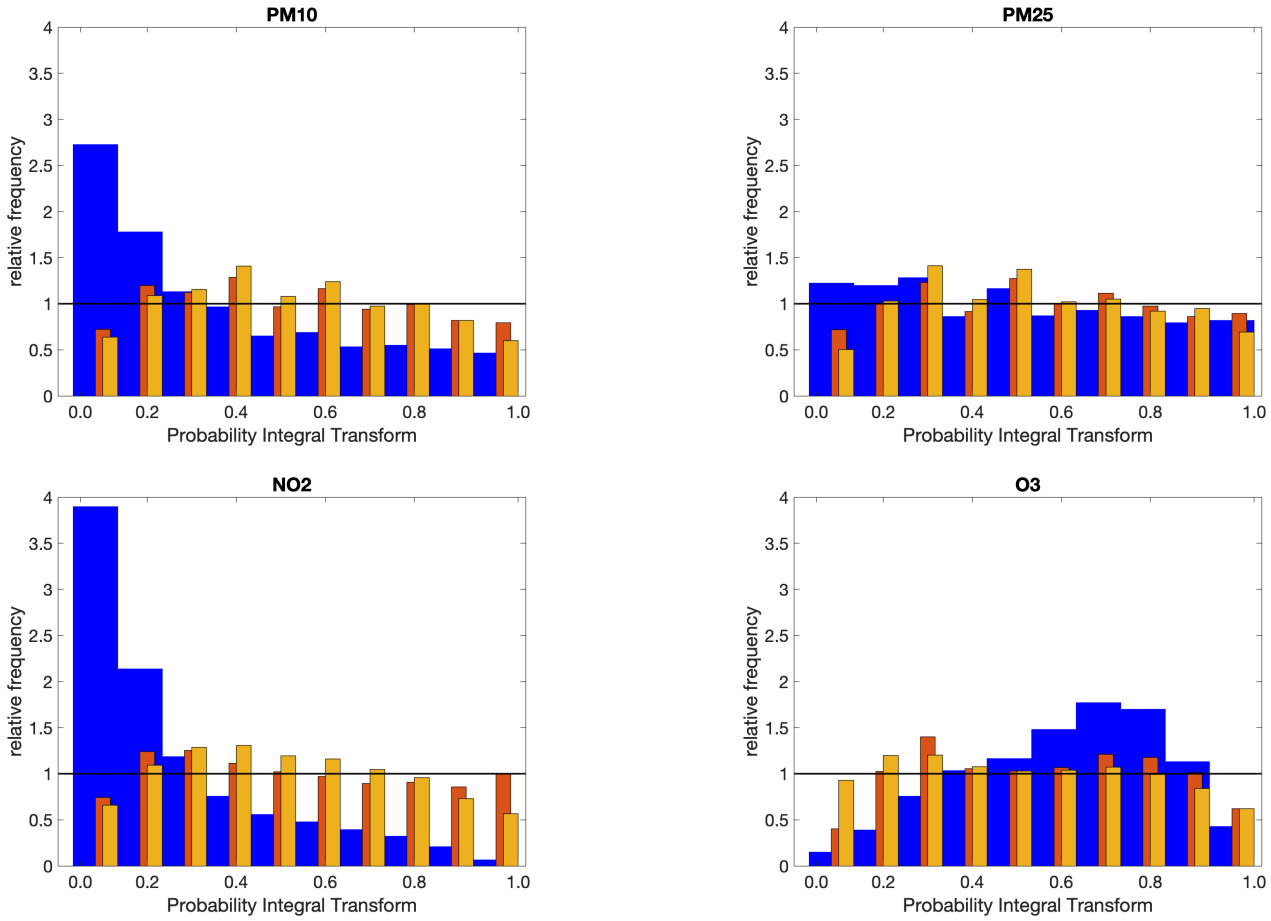

Figure S3. PIT for PM_{10} (upper-left panel), $PM_{2.5}$ (upper-right panel), NO_2 (lower-left panel) and O_3 (lower-right panel) for the prediction dataset. The blue bars correspond to the raw CAMS results, while the results after the application of the first and second stage to the validation data set are reported as orange and yellow bars, respectively. The orange and yellow bars have been slightly shifted and resized in width to not completely overlap the blue bars. The black horizontal lines have been drawn for reference: for a perfect reliable ensemble, the PIT should be flat, with a relative frequency equal to 1..

S6 Results from the raw CAMS data

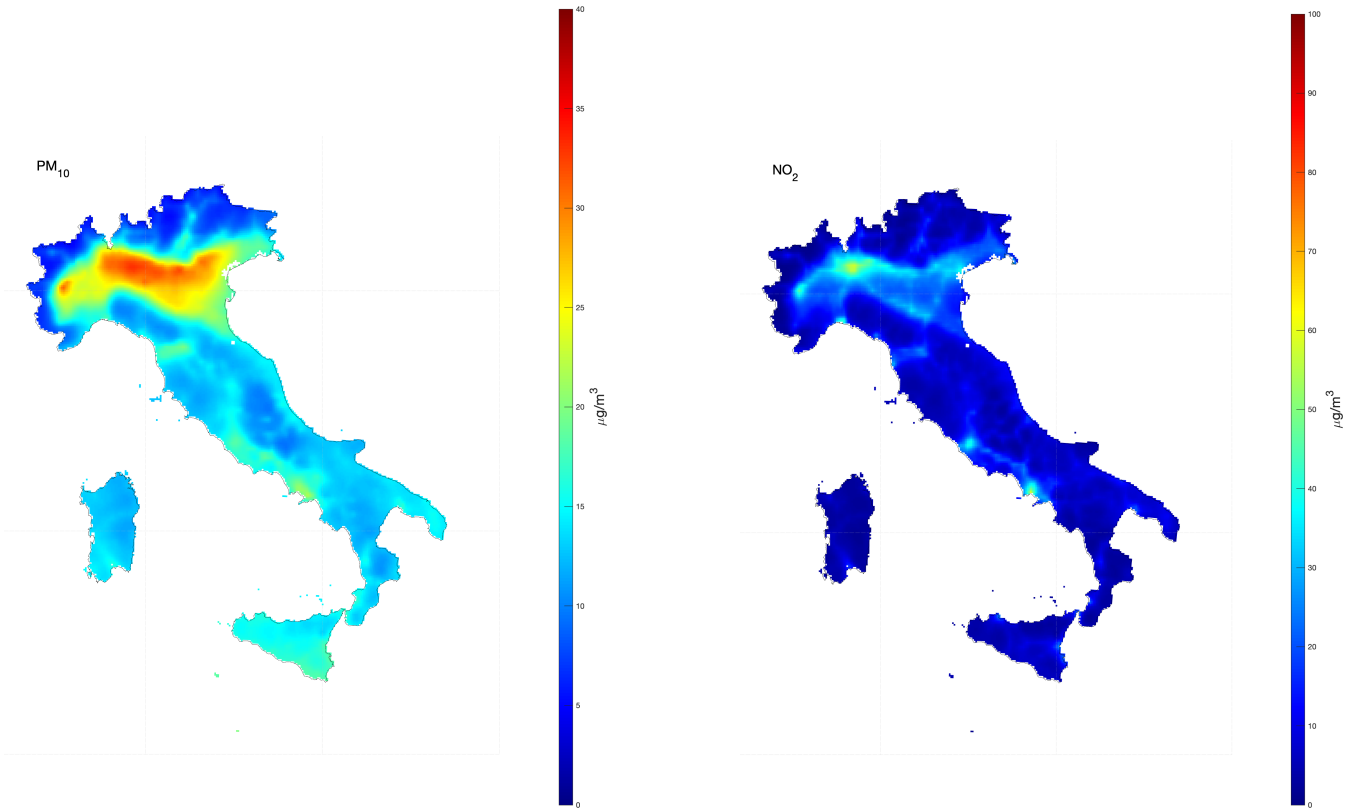

Figure S4. Median of PM_{10} concentration map (left) of daily means, and median NO_2 concentration map (right) of 1-hour daily maximum in 2022, from raw CAMS data and bi-linearly interpolated over a regular 4×4 km grid resolution.

S7 Results from the post-processing approach for $PM_{2.5}$ and O_3

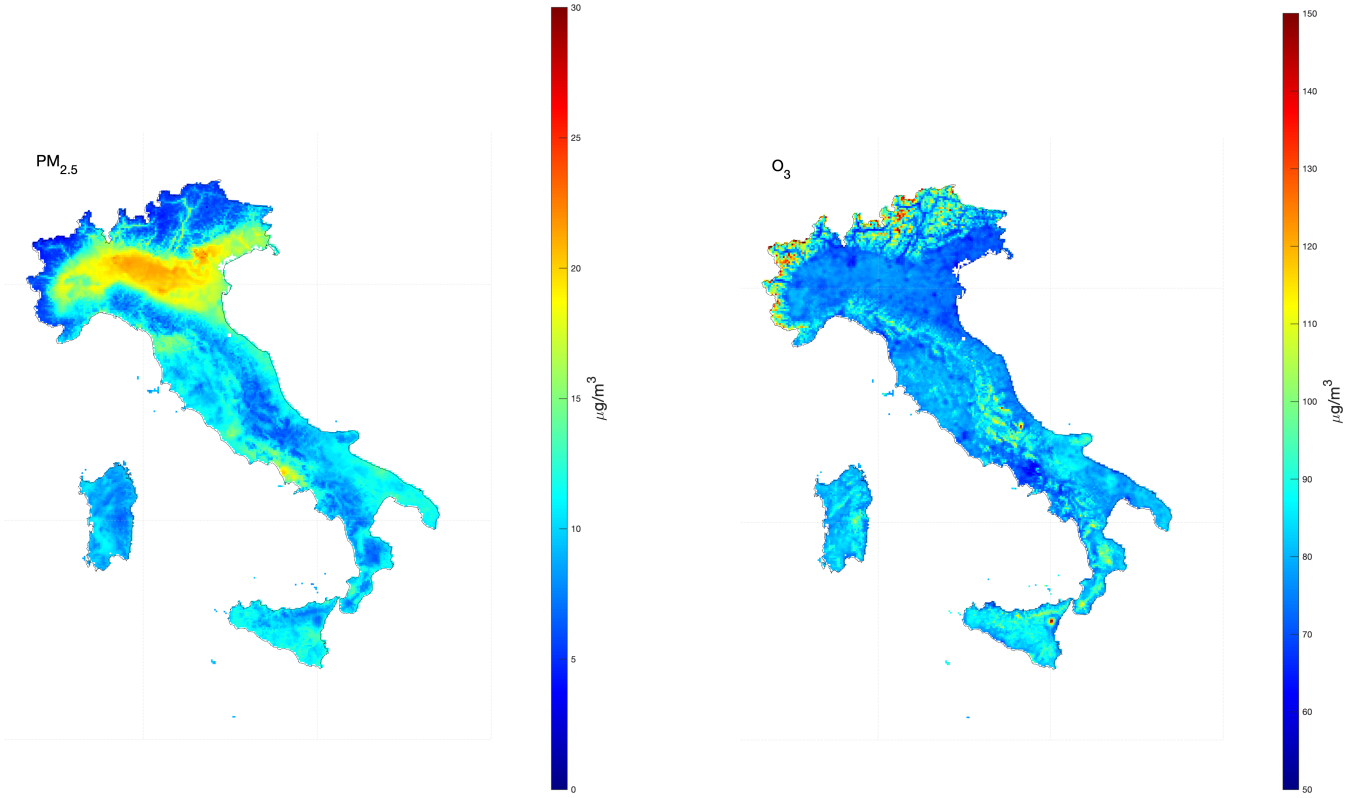

Figure S5. Median $PM_{2.5}$ concentration map (left) of daily means, and median O_3 concentration map (right) of highest 8-hour daily maximum in 2022, after the application of the second post-processing stage and estimated over a regular 4×4 km grid resolution.

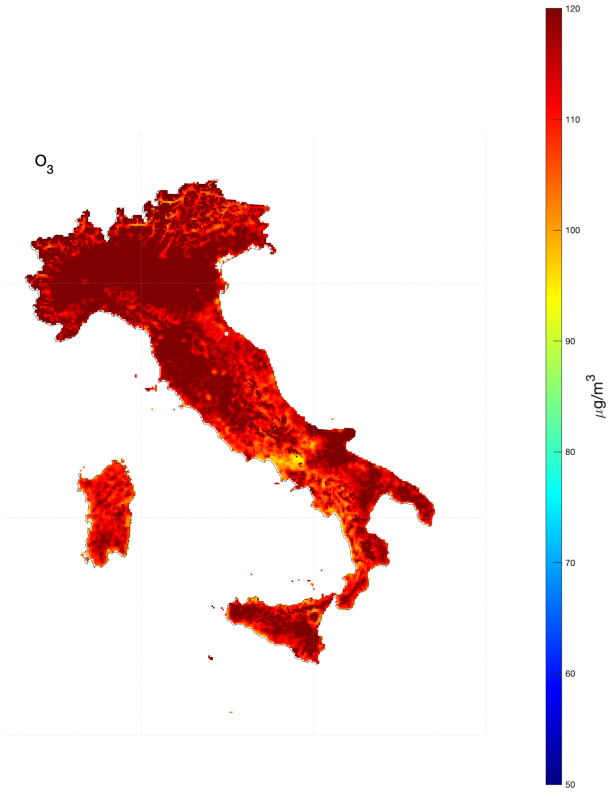

Figure S6. 95.1st percentile of O_3 highest 8-hour daily maximum after the application of the second post-processing stage and estimated over a regular 4×4 km grid resolution.

References

Bertrand, J.-M., Meleux, F., Ung, A., Descombes, G., and Colette, A.: Improving the European air quality forecast of Copernicus Atmosphere Monitoring Service using machine learning techniques, Atmospheric Chemistry and Physics Discussions, pp. 1–28, 2022.

Blangiardo, M., Cameletti, M., Baio, G., and Rue, H.: Spatial and spatio-temporal models with R-INLA, Spatial and spatio-temporal Epidemiology, 4, 33–49, 2013.

Cameletti, M., Lindgren, F., Simpson, D., and Rue, H.: Spatio-temporal modeling of particulate matter concentration through the SPDE approach, AStA Advances in Statistical Analysis, 97, 109–131, 2013.

Cressie, N. and Wikle, C. K.: Statistics for spatio-temporal data, John Wiley & Sons, 2015.

Fioravanti, G., Martino, S., Cameletti, M., and Cattani, G.: Spatio-temporal modelling of PM_{10} daily concentrations in Italy using the SPDE approach, Atmospheric Environment, 248, 118 192, 2021.

Gneiting, T., Raftery, A. E., Westveld, A. H., and Goldman, T.: Calibrated probabilistic forecasting using ensemble model output statistics and minimum CRPS estimation, Monthly Weather Review, 133, 1098–1118, 2005.

Gneiting, T., Balabdaoui, F., and Raftery, A. E.: Probabilistic forecasts, calibration and sharpness, Journal of the Royal Statistical Society: Series B (Statistical Methodology), 69, 243–268, 2007.

Konstantinoudis, G., Cameletti, M., Gómez-Rubio, V., Gómez, I. L., Pirani, M., Baio, G., Larrauri, A., Riou, J., Egger, M., Vineis, P., et al.: Regional excess mortality during the 2020 COVID-19 pandemic in five European countries, Nature Communications, 13, 482, 2022.

Lindgren, F., Rue, H., and Lindström, J.: An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach, Journal of the Royal Statistical Society: Series B (Statistical Methodology), 73, 423–498, 2011.

Murphy, A. H.: Skill scores based on the mean square error and their relationships to the correlation coefficient, Monthly Weather Review, 116, 2417–2424, 1988.

Scheuerer, M. and Möller, D.: Probabilistic wind speed forecasting on a grid based on ensemble model output statistics, The Annals of Applied Statistics, 9, 1328–1349, 2015.

Toth, Z., Talagrand, O., Candille, G., and Zhu, Y.: Probability and ensemble forecasts, in: Forecast verification: A practitioner's guide in atmospheric science, edited by Jolliffe, I. T. and Stephenson, D. B., chap. 7, pp. 137–163, John Wiley and Sons, First edn., 2003.

Wilks, D. S.: Univariate Ensemble Postprocessing, in: Statistical Postprocessing of Ensemble Forecasts, edited by Vannitsem, S., Wilks, D. S., and Messner, J. W., pp. 49–89, Elsevier, 2018.