https://doi.org/10.5194/egusphere-2024-4056
Preprint. Discussion started: 14 February 2025

Machine learning data fusion for high spatio-temporal resolution PM$_{2.5}$

Andrea Porcheddu[1], Ville Kolehmainen[1], Timo Lähivaara[1], and Antti Lipponen[2]

[1]Department of Technical Physics, University of Eastern Finland, Kuopio, Finland
[2]Finnish Meteorological Institute, Atmospheric Research Centre of Eastern Finland, Kuopio, Finland

Correspondence: Andrea Porcheddu (andrea.porcheddu@uef.fi)

Abstract. Understanding PM$_{2.5}$ variability at fine scale is crucial to assess urban pollution impact on the population and to inform the policy-making process. PM$_{2.5}$ in-situ measurements at ground level cannot offer gapless spatial coverage, while current satellite retrievals generally cannot offer both high-spatial and high-temporal resolution, with night-time estimation posing further challenges. This study tackles these difficulties, introducing an innovative deep learning data fusion method to estimate hourly PM$_{2.5}$ maps at 100 m resolution on urban areas. We combine low resolution geophysical model data, high resolution geographical indicators, PM$_{2.5}$ in-situ ground stations measurements and PM$_{2.5}$ retrieved at satellite overpass. To simultaneously treat spatial and temporal correlations in our data, we deploy a 3D U-Net based neural network model. To evaluate the model, we select the city of Paris, France, in the year 2019 as our study region and time. Quantitative assessment of the model is carried out using the ground station data with a leave-one-out cross-validation approach. Our method outperforms MERRA-2 PM$_{2.5}$ estimates, predicting PM$_{2.5}$ hourly (R^2 = 0.51, RMSE = 6.58 $\mu g/m^3$), daily (R^2 = 0.65, RMSE = 4.92 $\mu g/m^3$), and monthly (R^2 = 0.87, RMSE = 2.87 $\mu g/m^3$). The proposed approach and its possible future developments can be highly beneficial for PM$_{2.5}$ exposure and regulation studies at fine suburban scale.

1 Introduction

One of the key indicators in air quality monitoring and regulation is PM$_{2.5}$ which is the concentration of particulate matter (PM) with an aerodynamic diameter less than 2.5 μm in cubic meter of air ($\mu g/m^3$). PM$_{2.5}$ has different chemical compositions and it's emissions originate from different natural and anthropogenic sources such as fuel combustion, wildfires, and sea salt. From the epidemiological point of view, high PM$_{2.5}$ levels have been connected to many illnesses, such as stroke and cardiovascular and respiratory diseases (Pope and Dockery, 2006; Cohen et al., 2017). The pathogenicity of fine particulate matter pollution makes it one of the biggest environmental health risks as over 90% of the world's population lives in areas with annual mean PM$_{2.5}$ levels exceeding the new WHO 2021 air quality guideline of 5 $\mu g/m^3$ (Health Effects Institute, 2019).

PM$_{2.5}$ and other pollutants can be measured with high accuracy by in-situ ground station networks. However, the existing monitoring sites are sparsely located and mostly in developed countries, typically few stations in a large metropolitan area producing accurate measurements representing the conditions in the proximity of the ground stations. Despite some spatial interpolation techniques could be used (Deng, 2015) to estimate PM$_{2.5}$ over larger urban areas from the point-like measurements obtained by these ground stations, they do not alone permit accurate spatially distributed estimates for epidemiological studies and regulation at a suburban level scale. Aiming at a more appropriate spatial coverage and resolution, $PM_{2.5}$ can also be estimated by using airborne remote-sensing techniques, in particular satellite retrievals. In satellite remote sensing, $PM_{2.5}$ estimates are typically based on Aerosol Optical Depth (AOD), a quantity expressing electromagnetic radiation extinction through a column of air at a given wavelength. AOD is a columnar optical quantity while $PM_{2.5}$ is a concentration of particles at ground level. For AOD to $PM_{2.5}$ conversion, the estimation utilizes auxiliary measurement and model data such as aerosol vertical distribution and metereological variables (Chu et al., 2016; Tang et al., 2024). Nowadays many AOD satellite products exist with different spatial and temporal resolution. Different studies used low orbiting satellites (e.g. MODIS product (Levy et al., 2013)) that have one or two overpasses per day or geostationary satellites (e.g. AHI product (Bessho et al., 2016)) giving sub-hourly estimates. Instruments on low orbiting satellites have generally higher spatial resolution than geostationary ones. Although giving high spatial resolution $PM_{2.5}$ estimates, low orbiting satellites products have low temporal resolution (1-2 snapshots per day) and retrieving information on night-time aerosols is a challenging task for the development of geostationary satellites products. Reanalysis models such as MERRA-2 (Randles et al., 2017) and CAMS (Inness et al., 2019), and forecast models such as GEOS-CF (Keller et al., 2021) offer hourly $PM_{2.5}$ available globally. However, the spatial resolution of these $PM_{2.5}$ maps is low (tens of kilometers) for higher resolution studies such as distribution of pollution at suburban levels.

To estimate $PM_{2.5}$, we recently proposed a method (Porcheddu et al., 2024) leveraging the Sentinel-3 POPCORN AOD product (Lipponen et al., 2022). The POPCORN AOD is a post-process corrected version of Sentinel-3 SYNERGY land AOD, characterized by a high spatial resolution of 300 m and derived using a feed-forward neural network trained on AERONET-collocated data. This enhanced AOD product provides accurate spectral aerosol information for five regions of interest (Central Europe, Eastern USA, Western USA, Southern Africa, and India) for the year 2019, making it a valuable input for air quality estimation models. To post-process correct the MERRA-2 AOD-to-$PM_{2.5}$ conversion ratio, we deployed an ensemble of deep neural networks for a fusion of collocated ground station in-situ $PM_{2.5}$ data, MERRA-2 reanalysis model AOD and $PM_{2.5}$ data, spectral AERONET AOD, satellite-observed spectral top-of-atmosphere reflectances, and meteorology data. We also used various high-resolution geographical indicators representing, e.g., population density and land surface elevation. The deep learning model was used for estimation of $PM_{2.5}$ with 100 m resolution from low orbiting satellite images, producing 1-2 daily per overpass snapshots of high-resolution $PM_{2.5}$ data where AOD data was available.

In this study, we have two research questions. How could we obtain $PM_{2.5}$ maps offering large (e.g. metropolitan level) spatial coverage with both high spatial and temporal resolution? Considering satellite derived $PM_{2.5}$ maps where AOD data is missing, e.g. because of cloud covering, how can we estimate $PM_{2.5}$ at those locations? To address these questions, we propose a novel deep learning based data fusion method to produce hourly $PM_{2.5}$ estimates with 100 m resolution. We use a 3D U-Net architecture (Özgün Çiçek et al., 2016) to produce 24-hour sequences of hourly $PM_{2.5}$ maps. The model is trained to yield a small L_2-misfit with the $PM_{2.5}$ estimates obtained during satellite overpasses in our previous study (Porcheddu et al., 2024), as well as with available ground station data. As inputs, we utilize 24-hour sequences of geophysical model data (MERRA-2) providing low-resolution maps of meteorological and aerosol-related indicators (1-hour temporal resolution), and high-resolution geographical indicator maps (1-month temporal resolution). This allows the model to generate hourly $PM_{2.5}$

https://doi.org/10.5194/egusphere-2024-4056
Preprint. Discussion started: 14 February 2025

outputs for the entire 24-hour period covered by the inputs. The model is trained on data for the year 2019 in the city of Paris, France, and assessed against ground station data with a leave one out cross validation approach.

2 Data

This section describes the data used in the proposed deep learning based data fusion for high resolution $PM_{2.5}$. The proposed approach is tested using data from Paris, France, for the year 2019. Table A1 in the appendix lists all the features utilized as model inputs.

2.1 NOODLESALAD $PM_{2.5}$

NOODLESALAD $PM_{2.5}$ (Porcheddu et al., 2024) retrievals are obtained applying a deep learning based post-process correction approach to the MERRA-2 AOD-to-$PM_{2.5}$ conversion ratio. The post-process corrected AOD-to-$PM_{2.5}$ conversion ratio is utilized to map high resolution POPCORN SENTINEL-3 SYNERGY AOD estimate (Lipponen et al., 2022) to high resolution $PM_{2.5}$ estimate. The post-process correction of MERRA-2 AOD-to-$PM_{2.5}$ conversion ratio is carried out deploying an ensemble of fully-connected feed-forward neural networks and a fusion of surface in-situ $PM_{2.5}$ observations, MERRA-2 reanalysis model AOD and $PM_{2.5}$ data, spectral AERONET AOD, satellite-observed spectral top-of-atmosphere reflectances, meteorology data, and various high-resolution geographical indicators. The ensemble technique leads to a distribution of predictions for a single $PM_{2.5}$ estimate. The median of the ensemble is considered as the $PM_{2.5}$ estimate and the width of the distribution is regarded as an uncertainty related to the machine learning model training (model uncertainty). NOODLESALAD $PM_{2.5}$ offers a spatial resolution of 100 meters and is currently available for Sentinel-3A and 3B overpasses, covering Central Europe for the year 2019. The two Sentinel-3 satellites currently flying provide revisit times of less than two days for OLCI and less than one day for the SLSTR instrument at equator. Swath width of the OLCI instrument is 1270 km. SLSTR swath width is 1420 km for the nadir view and 750 km for the oblique view. In this work, we consider NOODLESALAD $PM_{2.5}$ retrievals in Paris, France, in 2019.

2.2 OpenAQ

OpenAQ (https://openaq.org/) is an open-access database for ground stations air quality data. In this study, we utilize OpenAQ as our source for surface in-situ $PM_{2.5}$ observations. OpenAQ offers pointwise air quality measurement data from thousands of stations. The temporal resolution of the data varies by station, with 1-hour and daily observations commonly available. Figure 1 shows a map of OpenAQ stations that provide hourly data within our region of interest. We discard $PM_{2.5}$ observations when they are greater than the calculated upper fence $Q3 + 6 \times (Q3 - Q1)$ (where $Q3$ and $Q1$ are respectively the third and first quartiles of the $PM_{2.5}$ distribution), regarding them as outliers. This step was carried out to filter extreme outliers, which can be caused by exceptional events or ground station malfunctions.

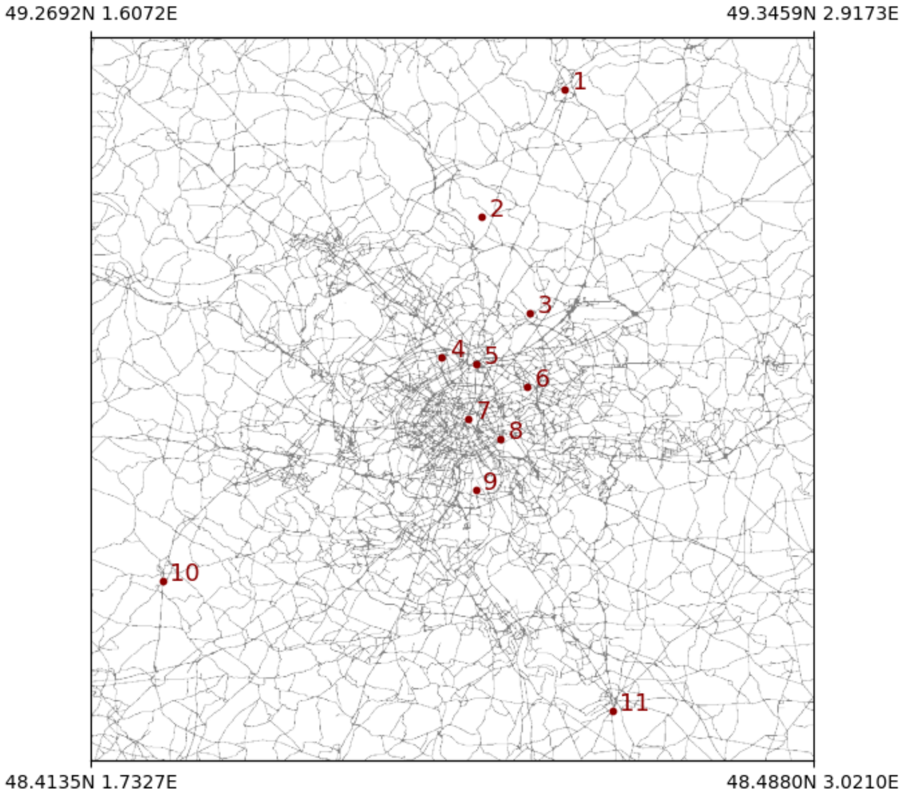


Figure 1. Map of OpenAQ stations in the region of interest (Paris, France).

2.3 MERRA-2

The Modern-Era Retrospective analysis for Research and Applications, Version 2 (MERRA-2), is NASA's reanalysis model (Randles et al., 2017). MERRA-2 provides model data for various variables in meteorology, aerosols and air quality. MERRA-2 has a spatial resolution of 0.5° x 0.625°, which is approximately 50 km^2 in the Central Europe region. The time-varying variables from MERRA-2 that we use have a temporal resolution of 1 hour, with both instantaneous and time-averaged values available depending on the variable and data product. Appendix A contains a list of all MERRA-2 variables that are used as inputs in the proposed approach.

In addition to the variables contained in the MERRA-2 data, we calculate certain input variables from the MERRA-2 meteorological and aerosol data. These data are defined as:

- **Relative humidity (RH) at the surface. Equation based on the Clausius-Clapeyron equation (see e.g. Michaelides et al., 2019):**

$$\mathrm{RH} = 0.263 \cdot \mathrm{PS} \cdot \mathrm{QLML} / \exp\left((17.67 \cdot (\mathrm{T2M} - 273.15)) / (\mathrm{T2M} - 29.65)\right)$$

https://doi.org/10.5194/egusphere-2024-4056
Preprint. Discussion started: 14 February 2025

- **Wind direction (WD10M) at 10 m:**

  $$\mathrm{WD10M} = \arctan\left(-\mathrm{V10M}/\mathrm{U10M}\right)$$

- **Wind speed (WS10M) at 10 m:**

  $$\mathrm{WS10M} = \sqrt{\mathrm{U10M}^2 + \mathrm{V10M}^2}$$

- **$\mathrm{PM}_{2.5}$ at surface:** (Buchard et al. (2016))

  $$\mathrm{PM}_{2.5} = (1.375 \cdot \mathrm{SO4SMASS} + 1.4 \cdot \mathrm{OCSMASS} + \mathrm{BCSMASS} + \mathrm{DUSMASS25} + \mathrm{SSSMASS25}) \cdot 10^9$$

- **AOD-to-$\mathrm{PM}_{2.5}$ ratio η:**

  $$\eta = \frac{\mathrm{PM}_{2.5}}{\mathrm{TOTEXTTAU}}$$

2.4 High-resolution geographical indicators

2.4.1 OpenStreetMap roads

OpenStreetMap is an open-source project that contains high spatial resolution map data. In our model, we utilize OpenStreetMap roads as a data source for inputs. Specifically, we calculate the distance to the nearest street or highway and use this measurement as one of our input variables. The distances are computed on a 100 m resolution grid. In OpenStreetMap, all paths, streets, and highways are categorized under 'highways'. However, we only consider certain sub-classes that include roads and highways accessible to car traffic, as these are potential sources of $\mathrm{PM}_{2.5}$ pollution (information from (OpenStreetMap, 2023)). Appendix A lists all the OpenStreetMap road types used to determine the distance to the nearest road.

2.4.2 NASA Black Marble Night Lights

NASA's Black Marble is a night light product derived from the Visible Infrared Imaging Radiometer Suite (VIIRS) day/night band (DNB) radiances captured during night-time. The DNB is extremely sensitive to light, allowing it to detect even very low-intensity lights on Earth's surface at night. Most of these night-time lights are attributed to human activities. Since the distribution of night lights closely reflects human presence, we use NASA's Black Marble Night Lights as a proxy for population density, incorporating it as an input in our models. We utilize Night Light data with a spatial resolution of 500 m, based on the annual data product VNP46A4 (Wang et al., 2020).

2.4.3 MODIS land cover type

We utilize the MODIS `MCD12Q1` land cover type data product (Sulla-Menashe and Friedl, 2018) to generate input variables that represent the distances to the nearest International Geosphere Biosphere Programme (IGBP) land cover types (Loveland and Belward, 1997; Belward et al., 1999). The MODIS `MCD12Q1` data product has a spatial resolution of 500 m. A complete list of the IGBP land cover types can be found in Appendix A.

https://doi.org/10.5194/egusphere-2024-4056
Preprint. Discussion started: 14 February 2025

2.4.4 Digital Elevation Model

We utilize the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) digital elevation model (DEM) to represent land surface elevation (Fujisada et al., 2011, 2012; NASA/METI/AIST/Japan Spacesystems, and US/Japan ASTER Science Team, 2019). The ASTER DEM provides a spatial resolution of 1 arcsecond, which is approximately 30 m.

3 Methods

Our objective is to estimate 3D (two spatial dimensions and time) $PM_{2.5}$ maps in the region of interest by fusion of satellite and ground station measurement data, model data and different indicators as inputs for the deep learning model. Since we are dealing with unstructured data in the form of images, a well-suited choice for the machine learning model is a Convolutional Neural Network (CNN) (LeCun et al., 1989). Furthermore, since both the input maps and the output maps represent the same region of interest, a U-Net model is an appropriate choice (Ronneberger et al., 2015). As the data is in 3D, we choose a variant of U-Net called 3D U-Net (Özgün Çiçek et al., 2016). The main difference between conventional U-Net and 3D U-Net is that the latter deploys 3D convolutions instead of 2D convolutions for processing 3D image data.

One must note that other network architectures could also be utilized. One possibility would be, e.g., to use a U-Net with convolutional Long Short-Term Memory (LSTM) layers (Shi et al., 2015). Convolutional LSTM layers behave as LSTM layers, with the key difference of performing their internal operations as convolutions, consequently being a possible choice for processing time series of 2D images. Nevertheless, we decided to use 3D U-Net as it was found computationally feasible and less memory intensive for processing the large data sets.

The input data consist of 4 dimensional arrays. The first dimension represents time and has size 24 in order to contain hourly information of a single day. The second dimension contains the different channels (i.e. the different input features) and the remaining two dimensions are the spatial dimensions with image size of 960x960. The output is a 3 dimensional array containing 24 hourly $PM_{2.5}$ maps in the region of interest as 960x960 images with pixel size 100 m x 100 m.

The model architecture has been implemented using the PyTorch framework (Paszke et al., 2019), a widely used library known for its flexibility and efficiency in developing deep learning models. A detailed schematic of the model architecture is provided in Fig. 2 to illustrate its structure and components.

The model consists of a contracting path (the encoder) and an expansive path (the decoder). On each level of the contracting path, 3D convolutions combined with ReLU activations and max pooling layers help in finding relevant features from the input maps, producing a new representation of the input with lower spatial and temporal resolution with a higher number of channels. The expansive path deploys 3D nearest neighbour upsampling and 3D convolutions followed by ReLU activations in order to recover a final output with the same spatial and temporal resolution of the input. Notice the skip connections linking each contracting path level to the corresponding expansive path level: from an intuitive point of view, these are useful to exploit the fine details contained in the input when generating the output. Finally, the output layer is a 3D 1x1x1 convolution followed by SoftPlus activations in order to constrain the output to be a positive definite array. When testing the network architecture,

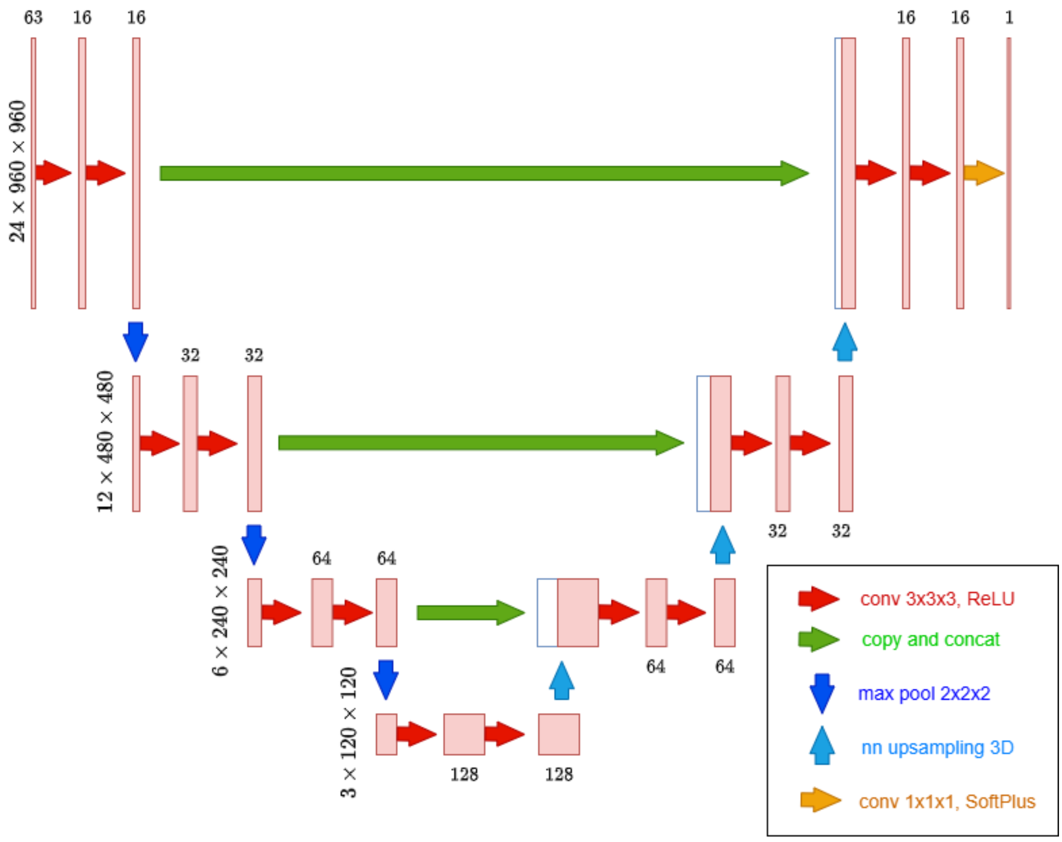


Figure 2. Visualization of the applied neural network architecture.

various kernel sizes, internal activation functions, and upsampling techniques were evaluated, but no significant differences in performance were observed.

3.1 Loss Function

We use the Mean Square Error (MSE) as the loss function in the supervised regression problem of fitting the 3D U-Net model to the training data. Ideally in clear sky conditions for a satellite overpass, we would have a full $PM_{2.5}$ map. In ideal conditions for a single ground station, we would have a full time series without missing data. However, in reality satellite overpass data can lack information at some pixels, e.g. cloud covering can hinder AOD retrievals, therefore hindering $PM_{2.5}$ estimates, and a ground station can malfunction, leading to missing data in the time-series of the station. Furthermore, on a single day we usually have significantly more $PM_{2.5}$ pixel values coming from the satellite overpasses than measured values from the ground stations. The ground stations $PM_{2.5}$ data are usually more accurate than satellite estimate data and they are our only data source available hourly. On the other hand, NOODLESALAD $PM_{2.5}$ is our data source providing information far from the ground stations at the time of satellite overpasses. Therefore, to take the missing data and highly different number of satellite versus ground station data available into account, we consider masking and weighting the data in the MSE loss function.

Let the batch size be denoted by B (with $B = 4$ for this particular case) and define $N = \{1,\dots,960\times 960\}$ as the set of all pixel indices. For each sample b in the batch, define $H_{st,b} \subseteq \{1,\dots,24\}$ as the set of hours during which ground stations measurements are available, and $H_{op,b} \subseteq \{1,\dots,24\}$ as the corresponding set of hours for satellite overpasses. Let $y_{b,h,i}$ represent the target measurement at sample b, hour h, and pixel index i, with $\hat{y}_{b,h,i}$ denoting the corresponding predicted value by the deep learning model. When a measurement is not available (due to lacking data from a ground monitoring station or failed satellite retrieval due to cloud covering) $y_{b,h,i}$ is encoded as 0, since this value does not naturally occur in the dataset ($PM_{2.5}$ is not zero in a realistic setting). On the other hand, $\hat{y}_{b,h,i}$ is always a positive real value by choice, since SoftPlus was chosen as activation function for our model output.

For each sample b and hour h, define:

$$S_{st,b,h} = \{i \in N \mid y_{b,h,i} \neq 0 \text{ and } h \in H_{st,b}\},$$
$$\tilde{S}_{op,b,h} = \{i \in N \mid y_{b,h,i} \neq 0 \text{ and } h \in H_{op,b}\},$$

representing pixel indices where ground station measurements ($S_{st,b,h}$) and satellite overpass measurements ($\tilde{S}_{op,b,h}$) are available. The sets $S_{st,b,h}$ and $S_{op,b,h}$ are defined separately for the ground stations and satellite overpass data. From the satellite overpass data, we create a subset $S_{op,b,h} \subset \tilde{S}_{op,b,h}$ of data with 25% of size of $\tilde{S}_{op,b,h}$ by uniform random sampling of the pixels to be used in the minimization. The random sampling is performed at each training step (for every update of the network parameters) and can be seen as an optimization technique analogue to batch shuffling. This undersampling is beneficial for training on overpass estimates, especially given the substantial pixel count (nearly 1 million when all pixels provide valid measurements at overpass times).

The average losses for the ground stations and overpass contributions are defined by summing over the respective sets of valid hours:

$$L_{st,b} = C_{st,b} \sum_{h\in H_{st,b}} \frac{1}{|S_{st,b,h}|} \sum_{i\in S_{st,b,h}} (y_{b,h,i} - \hat{y}_{b,h,i})^2$$

$$L_{op,b} = C_{op,b} \sum_{h\in H_{op,b}} \frac{1}{|S_{op,b,h}|} \sum_{i\in S_{op,b,h}} (y_{b,h,i} - \hat{y}_{b,h,i})^2$$

where $C_{st,b}$ and $C_{op,b}$ are factors depending on the sizes of the sets $H_{st,b}$ and $H_{op,b}$ (defined respectively as $|H_{st,b}|$ and $|H_{op,b}|$). If these sets are empty, $C_{st,b}$ and $C_{op,b}$ are equal to 0, otherwise they correspond to $\frac{1}{|H_{st,b}|}$ and $\frac{1}{|H_{op,b}|}$ respectively. Analogously, $|S_{st,b,h}|$ and $|S_{op,b,h}|$ represent the sizes of the sets $S_{st,b,h}$ and $S_{op,b,h}$.

We then define the sample-specific loss L_b as:

$$L_b = \begin{cases} \frac{L_{st,b}+L_{op,b}}{2}, & \text{if } L_{st,b} \neq 0 \text{ and } L_{op,b} \neq 0 \\ L_{st,b}, & \text{if } L_{op,b} = 0 \\ L_{op,b}, & \text{if } L_{st,b} = 0 \end{cases}$$

Finally, the overall loss for the batch is defined as:

$$\mathcal{L} = \frac{1}{B}\sum_{b=1}^{B} L_b$$

This formalization provides the necessary structure to include both ground station and satellite overpass target data within each batch.

3.2 Model training

For each data sample, one or two target maps correspond to the available satellite overpasses data (i.e. NOODLESALAD $PM_{2.5}$) while the others contain only ground stations data (therefore 11 pixels for each map when data is available from all the stations). In order to test the results from the network training, we used leave-one-out cross-validation (CV) (i.e. we removed one different station from each training and used the data left out of training for testing purposes). Therefore, we trained 11 different networks (one for each ground station in the region of interest). Furthermore, the dataset has been split into training set (approximately 80% of the data samples) used for the optimization of the network weights and validation set (approximately 20% of the data samples) used for early stopping of the optimization. The early stopping is a form of regularization useful to avoid overfitting of the network (Goodfellow et al., 2016). It consists in keeping track of both the training error and validation error with the objective of stopping the training when the validation error starts to increase (i.e. when the network stops to learn useful patterns and noise in the training set starts to play a significant role). While early stopping is not the only regularization technique applicable to train deep learning models, it offers a good trade-off between model performance and training time. We consider a patience parameter equal to 30, meaning that the training stops when no improvement on the validation loss is recorded over 30 epochs. Since using a small batch size introduces fluctuations in the loss during training, this choice of the patience parameter is reasonable, as a lower patience value could prematurely stop training and lead to underfitting.

We utilized the Adam algorithm (with learning rate equal to 0.0001) and the custom loss function described in 3.1 to optimize the network model parameters.

4 Results

We considered a leave-one-out CV approach, training 11 models, each time leaving one station out of the training as test station. The results of predicted $PM_{2.5}$ at the locations of the 11 AQ stations in Paris are shown in Fig. 3. Different fidelity metrics (per each trained model) were calculated to compare MERRA-2 estimates (orange bars) and our model predictions

https://doi.org/10.5194/egusphere-2024-4056
Preprint. Discussion started: 14 February 2025

(blue bars) to OpenAQ measurements (ground truth). Correlation R^2, Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) values are shown on the left, middle and right columns. These metrics are evaluated for hourly averages, daily averages and monthly averages (estimated using the hourly averages) on the top, middle and bottom row. The fidelity metrics averages show that our model clearly outperforms MERRA-2. R^2 CV averages for our model are 0.51 (hourly averages), 0.65 (daily averages) and 0.87 (monthly averages). R^2 CV averages for MERRA-2 are respectively 0.10, 0.18, and 0.42. RMSE CV averages for our model are 6.58 $\mu g/m^3$ (hourly averages), 4.92 $\mu g/m^3$ (daily averages) and 2.87 $\mu g/m^3$ (monthly averages). The same metrics averages for MERRA-2 are respectively 9.05 $\mu g/m^3$, 7.04 $\mu g/m^3$ and 4.08 $\mu g/m^3$. MAE CV averages for our model are 4.61 $\mu g/m^3$ (hourly averages), 3.59 $\mu g/m^3$ (daily averages) and 2.51 $\mu g/m^3$ (monthly averages). MAE CV averages for MERRA-2 are respectively 6.39 $\mu g/m^3$, 5.14 $\mu g/m^3$ and 3.30 $\mu g/m^3$. We can notice from Figure 3 that our model outperforms MERRA-2 on all hourly and daily value metrics, and in most monthly averaged with the exceptions that MERRA-2 has better monthly MAE for stations 5 and 11, better RMSE for station 11. The R^2 values still show a clear improvement of our model. For what regards station 5, the better RMSE and worse MAE are due to the fact that RMSE highlights outliers (i.e. MERRA-2 commits less but bigger mistakes). Anyway, the important R^2 values for both station 5 and station 11 show that our model correctly predicts the AQ trends better but is off mainly by a scaling factor. From Fig. 1, one would expect some differences in the performances at different ground stations, since the more isolated is the station, the less information from surrounding stations is present in the training data (we have ideally full $PM_{2.5}$ maps only once per day). Stations 1, 2, 10, and 11 are clearly positioned outside the city center of Paris. Looking at the metrics for these stations, only the location of station 11 seems to have somewhat different prediction accuracy by our model than the stations located in the city center. Although we considered a relatively small dataset to train and test our model, these results suggest it is not overfitting the training data.

Figure 4 shows daily cycle averages on the different seasons (top and middle rows) and monthly averages (bottom row) of 2019 at the test station 1. Black lines represent the ground station measurements, while orange lines and blue lines represent respectively MERRA-2 estimates and our model predictions. Focusing on the daily cycle averages, qualitatively our model and MERRA-2 seem comparable for the spring seasons (March, April, May). The difference is evident on the winter (December, January, February), autumn (September, October, November) and summer (June, July, August) seasons. Especially on the summer season, our model improves notably the accuracy and correlation over MERRA-2. The improvement can be seen quantitatively looking at the metrics on the bottom-right of Fig. 4 (here the estimates for all the seasons have been taken into account in the calculation of the evaluation metrics). Notice how the metrics values for MERRA-2 and our model have been encoded with the same colors of the legend. On the bottom-left of Fig. 4 we compare monthly averages estimates. Again, the difference between MERRA-2 and our model is evident. While MERRA-2 could seem to give a good approximation to the $PM_{2.5}$ annual average at the station, it is not able to capture the time series trend. Our model improves notably from this point of view, showing also improvements in accuracy. This is clear from the metrics for monthly averages, where the R^2 is more than 3 times higher for our model, while the RMSE and MAE are about half of the respective errors in the MERRA-2 estimates.

Figure 5 shows $PM_{2.5}$ seasonal averages maps by hour for the 2019 winter season (December, January, February) predicted by our model on the city of Paris. The dots represent ground stations measurements. The general time series trend reflects the $PM_{2.5}$ variations seen in Fig. 4 on the top-left panel. The $PM_{2.5}$ levels seem to decrease at day time, and raise again at night time. This behaviour can be expected and physically explained through boundary layer height variations. Spatial variations of $PM_{2.5}$ are reasonable and in agreement with what found predicting $PM_{2.5}$ levels at satellite overpass (Porcheddu et al., 2024): the city center and areas surrounding main highways are predicted as the most polluted areas. The maps shown in Fig. 5 are obtained considering station 1 as test station.

The maps shown in Fig. 6 represent $PM_{2.5}$ monthly averages for the year 2019 predicted by our model on the city of Paris. Dots represent ground stations measurementes. The general time series trend reflects the content of the bottom-left panel in Fig.4 as expected: $PM_{2.5}$ levels are higher in colder months, while lower in warmer months. Again, this temporal variation of $PM_{2.5}$ could be explained through boundary layer height variations and also residential heating plays an important role in winter. Spatial variations of $PM_{2.5}$ present the same structure already discussed before for Fig. 5. Again, the maps shown in Fig. 6 are obtained considering station 1 as test station. The agreement with the test station and training stations is generally good.

We employed the SHAP DeepExplainer to compute SHAP values and assess feature importance for the model predictions at station 1 (Lundberg and Lee, 2017). Due to computational constraints, the SHAP values were calculated using a smaller background dataset, with analysis conducted on a subset of 90 randomly selected days. Feature importances were determined by summing the normalized absolute SHAP values, and are shown in Fig.7. As expected, T2M (2-meter air temperature) emerged as one of the most significant predictors, consistent with prior findings. For instance, T2M influences the temporal variability of $PM_{2.5}$ through boundary layer dynamics and contains information about seasonal emission changes.

5 Conclusions

We developed a novel deep learning data fusion method to estimate hourly $PM_{2.5}$ at 100 m spatial resolution, utilizing low-resolution geophysical model data, high-resolution geographical indicators, satellite $PM_{2.5}$ retrievals and in-situ $PM_{2.5}$ ground measurements. A 3D U-Net based architecture was deployed to take into account both spatial and temporal correlations in the data at hand. The methodology was tested on data from Paris, France, for the year 2019.

The model outperforms MERRA-2 $PM_{2.5}$ estimates (our starting point, utilized as model input) on all the evaluation metrics considered. Our estimates are generally consistent with the $PM_{2.5}$ spatio-temporal variability assessed by ground stations measurements. Our method seems promising in answering our research questions: reliable gapless $PM_{2.5}$ maps at fine scale in absence of AOD data, due to absence of satellite overpass or due to failed AOD retrieval, are possible.

Further improvements could be obtained by various means. The method is flexible for what concern data sources, as different data sources could be utilized as inputs, and targets in the training process. In particular, different satellite $PM_{2.5}$ sources could be considered in the training. So far, we considered only NOODLESALAD $PM_{2.5}$. In future studies, we could take into account other satellite data at different satellite overpass times and with different spatial resolution. Considering that geostationary

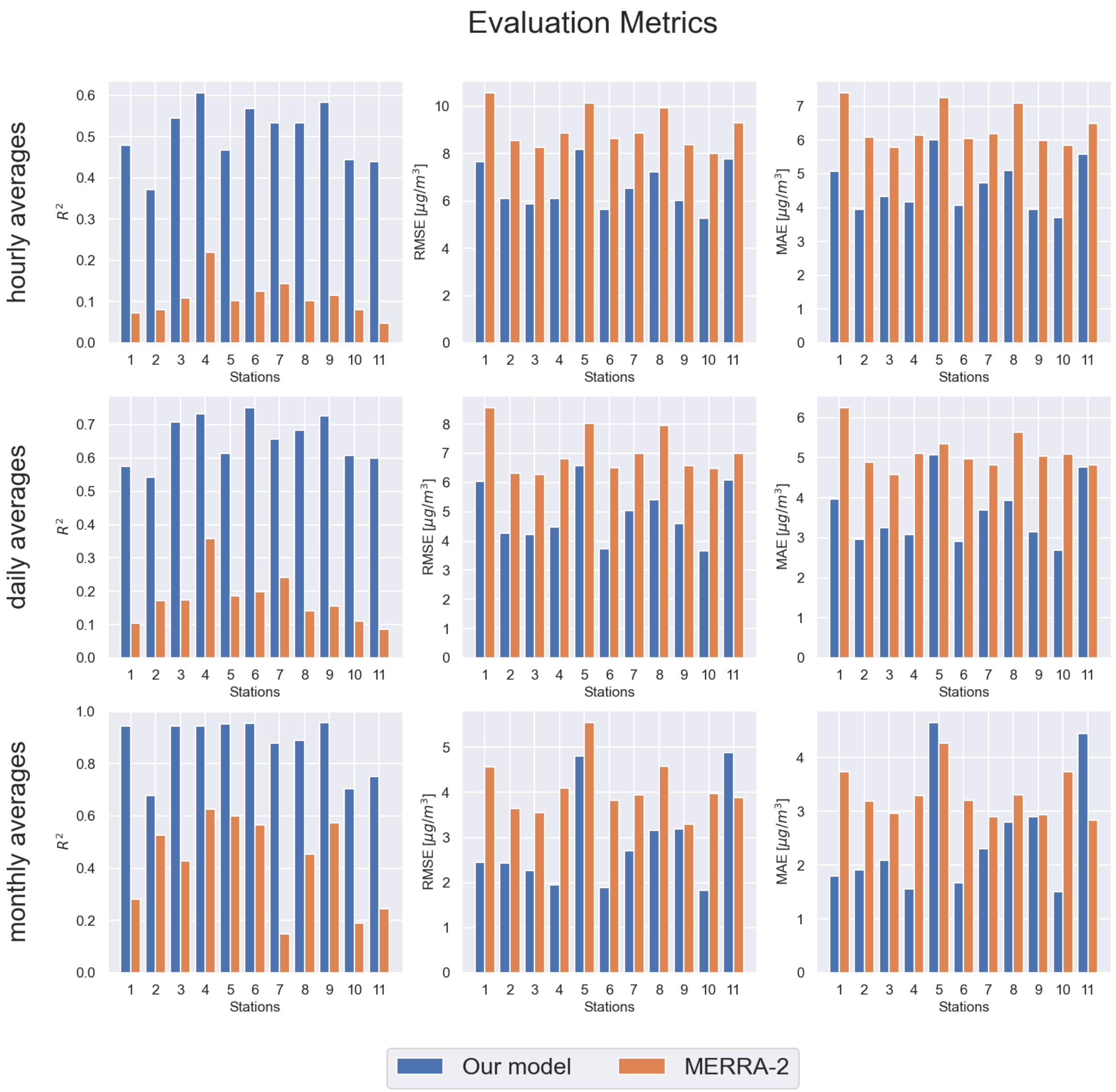
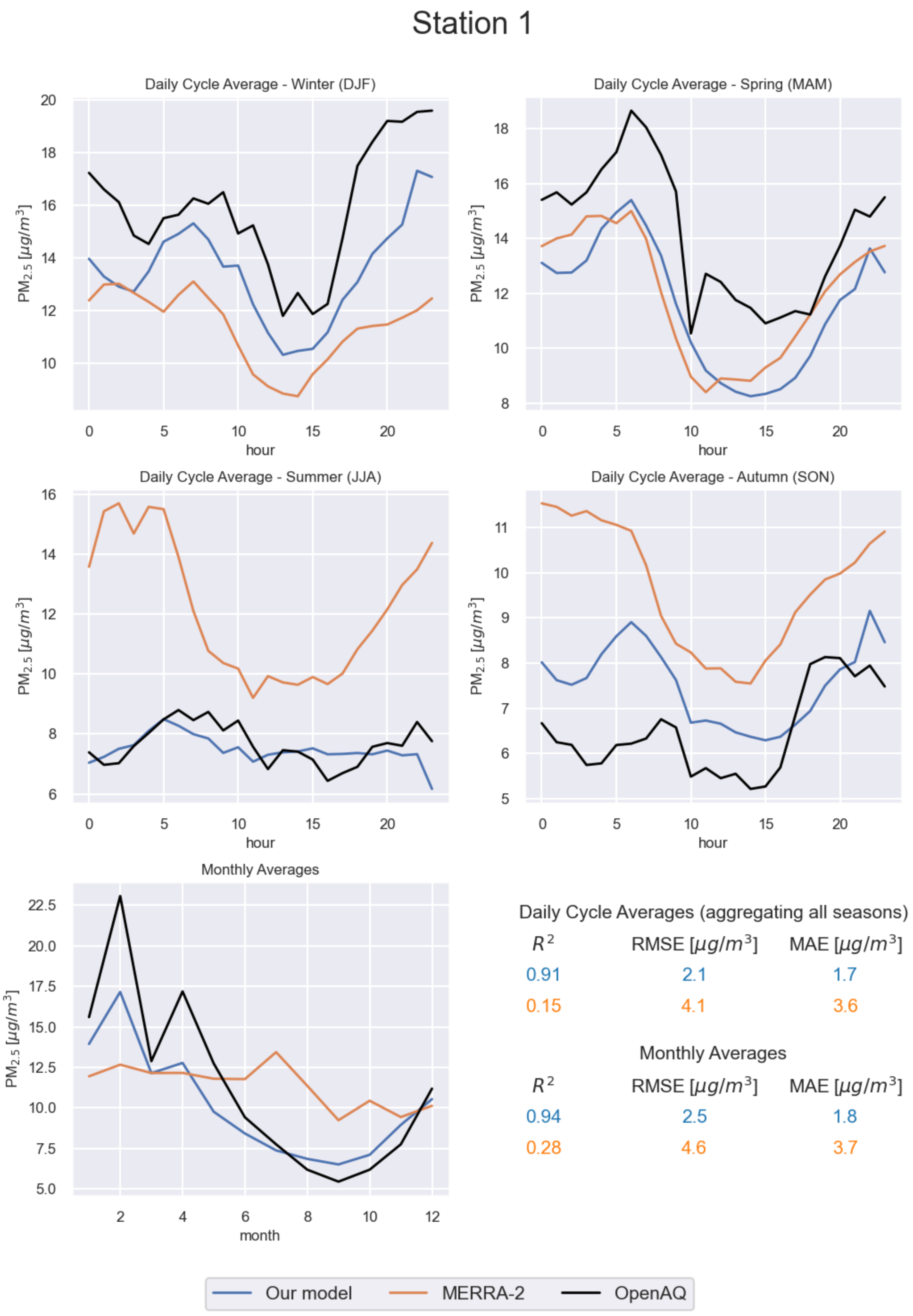


Figure 3. R^2 (left column), RMSE (middle column) and MAE (right column) evaluation metrics resulting from the leave-one-out cross-validation per each test station. The metrics have been calculated for hourly averages (top row), daily averages (middle row) and monthly averages (bottom row). Our model predictions (blue bars) and MERRA-2 estimates (orange bars) are compared to the ground truth data (OpenAQ).



Daily Cycle Averages (aggregating all seasons)

R^2	RMSE [$\mu g/m^3$]	MAE [$\mu g/m^3$]
0.91	2.1	1.7
0.15	4.1	3.6

Monthly Averages

R^2	RMSE [$\mu g/m^3$]	MAE [$\mu g/m^3$]
0.94	2.5	1.8
0.28	4.6	3.7

Figure 4. PM$_{2.5}$ daily cycle averages for the different seasons and monthly averages (at station 1). The black lines represent OpenAQ measurements, the orange lines represent MERRA-2 estimates and the blue lines are predicted by our model. The evaluation metrics comparing respectively MERRA-2 and our model to the ground truth (OpenAQ) are shown on the bottom-right.

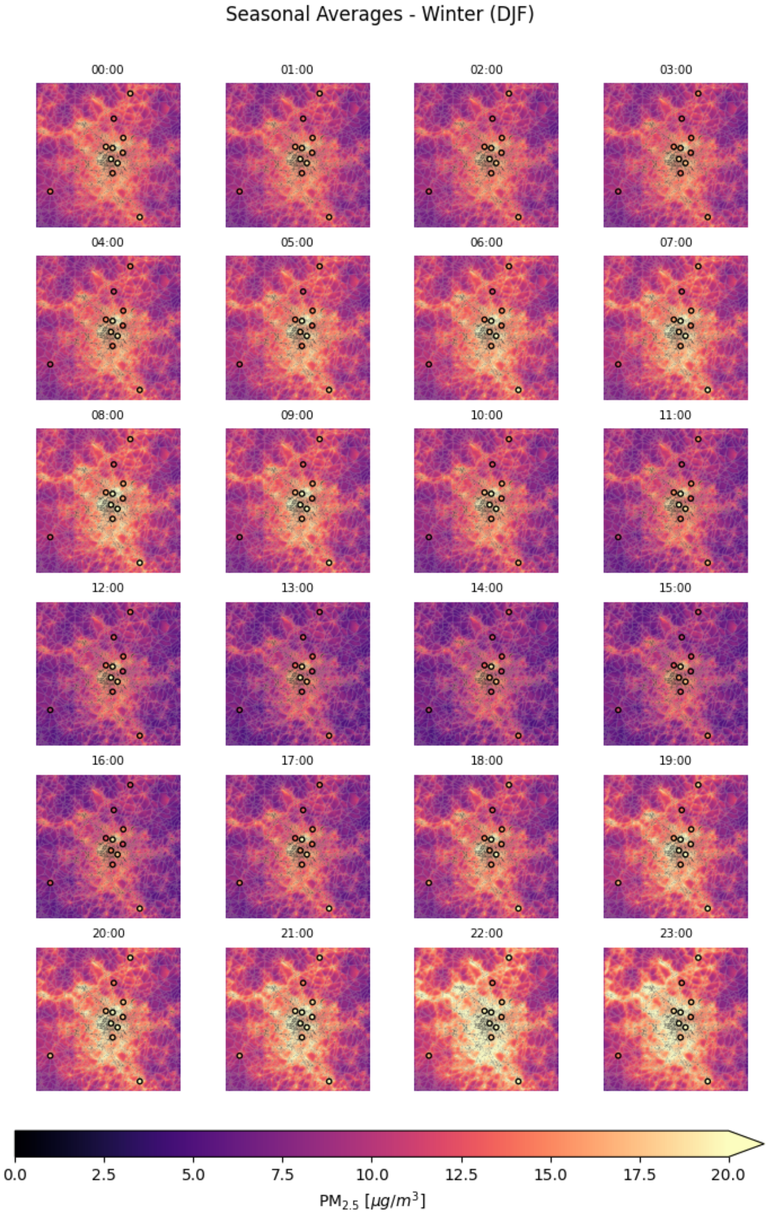


Figure 5. Predicted $PM_{2.5}$ seasonal averages maps by hour for the 2019 winter season (December, January and February) on the city of Paris. The dots represent ground stations measurements. These plots are obtained considering the model trained leaving out station 1.

https://doi.org/10.5194/egusphere-2024-4056
Preprint. Discussion started: 14 February 2025

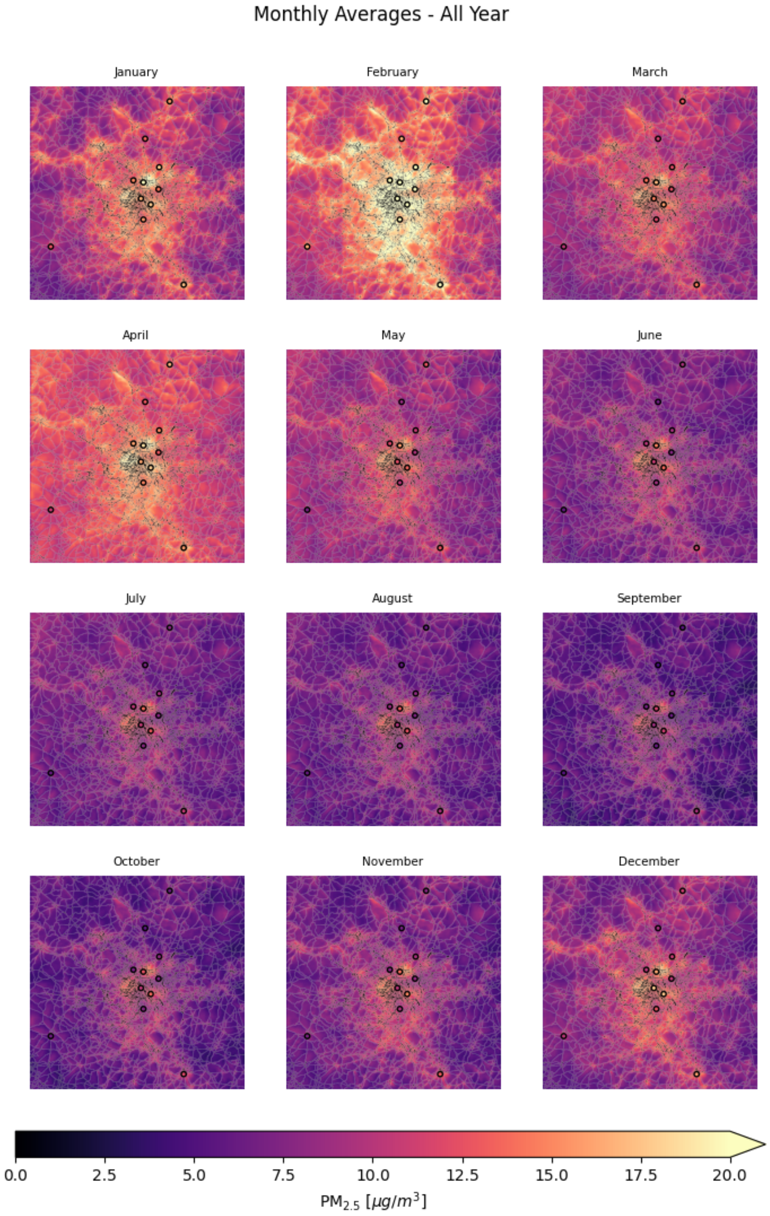


Figure 6. Predicted $PM_{2.5}$ monthly averages maps for the year 2019 on the city of Paris. The dots represent ground stations measurements. These plots are obtained considering the model trained leaving out station 1.

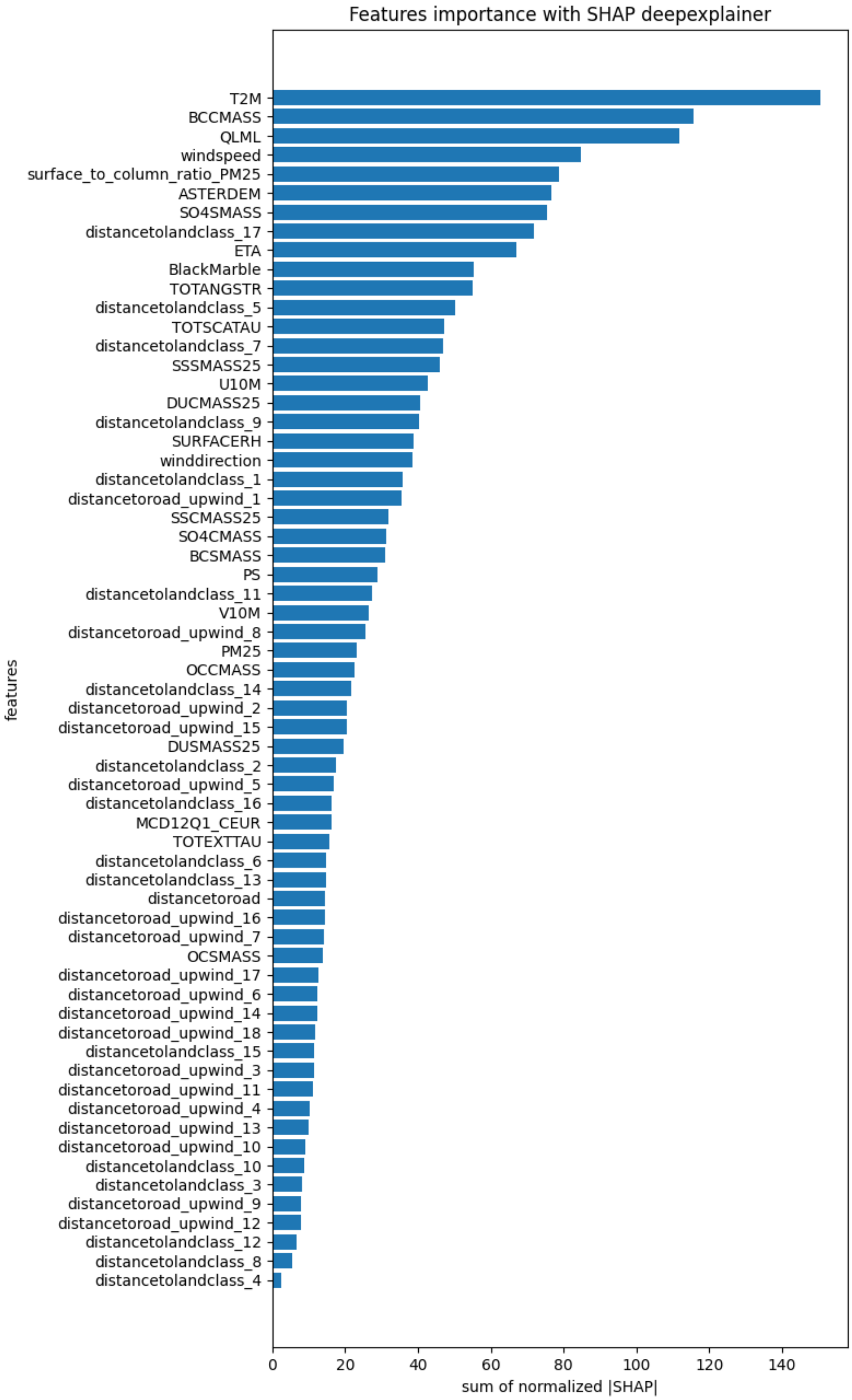


Figure 7. Feature importance calculated as sum of the normalized absolute SHAP values for predictions at station 1.

retrievals have high temporal resolution, we could also combine low orbiting instruments and geostationary instruments to integrate all the available information both on the spatial and temporal dimensions. Further, instead of relying solely on data, we could introduce physical constraints in our loss function, pushing the model training to the space of physical solutions assisted by differential equations. Physics Informed Neural Networks (PINNs) (Raissi et al., 2019) have shown promising results in many areas of science and they can be a practical solution to achieve a deep learning based assimilation model at fine scale. It is also important to address the issue of data imbalance. The $\mathrm{PM}_{2.5}$ distribution in the training data is inherently skewed toward lower values, posing a common challenge in training models. Additionally, satellite retrieval maps (NOODLESALAD $\mathrm{PM}_{2.5}$) are more susceptible to cloud cover during winter, causing seasonal imbalance to the training data. Expanding the dataset to include more locations and years could help mitigate these issues and improve model performance. In conclusion, given the encouraging results and possible future developments, we believe our methodology could be relevant for $\mathrm{PM}_{2.5}$ related exposure and regulation studies at finer (suburban level) scale.

U10M	V10M	PS
T2M	SO4SMASS	OCSMASS
BCSMASS	DUSMASS25	SSSMASS25
SO4CMASS	OCCMASS	BCCMASS
DUCMASS25	SSCMASS25	TOTEXTTAU
TOTANGSTR	TOTSCATAU	QLML
SURFACERH	PM25	surface_to_column_ratio_PM25
ETA	winddirection	windspeed
ASTERDEM	BlackMarble	distancetoroad
distancetoroad_upwind_1	distancetoroad_upwind_2	distancetoroad_upwind_3
distancetoroad_upwind_4	distancetoroad_upwind_5	distancetoroad_upwind_6
distancetoroad_upwind_7	distancetoroad_upwind_8	distancetoroad_upwind_9
distancetoroad_upwind_10	distancetoroad_upwind_11	distancetoroad_upwind_12
distancetoroad_upwind_13	distancetoroad_upwind_14	distancetoroad_upwind_15
distancetoroad_upwind_16	distancetoroad_upwind_17	distancetoroad_upwind_18
distancetolandclass_1	distancetolandclass_2	distancetolandclass_3
distancetolandclass_4	distancetolandclass_5	distancetolandclass_6
distancetolandclass_7	distancetolandclass_8	distancetolandclass_9
distancetolandclass_10	distancetolandclass_11	distancetolandclass_12
distancetolandclass_13	distancetolandclass_14	distancetolandclass_15
distancetolandclass_16	distancetolandclass_17	landclass

Table A1. Table of input features for the model.

305 *Code and data availability.* The OpenAQ data is open data and available for download at https://openaq.org/. The OpenStreetMap data is open data and available for download at https://www.openstreetmap.org/. All the NASA data (MERRA-2, MODIS, ASTER DEM) used in this work is open data and can be found and downloaded using the NASA Earthdata Search website at https://www.earthdata.nasa.gov/. The NASA Black Marble Night Lights data is available at https://blackmarble.gsfc.nasa.gov/. Data (including NOODLESALAD $PM_{2.5}$) and code will be available from the authors on a reasonable request.

310 Appendix A: Table with input features and lists of variables used from datasets

A1 MERRA-2 variables

We use the following meteorology-related variables from the MERRA-2 `M2T1NXSLV` dataset:

- **PS**: surface pressure (Pa)

- **T2M**: 2-meter air temperature (K)
- **U10M**: 10-meter eastward wind (m / s)
- **V10M**: 10-meter northward wind (m / s)

We use the following meteorology-related variables from the MERRA-2 `M2T1NXFLX` dataset:

- **QLML**: surface specific humidity (1)

We use the following aerosol and air quality related variables from the MERRA-2 `M2T1NXAER` dataset:

- **BCCMASS**: Black Carbon Column Mass Density (kg m^{-2})
- **BCSMASS**: Black Carbon Surface Mass Concentration (kg m^{-3})
- **DUCMASS25**: Dust Column Mass Density - PM 2.5 (kg m^{-2})
- **DUSMASS25**: Dust Surface Mass Concentration - PM 2.5 (kg m^{-3})
- **OCCMASS**: Organic Carbon Column Mass Density (kg m^{-2})
- **OCSMASS**: Organic Carbon Surface Mass Concentration (kg m^{-3})
- **SO4CMASS**: SO4 Column Mass Density (kg m^{-2})
- **SO4SMASS**: SO4 Surface Mass Concentration (kg m^{-3})
- **SSCMASS25**: Sea Salt Column Mass Density - PM 2.5 (kg m^{-2})
- **SSSMASS25**: Sea Salt Surface Mass Concentration - PM 2.5 (kg m^{-3})
- **TOTANGSTR**: Total Aerosol Angstrom parameter [470-870 nm] (1)
- **TOTEXTTAU**: Total Aerosol Extinction AOT [550 nm] (1)
- **TOTSCATAU**: Total Aerosol Scattering AOT [550 nm] (1)

A2 OpenStreetMap road types used to compute the distance to the closest road

We use the following road types to compute the distance to the closest road. The descriptions of the road types are obtained from OpenStreetMap (2023).

- **motorway**: A restricted access major divided highway, normally with 2 or more running lanes plus emergency hard shoulder. Equivalent to the Freeway, Autobahn, etc.
- **trunk**: The most important roads in a country’s system that aren’t motorways.

- **primary**: The next most important roads in a country's system.
- **secondary**: The next most important roads in a country's system.
- **tertiary**: The next most important roads in a country's system.
- **motorway_link**: The link roads (sliproads/ramps) leading to/from a motorway from/to a motorway or lower class highway. Normally with the same motorway restrictions.
- **trunk_link**: The link roads (sliproads/ramps) leading to/from a trunk road from/to a trunk road or lower class highway.
- **primary_link**: The link roads (sliproads/ramps) leading to/from a primary road from/to a primary road or lower class highway.
- **secondary_link**: The link roads (sliproads/ramps) leading to/from a secondary road from/to a secondary road or lower class highway.
- **tertiary_link**: The link roads (sliproads/ramps) leading to/from a tertiary road from/to a tertiary road or lower class highway.

A3 IGBP land cover types

IGBP classification contains the following land cover types:

- **1**: Evergreen needleleaf forests
- **2**: Evergreen broadleaf forests
- **3**: Deciduous needleleaf forests
- **4**: Deciduous broadleaf forests
- **5**: Mixed forests
- **6**: Closed shrublands
- **7**: Open shrublands
- **8**: Woody savannas
- **9**: Savannas
- **10**: Grasslands
- **11**: Permanent wetlands

- **12**: Croplands
- **13**: Urban and built-up
- **14**: Cropland/natural
- **15**: Snow and ice
- **16**: Barren
- **17**: Water bodies

Author contributions. **AP**: Conceptualization, Methodology, Software, Formal analysis, Writing — Original draft, Visualization **VK**: Conceptualization, Methodology, Formal analysis, Writing — Original draft, Supervision **TL**: Conceptualization, Methodology, Formal analysis, Writing — Original Draft, Supervision **AL**: Conceptualization, Methodology, Software, Formal analysis, Writing — Original draft, Supervision

Competing interests. The authors declare no competing interests.

Acknowledgements. This study was funded by the European Space Agency EO Science for Society programme via the NOODLESALAD project (contract number 4000137651/22/I-DT-lr). The research was also supported by the Research Council of Finland via the Finnish Centre of Excellence of Inverse Modelling and Imaging (project no. 353084), Flagship of Advanced Mathematics for Sensing Imaging and Modelling (grant no. 358944), and research project (grant no. 321761). The authors wish to acknowledge CSC – IT Center for Science, Finland, for computational resources.

Financial support. This research has been supported by the European Space Agency (grant no. 4000137651/22/I-DT-lr) and the Research Council of Finland (grant nos. 353084, 358944, and 321761).

https://doi.org/10.5194/egusphere-2024-4056
Preprint. Discussion started: 14 February 2025

References

Belward, A. S., Estes, J. E., and Kline, K. D.: The IGBP-DIS global 1-km land-cover data set DISCover: A project overview, Photogrammetric Engineering and Remote Sensing, 65, 1013–1020, 1999.

Bessho, K., Date, K., Hayashi, M., Ikeda, A., Imai, T., Inoue, H., Kumagai, Y., Miyakawa, T., Murata, H., Ohno, T., Okuyama, A., Oyama, R., Sasaki, Y., Shimazu, Y., Shimoji, K., Sumida, Y., Suzuki, M., Taniguchi, H., Tsuchiyama, H., Uesawa, D., Yokota, H., and Yoshida, R.: An Introduction to Himawari-8/9 - Japan's New-Generation Geostationary Meteorological Satellites, Journal of the Meteorological Society of Japan. Ser. II, 94, 151–183, https://doi.org/10.2151/jmsj.2016-009, 2016.

Buchard, V., Da Silva, A., Randles, C., Colarco, P., Ferrare, R., Hair, J., Hostetler, C., Tackett, J., and Winker, D.: Evaluation of the surface PM2.5 in Version 1 of the NASA MERRA Aerosol Reanalysis over the United States, Atmospheric Environment, 125, 100–111, 2016.

Chu, Y., Liu, Y., Li, X., Liu, Z., Lu, H., Lu, Y., Mao, Z., Chen, X., Li, N., Ren, M., Liu, F., Tian, L., Zhu, Z., and Xiang, H.: A Review on Predicting Ground PM2.5 Concentration Using Satellite Aerosol Optical Depth, Atmosphere, 7, https://doi.org/10.3390/atmos7100129, 2016.

Cohen, A. J., Brauer, M., Burnett, R., Anderson, H. R., Frostad, J., Estep, K., et al.: Estimates and 25-year trends of the global burden of disease attributable to ambient air pollution: an analysis of data from the Global Burden of Diseases Study 2015, The Lancet, 389, 1907–1918, 2017.

Deng, L.: Estimation of PM2.5 Spatial Distribution Based on Kriging Interpolation, in: Proceedings of the First International Conference on Information Sciences, Machinery, Materials and Energy, pp. 1791–1794, Atlantis Press, https://doi.org/10.2991/icismme-15.2015.370, 2015.

Fujisada, H., Urai, M., and Iwasaki, A.: Advanced methodology for ASTER DEM generation, IEEE transactions on geoscience and remote sensing, 49, 5080–5091, 2011.

Fujisada, H., Urai, M., and Iwasaki, A.: Technical methodology for ASTER global DEM, IEEE Transactions on Geoscience and Remote Sensing, 50, 3725–3736, 2012.

Goodfellow, I., Bengio, Y., and Courville, A.: Deep Learning, MIT Press, http://www.deeplearningbook.org, 2016.

Health Effects Institute: State of global air 2019, 2019.

Inness, A., Ades, M., Agustí-Panareda, A., Barré, J., Benedictow, A., Blechschmidt, A.-M., Dominguez, J. J., Engelen, R., Eskes, H., Flemming, J., Huijnen, V., Jones, L., Kipling, Z., Massart, S., Parrington, M., Peuch, V.-H., Razinger, M., Remy, S., Schulz, M., and Suttie, M.: The CAMS reanalysis of atmospheric composition, Atmospheric Chemistry and Physics, 19, 3515–3556, https://doi.org/10.5194/acp-19-3515-2019, 2019.

Keller, C. A., Knowland, K. E., Duncan, B. N., Liu, J., Anderson, D. C., Das, S., Lucchesi, R. A., Lundgren, E. W., Nicely, J. M., Nielsen, E., Ott, L. E., Saunders, E., Strode, S. A., Wales, P. A., Jacob, D. J., and Pawson, S.: Description of the NASA GEOS Composition Forecast Modeling System GEOS-CF v1.0, Journal of Advances in Modeling Earth Systems, 13, e2020MS002413, https://doi.org/https://doi.org/10.1029/2020MS002413, e2020MS002413 2020MS002413, 2021.

LeCun, Y., Boser, B., Denker, J., Henderson, D., Howard, R., Hubbard, W., and Jackel, L.: Handwritten Digit Recognition with a Back-Propagation Network, in: Advances in Neural Information Processing Systems, edited by Touretzky, D., vol. 2, Morgan-Kaufmann, https://proceedings.neurips.cc/paper_files/paper/1989/file/53c3bce66e43be4f209556518c2fcb54-Paper.pdf, 1989.

Levy, R. C., Mattoo, S., Munchak, L. A., Remer, L. A., Sayer, A. M., Patadia, F., and Hsu, N. C.: The Collection 6 MODIS aerosol products over land and ocean, Atmospheric Measurement Techniques, 6, 2989–3034, https://doi.org/10.5194/amt-6-2989-2013, 2013.

Lipponen, A., Reinvall, J., Väisänen, A., Taskinen, H., Lähivaara, T., Sogacheva, L., Kolmonen, P., Lehtinen, K., Arola, A., and Kolehmainen, V.: Deep-learning-based post-process correction of the aerosol parameters in the high-resolution Sentinel-3 Level-2 Synergy product, Atmospheric Measurement Techniques, 15, 895–914, 2022.

Loveland, T. R. and Belward, A.: The international geosphere biosphere programme data and information system global land cover data set (DISCover), Acta Astronautica, 41, 681–689, 1997.

Lundberg, S. M. and Lee, S.-I.: A Unified Approach to Interpreting Model Predictions, in: Advances in Neural Information Processing Systems 30, edited by Guyon, I., Luxburg, U. V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., and Garnett, R., pp. 4765–4774, Curran Associates, Inc., http://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions.pdf, 2017.

Michaelides, S., Lane, J., and Kasparis, T.: Effect of Vertical Air Motion on Disdrometer Derived Z-R Coefficients, Atmosphere, 10, 77, 2019.

NASA/METI/AIST/Japan Spacesystems, and US/Japan ASTER Science Team: ASTER Global Digital Elevation Model V003, distributed by NASA EOSDIS Land Processes DAAC, 2019.

OpenStreetMap: OpenStreetMap Wiki - Key:highway, https://wiki.openstreetmap.org/wiki/Key:highway, [Online; accessed 13-April-2023], 2023.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Köpf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., and Chintala, S.: PyTorch: An Imperative Style, High-Performance Deep Learning Library, https://arxiv.org/abs/1912.01703, 2019.

Pope, C. A. I. and Dockery, D. W.: Health Effects of Fine Particulate Air Pollution: Lines that Connect, Journal of the Air & Waste Management Association, 56, 709–742, https://doi.org/10.1080/10473289.2006.10464485, 2006.

Porcheddu, A., Kolehmainen, V., Lähivaara, T., and Lipponen, A.: Post-process correction improves the accuracy of satellite $PM_{2.5}$ retrievals, Atmospheric Measurement Techniques, 17, 5747–5764, https://doi.org/10.5194/amt-17-5747-2024, 2024.

Raissi, M., Perdikaris, P., and Karniadakis, G.: Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations, Journal of Computational Physics, 378, 686–707, https://doi.org/https://doi.org/10.1016/j.jcp.2018.10.045, 2019.

Randles, C., Da Silva, A., Buchard, V., Colarco, P., Darmenov, A., Govindaraju, R., Smirnov, A., Holben, B., Ferrare, R., Hair, J., et al.: The MERRA-2 aerosol reanalysis, 1980 onward. Part I: System description and data assimilation evaluation, Journal of climate, 30, 6823–6850, 2017.

Ronneberger, O., Fischer, P., and Brox, T.: U-Net: Convolutional Networks for Biomedical Image Segmentation, 2015.

Shi, X., Chen, Z., Wang, H., Yeung, D.-Y., kin Wong, W., and chun Woo, W.: Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting, https://arxiv.org/abs/1506.04214, 2015.

Sulla-Menashe, D. and Friedl, M. A.: User guide to collection 6 MODIS land cover (MCD12Q1 and MCD12C1) product, USGS: Reston, VA, USA, 1, 18, 2018.

Tang, D., Zhan, Y., and Yang, F.: A review of machine learning for modeling air quality: Overlooked but important issues, Atmospheric Research, 300, 107 261, https://doi.org/https://doi.org/10.1016/j.atmosres.2024.107261, 2024.

Wang, Z., Shrestha, R., and M.O., R.: VIIRS/NPP Lunar BRDF-Adjusted Nighttime Lights Yearly L3 Global 15 arc second Linear Lat Lon Grid [data set], https://doi.org/10.5067/VIIRS/VNP46A4.001, 2020.

Özgün Çiçek, Abdulkadir, A., Lienkamp, S. S., Brox, T., and Ronneberger, O.: 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation, 2016.