ENVIRONMENTAL RESEARCH
COMMUNICATIONS

PAPER • OPEN ACCESS

# Deep learning techniques applied to super-resolution chemistry transport modeling for operational uses

To cite this article: B Bessagnet *et al* 2021 *Environ. Res. Commun.* **3** 085001

View the article online for updates and enhancements.

You may also like

- A Comprehensive Study of a Film Maturation Process for Improving the Cycle Life of Silicon-Based Anodes
Cuauhtemoc Reale Hernandez, Zouina Karkar, Alix Tranchot et al.

- Synthesis and Electrochemical Studies of Ionic Liquid Electrolytes for Lithium-Air Batteries
Yuji Ono, Md Mijanur Rahman, Byambasuren Delgertsetseg et al.

- Impact of Micro-Porous Layer Properties on Fuel Cell Performance
Mehdi Andisheh-Tadbir, Zahra Tayarani, Mohamed El Hannach et al.

This content was downloaded from IP address 90.59.102.73 on 18/08/2022 at 09:49

Environmental Research Communications

PAPER

# Deep learning techniques applied to super-resolution chemistry transport modeling for operational uses

B Bessagnet[1,2,4], M Beauchamp[3,4], L Menut[1], R Fablet[3], E Pisoni[2] and P Thunis[2]

[1] LMD/IPSL, École Polytechnique, Institut Polytechnique de Paris, ENS, PSL Université, Sorbonne Université, CNRS, 91128 Palaiseau, France
[2] European Commission, Joint Research Centre (JRC), Ispra, Italy
[3] IMT Atlantique, Lab-STICC UMR CNRS, 655 Avenue du Technopôle, 29280 Plouzané, France
[4] co-first authors.

**E-mail:** bertrand.bessagnet@lmd.ipsl.fr **and** bertrand.bessagnet@ec.europa.eu

**Keywords:** Air pollution, deep learning, multi-layer perceptron, convolutional neural network, modeling

OPEN ACCESS

RECEIVED
20 April 2021

REVISED
21 July 2021

ACCEPTED FOR PUBLICATION
27 July 2021

PUBLISHED
4 August 2021

Original content from this work may be used under the terms of the Creative Commons Attribution 4.0 licence.

Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

## Abstract

Air quality modeling tools are largely used to assess air pollution mitigation and monitoring strategies. While neural networks (NN) were mostly developed based on observations to derive statistical models at stations, the use of Eulerian chemistry transport models (CTMs) was mainly devoted to air quality predictions over large areas and the evaluation of emission reduction strategies. In this study, we investigate deep learning architectures to create a metamodel of the process oriented CTM CHIMERE and significantly reduce the computing times required for super-resolution simulations. The key point is the selection of input variables and the way to implement them in the NN. We perform a quantitative evaluation of the proposed approaches on a real case-study. The best NN architecture displays very good performances in terms of prediction of pollutant concentrations observed at stations with respect to the raw super-resolution CHIMERE simulation, with a correlation coefficient above 0.95. The best NN is also able to display better performances when compared to observations than the raw high resolution simulation. Currently the model is designed to be used for air quality forecasting and requires improvement for the definition of air quality management strategies.

## 1. Introduction

Currently, about 55% of the world's population lives in urbanized areas, and this number is expected to increase by 68% by 2050 [1]. As mentioned in the most recent European Environment Agency annual report, around 25% of the European urban population is exposed to air quality exceeding the European Union air quality standards, and air pollution is the leading preventable risk factors for premature death in Europe, being responsible for 400,000 deaths each year [2]. The simulation of Air Quality in urban areas remains a major challenge, in particular to assess the exposure of citizens to pollutants, identify the sources of pollution and adapt the strategies to lower the air pollutant concentrations. Operational modeling tools are expected to be more robust and computationally-efficient to quickly simulate air quality and propose adequate measures to monitor, curb and control air pollution. At the regional scale, operational forecasting systems mainly rely on deterministic models that perform simulations with average resolutions of 10 km in the case of a European domain and 3km to 1 km for a regional to a urban domain. A chemistry transport model like CHIMERE [3] is suitable to work at such resolutions. This type of models is used in well-known platforms such as the COPERNICUS ensemble forecast [4], the French national forecast PREV'AIR [5], or the regional forecasts of air quality monitoring associations in France such as Airparif for the Paris region [6].

Machine learning techniques have also emerged as relevant solutions to forecast air quality at stations and using observations (meteorology, concentrations, emissions, landcover, etc.) as predictor variables [7–15]. Recently, deep learning schemes based on neural networks (NN) [16] has become more and more popular with increasing computer power and training data availability. Novel approaches using wavelet artificial network

© 2021 The Author(s). Published by IOP Publishing Ltd

techniques have been recently used for short term air quality forecasting [17]. They seem now relevant and possible to use in the air quality community. These studies show the need to evaluate: (i) the influence of the length of training data utilized on the overall NN model performance, (ii) the significance of the selected predictors and utilized model structure on the complexity and overall NN model performance, (iii) the links between the selected data normalization scheme and transfer function used and (iv) the influence of the initialization schemes for the weighting, bias and other training parameters on the overall NN model performances.

Some very recent works propose to embed NN techniques within CTMs, the objective being to get the best of state-of-the-art physics on-board CTMs and NN approaches [18, 19]. For instance, in [18], neural network schemes were used and evaluated to emulate process-oriented modeling outcomes. Especially, they designed a simple recurrent 3-layer NN to reproduce daily mean concentrations of some pollutants over Europe as predicted by the CMAQ (Community Multiscale Air Quality model) model. They found the trained NN may estimate air pollutant concentrations several orders faster than the original model and with reasonably small errors. They designed a simple recurrent 3-layer NN to reproduce daily mean concentrations of some pollutants over Europe as simulated by the Community Multiscale Air Quality model (CMAQ). Convolutional Neural Networks (CNN) have also been recently used for bias corrections to improve a 7 days air quality forecast issued from a chemistry transport model [20]. Other recent works showed the possible use of NN to act as drop-in chemical solvers with orders-of-magnitude performance gain [21, 22], though error propagation over long time periods needs to be addressed.

Following these previous works, this study proposes a novel approach to design a NN-based emulator of high-resolution air pollution simulations which are highly time-consuming. In [23], they showed that a very simple regression with an adequate selection of input variables may relevantly emulate a CTM and mimic the behavior with identical performances. Though, this preliminary study was made on daily datasets over a large period. Here, we introduce a new CTM-NN strategy and base the training on a shorter period with hourly data. Compared with previous works, the novelty of our approach lies in the input data selection and the designed NN architectures. The use of an hourly dataset is also new compared to the previous study allowing a shorter period for the training phase and get a better representation of the evolution of dispersion conditions. We report a quantitative evaluation on a representative test case, which demonstrates the relevance of the proposed CTM-NN approach. At this stage an evaluation on $PM_{10}$, $PM_{2.5}$ and $NO_2$, ozone is not addressed and it will be in a follow-up study by introducing more physics in the NN approach. We further discuss possible applications and improvement for operational uses.

## 2. Methods

### 2.1. The original principles

In a previous study [23], a metamodel of the Chemistry Transport Model CHIMERE was realized to downscale a low $0.5^o \times 0.25^o$ horizontal resolution simulation to a higher resolution of $0.1^o \times 0.05^o$. Using of air quality models is highly computationally-demanding mainly because of the CFL (Courant-Friedrichs-Lewy) condition. This condition is mandatory for convergence when partial differential equations are solved by finite differencing. The CFL for a grid box as presented in equation (1) is a ratio between the mean wind speed $v_{x_i}$, the time-step $\Delta t$ and the cell size in all three directions $i$ as $\Delta x_i$. More the wind speed is high, more the time-step has to decrease. The wind speed varying every minute, the best way to optimize the numerical cost of a simulation is to adapt the time-step.

$$CFL = \Delta t \sum_{i=1}^{3} \frac{v_{x_i}}{\Delta x_i} \quad (1)$$

The first version of this methodology was based on a training process over a 6-months period and applied over the next 6 months. This simple and straightforward methodology provided means to assess the extent to which the proposed approach is able to capture the main spatial and temporal patterns of the main pollutants daily concentrations. The gain in computing time was very impressive because the costly step including all physics and chemical processes was by-passed. The assessment of performances observations were close to those obtained with the raw model high resolution simulations with CHIMERE. In the following, uppercase letters refer to variables at low/coarse resolution (LR) and lowercase for the high resolution simulation (HR). The bases of the methodology is described hereafter.

In [23, 24], a given high resolution grid cell was assumed to behave as a 'city'. Based on the atmospheric diffusion theory, a simple methodology have been hypothesized to evaluate urban increments of concentrations due to a city [25]. Thus, under neutral atmospheric conditions, the vertical diffusion of a non-reactive pollutant from a continuous point source is described in a general form through the relationship as follow in equation (2)

assuming a Gaussian dispersion in a box model approach :

$$\sigma_z^2 = \frac{2kx}{u} \tag{2}$$

with $\sigma_z^2$ indicating the variance of the vertical diffusion after a distance $x$ from the source, $k$ as the eddy diffusivity and $u$ as the wind speed.

As described in [23, 24], a generalization of this equation has been hypothetized to evaluate the concentration difference $\Delta c$ between a fine grid and a coarse grid simulation of a primary pollutant $p$ concentration influenced by low level sources of primary pollutants which can be finally expressed as:

$$\begin{aligned} c &= C + \Delta c \\ \Delta c &= \alpha \frac{1}{\sqrt{ku}}(e\sqrt{d} - E\sqrt{D}) + \beta \\ d &\propto \sqrt{\delta x \delta y} \\ D &\propto \sqrt{\delta X \delta Y} \end{aligned} \tag{3}$$

- $c$ and $C$ ($\mu$g m$^{-3}$) are the concentrations at the high resolution (HR) and the low resolution (LR), respectively. All concentrations are interpolated from the coarse grid to the fine grid.
- $k$ is the vertical mixing coefficient (m$^2$ s$^{-1}$) over the fine mesh.
- $u$ is the 10 m horizontal wind speed (m s$^{-1}$) over the fine mesh.
- $\delta$X, $\delta$Y, $\delta$x, $\delta y$ are the coarse longitude, latitude and the fine longitude, latitude increments of the meshes (in degree), respectively. In this study they are constant, but we have considered them, because for other domains they can vary.
- $d$ and $D$ are the characteristic lengths respectively derived from the previous parameters for the fine and coarse meshes.
- $e$ and $E$ ($\mu$g m$^{-2}$ s$^{-1}$) are defined as the high and coarse resolution low-level emission fluxes interpolated over the fine grid. The sum of primary $PM_{2.5}$ and $PM_{10}$ emissions is considered while for NO and $NO_2$ the total NOx emissions is taken into account. The emissions of the two first levels (approximately below 30 m) are considered.
- $\alpha$ and $\beta$ are regression coefficients. They embed indirectly missing geographical, physical and chemical processes, partly lost during the simplification process. It is noteworthy that $\beta$ here has not the same meaning than in [24, 25], it represents here a residual value issued from the regression method and is expected close to 0.

An illustration of this concept is displayed in figure 1. As explained in the next sections, the previous equations list the relevant variables which are selected as inputs for the NN.

### 2.2. CHIMERE ouputs as input for the Neural Networks

For this new study, the CHIMERE configuration used to create the input data for our neural network strategies is briefly described here, and the reader can refer to the usual reference CHIMERE publications [3, 26–28] for details. Particulate matter includes primary particulate matter, secondary inorganic species such as nitrate, sulfate, ammonium and organic aerosol resulting from the oxidation of anthropogenic and biogenic precursors. Gas-particle partitioning of the condensable oxidation products [28] is also taken into account. Biogenic emissions are computed with the model MEGAN version 2.1 [29]. Sea-salt and mineral dust emissions issued from desert and agricultural areas are also implemented with recent parameterizations. Particle sizes range from 10nm to 40 $\mu$m over 10 bins.

The WRF simulation (WRF 3.7.1 version) used for the meteorology is nudged with NCEP (National Centers for Environmental Prediction) final analysis from GFS (Global Forecast System) meteorological fields at $1^o \times 1^o$. The 6-hourly time resolution version is used at the coarsest model initial and domain boundaries (ds083.2 dataset [30]).

For this study, the set-up and domains are rigorously the same depicted in a previous work aiming at simulating the air quality at fine resolution over the French Alps [31] providing CTM outputs for the meteorology and concentrations fields. Two different domains are defined as: (i) EUR01 covering a large part of Western Europe and (ii) over ALP0033 for the whole French Alps with detailed characteristics are detailed in table 1. The Alps domain (APL0033) encompasses an area from the Lyon municipality on the West part to the

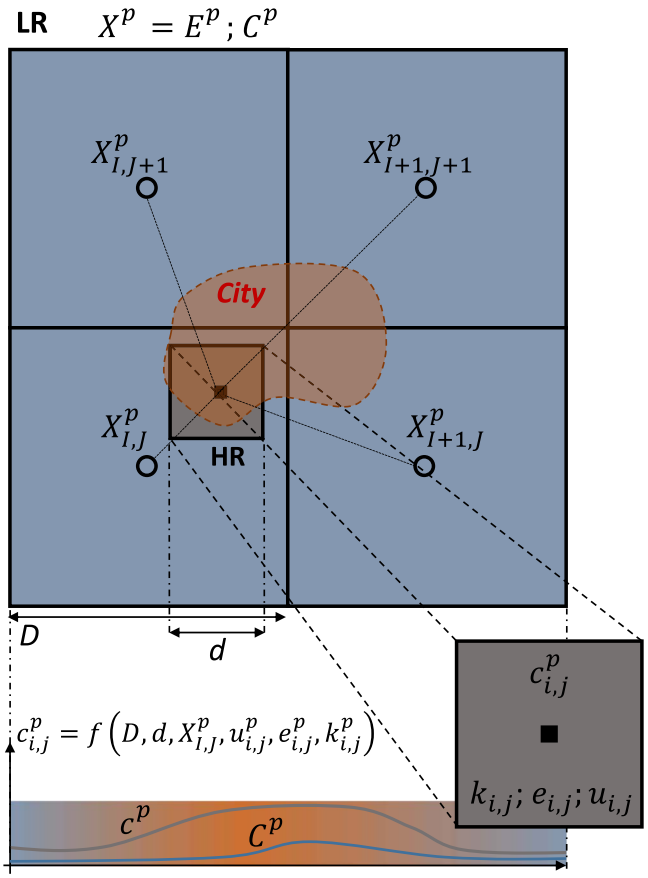



**Figure 1.** Illustration of the relation between high and low resolutions of all variables with the Δ concept. The presence of a city is responsible of an increase $\Delta c$ of primary pollutant concentrations enhanced by local meteorology and emissions. The coarse resolution variables C and E as used in input for the NN are reinterpolated over the fine mesh APL0033 here.

**Table 1.** Domain specifications for the Boundary Conditions (BOUN) and Initialization (INIT).

| Domain Name | EUR01 | ALP0033 |
|---|---|---|
| Area | Western Europe | French Alps |
| Coverage | 11.85°W-33.25°N/33.85°E-60.85°N | 4.417°E-44.317°N/7.783°E-46.583°N |
| Number of grid points | 459 × 278 | 103 × 70 |
| Resolution | 1/10°×1/10° | 1/30° × 1/30° |
| Number of CHIMERE levels | 15 | 20 |
| Chemical BOUN-INIT | LMDzINCA GOCART | EUR01 |
| Met. BOUN-INIT | NCEP/GFS | EUR01 |
| Met. Nudging | Yes | No |

Leman Lake on the north to the Piemonte region in Italy on the East Part with a resolution of about 3 km (figure 2). Grenoble and the Arve valley from Geneva to Chamonix are known to be air pollution hot spots in France due to their location in deep valleys with frequent stagnant cold meteorological conditions in wintertime. The HR simulation is performed over the ALP0033 and the LR simulation over the EUR01 domain and interpolated over the ALP0033 so that the ALP0033 mesh will be the working grid for the neural networks. The simulation was performed from 2013-11-15 01:00 UTC to the 2021-12-21 00:00 UTC with a spin-up period from the begin of November to ensure a good initialization. Therefore 864 hours are available to train (432 hours) and evaluate (432 hours) the various neural network strategies. The available observation dataset for the model evaluation is reported in [31] including only rural and urban background stations (figure 2).

### 2.3. NN-based architecture in this study

The basic concept of our approach is presented in figure 3. We aim to replace the high resolution simulation HR by a neural-network-based model to save computing time so that we could produce quick scenarios analyses or air quality predictions. The key idea is to exploit the low resolution in the inputs of the NN to convey an

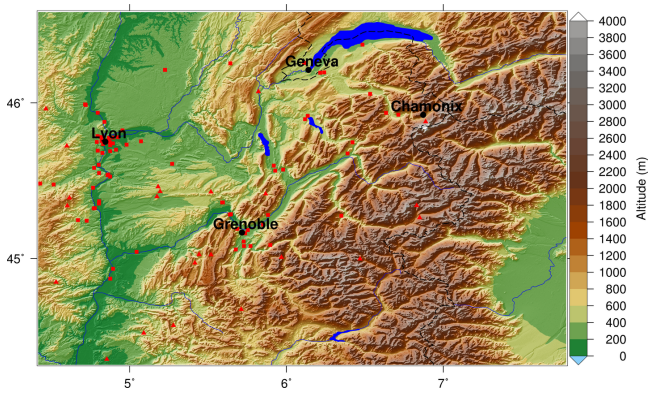



**Figure 2.** The target domain over the Alps (ALP0033). Red triangle and square symbols are the locations of rural and urban air quality monitoring sites used for the evaluation respectively. The dark blue color is used for the inland water bodies.

information on chemistry and long range transport processes that are of major importance for air pollution issues.

In [23], the downscaling from coarse/low (LR) to high (HR) resolution is performed by $N$ pixel-grid linear regressions of the increment $\Delta c$ based on equation (3), where $N$ denotes the number of grid cells in the HR target ALP0033. In this work, we consider an extension of this preliminary work with more sophisticated deep learning-based models. We aim to train as $N$ location-specific independent super-resolution (SR) operators $\Phi = \{\Phi_i\}, i = 1, \cdots, N$ where $c_i^\star = \Phi_i(C_i, \Lambda_i)$, $c_i^\star$ and $C_i$ respectively denote the estimation of the high and coarse resolution (this latter been interpolated over the fine mesh ALP0033) in grid cell/pixel $i$, and $\Lambda_i = (d, D, e, E, k, u)_i$ denotes as used in equation (3) the set of additional covariates at the same location over the high resolution mesh ALP0033. The first natural generalization of [23] is to involve independent pixel-grid multi-layers perceptron (MLP) for SR operator $\Phi$, see figure 4, instead of a linear regression. Here, we use a MLP architecture with two hidden layers of 16 and 8 neurons with ReLU (Rectified Linear Unit) activation and a final linear mapping to $c_i^\star$. Instead of constraining the inputs of the network to variable $\frac{1}{\sqrt{ku}}(e\sqrt{d} - E\sqrt{D})$ used in the linear regression, the input data of the MLP with all the potential covariates are namely $C$, $k$, $u$, $d$, $D$, $e$ and $E$ and let the training extract the relevant features through the weight parameters of the first two hidden layers of the MLP. This choice provides more flexibility. In this configuration, the number of parameters for each submodels $\Phi_i$ is 289. In the end, on the global 69x102 ALP0033 domain, the total number of parameters is 2 033 982. The test case here differs from [23] where a 12 months period was used for the training and validation and only the daily values where exploited. In this new study a shorter period of 36 days is used and the training and validation processes are performed on an hourly basis.

Because the working frame is a two-dimensional fields discretized on a regular grid, additional options may be envisaged for SR operator $\Phi$ by considering the use of convolutional neural networks (CNN) [16]. The underlying idea is to exploit the potential spatial relationships within local neighborhoods to ease the learning of $\Phi$ which will be a global SR operator and not an aggregation of $N$ independent pixel-grid SR operators. This framework can be extended to spatio-temporal CNN even though we believe that for this super-resolution task, a spatial formulation of the problem is satisfactory enough. This is still an interesting idea though if the same work has to be achieved on data with high missing data rates, such as remote sensing and/or *in situ* dataset [32]. Here, a simple CNN architecture is first considered in which the inputs are the same covariates used in the MLP, but stacked here in addition of the coarse resolution $C$ as supplementary channels. The CNN architecture comprises a first hidden layer with 128 Conv2D $3 \times 3$ filters + ReLU activation (Rectified Linear Unit) with a batch normalization, followed by a second hidden layer with 64 $5 \times 5$ filters + ReLU activation. The final layer maps the outputs of the second hidden layer to the required HR resolution by a single linear Conv2D $3 \times 3$ filter, see figure 5. The total number of parameters for this CNN is 84 737.

We also draw from the learning-based super-resolution literature to evaluate more sophisticated CNN architecture. Here, we consider the Deep Residual Channel Attention Networks (RCAN) architecture [33], which is among the state-of-the-art deep learning models for the super-resolution of natural images. In this setup, the so-called residual-in-residual (RIR) structures are the elementary building blocks of the deep neural network architecture. RIR blocks are typically made up of several residual groups (RG) with long skip connections, while each residual group itself contains a predefined number of residual cells with short skip connections. The underlying idea lies in the fact that RIR blocks allows to bypass low-frequency information through multiple skip connections, thus allowing the main network to focus on the learning of high-frequency information. The RCAN architecture also involves a channel attention-based mechanism [34, 35] to rescale

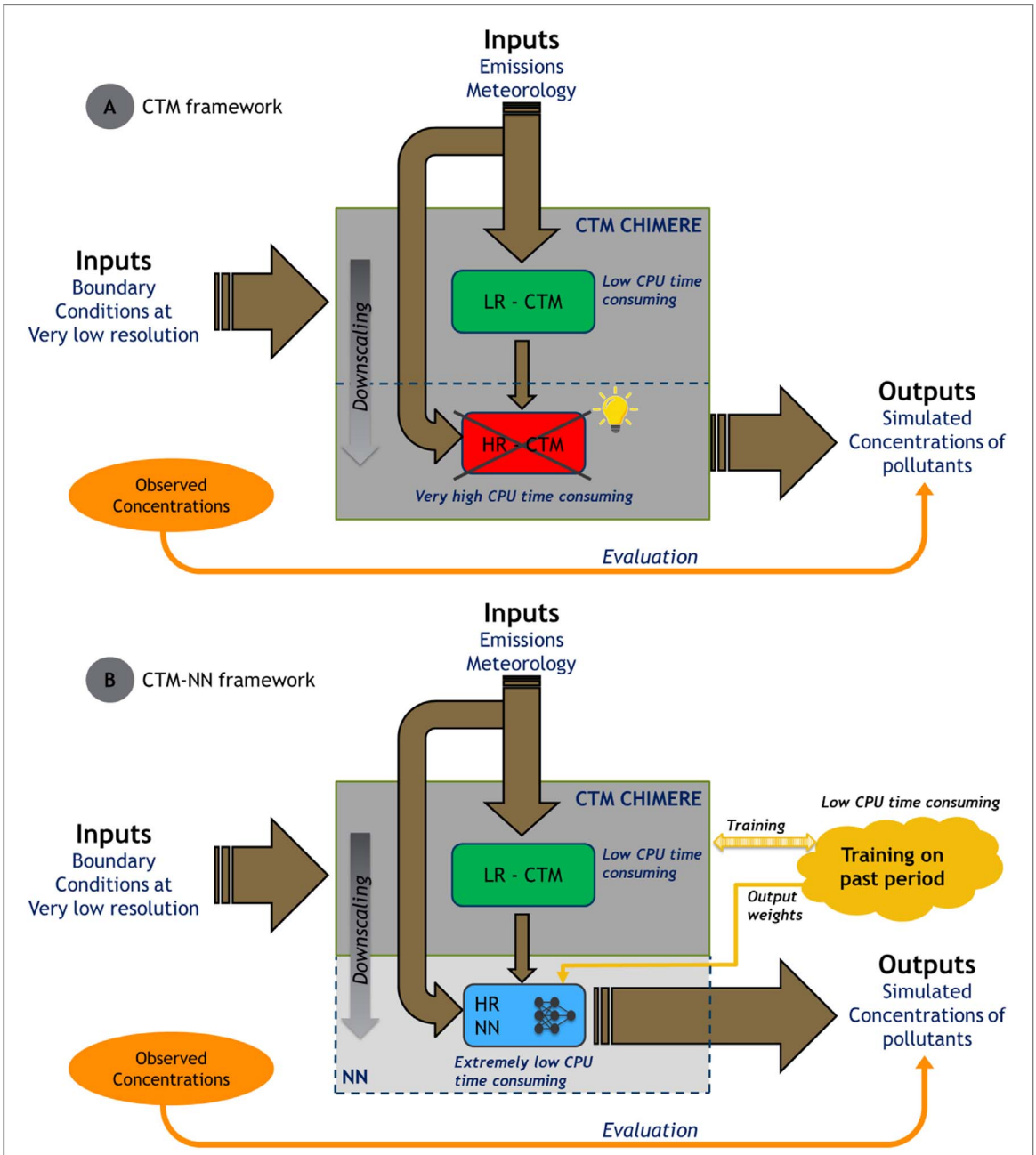



**Figure 3.** Conceptual scheme of the methodology. Panel (A) at the top is the presentation of the usual concept of air quality model with two domains at low and high resolutions (LR and HR respectively). Panel (B) is the new concept using a NN to skip the most time consuming HR-CTM simulation.

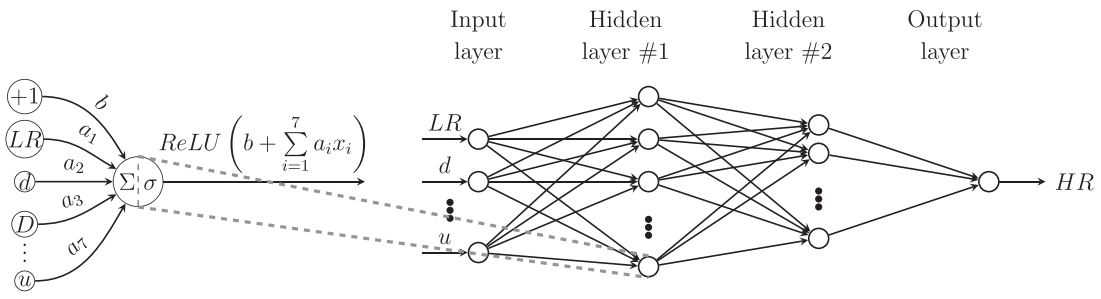



**Figure 4.** Pixel-independent Multi Layer Perceptron (MLP).

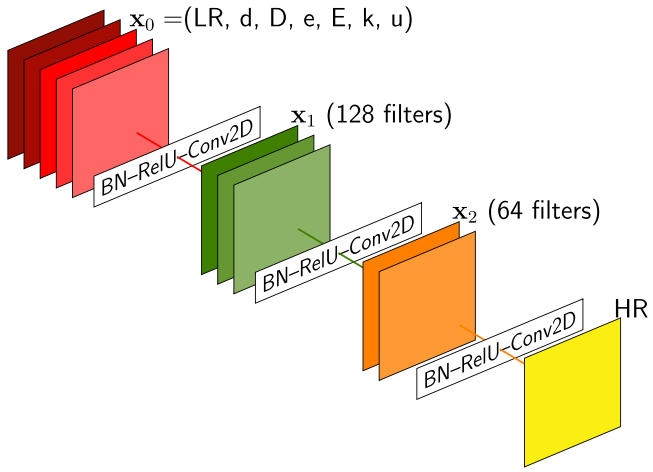



**Figure 5.** Convolutional Neural Network (CNN) architecture.

channel-wise features through interdependencies among channels. The total number of parameters in our RCAN configuration is 299,393.

The so-called back-propagation strategy to calculate the fitting coefficient of the NN is the essence of neural network training. It is the practice of fine-tuning the weights of a neural net based on the error rate (i.e. loss) obtained in the previous epoch (i.e. iteration). Proper tuning of the weights ensures lower error rates, making the model reliable by increasing its generalization. Regarding the training phase, we use for all architectures as training loss function $\mathcal{L}$ the root mean squared error between the 'true' high resolution ($c_i$) and the output of the neural network ($c_i^\star$):

$$\mathcal{L} = \frac{1}{N}\sum_{i=1}^{N}(c_i^\star - c_i)^2$$

It is noteworthy that it could be interesting in future works to assess if other loss functions can be of interest, for instance to constrain the SR operator $\Phi$ to behave better during pollution episodes. Regarding the training strategy, we use an Adam optimizer with a batch size of 4 through 100 epochs on a Microsoft Azure Virtual Machine (VM) powered by NVIDIA Tesla K80 with a GPU memory of 12GiB. The training time of a single MLP is only of 15 seconds, but without any parallelization, the whole training on the 7038 grid cells of the ALP0033 domain is about 26 hours. The same number of epochs is used for MLP, CNN and RCAN, though the RMSE loss function stabilizes between 50 and 100 cycles depending of the architecture. No major overfitting has been observed for this number of epochs though this is the case if increasing the number of cycles up to 200. The CNN-based architectures provide faster training procedures with only 50 minutes for the basic CNN and around 6 hours for RCAN. In the end, if the training time can significantly differ according to the NN architecture, their application on new datasets only takes a few seconds. We remind that it can take several hours for a full computation with a CTM at high resolution depending on the resolution and the domain size. All NN models are implemented using keras framework [36].

## 3. Results

In this section, we report the results obtained for $NO_2$ and $PM_{2.5}$ with the three NN-based super resolution strategies. We use the first half of the dataset for training: 432 hours ranging from 2013-11-15 01:00 UTC to 2013-12-03 00:00 UTC and the other half for validation, ending on 2013-12-21 00:00 UTC. Complementary results are provided in appendix A (figure A1, A2, A3 and A4) for $PM_{10}$ and the evaluation on an hourly basis in tables B1, B2 and B3 of appendix B. Figure C5 of appendix C provides an evaluation on $\Delta c$ for each NN. The definition of evaluation metrics is provided in appendix D.

Typical patterns of high concentrations are observed along the road traffic network and urbanized areas with the original high resolution (HR). $PM_{2.5}$ concentrations maps are smoother since PM emissions are more spread over rural areas particularly due to wood burning and long range transport of such species. Figures 6(a) and 6(b) respectively show the $NO_2$ and $PM_{2.5}$ mean concentrations over the validation period. CHIMERE high-resolution (top left panel) is the target that the super resolution NN operator aims at reconstructing, starting from the coarse resolution (top right panel) and additional covariates as input datasets. On the bottom panel, the maps obtained with the three NN-based super resolution architectures are displayed: from left to right MLP, CNN and RCAN. They all exhibit a significant improvement as compared to the coarse simulation. The main patterns are well reproduced by the three NN

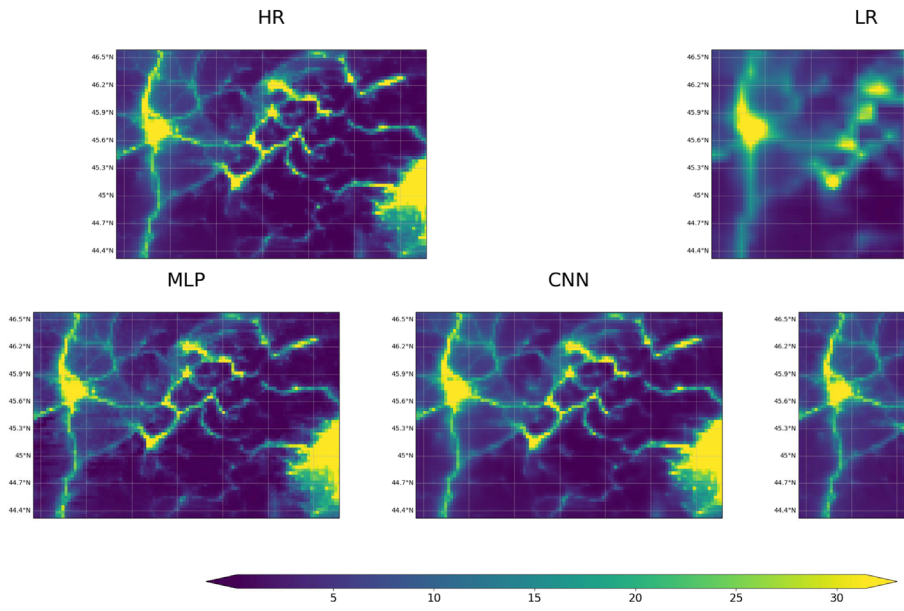



**Figure 6.** $NO_2$ (a) and $PM_{2.5}$ (b) mean concentrations ($\mu g\ m^{-3}$) over the validation period for CHIMERE high-resolution and coarse resolution (top line) and 3 NN-based super resolution architectures (bottom line): from left to right, pixel-based independent multi-layer perceptron (MLP), Convolutional neural network (CNN) and Residual channel attention network (RCAN).

architectures while the coarse resolution acts like a smoother of both $NO_2$ and $PM_{2.5}$ pollution level in these areas of interest.

Because the $PM_{2.5}$ increment of concentrations between the high and coarse resolutions is less easily explained by the high resolution emission covariates, the SRNN applied to $PM_{2.5}$ coarse resolution is slightly less efficient in comparison to the results obtained for $NO_2$. It is especially noticeable in the southeast quarter of the domain where the convolutional-based SRNN $NO_2$ downscaling is very similar to the high resolution. Meanwhile, it is not

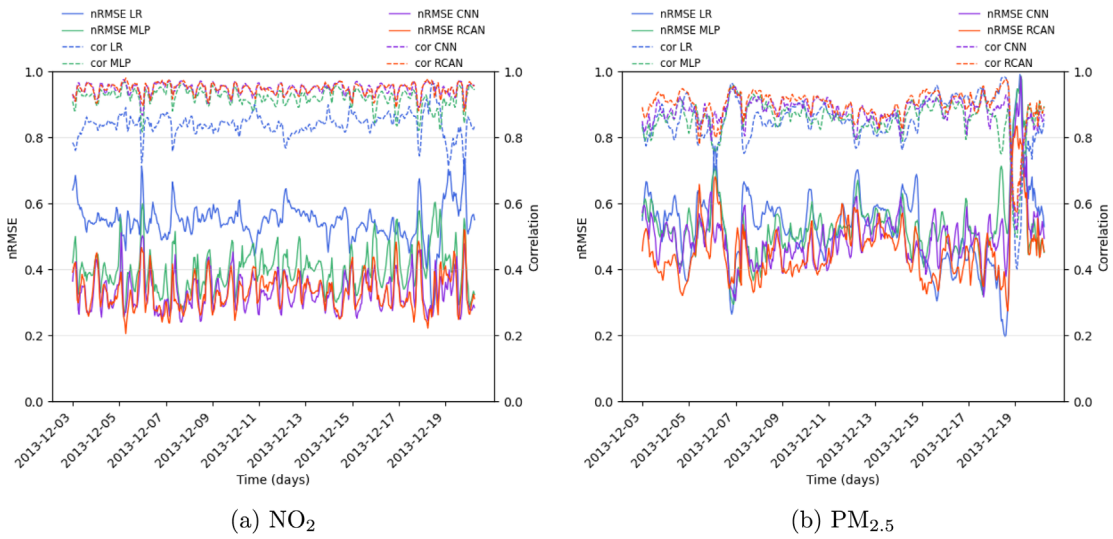



**Figure 7.** $NO_2$ (a) and $PM_{2.5}$ (b) nRMSE and correlation time series (a) of coarse resolution, pixel-based independent multi-layer perceptron (MLP), Convolutional neural network (CNN) and Residual channel attention network (RCAN).

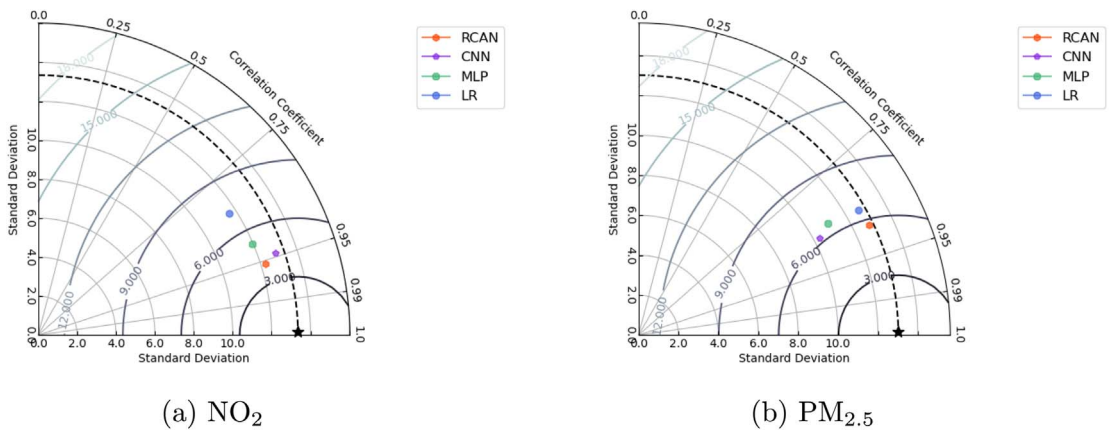



**Figure 8.** $NO_2$ (a) and $PM_{2.5}$ (b) summary statistics over the validation period as Taylor diagrams for low resolution simulation (LR), multi-layer perceptron (MLP), Convolutional neural network (CNN) and Residual channel attention network (RCAN) versus the high resolution (HR) considered as the 'truth'.

straightforward to say if the modifications of the $PM_{2.5}$ coarse resolution proposed by the SRNNs really bring the simulation closer to the high resolution in this area of the ALP0033 domain.

If the gain provided by the super resolution solution presented here is obvious when looking at the maps, we present in figure 7(a) and 7(b) error statistics between the HR simulations and the various neural networks as Normalized Root Mean Square Error as nRMSE (solid lines) and correlation (dashed lines). These statistics are displayed on an hourly basis in order to identify which NN architecture behaves best along the validation period. Each grid point provide an hourly HR 'truth' and SR modeled output. CNN and RCAN architectures seem to increase the already significant gain of the MLP architecture. In terms of correlation, RCAN is slightly better than the basic CNN architecture but the $NO_2$ variability of the latter, see the CNN-based standard deviation on the Taylor diagram provided in figure 8(a), is closer to the HR truth. Correlations are usually higher than 0.95 for $NO_2$ concentrations for the three NN architectures. The hourly frequency results B are similar with again the best performances for the RCAN. It is noteworthy that even the evaluation on $\Delta c$ for each NN (Appendix C) shows the ability of the NN to mimic the HR model for the $\Delta c$ with a correlation up to 0.95 for $NO_2$ by the RCAN approach. Regarding the statistics related to $PM_{2.5}$ concentrations, they all indicate RCAN (see figure 8(b)) is the best super resolution approach among the three architectures evaluated. On this specific pollutant, it is also interesting to note that RCAN is the single super resolution architecture able to deal with the abrupt change of performance of the coarse resolution in the last hours of December 18 (see figure 7(b)): even if the nRMSE and correlation with HR become worse than earlier in the validation period, they are the best

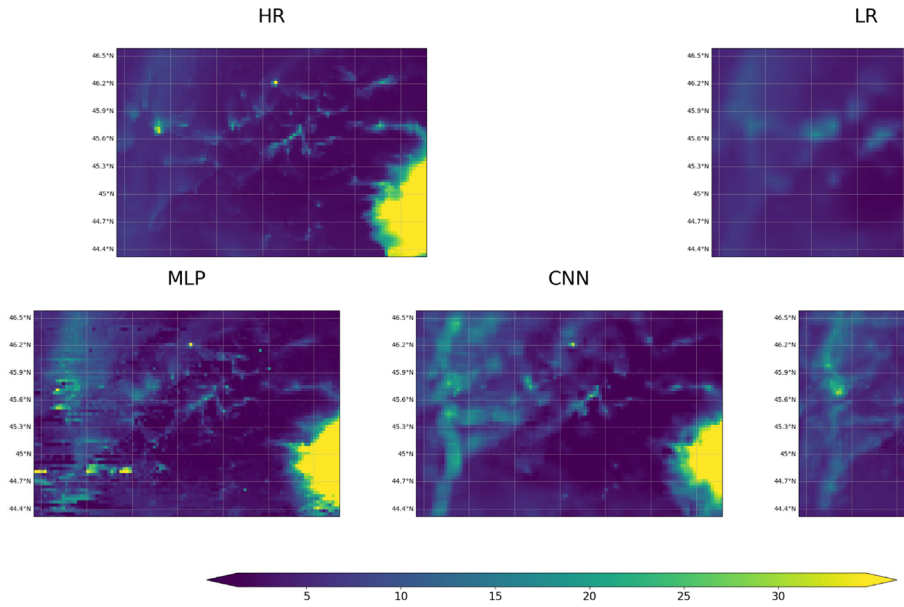
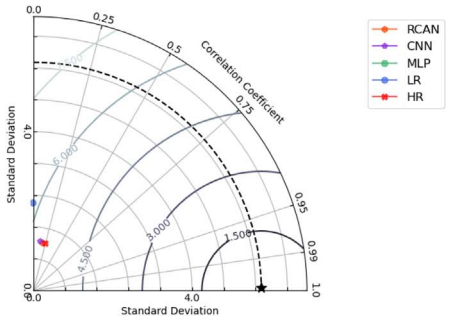



**Figure 9.** $PM_{2.5}$ concentrations on 2013-12-18 18:00 UTC for CHIMERE high-resolution and coarse resolution (top line) and the three NN-based super resolution architectures (bottom line): from left to right, pixel-based independent multi-layer perceptron (MLP), Convolutional neural network (CNN) and Residual channel attention network (RCAN).

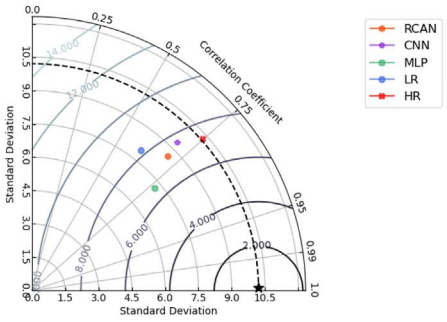
trade-off while both the MLP and in a lesser way the CNN do not capture this singularity in the validation period. The same conclusions hold for $PM_{10}$ (see appendix A). Regarding this specific issue, the $PM_{2.5}$ mapping are also provided in figure 9 on 2013-12-18 18:00 UTC as complementary information. It is clear that the impact of the emission covariates is here largely overestimated in this situation with low $PM_{2.5}$ pollution levels along the roads. This issue directly relates to the learning issue of optimizing the RMSE loss function in average over this training period. It might be more efficient to consider other loss functions and training datasets to address the issue of learning how the HR behaves in specific conditions for operational applications.

As in [23] it is interesting to compare the HR, LR and the three NN simulation outputs with real observation data. For the observations available in the ALP0033 domain along the validation period, we provide in figures 10 and 11 additional Taylor diagrams for the comparison of HR, LR, and the three SRNN to daily-averaged rural and urban observations. Complementary statistics (average bias, RMSE and correlation) are also given in tables 10(c) and 11(c), as well as the original hourly statistics in appendix B: they lead to the same conclusions except that the MLP architecture might improve in terms of RMSE and correlation at the expense of a degraded variability. Regarding the $NO_2$, because the high-resolution behaves relatively poorly on the rural observations with an average correlation lower than 0.25, it is not surprising to see the SRNNs doing the same. For urban areas, we report a greater similarity to the HR ground truth for all NN models with an average correlation of 0.75 for SR-MLP model and slightly lower performance for SR-CNN and SR-RCAN ones (0.70 for the basic CNN and 0.71 for RCAN). Regarding the poor concordance between rural observations and the HR ground truth, it suggests the importance of using data assimilation in high-resolution CTMs. On this particular issue, our SRNN models may be used as fast surrogate simulations in a model-based ensemble data assimilation framework [37–39]. They can also be extended to a fully NN-based data assimilation scheme, where the end-to-end learning strategy consists in using both the coarse resolution (with the HR covariates) and the observations to feed a neural network whose target is the anomaly between the observations and the coarse resolution [32, 40]. Also, these formulations have the advantage of addressing both interpolation, reconstruction and forecasting issues where only a combination of LR and HR covariates, possibly irregularly-sampled, are available.

Performances on $PM_{2.5}$ concentrations for the HR resolution at rural stations is better than for $NO_2$ with an average correlation of 0.70 and RMSE of 5.62 $\mu g\ m^{-3}$; for this pollutant, the CNN-based architectures are even closer to the observations than HR with an improvement of the correlation up to 0.82 with the basic CNN and a similar RMSE with RCAN. Even at urban sites, RCAN behaves better than HR with similar correlations but lower RMSE and biases. This supports the use of such a super resolution approach as surrogate model in operational applications.




(a) All rural




(b) All urban

| Species | Method | Typology | # stations | Obs. | Mod. | Bias | RMSE | Cor. |
|---|---|---|---|---|---|---|---|---|
| $NO_2$ | HR | RUR | 7 | 13.10 | 4.73 | -8.37 | 10.06 | 0.24 |
| | LR | | | | 5.54 | -7.56 | 9.76 | -0.01 |
| | MLP | | | | 4.19 | -8.91 | 10.83 | -0.12 |
| | CNN | | | | 4.81 | -8.29 | 10.07 | 0.13 |
| | RCAN | | | | 4.58 | -8.53 | 10.23 | 0.19 |
| | HR | URB | 40 | 47.83 | 33.37 | -14.46 | 16.19 | 0.75 |
| | LR | | | | 24.45 | -23.39 | 24.79 | 0.61 |
| | MLP | | | | 30.16 | -17.67 | 18.85 | 0.77 |
| | CNN | | | | 31.14 | -16.69 | 18.35 | 0.70 |
| | RCAN | | | | 29.78 | -18.05 | 19.47 | 0.71 |

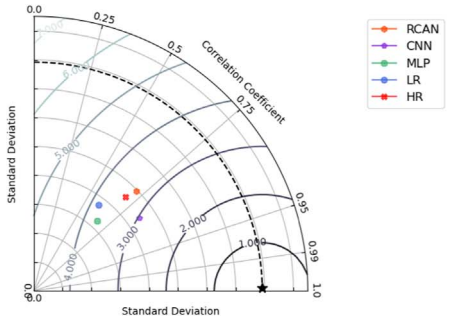
(c) Daily statistics

**Figure 10.** $NO_2$ daily statistics of high resolution (HR), coarse resolution (LR), pixel-based independent multi-layer perceptron (MLP), Convolutional neural network (CNN) and Residual channel attention network (RCAN) for comparison with all rural observations (a) and all urban observations (b) as Taylor diagrams summarized in table (c). Observations, model outputs, biases and RMSE are expressed in $\mu g\ m^{-3}$.

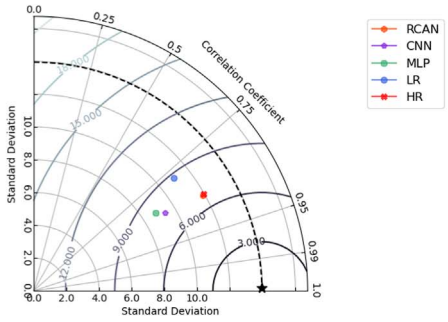
## 4. Discussion

The best NN architecture RCAN is able to reproduce the behavior of the raw HR simulation with satisfactory performances. For some pollutants the NN model is even able to provide better results probably by smoothing some aberrant values calculated by the CTM during very stable situations leading to unrealistic concentration peaks. With minor improvements on ozone chemistry and the use of observational data to constrain the system similarly to MOS (Model Output Statistic) techniques or using a CNN approach for bias corrections [20], our approach can be quickly deployed for air quality forecasting. Once the training is performed the forecasting chain could deliver a forecast in a few second instead of hours. Then, it should be further investigated if a generic NN-based model is sufficient or if it has to be adapted to specific meteorological conditions. In the latter, a training strategy must be investigated with probably a moving 15 days or a monthly update of the learning process by taking the last 15 to 30 days for instance. This would have the advantage of training the NN with similar meteorological conditions.

More interesting, expectations lie in the field of air quality modeling for policy making and impact assessment. These NN approaches can be complementary of statistical analysis embedded in surrogate models like the Screening for High Emission Reduction Potential on Air—SHERPA [41, 42] developed to support the design of air quality plans in the context of the EU Air Quality directive [43] by the member states. The approach proposed in SHERPA is based on the cell-per-cell relationships linking the concentration at a grid cell *i* to the emissions in the surrounding cells. It builds on the concept of Geographically Weighted Regression (GWR) as used in [44]) or local modeling approaches [45], a family of approaches that uses 'bell-shaped' kernel functions to establish weighted, local regressions between input and output variables. SHERPA is designed to evaluate the impact of an emission reduction for a given activity sector and area to a selected location through the mathematical representation of Source Receptor Relationships (SRR). The SHERPA model works so far on yearly and season averaged concentrations. Working over large time-averaged periods smooths the results and limits the impact of non-linearities induced by complex high-frequency phenomena and interactions between chemical species. SHERPA requires a minimum number of simulations with targeted




(a) All rural




(b) All urban

| Species | Method | Typology | # stations | Obs. | Mod. | Bias | RMSE | Cor. |
|---|---|---|---|---|---|---|---|---|
| $PM_{2.5}$ | HR | RUR | 3 | 12.48 | 7.85 | -4.49 | 5.62 | 0.70 |
| | LR | | | | 6.30 | -6.06 | 7.15 | 0.60 |
| | MLP | | | | 6.49 | -5.89 | 6.95 | 0.67 |
| | CNN | | | | 6.72 | -5.64 | 6.33 | 0.82 |
| | RCAN | | | | 7.75 | -4.58 | 5.62 | 0.72 |
| | HR | URB | 12 | 35.68 | 24.30 | -11.36 | 13.29 | 0.87 |
| | LR | | | | 21.30 | -14.37 | 16.82 | 0.78 |
| | MLP | | | | 21.03 | -14.64 | 16.70 | 0.84 |
| | CNN | | | | 21.34 | -14.32 | 16.21 | 0.86 |
| | RCAN | | | | 25.44 | -10.23 | 12.31 | 0.87 |

(c) Daily statistics

**Figure 11.** $PM_{2.5}$ daily statistics of high resolution (HR), coarse resolution (LR), pixel-based independent multi-layer perceptron (MLP), Convolutional neural network (CNN) and Residual channel attention network (RCAN) for comparison with all rural observations (a) and all urban observations (b) as Taylor diagrams summarized in table (c). Observations, model outputs, biases and RMSE are expressed in $\mu g\ m^{-3}$.

emission reductions scenarios and the goal is to increase its resolution for a better representation of the local scale. Definitively, our approach paves the way for producing fast scenario simulations at high resolution to feed this type of models, but care must be taken to ensure that a minimum of physics is embedded to deal with non-linearities as previously mentioned. For instance, the case of Ozone but also the formation of secondary particles like the ammonium nitrate can be highlighted. Our approach could also inspire new developments in SHERPA-like models benefiting from more recent developments in machine learning techniques on image processing and analyses. Moreover, our developments can be easily adapted to any other CTM outputs.

Future works may focus on how to integrate physical constraints in the neural network to improve these first encouraging results on a very complex area with steep slopes enhancing local effects. However, this approach deserved to be tested over a larger domain and a finest resolution with very different chemical and meteorological regimes. It is especially relevant for $O_3$, in [23], a special treatment for such a secondary pollutant is proposed based on the two main equations of the ozone chemistry involving NOx and Ozone. If the correlations for $NO_2$ are even better for the RCAN architecture compared to CHIMERE *versus* observations, the discrepancies in terms of bias can be a consequence of the local interactions with Ozone that are not considered in our methodology. This type of physical processes can be easily implemented in an efficient NN-based scheme as a way of forcing the consistency between the super resolution outputs for $NO_2$, NO, $O_3$ and probably the Volatile Organic Compounds. This directly relates to one of the main branch of physically-guided neural networks that aims at designing specific NN architectures to embed the physics in the modeling system. An other option would be to keep similar CNN and attention-based architectures proposed in the paper while adding additional constraints on the physics in the loss function: it is an active field of research in what is called 'physically-informed neural networks' [46–48]. At last, the use of satellite data like aerosol optical depth with ground observations as input data for a NN allows to create an adequate model to predict Super-Resolution PM2.5 concentrations as it has been reported over Beijing [49, 50]. Introducing observational data in our approach is another way to improve a forecasting system based on a metamodel.

## 5. Conclusion

In this study, we have developed and evaluated the relevance of neural-network super-resolution approaches to downscale coarse chemistry transport modeling simulations focusing on three criteria air pollutants: $NO_2$, $PM_{2.5}$ and $PM_{10}$. These learning-based techniques take advantage of (i) fast coarse simulation outputs from the CHIMERE CTM embedding complex mathematical representations of physics and chemistry and particularly the long-range transport of pollutants, and (ii) more local features which are retrieved by neural network approaches. The reported quantitative and qualitative evaluation against both the high resolution reference simulation and real observation datasets support the relevance of the neural-network-based downscaling for the operational monitoring and forecasting of air quality. Future directions at short and medium terms are identified to use this kind of techniques handling non linearities related to secondary pollutants, interactions between species and emission reduction strategies.

At last, let us remind that using neural network is complementary of developing more and more complex physical models. They are good instruments to simplify complex models for operational uses (by catching the main patterns) and offering the possibility to develop more and more sophisticated deterministic models representing the 'real world'.

## Acknowledgments

This research was funded by the DGA (French Directorate General of Armaments; grant no. 2018 60 0074) in the framework of the NETDESA project. It was also supported by ANR Projects OceaniX. It benefited from HPC and GPU resources from Azure (Microsoft EU Ocean awards) and from GENCI-IDRIS (Grant 2020-101030).

## Data availability statement

The data that support the findings of this study are available upon reasonable request from the authors.

## Appendix A. Error statistics and performances for $PM_{10}$

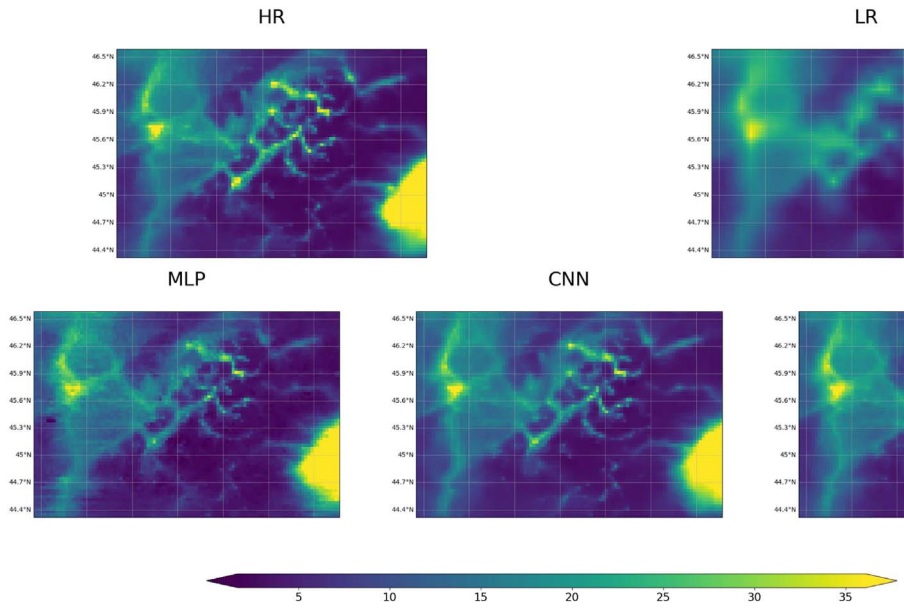



**Figure A.1.** $PM_{10}$ mean concentrations over the validation period for CHIMERE high-resolution and coarse resolution (top line) and three NN-based super resolution architectures (bottom line): from left to right, pixel-based independent multi-layer perceptron (MLP), Convolutional neural network (CNN) and Residual channel attention network (RCAN)

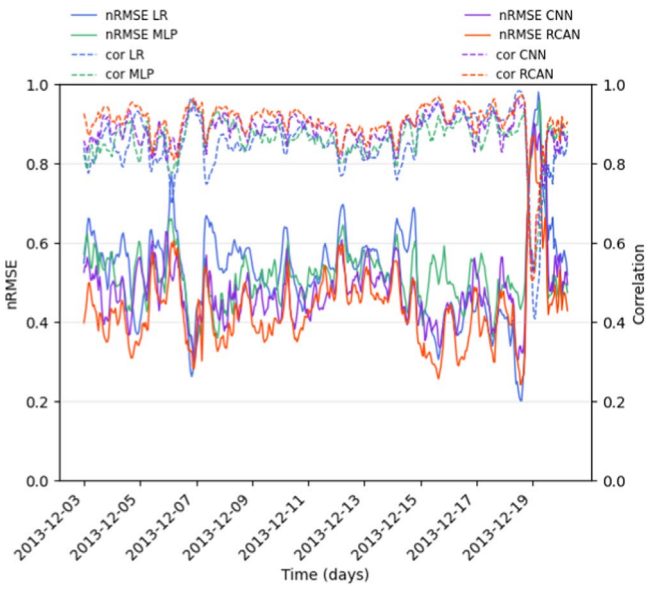



**Figure A.2.** $PM_{10}$ nRMSE and correlation time series (a) of coarse resolution, pixel-based independent multi-layer perceptron (MLP), Convolutional neural network (CNN) and Residual channel attention network (RCAN) and summary statistics over the validation period as Taylor diagrams (b).

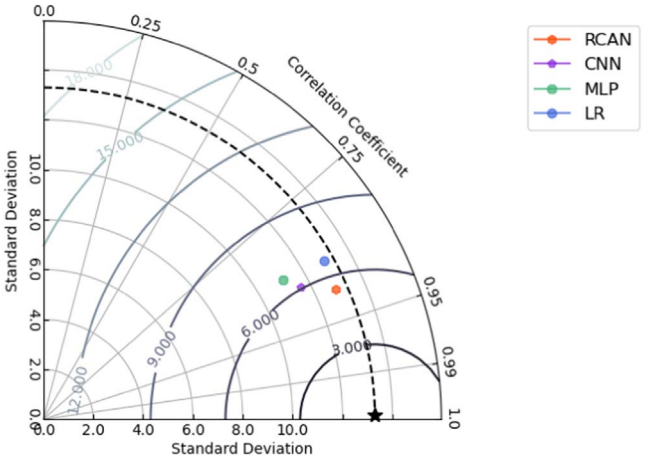
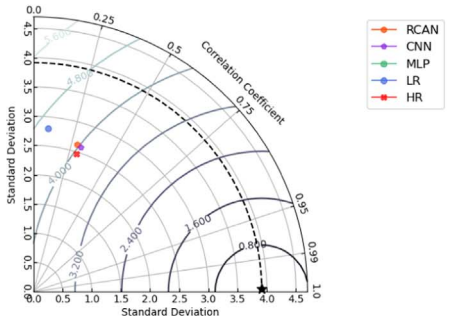
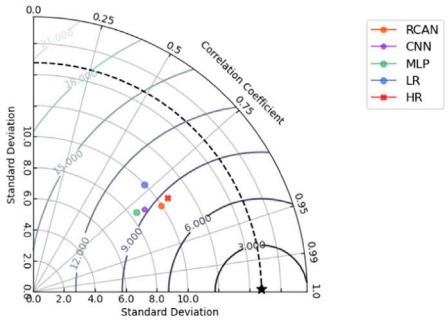



**Figure A.3.** $PM_{10}$ summary statistics over the validation period as Taylor diagrams for low resolution simulation (LR), multi-layer perceptron (MLP), Convolutional neural network (CNN) and Residual channel attention network (RCAN) versus the high resolution (HR) considered as the 'truth'.




(a) All rural




(b) All urban

| Species | Method | Typology | # stations | Obs. | Mod. | Bias | RMSE | Cor. |
|---|---|---|---|---|---|---|---|---|
| $PM_{10}$ | HR | RUR | 4 | 11.24 | 8.15 | -3.09 | 5.01 | 0.30 |
| | LR | | | | 7.66 | -3.58 | 5.83 | 0.09 |
| | MLP | | | | 4.93 | -6.31 | 8.78 | -0.24 |
| | CNN | | | | 8.18 | -3.06 | 5.01 | 0.31 |
| | RCAN | | | | 8.31 | -2.93 | 4.99 | 0.28 |
| | HR | URB | 35 | 43.75 | 23.52 | -20.23 | 21.97 | 0.82 |
| | LR | | | | 20.13 | -23.62 | 25.75 | 0.72 |
| | MLP | | | | 20.59 | -23.15 | 25.05 | 0.79 |
| | CNN | | | | 22.25 | ,-21.50 | 23.40 | 0.80 |
| | RCAN | | | | 23.00 | -20.75 | 22.49 | 0.83 |

(c) Daily statistics

**Figure A.4.** $PM_{10}$ daily statistics of high resolution (HR), coarse resolution (LR), pixel-based independent multi-layer perceptron (MLP), Convolutional neural network (CNN) and Residual channel attention network (RCAN) for comparison with all rural observations (a) and all urban observations (b) as Taylor diagrams summarized in table (c). Observations, model outputs, biases and RMSE are expressed in $\mu g\ m^{-3}$.

## Appendix B. Comparison with the hourly observations

**Table B.1.** $NO_2$ hourly statistics of high resolution (HR), coarse resolution (LR), pixel-based independent multi-layer perceptron (MLP), Convolutional neural network (CNN) and Residual channel attention network (RCAN). Observations, model outputs, biases and RMSE are expressed in $\mu g\ m^{-3}$.

| | | | | Obs. | Mod. | Bias | RMSE | Cor. |
|---|---|---|---|---|---|---|---|---|
| Species | Method | Typology | # stations | | | | | |
| $NO_2$ | HR | RUR | 7 | 13.64 | 4.71 | −8.92 | 11.12 | 0.35 |
| | LR | | | | 5.54 | −8.09 | 10.82 | 0.16 |
| | MLP | | | | 4.54 | −9.09 | 11.45 | 0.20 |
| | CNN | | | | 5.04 | −8.59 | 10.88 | 0.33 |
| | RCAN | | | | 4.93 | −8.70 | 11.07 | 0.25 |
| | HR | URB | 40 | 48.67 | 33.62 | −15.05 | 19.67 | 0.65 |
| | LR | | | | 24.60 | −24.06 | 26.90 | 0.61 |
| | MLP | | | | 30.27 | −18.39 | 21.40 | ,0.69 |
| | CNN | | | | 32.24 | ,-16.42 | 20.07 | 0.69 |
| | RCAN | | | | 28.31 | −20.36 | 23.23 | 0.68 |

**Table B.2.** $PM_{2.5}$ hourly statistics of high resolution (HR), coarse resolution (LR), pixel-based independent multi-layer perceptron (MLP), Convolutional neural network (CNN) and Residual channel attention network (RCAN). Observations, model outputs, biases and RMSE are expressed in $\mu g\ m^{-3}$.

| | | | | Obs. | Mod. | Bias | RMSE | Cor. |
|---|---|---|---|---|---|---|---|---|
| Species | Method | Typology | # stations | | | | | |
| $PM_{2.5}$ | HR | RUR | 3 | 13.20 | 7.95 | −4.83 | 8.16 | 0.30 |
| | LR | | | | 6.35 | −6.58 | 8.97 | 0.42 |
| | MLP | | | | 6.54 | −6.52 | 8.86 | 0.46 |
| | CNN | | | | 6.85 | −5.95 | 8.48 | 0.43 |
| | RCAN | | | | 7.91 | −4.88 | 7.71 | 0.45 |
| | HR | URB | 12 | 36.65 | 24.84 | −11.81 | 15.00 | 0.80 |
| | LR | | | | 21.71 | −14.94 | 18.44 | 0.71 |
| | MLP | | | | 21.41 | −15.23 | 18.31 | 0.76 |
| | CNN | | | | 21.70 | ,−14.94 | 17.83 | 0.78 |
| | RCAN | | | | 25.96 | −10.69 | 14.23 | 0.79 |

**Table B.3.** $PM_{10}$ hourly statistics of high resolution (HR), coarse resolution (LR), pixel-based independent multi-layer perceptron (MLP), Convolutional neural network (CNN) and Residual channel attention network (RCAN). Observations, model outputs, biases and RMSE are expressed in $\mu g\ m^{-3}$.

| | | | | Obs. | Mod. | Bias | RMSE | Cor. |
|---|---|---|---|---|---|---|---|---|
| Species | Method | Typology | # stations | | | | | |
| $PM_{10}$ | HR | RUR | 4 | 11.44 | 8.24 | −3.20 | 6.24 | 0.25 |
| | LR | | | | 7.73 | −3.71 | 6.75 | 0.18 |
| | MLP | | | | 4.84 | −6.59 | 9.82 | -0.12 |
| | CNN | | | | 8.30 | −3.14 | 6.18 | 0.27 |
| | RCAN | | | | 8.42 | −3.01 | 6.05 | 0.28 |
| | HR | URB | 35 | 44.87 | 23.95 | −20.92 | 23.38 | 0.77 |
| | LR | | | | 20.41 | −24.45 | 27.29 | 0.68 |
| | MLP | | | | 20.91 | −23.96 | 26.60 | 0.73 |
| | CNN | | | | 22.51 | ,-22.36 | 24.93 | 0.76 |
| | RCAN | | | | 23.41 | −21.46 | 23.92 | 0.78 |

## Appendix C. Statistics on Δ

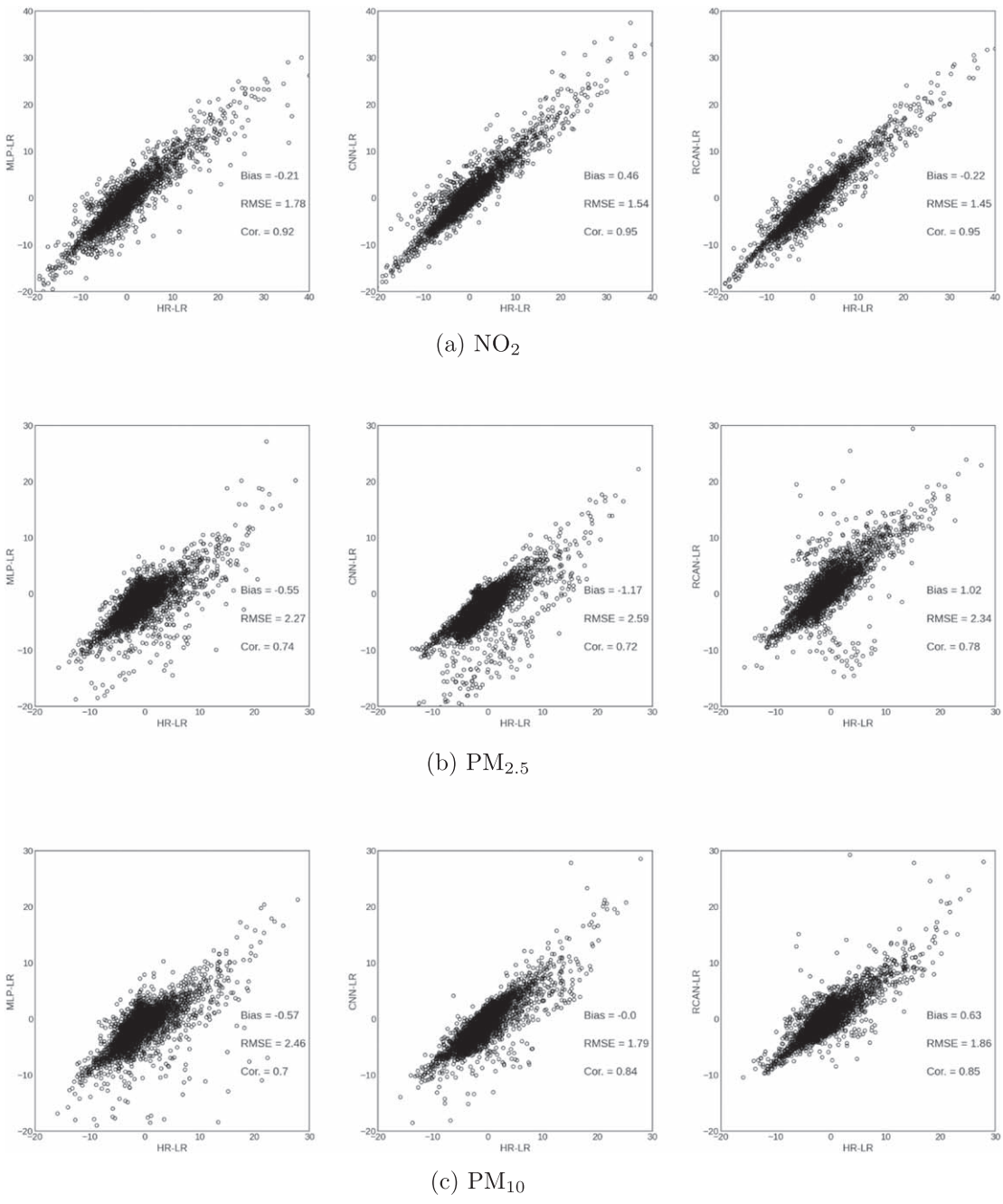



**Figure C.5.** Scatterplots between the 'true' increment HR-LR and the SR-based increments: from left to right, MLP-LR, CNN-LR and RCAN-LR for $NO_2$ (a), $PM_{2.5}$ (b) and $PM_{10}$ (c) expressed in $\mu g\ m^{-3}$.

## Appendix D. Definition of error statistics

The mean bias (MB), Pearson correlation (R), the root mean square error (RMSE) and its 'normalized' version (nRMSE) are defined herebelow for $n$ the number of data, $M$ the model predicted value and $O$ the corresponding observation or HR 'truth'.

$$\bar{V} = \frac{1}{n}\sum_{i=1}^{n} V_i \quad V = O,\ M \tag{4}$$

$$MB = \frac{1}{n}\sum_{i=1}^{n}(M_i - O_i) \tag{5}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(M_i - O_i)^2} \tag{6}$$

$$nRMSE = \frac{\sqrt{\sum_{i=1}^{n}(M_i - O_i)^2}}{\sqrt{\sum_{i=1}^{n}(O_i - \bar{O})^2}} \tag{7}$$

$$R = \frac{\sum_{i=1}^{n}(M_i - \bar{M})(O_i - \bar{O})}{\sqrt{\sum_{i=1}^{n}(M_i - \bar{M})^2}\sqrt{\sum_{i=1}^{n}(O_i - \bar{O})^2}} \tag{8}$$

## ORCID iDs

B Bessagnet https://orcid.org/0000-0003-2062-4681

## References


[1] UN 2019 World Urbanization Prospects: The 2018 Revision Tech. Rep. ST/ESA/SER.A/420 Department of Economic and Social Affairs, Population Division, United Nations New York, USA
[2] EEA 2019 Air quality in Europe: 2019 report (LU: Publications Office)
[3] Mailler S *et al* 2017 *Geoscientific Model Development* **10** 2397–423
[4] Marécal V *et al* 2015 *Geoscientific Model Development* **8** 2777–813
[5] Rouïl L *et al* 2009 *BAMS* **90** 73–83
[6] Host S, Honoré C, Joly F, Saunal A, Tertre A L and Medina S 2020 *Environ. Res.* **185** 109405
[7] Feng R, Zheng H J, Gao H, Zhang A R, Huang C, Zhang J X, Luo K and Fan J R 2019 *J. Clean. Prod.* **231** 1005–15
[8] Alimissis A, Philippopoulos K, Tzanis C and Deligiorgi D 2018 *Atmos. Environ.* **191** 205–13
[9] Kaya K and Gündüz Öüğüdücü C 2020 *Sci. Rep.* **10** 3346
[10] Liu T, Ying Y, Xu Y, Ke D and Su K 2018 *Fine-Grained Air Quality Prediction using Attention Based Neural Network 2018 International Joint Conference on Neural Networks (IJCNN) (Rio de Janeiro: IEEE)* 1–6
[11] Shams S R, Jahani A, Kalantary S, Moeinaddini M and Khorasani N 2021 *Sci. Rep.* **11** 1805
[12] Iskandaryan D, Ramos F and Trilles S 2020 *Applied Sciences* **10** 2401
[13] Turabieh H, Sheta A, Braik M and Kovač-Andrić E 2021 A Layered Recurrent Neural Network for Imputing Air Pollutants Missing Data and Prediction of $NO_2$, $O_3$, $PM_{10}$, and $PM_{2.5}$ *Forecasting in Mathematics - Recent Advances, New Perspectives and Applications* ed A Abou Jaoude (IntechOpen)
[14] Cabaneros S M, Calautit J K and Hughes B R 2019 *Environ. Modell. Softw.* **119** 285–304
[15] Mokhtari I, Bechkit W, Rivano H and Yaici M R 2021 *IEEE Access* **9** 14765–78
[16] LeCun Y, Bengio Y and Hinton G 2015 *Nature* **521** 436–44
[17] Guo Q, He Z, Li S, Li X, Meng J, Hou Z, Liu J and Chen Y 2020 *Aerosol Air Qual. Res.* **20** 1429–39
[18] Vlasenko A, Matthias V and Callies U 2021 *Atmos. Environ.* **254** 118236
[19] Shukla K, Dadheech N, Kumar P and Khare M 2021 *Chemosphere* **272** 129611
[20] Sayeed A, Lops Y, Choi Y, Jung J and Salman A K 2021 *Atmos. Environ.* **253** 118376
[21] Kelp M M, Tessum C W and Marshall J D 2018 Orders-of-magnitude speedup in atmospheric chemistry modeling through neural network-based emulation *arXiv: Atmospheric and Oceanic Physics* arXiv:1808.03874
[22] Kelp M M, Jacob D J, Kutz J N, Marshall J D and Tessum C W 2020 Toward Stable, General Machine-Learned Models of the Atmospheric Chemical System *Journal of Geophysical Research: Atmospheres* **125** e2020JD032759
[23] Bessagnet B, Couvidat F and Lemaire V 2019 *Environ. Modell. Softw.* **116** 100–9
[24] Amann M *et al* 2011 *Environ. Modelling Softw.* **26** 1489–501
[25] Amann M, Cofala J, Gzella A, Heyes C, Klimont Z and Schöpp W 2007 Estimating concentrations of fine particulate matter in urban background air of european cities *Tech. Rep. Interim Report IR-07-001 IIASA* 1–50 http://pure.iiasa.ac.at/id/eprint/8454/
[26] Briant R, Tuccella P, Deroubaix A, Khvorostyanov D, Menut L, Mailler S and Turquety S 2017 *Geoscientific Model Development* **10** 927–44
[27] Tuccella P, Menut L, Briant R, Deroubaix A, Khvorostyanov D, Mailler S, Siour G and Turquety S 2019 *Atmosphere* **10** 20
[28] Couvidat F, Bessagnet B, Garcia-Vivanco M, Real E, Menut L and Colette A 2018 *Geoscientific Model Development* **11** 165–94
[29] Guenther A B, Jiang X, Heald C L, Sakulyanontvittaya T, Duhl T, Emmons L K and Wang X 2012 *Geoscientific Model Development* **5** 1471–92

[30] NCAR/UCAR 2000 NCEP FNL operational model global tropospheric analyses, continuing from july 1999, research data archive at the national center for atmospheric research, computational and information systems laboratory https://doi.org/10.5065/D6M043C6
[31] Bessagnet B, Menut L, Lapere R, Couvidat F, Jaffrezo J L, Mailler S, Favez O, Pennel R and Siour G 2020 *Atmosphere* **11** 565
[32] Beauchamp M, Fablet R, Ubelmann C, Ballarotta M and Chapron B 2020 *Remote Sensing* **12** 3806
[33] Zhang Y, Li K, Li K, Wang L, Zhong B and Fu Y 2018 *Image Super-Resolution Using Very Deep Residual Channel Attention Networks* ed V Ferrari, M Hebert, C Sminchisescu and Y Weiss vol 11 211 Computer Vision – ECCV 2018 (Cham: Springer) pp 294–310
[34] Bahdanau D, Cho K and Bengio Y 2016 Neural machine translation by jointly learning to align and translate *arXiv:Computation and Language* arXiv:1409.0473
[35] Luong M T, Pham H and Manning C D 2015 Effective approaches to attention-based neural machine translation *arXiv: Computation and Language* arXiv:1508.04025
[36] Chollet F *et al* 2015 Keras, the python deep learning api *Github* https://github.com/fchollet/keras
[37] Lguensat R, Viet P H, Sun M, Chen G, Fenglin T, Chapron B and Fablet R 2019 *Remote Sensing* **11** 858
[38] Ouala S, Fablet R, Herzet C, Chapron B, Pascual A, Collard F and Gaultier L 2018 *Remote Sensing* **10** 1864
[39] Brajard J, Carrassi A, Bocquet M and Bertino L 2021 *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **379** 20200086
[40] Fablet R, Drumetz L and Rousseau F 2020 *arXiv: Machine Learning* arXiv:2006.03653 [physics]
[41] Clappier A, Pisoni E and Thunis P 2015 *Environ. Modelling Softw.* **74** 66–74
[42] Pisoni E, Clappier A, Degraeuwe B and Thunis P 2017 *Environ. Modelling Softw.* **90** 68–77
[43] European Union 2008 *Directive 2008/50/EC of the European Parliament and of the Council of 21 May 2008 OJ L 152* 1–44
[44] Páez A, Farber S and Wheeler D 2011 *Environment and Planning A: Economy and Space* **43** 2992–3010
[45] Lloyd C 2011 *Local Models for Spatial Analysis* II edn (Boca Ranton, FL: Taylor and Francis)
[46] Raissi M 2018 *Journal of Machine Learning Research* **19** 1–24
[47] Jagtap A D and Karniadakis G E 2020 *Communications in Computational Physics* **28** 2002–41
[48] Raissi M, Perdikaris P and Karniadakis G E 2019 *J. Comput. Phys.* **378** 686–707
[49] Yan X, Zang Z, Jiang Y, Shi W, Guo Y, Li D, Zhao C and Husi L 2021 *Environ. Pollut.* **273** 116459
[50] Yan X, Zang Z, Luo N, Jiang Y and Li Z 2020 *Environ. Int.* **144** 106060