



Deep learning-based downscaling of PM_{2.5}, NO₂ and O₃ over Europe across emission-reduction scenarios

Anthony Rey-Pommier ¹, Enrico Pisoni ¹, Maxime Beauchamp ^{2,3}, Bertrand Bessagnet ⁴, Alexander de Meij ⁵, Stefano Zauli-Sajani ¹, and Philippe Thunis ^{1,*}

¹European Commission, Joint Research Centre (JRC), Ispra, Italy

²Fondazione CMCC, Bologna, Italy

³IMT Atlantique, Lab-STICC OSE, INRIA Odyssey, Plouzané, France

⁴LMD/IPSL, École Polytechnique, Institut Polytechnique de Paris, ENS, PSL Université, Sorbonne Université, CNRS, France

⁵MetClim, Varese, Italy

*Retired with activated senior agreement

Correspondence: Anthony Rey-Pommier (anthony.rey-pommier@ec.europa.eu)

Abstract. Atmospheric pollutants pose significant health risks in Europe, where many cities exceed WHO and EU air quality standards. To support informed air quality policies, simulations of air pollution levels are required. They are generally provided by Chemical Transport Models, but these can be computationally expensive at high spatial resolution. In this context, deep learning methods have emerged as alternatives to full simulations. Using low and high resolution EMEP simulations, we investigate how different training strategies affect the robustness of a Residual Channel Attention Network for downscaling hourly particulate matter, nitrogen dioxide and ozone ground concentration fields from a $0.1^\circ \times 0.05^\circ$ to a $0.02^\circ \times 0.01^\circ$ resolution over Europe. The geographic extent and temporal sampling of the training data appear to be key determinants for the performance of the downscaling process: training the network on days that are evenly distributed throughout the year generally yields a better performance than training it on the same number of days but restricted to the winter season, and, similarly, the choice of a larger spatial domain during training increases the model performance in comparison with a smaller spatial domain. The optimised model achieves better performance in reconstructing fine scale spatial gradients than a bilinear interpolation of low resolution simulations. It generalises to multiple emission reduction scenarios when the link between the perturbed emissions and the target pollutant is strong and direct, while weaker and more indirect responses are less reliably reconstructed. An optimal compromise between temporal resolution and downscaling accuracy is also identified, corresponding to approximately one day of temporal averaging. These results highlight the reliability of neural network-based downscaling techniques, to support their integration into operational air quality assessment and scenario planning tools as complements to full simulations.

1 Introduction

Air pollution remains one of the major challenges for public health, responsible for more than 6.7 million premature deaths globally in 2019 (World Health Organization, 2024) and ranking as the second leading determinant for mortality in 2021 (Health Effects Institute, 2024). Among air pollutants, fine particulate matter (PM_{2.5}), nitrogen dioxide (NO₂) and ozone (O₃)



are of greatest concern, due to their impacts on cardiovascular and respiratory systems (European Environment Agency, 2020). $\text{PM}_{2.5}$ and NO_2 are mostly emitted from combustion processes in vehicles, power plants and residential heating systems (Maroa and Inambao, 2020; McDuffie et al., 2021), and $\text{PM}_{2.5}$ can also be formed via secondary processes from gaseous precursors such as sulphur dioxide or ammonia. O_3 formation follows an indirect pathway as well, through the transformation of nitrogen oxides and organic compounds under solar radiation (Bowman and Seinfeld, 1994). Despite substantial air quality improvements over the past decades, partially due to air quality legislation in developed countries (Turnock et al., 2016; Ma et al., 2019), concentrations of these three pollutants remain high, particularly in densely populated regions and industrial areas. In Europe, many urban areas experience concentrations that still exceed the guidelines of the World Health Organization (World Health Organization, 2021) or the European Union (EU) air quality standards (European Union, 2024).

Air pollution levels are determined by the transport and the chemical transformations of pollutant species emitted from various sources, which are deeply influenced by meteorological conditions and local geographical characteristics. Strong concentration gradients can thus occur over distances of only a few hundred meters due to variations in traffic intensity, urban configuration, and atmospheric mixing conditions. To support optimal air quality policies aiming to reduce human exposure to dangerous pollution levels, high-resolution air quality data are therefore required. To produce such fields over large spatial scales, Chemical Transport Models (CTMs) currently represent the reference approach. By solving the governing equations describing emission, transport, chemical transformation, deposition and removal processes, they provide a mechanistic representation of pollutant dynamics. They are widely used for operational air quality forecasting, regulatory assessments and scenario analysis. CTMs perform reliably at horizontal resolutions of a few kilometres, where subgrid-scale processes can be represented with parameterisations. If the spatial resolution is refined to the order of a few hundred metres, the importance of phenomena like street canyon flows (Vardoulakis et al., 2003) and local plume dispersion becomes dominant. At these scales, the conventional parameterisations of turbulent mixing, convection and chemistry become less valid, and alternative approaches like computational fluid dynamics modelling (Anderson and Wendt, 1995) become more appropriate. However, even when CTMs are operated at intermediate resolutions, resolving finer-scale pollution structures comes with substantial computational costs, especially when multiple emission scenarios need to be investigated to design air quality policies. Consequently, while CTMs remain the standard method for city-scale air quality assessments, their application at spatial resolutions below a few kilometres is limited.

An alternative to these computationally expensive high resolution CTM simulations consists in performing less computationally expensive simulations at a coarser resolution before reconstructing the corresponding fine-scale concentration fields. This process is commonly referred to as spatial downscaling or super-resolution (SR) mapping. It aims to recover sub-grid spatial variability while preserving the large-scale information contained in original low resolution simulations. Several downscaling approaches have consequently been developed for air quality applications. For instance, proxy-based methodologies can be developed to refine CTM concentration fields by separating concentration components, redistributing emission-driven fractions with high-resolution activity data, and applying meteorology-dependent Gaussian diffusion to reconstruct fine-scale patterns (Mu et al., 2022; Bessagnet et al., 2023). In recent years, deep learning methods have also been increasingly used as alternative



downscaling tools. These statistical methods identify the main pollution patterns and learn non-linear relationships between low resolution simulations and fine-scale spatial structures. In particular, convolutional neural networks, which are well-adapted for the reconstruction of high-resolution images from degraded low resolution inputs, have gained attention in climate and Earth system sciences for downscaling meteorological and climate variables such as precipitation or temperature (Wang et al., 2021; Soares et al., 2024). These approaches demonstrated that neural networks can efficiently reproduce fine-scale spatial patterns found in high-resolution fields, while reducing the computational cost associated with high-resolution simulations. In addition to computational advantages, neural network based approaches may also significantly reduce data storage requirements, as only lower resolution simulations need to be archived while high-resolution fields can be reconstructed when required (Soares et al., 2024).

The application of methods used for conventional image super-resolution problems to atmospheric composition fields raises several challenges. Atmospheric pollutant fields are not only characterised by spatial structures but also by strong temporal variability driven by meteorological conditions and chemical regimes. A model trained on a limited range of conditions may fail to generalise to unseen situations, such as meteorological conditions associated with different seasons or climates that can alter atmospheric transport and chemical regimes, while different emission patterns may induce new interactions between pollutants. Such changes in atmospheric dynamics and emission conditions can lead to fine-scale spatial patterns that differ substantially from those represented in the training data, potentially limiting the ability of the downscaling model to accurately reconstruct them. This question is particularly relevant for operational air quality applications, where downscaling models should generalise to new meteorological conditions and emission scenarios without requiring additional training. The selection of the training data to maximise model generalisation is thus of fundamental importance, and is a subject on which insufficient research has been produced so far. In particular, it is unclear whether model performance is primarily controlled by the spatial diversity of training samples, the temporal representativeness of meteorological conditions, or the range of pollutant concentrations covered in the training dataset.

In this study, we further develop a deep learning framework for the statistical downscaling of hourly air quality fields over Europe. The present study draws on the work of Bessagnet et al. (2021) and Beauchamp et al. (2026), which demonstrated the effectiveness of deep learning for downscaling chemistry transport model outputs and showed that neural super-resolution architectures can enhance spatial detail while preserving key pollutant structures. Drawing on this methodological foundation, we investigate how the choice of training samples influences downscaling performance and assess the ability of the framework to generalise to simulations outside the training distribution, including scenarios with modified emission levels, precursor species, and activity sectors. The model is based on a Residual Channel Attention Network (RCAN), an architecture originally developed for image super-resolution tasks (Zhang et al., 2018) that combines residual learning and channel attention mechanisms to extract relevant spatial features and reconstruct fine-scale structures from lower-resolution inputs. The network downscales $\text{PM}_{2.5}$, NO_2 , and O_3 ground concentrations from a $0.1^\circ \times 0.05^\circ$ grid to a $0.02^\circ \times 0.01^\circ$ grid over a large part of Europe, approaching the spatial scales at which local atmospheric processes become increasingly important. Training and val-



validation are performed using low- and high-resolution simulations from the EMEP chemical transport model (Simpson et al.,
90 2012) as input-target pairs.

This article is structured as follows: Section 2 describes the domains used in this study, the meteorological and the emission datasets with which EMEP simulations are run, and the neural network architecture, before presenting the training and validation strategies. Section 3 presents the results of the optimisation of spatial and temporal training configurations, evaluates model generalisation for different emission scenarios on the same domain, analyses the ability of the model to reconstruct
95 fine-scale structures and the persistence of the artefacts it generates. Section 4 discusses the implications and limitations of the proposed approach.

2 Data and methods

2.1 Working domains

For this study, we use a domain of reference that spans from 12.35°W to 27.35°E in longitude and from 37.52°N to 52.92°N
100 in latitude, i.e. most of Western and Central Europe. Most of the climate regimes in Europe are represented, with pronounced meteorological variability between regions. In addition to this diversity, the industrial and urban centres that are covered are highly diverse, exhibiting markedly different emission structures that reflect the impact of distinct industrial development and environmental policies throughout the continent. As a result, various chemical regimes are experienced within the domain.

In addition to this domain of reference, a domain of smaller size, referred to as the “restricted domain” and covering lon-
105 gitudes from 0.67°W to 17.07°E and latitudes from 40.26°N to 52.20°N, is used for the downscaling optimisation strategy highlighted in Section 2.6. It is approximately one-third the size of the domain of reference, but deliberately includes oceanic, continental and Mediterranean climates, major urban agglomerations (such as Paris, Barcelona, Rome or Frankfurt), North Sea coastal zones, mountainous areas and flat hinterlands. The restricted domain was selected so that the statistical distribution of all predictors used in the downscaling framework matches that of the full domain. The boundaries of the two domains are
110 displayed in the Supplementary Materials (Figure S1), as well as their corresponding statistical distributions for all target and input variables used in this study (Figure S2).

2.2 Meteorological data

We use the meteorological data from the Integrated Forecast System (IFS) of the European Centre for Medium-Range Weather Forecasts (ECMWF), which provides six-hourly meteorological fields at a horizontal resolution of $0.25^\circ \times 0.25^\circ$ for the whole
115 globe. The IFS assimilates a suite of observations, including surface stations, radiosondes, aircraft reports, satellite radiances and scatterometer winds, which yields a physically consistent representation of temperature, wind, pressure, cloud properties, geopotential height, eddy diffusivity and specific humidity. These variables together drive the advection, dispersion and depo-



sition processes. Meteorological fields are accessible in open access at <https://www.ecmwf.int/en/forecasts/datasets/open-data> (last access: 08 September 2026).

120 2.3 Emissions data

We use anthropogenic emissions data that were constructed on the basis of the bottom-up gridded dataset CAMS-REG-ANT (version 8.0). This regional emission inventory covers European countries, as well as the Asian part of Türkiye and the western part of Russia. Its domain spans from 30°W to 60°E and from 30°N to 72°N (Kuenen et al., 2022). It includes major air pollutants and greenhouse gases. The inventory is organised into 11 sectors: power plants, industrial sources, other stationary
 125 combustion sources, fugitive emissions, solvents, road transport, aviation, off-road activities, waste, agriculture (livestock), and other agricultural activities. These sectors are constructed by merging multiple local datasets to create sector-specific spatial proxies at a resolution of $0.1^\circ \times 0.05^\circ$. The version used in this study includes a fully reprocessed time series for point sources and an updated spatial distribution for road transport, improving the allocation of emissions to different road types and enhancing the accuracy of road networks across Europe. To generate emissions at finer resolution, we use the methodology
 130 developed by Bessagnet et al. (2023), with which CAMS-REG-ANT emissions are spatially downscaled to a $0.02^\circ \times 0.01^\circ$ grid using activity-specific proxy datasets. Within each cell of the $0.1^\circ \times 0.05^\circ$ grid, emissions are redistributed onto the high-resolution grid according to normalised proxy fields, ensuring mass conservation.

2.4 EMEP simulations

The EMEP MSC-W (called EMEP henceforth) model is a CTM that was established by the European Monitoring and Eval-
 135 uation Programme for Transboundary Long Range Transported Air Pollutants (Simpson et al., 2012). It is an Eulerian model focusing on the European continent that simulates the atmospheric transport of particulate matter, photo-oxidants, inorganic and organic gases and pollutants, as well as wet and dry deposition. Twenty vertical levels are used, with the first level around 45 m. To drive its simulations, meteorological input data are provided at a daily resolution, while emission input data are provided as annual budgets. Initial and boundary conditions are specified within the EMEP framework, with boundary con-
 140 ditions for some species, such as ozone, Saharan dust, and sea salt, based on results from external global models. For dust, the model also accounts for emissions transported from deserts, semi-arid regions, agricultural and bare lands. In contrast, the concentrations of sulphate nitrate, ammonium, sulphur dioxide, and other species are prescribed internally as fixed background values within the model. Initial ozone levels are derived from observations. For other air quality parameters, background and initial conditions are set within the model using functions based on observations (Tarrasón et al., 2004). The formation of
 145 secondary aerosols involves intricate chemical and physical pathways, such as the oxidation of sulphur dioxide into sulphate particles, the reaction of nitrogen oxides to form nitrate aerosols, and the transformation of volatile organic compounds into organic aerosols. More detailed information on the boundary conditions, meteorological drivers, land cover, model physics, and chemistry is provided in de Meij et al. (2022) and in the EMEP Status Report 2023 (Fagerli et al., 2023). It includes the validation of the model through comparisons against measurements of $\text{PM}_{2.5}$, NO_2 , and O_3 .



150 For this study, European air quality simulations using version 5.5 of EMEP are driven by ECMWF meteorological fields for the year 2021, described in Section 2.2, and CAMS-REG-ANT anthropogenic emissions for the year 2022, described in Section 2.3. The simulations are performed at two horizontal resolutions: high resolution (HR) simulations correspond to a $0.02^\circ \times 0.01^\circ$ grid ($\sim 2.2 \times 0.8 \text{ km}^2$ over European mean latitudes), while low-resolution (LR) simulations correspond to a $0.1^\circ \times 0.05^\circ$ grid ($\sim 11.1 \times 4.0 \text{ km}^2$). To match the EMEP source code structure, both the LR and HR emission datasets are re-organised according to the Gridded Nomenclature For Reporting (GNFR), which is an aggregated version of the Nomenclature For Reporting to the EU (UNECE, 2023). Similarly, IFS data are downscaled from its original $0.25^\circ \times 0.25^\circ$ by bilinear interpolation on both LR and HR grids. The simulations are limited to the domain of reference and provide hourly values for ground concentration fields. They are performed in pairs, which means that each simulation performed at high-resolution is mirrored with the corresponding simulation at low-resolution, with the input data interpolated accordingly. To cover different chemical regimes, 29 simulations are performed on the domain of reference, for different periods of the year and with different abatement levels on emissions. They include:

- Eight Base Case (BC) simulations, where emissions are not reduced, covering January 01-28, February 08-12, April 01-07, May 08-12, July 01-07, August 08-12, October 01-07 and November 08-12;
- Four simulations where all emissions are halved (-50%ALL) on the entire domain of reference, covering January 01-05, April 01-05, July 01-05 and October 01-05. Cumulating the reduction of all species allow reaching minimal concentrations for $\text{PM}_{2.5}$ and NO_2 for each season;
- Twelve precursor simulations, where emissions of different precursors are halved on the entire domain of reference: ammonia (NH_3), nitrogen oxides (NO_x) and primary particulate matter (PPM), sulphur dioxide (SO_2) and volatile organic compounds (VOCs). For each of these three scenarios, simulations cover February 08–12, May 08–12, August 08–12 and November 08–12. These scenarios are named -50% NH_3 , -50% NO_x &PPM and -50%VOC& SO_2 respectively. The joint reduction of NO_x and PPM, as well as SO_2 and VOCs, is done because both pairs influence $\text{PM}_{2.5}$ through complementary pathways: PPM contributes directly to particulate concentrations while NO_x forms secondary nitrate aerosols, and SO_2 produces sulphate aerosols while VOCs generate secondary organic aerosols. Grouping these precursors enables the assessment of the main $\text{PM}_{2.5}$ formation pathways while limiting the number of computationally expensive simulations;
- Four sector simulations, where a 35% reduction on the entire domain of reference is applied to emissions from the transport sector (-35%Transport) or the residential sector (-35%Residential), which are the main anthropogenic sectors responsible for $\text{PM}_{2.5}$ pollution in Europe (Thunis et al., 2025). In these two cases, February 08-12 and August 08-12 are covered;
- One simulation, where emission abatements are applied selectively for different regions within the domain of reference for February 08-12. It includes simultaneous reductions of residential emissions (-50%) in the urban areas of Bordeaux,



France, Naples, Italy and Prague, Czechia, transport emissions (-50%) in the urban areas of Valencia, Spain, Frankfurt, Germany and Katowice, Poland, and total emissions (-35%) in the regions of Brittany, France, Tuscany, Italy and Macroregion One, Romania. The urban areas are defined according to OECD (2012) while the regions correspond to the first level of the Nomenclature of Territorial Units for Statistics (NUTS1). This scenario, named REGFUA, allows exploring non-uniform emission reductions at different spatial scales.

The overview of the simulations is presented in Table 1. Labels T-W and T-Y denote simulations used for training the Winter- and Year-Models respectively, E denotes simulations used for model evaluation during sampling optimisation, and G denotes simulations used for evaluation after optimisation.

Simulation	JAN 01-28	JAN 01-05	FEB 08-12	APR 01-05	MAY 08-12	JUL 01-05	AUG 08-12	OCT 01-05	NOV 08-12
Base Case (BC, no reduction)	T-W	T-Y	E	T-Y	E	T-Y	E	T-Y	E
-50%ALL		E		E		E		E	
-50%NH ₃			G		G		G		G
-50%NO _x & PPM			G		G		G		G
-50%VOCs & SO _x			G		G		G		G
-35%Transport			G				G		
-35%Residential			G				G		
REGFUA			G						

Table 1. List of EMEP simulations performed at high- and low-resolution. T-W corresponds to simulations used to train the Winter-Model while T-Y corresponds to simulations used to train the Year-Model, which are described in Section 2.6. E corresponds to simulations used to estimate the performance of the super-resolution model for the sampling optimisation, which are used in Sections 3.1 and 3.2. G corresponds to simulations used to estimate its performance after optimisation, which are used in Sections 3.3 to 3.6.

2.5 Network configuration and selection of predictors

Here, we focus on deterministic reconstruction. Downscaling air pollutant concentrations can be formulated as a deterministic reconstruction problem (Mu et al., 2022; Nuterman et al., 2021), in which fine-scale patterns are constrained by high-resolution emissions and meteorology. The objective is to estimate a single physically consistent high-resolution concentration field from coarse CTM simulations, rather than to represent the conditional distribution of plausible high-resolution states. Therefore, we do not consider generative architectures (Goodfellow et al., 2020) or diffusion models (Song et al., 2020). Instead, RCAN was selected because its residual and channel-attention mechanisms are well suited to preserving and reconstructing spatial features in continuous fields, such as pollutant concentration fields, which exhibit image-like spatial structures, including well-defined gradients. RCAN enhances small-scale information associated with emissions and topography while maintaining the large-scale patterns represented by the EMEP model. The network receives several predictors and first projects them onto sixteen latent feature channels through an initial convolution. These feature maps are then processed by five residual groups,



each of which contains ten residual channel attention blocks (RCABs). These are composed of two 3×3 convolutions followed by a channel attention module. The attention module performs global-average pooling and passes the resulting descriptor through a pair of 1×1 convolutions. The first reduces the channel dimension by a factor of eight, while the second restores it. A ReLU activation is applied between the two convolutions, and the resulting descriptor is transformed using a sigmoid function to produce channel-wise weights that are applied to the original feature map. Local skip connections add the block input to its output, and a group level skip connection adds the input of each residual group to its output, thereby stabilising the optimisation of a deep architecture. After the residual groups, a 3×3 convolution restores the original projection (global skip) and the resulting feature map is up-scaled by a convolution that expands the channel dimension, after which a final convolution reduces the representation to a single high-resolution concentration field ($\text{PM}_{2.5}$, NO_2 , or O_3). The 20 predictors, listed in Table 2, are hourly fields gridded at high-resolution over the domain of reference introduced in Section 2.1. They combine low-resolution surface $\text{PM}_{2.5}$, NO_2 , SO_2 , NH_3 , O_3 and non-methane VOC fields with 10 m wind speed, eddy diffusivity, and anthropogenic emission fluxes for NH_3 , NO_x , $\text{PM}_{2.5}$, SO_2 and VOCs at both low and high resolutions. Low-resolution EMEP fields are bilinearly interpolated onto the high-resolution grid before being provided as inputs to the network.

Variable name	Quantity	Source
SURF_ug_PM25_rh50 or SURF_ug_NO2 or SURF_ug_O3	Ground $\text{PM}_{2.5}$ concentration (HR) or Ground NO_2 concentration (HR) or Ground O_3 concentration (HR)	EMEP output (HR)
ws_10m	10 metre-wind speed (HR)	IFS data adapted for EMEP and bilinearly interpolated onto HR grid
Kz_m2s	Eddy diffusivity (HR)	IFS data adapted for EMEP and bilinearly interpolated onto HR grid
Sec_Emis_mgm2_nh3	NH_3 anthropogenic emissions (HR)	CAMS-REG-ANT_v8.0 reorganised according to Bessagnet et al. (2023)
Sec_Emis_mgm2_nox	NO_x anthropogenic emissions (HR)	CAMS-REG-ANT_v8.0 reorganised according to Bessagnet et al. (2023)
Sec_Emis_mgm2_pm25	$\text{PM}_{2.5}$ anthropogenic emissions (HR)	CAMS-REG-ANT_v8.0 reorganised according to Bessagnet et al. (2023)
Sec_Emis_mgm2_sox	SO_x anthropogenic emissions (HR)	CAMS-REG-ANT_v8.0 reorganised according to Bessagnet et al. (2023)
Sec_Emis_mgm2_voc	VOC anthropogenic emissions (HR)	CAMS-REG-ANT_v8.0 reorganised according to Bessagnet et al. (2023)
ws_10m_lr	10 metre-wind speed (LR)	IFS data adapted for EMEP and bilinearly interpolated onto HR grid
Kz_m2s_lr	Eddy diffusivity (LR)	IFS data adapted for EMEP and bilinearly interpolated onto HR grid
Sec_Emis_mgm2_nh3_lr	NH_3 anthropogenic emissions (LR)	CAMS-REG-ANT_v8.0 (LR) adapted for EMEP and bilinearly interpolated onto HR grid
Sec_Emis_mgm2_nox_lr	NO_x anthropogenic emissions (LR)	CAMS-REG-ANT_v8.0 (LR) adapted for EMEP and bilinearly interpolated onto HR grid
Sec_Emis_mgm2_pm25_lr	$\text{PM}_{2.5}$ anthropogenic emissions (LR)	CAMS-REG-ANT_v8.0 (LR) adapted for EMEP and bilinearly interpolated onto HR grid
Sec_Emis_mgm2_sox_lr	SO_x anthropogenic emissions (LR)	CAMS-REG-ANT_v8.0 (LR) adapted for EMEP and bilinearly interpolated onto HR grid
Sec_Emis_mgm2_voc_lr	VOC anthropogenic emissions (LR)	CAMS-REG-ANT_v8.0 (LR) adapted for EMEP and bilinearly interpolated onto HR grid
SURF_ug_PM25_rh50_lr	Ground $\text{PM}_{2.5}$ concentration (LR)	EMEP output (LR), bilinearly interpolated onto HR grid
SURF_ug_NO2_lr	Ground NO_2 concentration (LR)	EMEP output (LR), bilinearly interpolated onto HR grid
SURF_ug_NH3_lr	Ground NH_3 concentration (LR)	EMEP output (LR), bilinearly interpolated onto HR grid
SURF_ppbC_NMVOC_lr	Ground NMVOC concentration (LR)	EMEP output (LR), bilinearly interpolated onto HR grid
SURF_ug_SO2_lr	Ground SO_2 concentration (LR)	EMEP output (LR), bilinearly interpolated onto HR grid
SURF_ug_O3_lr	Ground O_3 concentration (LR)	EMEP output (LR), bilinearly interpolated onto HR grid

Table 2. List of target variable (first line) and predictors for the downscaling model.



Before training, each predictor is standardised, with the mean and standard deviation being supplied in the configuration file.

215 The data are partitioned into spatial patches of 240×240 grid cells, with each patch corresponding to a single hourly time step. During training, patches are sampled with a stride of five cells in latitude and longitude, while during inference a larger stride of two hundred cells is used to reduce overlap. A Hanning-type weighting mask is applied to every patch in the weighted mean-squared-error (MSE) loss. By assigning greater weight to the centre of each patch, the loss focuses on the region of interest and reduces artefacts that can arise at patch borders. Training and validation are carried out on a single NVIDIA A100-SXM4

220 80 GB GPU using the Adam optimiser with an initial learning rate of 1×10^{-3} . Batches of sixteen patches are processed per optimisation step and gradient accumulation over sixteen steps yields an effective batch size of 256. The model is trained for 200 epochs, and the weighted loss and weighted root-mean-square error (RMSE) are logged for each epoch. The model state with the lowest validation RMSE is retained. During inference, the domain of reference is tiled into overlapping patches, which are processed independently, and the predictions are re-assembled to produce the final high-resolution field. A complete

225 training run (approximately 1.2 million training patches) requires roughly seven hours on the GPU, after which generating a single hourly field over the entire European domain takes only a few seconds.

2.6 Sampling optimisation

The downscaling framework is optimised in two successive steps by quantifying the influence of temporal and spatial sampling on the ability of the model to generalise to unseen situations.

230 The first optimisation step examines the influence of the temporal sampling strategy. Two RCAN models are trained on the reference domain using different subsets of the same year. The first model, referred to as the Winter-Model, is trained on the first 24 days of January 2021 and validated on the remaining four days of the month. January was selected because it is typically characterised by low temperatures, high pollutant concentrations and pronounced spatial gradients, providing a large diversity of pollution levels despite representing only winter conditions. The second model, referred to as the Year-Model, is trained

235 on the first six days of January, April, July and October 2021, while the seventh day of each month is used for validation. Although both models are trained and validated on the same total number of days, the Year-Model samples a much broader range of meteorological conditions and chemical regimes by spanning the four seasons. The corresponding learning curves are shown in the Supplementary Materials (Figure S3). Comparing these two strategies therefore allows us to determine whether model generalisation benefits more from exposing the network to intense winter pollution episodes or to a wider diversity

240 of atmospheric conditions. The two models are tested on independent Base Case simulations (February 08–12, May 08–12, August 08–12 and November 08–12) and on independent -50%ALL simulations (January 01–05, April 01–05, July 01–05 and October 01–05) over the domain of reference.

The second optimisation step investigates the influence of the spatial sampling strategy. The temporal sampling strategy that provides the best performance in the first optimisation step is retained, while training is conducted over the restricted

245 domain introduced in Section 2.1. Although this domain represents only about one-third of the reference domain, it was



specifically selected so that the statistical distributions of all predictors closely match those of the full European domain, while still encompassing a wide range of climatic conditions, emission patterns and chemical regimes. This allows assessing whether a training domain that is geographically smaller but statistically representative is sufficient for continental-scale generalisation. Two experiments are performed. In Experiment A, the original training hyperparameters are retained, so that the restricted domain provides approximately one-third as much training data. In Experiment B, the number of batches per epoch is tripled, allowing the network to process approximately the same effective number of training samples as when the reference domain is used. The corresponding learning curves are shown in the Supplementary Materials (Figure S4). Both experiments are tested on the domain of reference using the same Base Case and -50%ALL simulations, and their performance is compared with that of the model trained on the full reference domain.

3 Results

3.1 Temporal optimisation of training data

The performance of the network following the Winter-Model and Year-Model training strategies is assessed using the normalised RMSE and coefficient of determination with respect to the high-resolution reference simulation. These metrics are calculated for all grid cells and time steps unless mentioned otherwise. Figure 1 presents this evaluation for $\text{PM}_{2.5}$, NO_2 and O_3 for all simulations mentioned in Section 2.6. The effect of not downscaling ground concentrations is also assessed by calculating the performance metrics between high- and bilinearly interpolated low-resolution concentration fields, which constitute the main input to the downscaling model. For all pollutants, evaluation period or emission scenario, both downscaling models substantially outperform the bilinear interpolation of low-resolution fields, demonstrating that the super-resolution model successfully reconstructs a large fraction of the fine-scale spatial variability lost in the low-resolution simulations. The reduction in normalised RMSE is particularly pronounced for NO_2 , for which the error decreases by approximately 25–45% in comparison with the interpolated fields, while improvements of around 10–25% are obtained for $\text{PM}_{2.5}$ and O_3 . Overall, the model trained on samples distributed throughout the year generally yields the lowest errors. This behaviour is particularly evident for $\text{PM}_{2.5}$, for which this Year-Model consistently outperforms the Winter-Model for every evaluation period and for both Base Case and -50%ALL simulations. These improvements remain modest and never exceed 0.03 points, but their systematic nature indicates that exposing the network to a broader range of meteorological conditions and chemical regimes improves its ability to generalise to unseen EMEP simulations on the same domain. Besides, the largest differences are observed for the -50%ALL scenarios, suggesting that seasonal diversity in the training dataset also improves the robustness of the model to changes in emission levels. Similarly, the Year-Model generally performs better than the Winter-Model for O_3 , with lower normalised RMSE values obtained for most evaluation periods, particularly during summer and for the emission-reduction scenarios. This behaviour is consistent with the strong seasonal variability of ozone photochemistry, which is better represented when the training dataset spans multiple seasons. For NO_2 however, the differences between the two sampling strategies remain relatively small and no systematic advantage emerges. This limited sensitivity to the temporal sampling strategy may be due to the



stronger influence of local emission patterns on NO_2 concentrations, which are already well represented in the winter training period. These trends are also observed when the metric is changed from normalised RMSE to the coefficient of determination, whose corresponding figure is displayed in the Supplementary Materials (Figure S5).

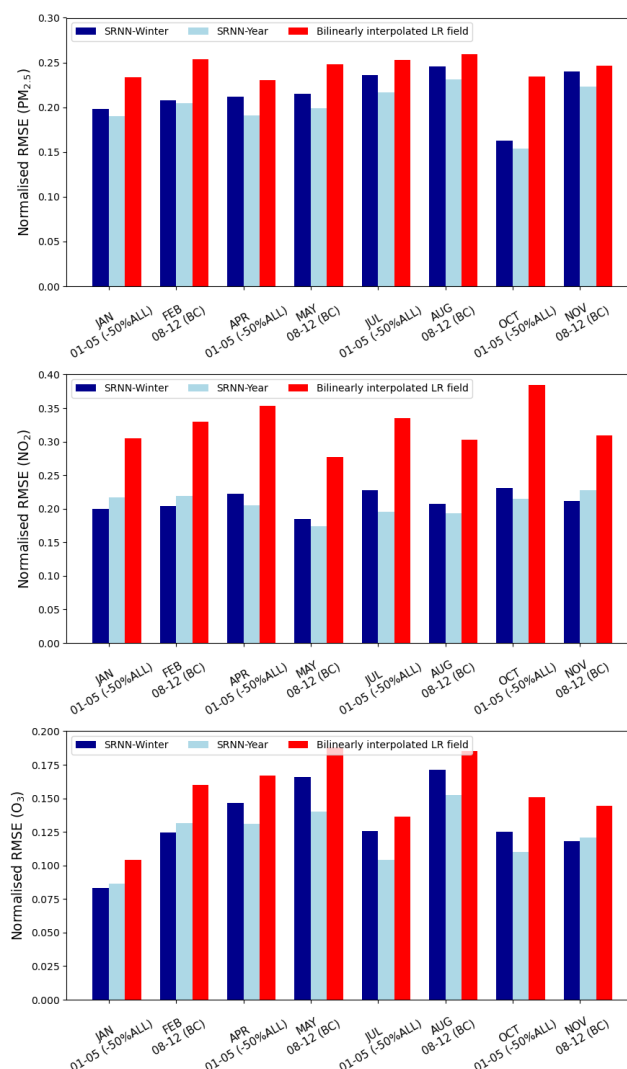


Figure 1. Comparison of the temporal sampling strategies used for training the downscaling model, using normalised RMSE values for $\text{PM}_{2.5}$ (top), NO_2 (middle) and O_3 (bottom) over independent Base Case (BC) and simulations where all pollutant emissions are halved (-50%ALL). Dark blue bars correspond to the model trained exclusively on winter data (Winter-Model), light blue bars to the model trained on samples distributed across the four seasons (Year-Model), and red bars to the bilinearly interpolated low-resolution concentration fields used as input for both downscaling models.



The observed differences between the enhancements done by the downscaling model can be interpreted considering the characteristic spatial scales and major controlling processes of each pollutant. Here, low general improvements over bilinearly interpolated low-resolution fields are obtained for $\text{PM}_{2.5}$ and O_3 , with an average improvement over the different tests of 0.044 and 0.032 respectively for the Year-Model. For $\text{PM}_{2.5}$, this low value is obtained because $\text{PM}_{2.5}$ concentrations result from the combined contributions of primary emissions, secondary aerosol formation, but also long-range transport. $\text{PM}_{2.5}$ is thus not rapidly removed chemically, and its regional concentration field generally exhibits spatial gradients that are observed on a spatial scale typically spanning several tens of kilometres. As a consequence, low-resolution simulations already capture a significant fraction of the large-scale spatial variability, limiting the improvement by the statistical downscaling. Similarly, the relatively low improvement for O_3 can be attributed to its relatively long atmospheric lifetime and the importance of regional transport, which result in spatially coherent concentration fields that are already substantially captured at low resolution. Conversely, the short atmospheric lifetime of NO_2 and its strong dependence on local anthropogenic emissions result in highly localised concentration gradients around major roads, urban centres, industrial facilities and shipping corridors. Such gradients are substantially smoothed out in coarse resolution simulations, and the downscaling process is able to reconstruct part of the missing fine-scale variability by exploiting the spatial relationships between emissions, meteorological fields and neighbouring grid cells, resulting in a performance improvement higher than for $\text{PM}_{2.5}$ and O_3 , with an average improvement over bilinearly interpolated low-resolution fields of 0.114.

The results of this first optimisation step indicate that increasing the seasonal diversity of the training dataset generally improves the generalisation capability of the downscaling model to EMEP simulations on the same domain, without compromising its performance during winter days. The Year-Model training strategy is thus retained for the rest of this study.

3.2 Spatial optimisation of training data

The influence of the geographic extent is assessed by comparing the results of Experiments A and B, in which the network is trained on the restricted domain, with those obtained using the network trained on the domain of reference. Training and validation periods correspond to those of the Year-Model strategy following the results presented in Section 3.1, i.e. six days by season. Figure 2 compares the performance of the three training strategies for each pollutant. All four Base Case and four –50%ALL simulations are used again for this assessment. In Experiment A, where the original training hyperparameters and the sampling configuration are retained, the use of the restricted training domain leads to a systematic degradation of the downscaling performance for the three pollutants in comparison with a model trained on the domain of reference. For $\text{PM}_{2.5}$, the increase in normalised RMSE is particularly pronounced, with errors exceeding those of the bilinearly interpolated low-resolution fields for all but one test. For NO_2 , a marked degradation is also observed, although the downscaling performance generally remains comparable to or slightly better than that of the bilinearly interpolated low-resolution field. The deterioration is less pronounced for O_3 , for which both experiments generally outperform the low-resolution interpolated fields. Increasing the number of batches per epoch in Experiment B slightly improves the downscaling performance for $\text{PM}_{2.5}$, but reduces it for NO_2 and O_3 , suggesting that a significant fraction of the degradation observed in Experiment A comes from the geographical



restriction itself. These trends are also observed when the metric is changed from normalised RMSE to the coefficient of
315 determination, whose corresponding figure is displayed in the Supplementary Materials (Figure S6).

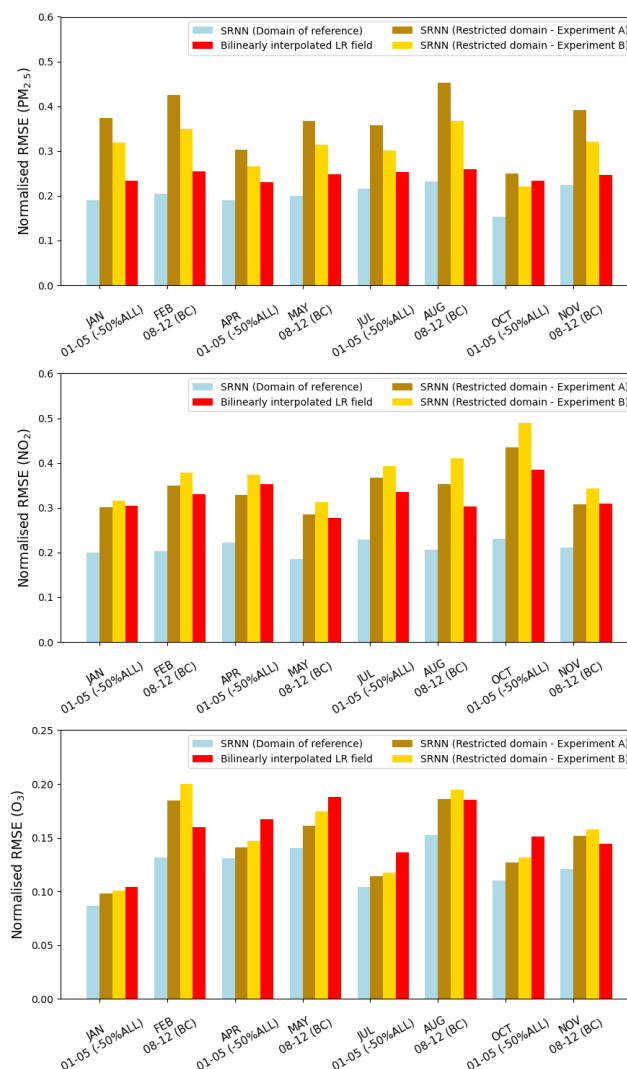


Figure 2. Comparison of the spatial sampling strategies used for training the RCAN downscaling model using normalised RMSE for $PM_{2.5}$ (top), NO_2 (middle) and O_3 (bottom) over independent Base Case (BC) and simulations where all pollutant emissions are halved (-50%ALL). Light blue bars correspond to the network trained on samples distributed across the four seasons and on the domain of reference, and the yellow bars correspond to the same network trained on the restricted domain while maintaining the volume of training data (light yellow) or not (dark yellow). Red bars correspond to the bilinearly interpolated low-resolution concentration fields used as input to the downscaling models.



Because the statistical distribution of predictors is similar for both domains on the training period, the performance degradation in Experiments A and B cannot be attributed to the network being exposed to an insufficiently wide range of predictor values. Instead, this suggests that reproducing the marginal distributions of the predictors is not sufficient to achieve the same level of generalisation. Here, the spatial distribution of those predictors differs between the two domains because of the variety of topography, urban structures, land-use patterns and synoptic meteorological conditions. A possible explanation for the observed degradation is that the larger geographical domain exposes the network to a wider diversity of spatial configurations linking emissions, meteorology and topography, resulting in higher performances when the network is trained on the domain of reference. This interpretation is consistent with the pollutant-dependent behaviour observed in Figure 1. $\text{PM}_{2.5}$ exhibits the largest degradation when the training domain is reduced, reflecting its dependence on coupled processes including primary emissions, secondary aerosol formation, regional transport and meteorology, whose interactions occur over a wide range of spatial scales. Conversely, NO_2 remains comparatively less sensitive to the geographical extent of the training domain because its fine-scale variability is primarily controlled by local emission sources, whose spatial characteristics are already well represented within the restricted domain. Finally, the degradation observed for O_3 is also less pronounced than for $\text{PM}_{2.5}$, with a performance similar between the super-resolved fields obtained by a training on the reduced domain and the bilinearly interpolated fields.

The results of this second optimisation step indicate that the spatial diversity of the training dataset is also an factor of improvement for the generalisation capability of the downscaling model. The training strategy combining the Year-Model on the domain of reference is thus retained for the rest of this study.

3.3 Generalisation to emission perturbation scenarios

Beyond the optimisation of the training configuration, an important requirement for applying a super-resolution neural network to air quality assessment is its ability to generalise to simulations performed over the same domain but under conditions not represented in the training dataset. These conditions can differ in terms of modified emission levels, precursor contributions or dominant source sectors. We evaluate this generalisation capability by testing whether the network can reconstruct high-resolution concentration fields associated with different chemical pathways linked to reductions in NH_3 , NO_x & PPM, VOCs & SO_x emissions. These perturbations modify the chemical regime using nonlinear paths (Thunis and Clappier, 2014; Lu et al., 2024). We also assess the ability of the network to preserve the patterns associated with different emission sources using scenarios with transport emission reductions, mainly affecting urban areas and road corridors, as well as residential emission reductions, mostly associated with population distribution.

These results are presented in Figure 3. Normalised RMSEs obtained by the neural network are compared with bilinear interpolation of the low-resolution EMEP simulations, providing a baseline for quantifying the added value of the downscaling process. For all tested emission scenarios and for all three target pollutants, the downscaled fields consistently outperform bilinearly interpolated fields. Indeed, the scores of bilinearly-interpolated LR fields with the HR reference are systematically



above 0.23 for $\text{PM}_{2.5}$, 0.27 for NO_2 and 0.10 for O_3 , with averages of 0.25, 0.32 and 0.16 respectively. A notable performance spread is observed across emission reduction scenarios. Conversely, the neural network reaches errors systematically lower than 0.24 for all scenarios and species, and with an average of 0.21 for $\text{PM}_{2.5}$, 0.21 for NO_2 , and 0.13 for O_3 . These error reductions are particularly clear for the NH_3 and the VOCs & SO_x emission reductions, which alter secondary inorganic and organic aerosol formation via coupled chemical pathways and modify the spatial distribution of chemically driven gradients. Here, this stable performance of the SRNN indicates a better reconstruction of fine-scale spatial variability than bilinearly interpolated LR fields, highlighting the reconstruction of concentration fields that are overall similar to those generated by HR simulations across different chemically perturbed conditions. These trends remain valid when using the coefficient of determination instead of the normalised RMSE, whose scores can be found in the Supplementary Materials (Figure S7).

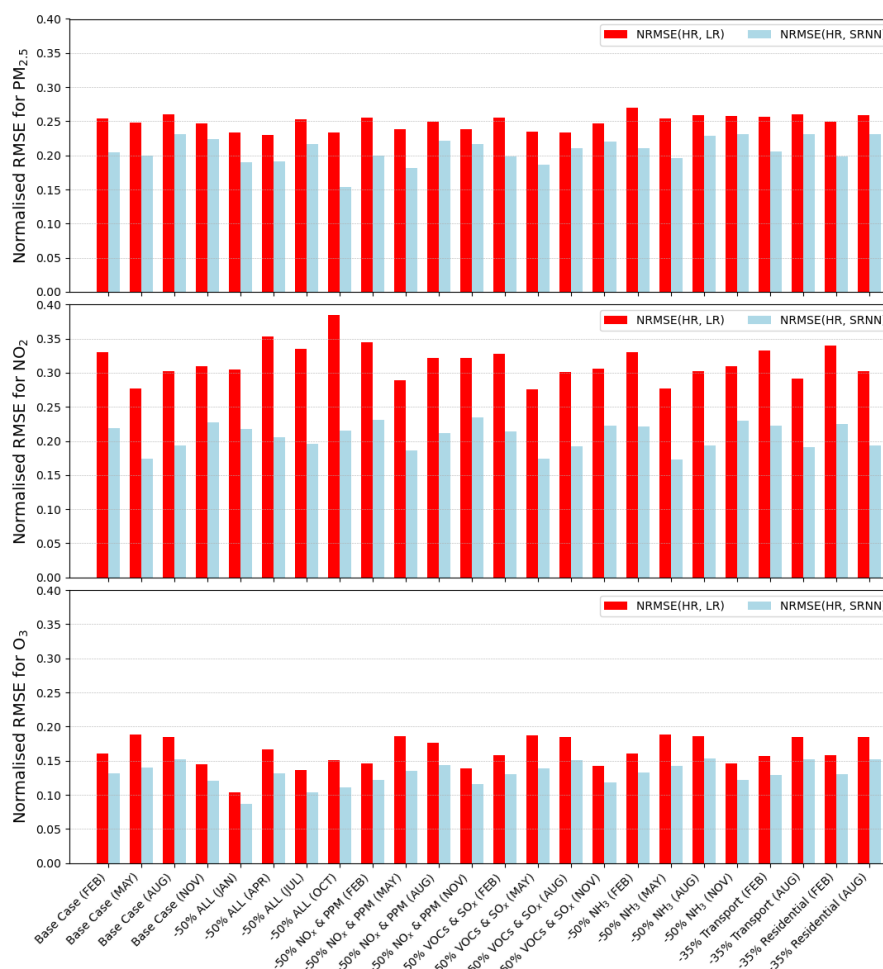


Figure 3. Normalised RMSE between high-resolution (HR) EMEP simulations and the corresponding super-resolved (blue) or bilinearly interpolated low-resolution (red) concentration fields for $\text{PM}_{2.5}$ (top), NO_2 (middle), and O_3 (bottom). Results are shown for all emission perturbation scenarios. Within each category, scenarios are ordered chronologically according to the simulation period.



Although this higher performance demonstrates the utility of the super-resolution framework, it does not reveal whether the model accurately reproduces the magnitude and spatial distribution of the strongest fine-scale responses to emission changes. These details require additional focus and cannot be fully characterised by metrics such as the normalised RMSE or the coefficient of determination. Indeed, a model may accurately reconstruct the concentration fields of both a reference and a perturbed scenario at large scale while failing to reproduce the fine-scale component of their response that is unresolved in the low-resolution simulations, particularly in areas with strong emission gradients such as city centres and industrial areas. Focusing on concentration differences between the Base Case and perturbed scenarios isolates the response to the emission reduction from the background concentration field and therefore provides a more direct assessment of the ability of the neural network to reproduce policy-relevant changes. This is particularly pertinent when emission changes involve nonlinear chemical processes, for which the spatial distribution of concentration responses may differ substantially between resolutions. We use the high- and low-resolution simulated concentrations for the simulated Base Case and perturbed scenario, noted as $C_{HR,BC}$, $C_{LR,BC}$, $C_{HR,S}$, $C_{LR,S}$, together with the corresponding super-resolved fields $C_{SRNN,BC}$ and $C_{SRNN,S}$, and define the scenario responses as $\Delta C_{HR} = C_{HR,BC} - C_{HR,S}$, $\Delta C_{LR} = C_{LR,BC} - C_{LR,S}$ and $\Delta C_{SRNN} = C_{SRNN,BC} - C_{SRNN,S}$. The fine-scale response enhancements are then defined as the response improvement generated by the transition from low- to high-resolution data, noted $\delta C_{HR} = \Delta C_{HR} - \Delta C_{LR}$ for the EMEP reference and $\delta C_{SRNN} = \Delta C_{SRNN} - \Delta C_{LR}$ for the downscaled fields. Figure 4 presents quantile–quantile plots comparing these two fine-scale response enhancements for the different emission-reduction scenarios. Particular attention is given to the lower and upper tails of the distributions, represented by the lightest and darkest shades in the figure respectively, as these correspond to pixels where the use of HR data is the most relevant for assessing scenario-induced concentration changes. This second analysis reveals three distinct behaviours in the ability of the SRNN to reproduce emission-induced concentration responses.

The first behaviour corresponds to cases where the concentration change enhancement at high resolution is correctly reconstructed by the neural network in magnitude, indicating that the network successfully captures the relationship between emission perturbations and concentration changes. In quantile-quantile plots, this behaviour is characterised by a close agreement in the amplitude of the extreme values of δC_{HR} and δC_{SRNN} for both the lower and upper tails of their distributions. This is mainly observed for scenarios involving direct modifications of dominant emission sources, such as transport and residential emission reductions affecting $PM_{2.5}$, NO_2 and O_3 , as well as NO_x & PPM reductions affecting NO_2 . In these cases, highest δC_{HR} and δC_{SRNN} values are both found where emissions generally reach a local maximum, i.e. in city cores and along major roads, whereas lowest values are found for pixels surrounding these hotspots, where concentration gradients are high. These spatial patterns are strongly linked to the high-resolution emission fields provided as predictors, allowing the network to preserve both the amplitude and spatial localisation of the induced concentration changes.

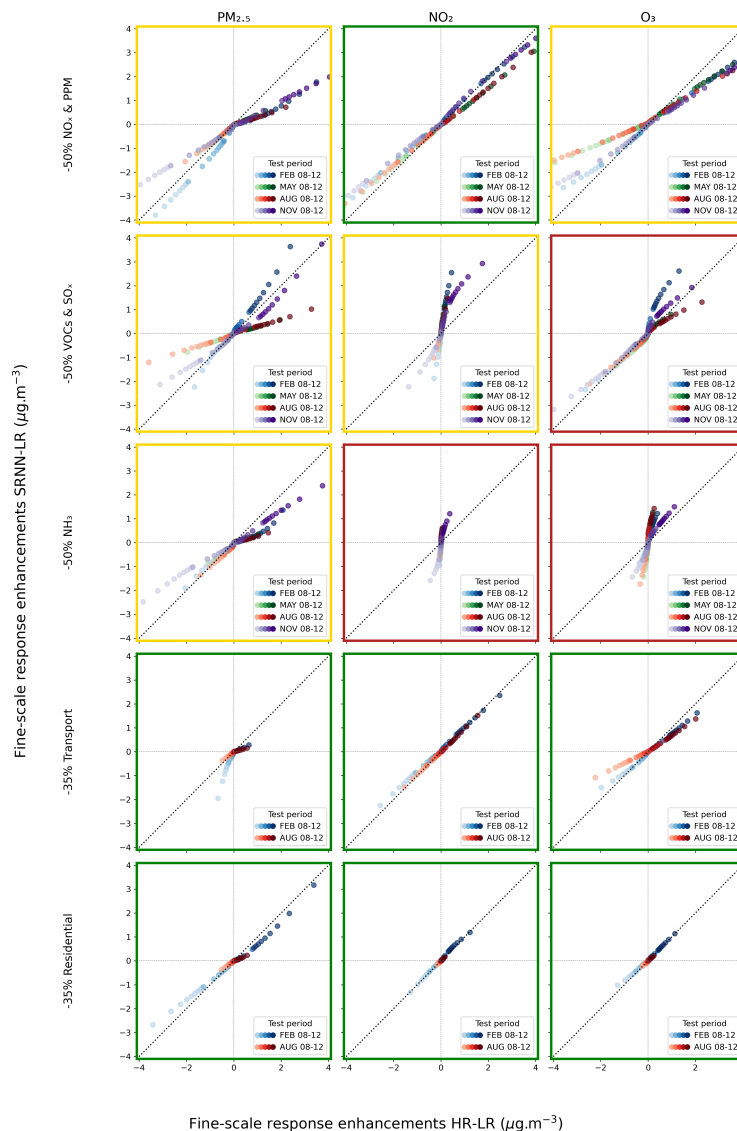


Figure 4. Quantile–quantile comparison of the fine-scale response enhancements obtained from the high-resolution references and the SRNN reconstructions. The plotted quantities represent the additional concentration changes revealed by the transition from low to high resolution, relative to the Base Case. For each panel, the x-axis represents $\delta C_{HR} = \Delta C_{HR} - \Delta C_{LR}$, i.e. the concentration change enhancement obtained from the high-resolution EMEP simulations relative to the corresponding low-resolution simulations, while the y-axis represents $\delta C_{SRNN} = \Delta C_{SRNN} - \Delta C_{LR}$, i.e. the corresponding response reconstructed by the SRNN. The left, middle and right columns correspond to $PM_{2.5}$, NO_2 and O_3 respectively. Each row corresponds to an emission-reduction scenario, and the different colour scales indicate the evaluated months. For each combination of pollutant, scenario and period, specific percentiles are displayed, with the nine lightest points corresponding to the lower tail of the distribution (P0.1–P0.9), followed by P1–P9, P10–P90, P91–P99 and the upper tail for the nine darkest points (P99.1–P99.9). The dotted line represents the 1:1 relationship (perfect agreement between HR EMEP response and SRNN reconstruction). The green, yellow or red frames correspond to the attribution of the first, second and third observed downscaling behaviours identified in Section 3.3.



The second behaviour is that of an over-smoothed reconstruction of concentration change enhancement, characterised by a good agreement in the spatial localisation of the extreme values of δC_{HR} and δC_{SRNN} , but with an attenuation of their amplitude in the SRNN reconstruction. The SRNN thus preserves the general spatial organisation of the response but underestimates the amplitude of the strongest concentration changes. The most extreme fine-scale response enhancements are therefore shifted towards less extreme values. Here, this behaviour is clearly observed for NH_3 , NO_x & PPM and VOCs & SO_x reductions affecting $PM_{2.5}$, as well as the NO_x & PPM reduction affecting O_3 . These situations include aerosol formation or photochemical processes that depend on nonlinear interactions between precursors, meteorology and chemical regimes that are modified. This underestimated response was already identified during the initial development of the framework by Beauchamp et al. (2026), which only studied the generalisation abilities on reductions of PPM or all pollutants at once on $PM_{2.5}$ and NO_2 winter concentrations, and reached ratios of super-resolution to high-resolution values of 0.68 and 0.52 for the distribution tails (percentiles 91-99 and 1-9 respectively). Here, in all 16 scenarios mentioned above, this ratio varies between 0.34 and 1.14 (mean: 0.72) for the extreme-lower tail (percentiles 0.1 to 0.9) and between 0.26 and 0.72 (mean: 0.56) for the extreme-upper tail (percentiles 99.1 to 99.9). Moreover, the locations of these strongest changes are similar for δC_{HR} and δC_{SRNN} values, which confirms that the model captures the main spatial organisation of these responses despite their reduced amplitude. An overestimation of the strongest concentration changes occurs only for the VOCs & SO_x reduction in February affecting $PM_{2.5}$, with ratios reaching 1.10 for the extreme-lower tail and 1.41 for the extreme-upper tail.

The third behaviour corresponds to cases where neither the amplitude nor the spatial localisation of the extreme values of δC_{HR} and δC_{SRNN} is correctly reproduced by the SRNN. This behaviour is particularly observed for NH_3 reduction affecting NO_2 and O_3 , and, to a certain extent, for VOCs & SO_x reduction affecting O_3 . In these cases, high δC_{SRNN} values are found where no clear high δC_{HR} values are located. This difference in spatial localisation is quantified by the spatial correlations between the HR and SRNN frequency maps of extreme additional concentration responses, which are presented in the Supplementary Materials (Table S1). Spatial correlations are generally high for the NO_x & PPM, Transport, and Residential scenarios, indicating a good preservation of the spatial organisation of the strongest additional responses. In contrast, substantially lower correlations are obtained for the VOCs & SO_x and NH_3 reductions, particularly for the upper-tail responses in the case of O_3 . This failure to reproduce both the magnitude and the spatial organisation of the most extreme fine-scale response enhancements contrasts with the first two behaviours.

To illustrate these differences in the spatial localisation of the extreme responses, Figure 5 shows the temporal frequency with which each grid cell exhibits very low or very high fine-scale response enhancements during the simulation period. For each time step, the 10th and 90th percentiles of the fine-scale NO_2 response enhancement distribution are first determined, and the frequency with which each grid pixel falls below the 10th percentile or above the 90th percentile is then calculated over the entire simulation period. Results are shown for the -50% NO_x & PPM scenario in November, which is representative of the first behaviour, and the -50% NH_3 scenario in November, which is representative of the third behaviour. For the -50% NO_x & PPM scenario, the areas that frequently exhibit extreme additional concentration responses are broadly consistent between the HR simulation and the SRNN reconstruction, particularly around major emission hotspots, suggesting that the network



correctly captures the spatial localisation of the response. In contrast, for the NH_3 reduction, the SRNN frequently produces extreme fine-scale response enhancements in several areas where corresponding extreme responses are absent or much less frequent in the HR reference.

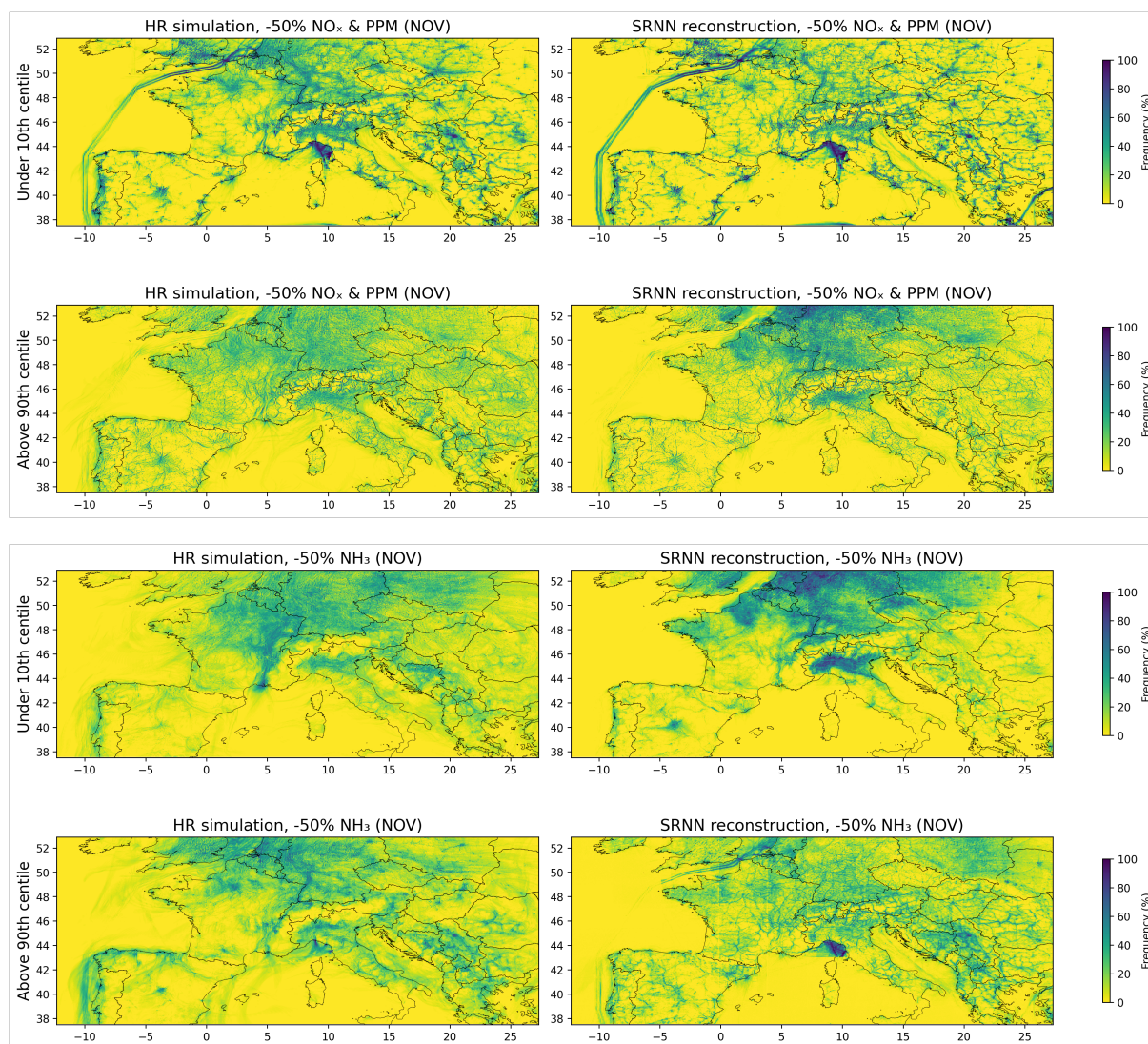


Figure 5. Frequency of extreme fine-scale NO_2 response enhancements brought by the high-resolution (HR) EMEP simulations (left) or the corresponding super-resolution (SRNN) reconstructions (right) from low-resolution simulations in November. Each grid cell represents the frequency (as percentage of time steps in the 5-day scenario) with which the additional NO_2 concentration response δC_{HR} or δC_{SRNN} falls below the spatial 10th percentile (first and third rows) or above the spatial 90th percentile (second and fourth rows) of the corresponding hourly fine-scale response enhancement field. Results are shown for the -50% NO_x & PPM scenario (first and second rows) and the -50% NH_3 scenario (third and fourth rows), both in November.



3.4 Analysis of small-scale details around urban areas

Another reconstruction aspect that cannot fully be assessed using domain-wide statistical metrics is the quantification of reconstruction artefacts where spatial structures are relevant for local air quality assessment. In particular, urban environments are characterised by strong concentration gradients associated with traffic emissions, residential heating, urban morphology and local meteorological conditions, which are smoothed out in coarse-resolution simulations. We therefore examine the spatial distribution of reconstructed fields at the scale of individual urban areas for all scenarios, including the REGFUA scenario described in Section 2.4. In this scenario, residential emissions in the urban areas of Bordeaux (France), Naples (Italy) and Prague (Czechia) are reduced by 50%, transport emissions in those of Valencia (Spain), Frankfurt (Germany) and Katowice (Poland) are reduced by 50%, and total emissions in the regions of Brittany (France), Tuscany (Italy) and Macroeconomy One (Romania), whose regional capitals are Rennes, Florence and Cluj-Napoca respectively, are reduced by 35%. The differences between reconstructed concentration fields and high-resolution reference are evaluated on a 60×45 -cell domain ($\sim 100 \times 50$ km²) centred around these nine urban areas. For each city and emission scenario, we calculate at each grid cell the frequency at which the relative absolute difference $|C_{HR} - C_{SRNN}|/C_{HR}$ between the reconstructed and reference concentration exceeds 25%. This threshold is used as a pragmatic criterion to distinguish substantial local discrepancies from smaller reconstruction differences. This allows identifying locations where discrepancies persist over time. This assessment is performed only for the scenarios in February, which is the month common to all simulated scenarios. In the Supplementary Materials (Figure S8), the resulting frequency maps are displayed for PM_{2.5}, while Figure 6 focuses on the example of Prague, Czechia, for the REGFUA scenario. Discrepancies exceeding 25% are generally localised rather than uniformly distributed across the urban areas. Most domains exhibit relatively low frequencies, with 90% of the cells exceeding this threshold less than 2 out of 120 time steps around Rennes and 21 out of 120 time steps around Naples. This denotes the presence of slight reconstruction artefacts, while higher frequencies that characterise persistent artefacts are concentrated in specific parts of the domains and often form spatially coherent structures. This is the case in Prague, where discrepancies concern cells with frequencies higher than 50/120 time steps outside the urban area, above a zone that corresponds to a coal mining site around (50.5°N, 13.5°E) visible on Figure 6. This is also the case in and Katowice for another mining site around (50.0°N, 18.6°E). By contrast, several other cities show predominantly low-frequency discrepancies over most of their domains. The frequency of persistent or non-persistent discrepant cells does not appear to change across scenarios, including the local and regional emission-reduction scenario for which there is no systematic increase in reconstruction errors for all nine urban areas. This suggests that the SRNN can retain fine-scale information under spatially heterogeneous emission perturbations, although some local structures remain difficult to reconstruct accurately. Finally, it must be noted that the spatial patterns observed for PM_{2.5} are broadly similar for NO₂ and O₃, although mean frequencies are generally lower in the case of O₃. These results complement the domain-wide evaluation presented in Section 3.3 by showing that good overall statistical performance does not necessarily imply homogeneous accuracy at the urban scale. The presence of localised discrepancies does not indicate a general failure of the downscaling approach, as most of the urban domains remain characterised by relatively low frequencies of large relative errors. The analysis therefore highlights the ability of the SRNN to recover the spatial heterogeneity of PM_{2.5} concentrations while identifying specific loca-



tions where the reconstruction remains less reliable. Such information is particularly relevant for applications involving urban exposure assessment or spatially targeted emission-reduction strategies.

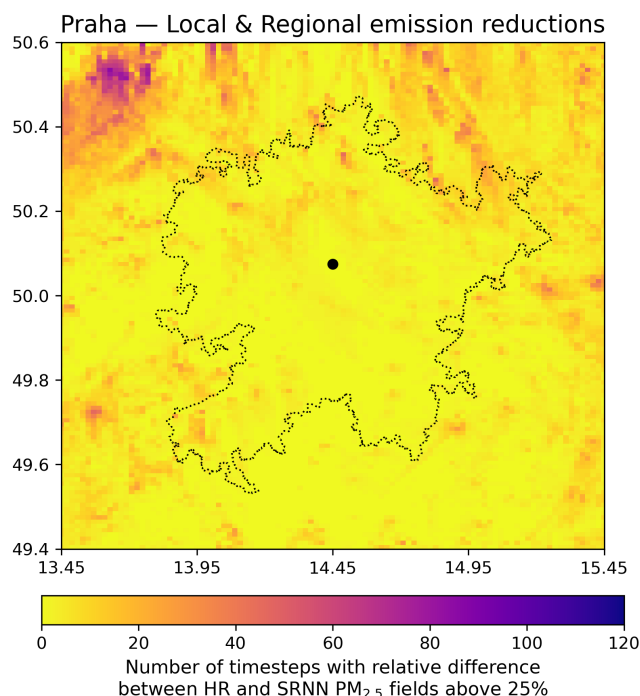


Figure 6. Frequency of large relative differences between the high-resolution EMEP simulations and the corresponding SRNN reconstructions of $\text{PM}_{2.5}$ concentration around the urban area of Prague (Czechia) for the REGFUA scenario in February. For each grid pixel, the colour indicates the number of time steps for which the absolute relative difference between the SRNN and HR concentrations exceeds 25%. The black dotted line indicate the boundaries of the functional urban area and the black point the pixel of highest mean concentration in the scenario.

460 3.5 Influence of temporal averaging on downscaling performance

The non-persistent artefacts studied in the previous subsection can be progressively smoothed out when the downscaled fields are averaged over longer periods. However, this reduction comes at the expense of temporal detail, since averaging also removes genuine short-term variability in pollutant concentrations. To quantify this trade-off, we calculate the evolution of the normalised RMSE between the temporally averaged super-resolution fields and the corresponding high-resolution reference as a function of the averaging horizon. For each time step in an emission-reduction scenario, the bilinearly interpolated low-resolution and super-resolution concentration fields are averaged from the beginning of the evaluation period up to the current time step. The resulting temporally averaged fields are then compared with the corresponding temporally averaged HR reference. The results are shown in Figure 7.

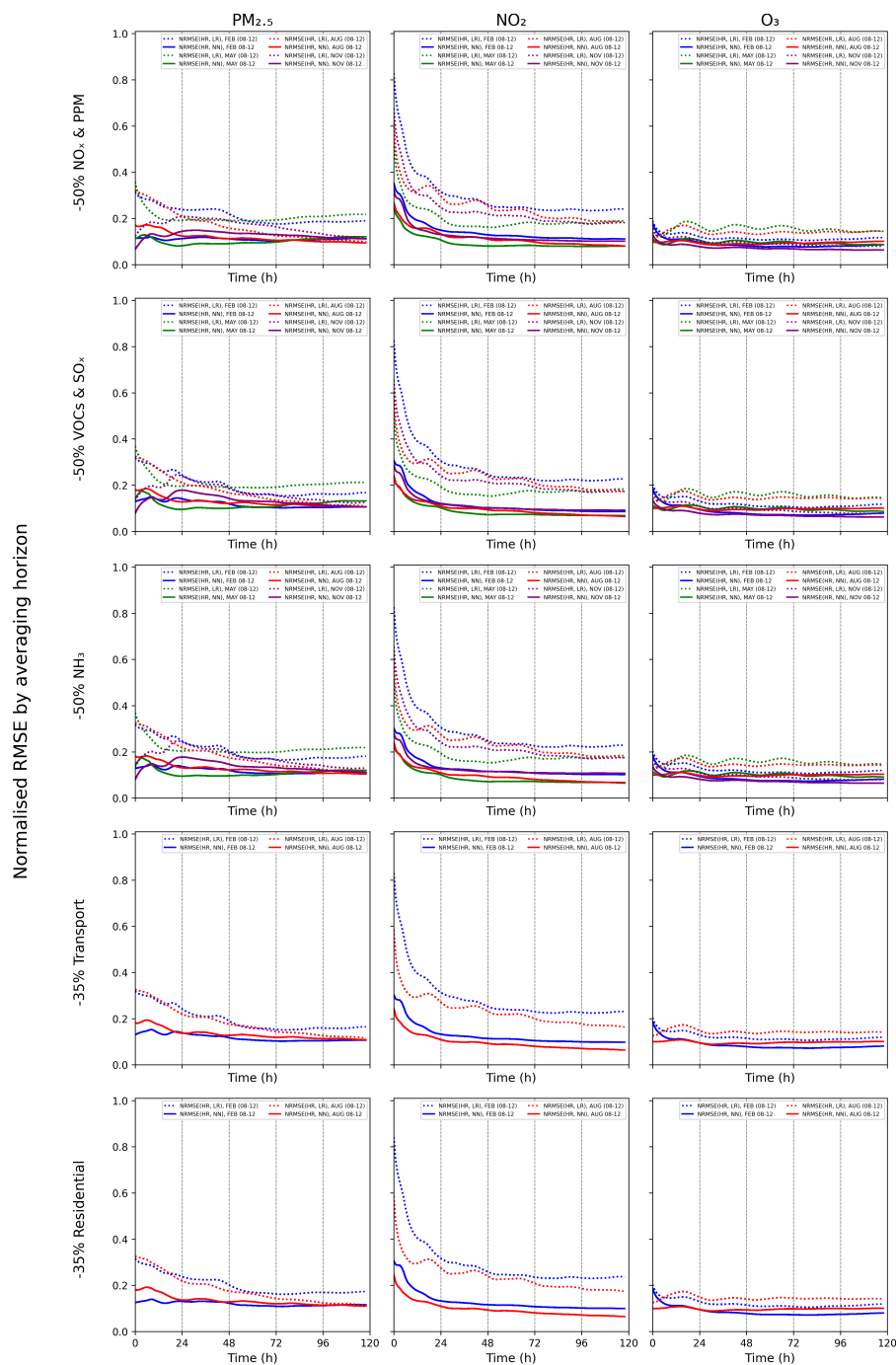


Figure 7. Evolution of normalised RMSE between the low-resolution (dotted lines) or super-resolution (solid lines) concentration fields and the high-resolution reference as a function of the averaging horizon. The left, middle and right columns correspond to the $PM_{2.5}$, NO_2 and O_3 evaluations, respectively. Each row corresponds to an emission-reduction scenario.



As the averaging horizon increases, the spatial artefacts created by the neural network are progressively smoothed. At the same time, temporal averaging also removes part of the fine-scale pollution structures present in the HR reference at the hourly scale. These two effects lead to a decrease of the error with increasing averaging horizon. This decrease is consistently observed for $\text{PM}_{2.5}$, NO_2 and O_3 across all considered evaluation periods and emission scenarios. This trend is also obtained when the averaging is initiated at any other time step, indicating that the observed trend does not depend on the initial point of the averaging period. The strongest reduction in normalised RMSE occurs during the first few tens of hours, after which the improvement progressively levels off. Beyond a certain temporal scale, fine-scale structures that are absent from the LR input or that are not perfectly reconstructed by the neural network are thus smoothed. The resulting error plateau therefore reflects a trade-off between reducing short-term reconstruction errors and progressively removing temporal and spatial details that are present in the HR reference. A one-day averaging horizon can therefore be considered a useful compromise between retaining temporal resolution and reducing discrepancies between the downscaled and HR fields.

4 Discussion

This study explores how the composition of the training dataset affects the performance and generalisation ability of a deep learning-based method for downscaling low-resolution ground-level concentrations of $\text{PM}_{2.5}$, NO_2 and O_3 over Europe. The results show that the diversity of the training data plays an important role in determining model performance. In terms of temporal sampling, training across the four seasons generally outperforms an equivalent amount of winter-only training data. Similarly, sampling across a larger training domain yields better performance than training on a more restricted domain, even when the statistical distributions of predictors are deliberately selected to match those of the larger domain. This indicates that model performance depends not only on the amount of training data or on the range of predictor values, but also on the diversity of spatial configurations and local relationships represented in the training samples. This is particularly relevant for convolutional neural networks, which exploit spatial relationships between neighbouring predictors rather than treating each variable independently. Here, $\text{PM}_{2.5}$ and O_3 benefit more from sampling throughout the year than from throughout winter only, because their spatial variability is strongly influenced by meteorological and chemical regimes. In contrast, NO_2 benefits more from super-resolution in general, because its local gradients are substantially finer than the low-resolution grid, leaving considerable fine-scale information to be reconstructed, while a larger fraction of $\text{PM}_{2.5}$ and O_3 variability occurs at regional scales already represented at low resolution.

The optimised model also generalises to a wide range of emission-reduction scenarios that differ from those represented during training, including changes in emission levels, precursor species and activity sectors, without requiring retraining. This suggests that the learned downscaling relationships are not restricted to a specific emission configuration. However, the results also reveal an important distinction between reconstructing the concentration fields and reconstructing their responses to emission perturbations. The model can reconstruct plausible fine-scale concentration structures for a wide range of scenarios, while some of the corresponding emission-induced changes are less reliably reproduced. This distinction is particularly relevant



for policy applications, where the quantity of interest is often the change in concentration resulting from a mitigation measure rather than the absolute concentration itself. Here, the ability to reproduce the fine-scale component of the concentration response appears to depend on how strongly and directly the target pollutant responds to the perturbed emissions. The optimised neural network reproduces many of the strongest fine-scale responses in both location and amplitude, particularly when the affected pollutant is closely linked to the perturbed emission source. However, weaker responses associated with indirect and nonlinear chemical interactions are more difficult to reconstruct. In most cases, the model identifies the location of these fine-scale responses but sometimes underestimates their amplitude. In extreme cases, when the emission perturbation produces only a very weak fine-scale response in the target pollutant, the network may produce strong local responses that do not coincide with those simulated by the high-resolution EMEP model. For complex emission scenarios, the suitability of the downscaling model should therefore be assessed against the expected magnitude and spatial localisation of the concentration changes induced by the emission perturbations.

Despite these limitations, the model shows strong potential for rapidly screening multiple emission scenarios without requiring high-resolution CTM simulations. This is particularly relevant for policy applications, where many emission-reduction strategies may need to be explored across different pollutants, sectors and geographical areas. The SRNN does not replace fully-fledged simulations in representing atmospheric chemistry or responses to an emission perturbation, but it provides a computationally inexpensive way of reconstructing fine-scale spatial structure from low-resolution simulations that already account for these processes, with a few seconds required to downscale a single time step of low-resolution inputs.

Finally, the generalisation capacity has been assessed for different scenarios, but with tests remaining within the domain of reference and using EMEP simulations for the year 2021. It is therefore unclear whether the same performance would be obtained for other regions, other years or on simulations of other CTMs. In addition, the present evaluation measures the ability of the SRNN to reproduce the fine-scale information contained in high-resolution EMEP simulations rather than observations. Future studies should therefore assess generalisation across independent meteorological years, geographical regions and CTMs, and evaluate the ability of the network to reproduce measured concentrations. Training strategies specifically designed to improve the reconstruction of weak, chemically mediated responses could also be investigated.

525 5 Conclusions

In this study, we investigated the ability of machine learning-based approach to downscale hourly $\text{PM}_{2.5}$, NO_2 and O_3 concentrations over Europe, with particular emphasis on the influence of the spatial and temporal composition of the training dataset and on the generalisation to emission-reduction scenarios. Training across the four seasons improves generalisation compared with an equivalent winter-only dataset, particularly for $\text{PM}_{2.5}$ and O_3 , while the benefit is limited for NO_2 . Similarly, training over a larger geographical domain improves performance even when the predictor distributions match those of a restricted domain. These results indicate that, for convolutional architectures, the diversity of spatial configurations and local predictor relationships in the training data is important beyond the breadth of the underlying statistical distributions.



Based on the results above, the optimised training strategy combines year-round sampling over a geographical domain covering most of Europe. The resulting model correctly reconstructs absolute concentration fields at high resolution. It consistently outperforms bilinearly interpolated low-resolution fields by successfully reconstructing a substantial part of the fine-scale spatial variability represented in the EMEP simulations. It also generalises to a range of emission-reduction scenarios involving different emission levels, precursor species and activity sectors without requiring retraining. However, accurate reconstruction of absolute concentration fields does not necessarily imply equally accurate reconstruction of their fine-scale responses to emission perturbations. The strongest fine-scale components of the concentration response are well reproduced when the affected pollutant is directly linked to the perturbed emission source, whereas weaker responses mediated through indirect chemical interactions can be attenuated or incorrectly localised. Temporal averaging provides a complementary way to reduce reconstruction artefacts, which are mostly non-persistent: errors decrease rapidly over the first few tens of hours and progressively level off, with approximately one day providing a useful compromise between reducing short-term reconstruction errors and retaining temporal detail.

Overall, these results demonstrate the potential of neural-network-based downscaling as a computationally efficient approach for rapidly screening high-resolution air quality scenarios. Future studies should therefore assess the generalisation of such architectures across independent meteorological years, geographical regions and chemical transport models, and evaluate their ability to reproduce both measured concentrations and emission-induced concentration changes across a wide range of response magnitudes.

Code and data policy. The SRNN code, configuration files, pretrained model checkpoint, input data for the scenario with halved emissions ("-50%ALL") on a reduced domain and period, and evaluation notebooks are archived at <https://doi.org/10.5281/zenodo.23376831> (Rey-Pommier et al., 2026). The repository provides the resources required to reproduce the downscaling of NO₂ concentrations and evaluate model performance.

Author contributions. The conceptualization was done by MB, BB, EP, SZS and PT. The neural network architecture was developed by MB. The EMEP simulations were performed by AdM. Data curation was done by ARP. The training and testing of the neural network were done by ARP. The formal analysis and visualization were done by ARP. The article was written by ARP. All authors contributed to the writing, reviewing, and editing of the paper. They have read and agreed to the published version of the manuscript.

Competing interests. The contact author has declared that none of the authors has any competing interests.



560 References

- Anderson, J.D.; Wendt, J. *Computational Fluid Dynamics*; McGraw-Hill: New York, NY, USA, **1995**; Volume 206, p. 332.
- Beauchamp, M.; Bessagnet, B.; Pisoni, E.; Rey-Pommier, A.; de Meij, A.; Thunis, P. Enhancing air quality simulations with neural down-scaling architectures. *Atmospheric Science Letters* **2026**, *27*, e70003.
- Bessagnet, B.; Beauchamp, M.; Menut, L.; Fablet, R.; Pisoni, E.; Thunis, P. Deep learning techniques applied to super-resolution chemistry
 565 transport modeling for operational uses. *Environmental Research Communications* **2021**, *3*, 085001.
- Bessagnet, B.; Pisoni, E.; de Meij, A.; Létinois, L.; Thunis, P. A simple and fast method to downscale chemistry transport model output fields from the regional to the urban/district scale. *Environmental Modelling & Software* **2023**, *164*, 105692.
- Bowman, F.M.; Seinfeld, J.H. Fundamental basis of incremental reactivities of organics in ozone formation in VOC/NO_x mixtures. *Atmospheric Environment* **1994**, *28*, 3359–3368.
- 570 de Meij, A.; Astorga, C.; Thunis, P.; Crippa, M.; Guizzardi, D.; Pisoni, E.; Valverde, V.; Suarez-Bertoa, R.; Oreggioni, G.D.; Mahiques, O.; Franco, V. Modelling the impact of the introduction of the EURO 6d-TEMP/6d regulation for light-duty vehicles on EU air quality. *Applied Sciences* **2022**, *12*, 4257.
- European Environment Agency. Air Quality in Europe – 2020 Report; EEA Report No. 09/2020; European Environment Agency: Copenhagen, Denmark, **2020**.
- 575 European Union. Directive (EU) 2024/2881 of the European Parliament and of the Council of 23 October 2024 on Ambient Air Quality and Cleaner Air for Europe (Recast). *Official Journal of the European Union* **2024**, 20 November 2024, PE/88/2024/REV/1.
- Fagerli, H.; Benedictow, A.; van Caspel, W.; Gauss, M.; Ge, Y.; Jonson, J.E.; Klein, H.; Nýfri, A.; Simpson, D.; Tsyro, S.; et al. *Transboundary Particulate Matter, Photo-Oxidants, Acidifying and Eutrophying Components*; Norwegian Meteorological Institute: Oslo, Norway, **2023**.
- Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial networks.
 580 *Communications of the ACM* **2020**, *63*, 139–144.
- Health Effects Institute. State of Global Air 2024: Special Report; Health Effects Institute: Boston, MA, USA, **2024**.
- Kuenen, J.; Dellaert, S.; Visschedijk, A.; Jalkanen, J.-P.; Super, I.; Denier van der Gon, H. CAMS-REG-v4: A state-of-the-art high-resolution European emission inventory for air quality modelling. *Earth System Science Data* **2022**, *14*, 491–515.
- Lu, N.; Zhang, L.; Wang, X.; Liu, Z.; Li, D.; Xu, J.; Tan, H.; Zhou, M.; Henze, D.K. Assessing the nonlinearity of wintertime PM_{2.5} formation
 585 in response to precursor emission changes in North China with the adjoint method. *Environmental Research Letters* **2024**, *19*, 084048.
- Ma, Z.; Liu, R.; Liu, Y.; Bi, J. Effects of air pollution control policies on PM_{2.5} pollution improvement in China from 2005 to 2017: A satellite-based perspective. *Atmospheric Chemistry and Physics* **2019**, *19*, 6861–6877.
- Maroia, S.; Inambao, F. The NO_x Formation Routes. In *Biodiesel, Combustion, Performance and Emissions Characteristics*; Springer International Publishing: Cham, Switzerland, **2020**; pp. 7–15.
- 590 McDuffie, E.E.; Martin, R.V.; Spadaro, J.V.; Burnett, R.; Smith, S.J.; O'Rourke, P.; Hammer, M.S.; van Donkelaar, A.; Bindle, L.; Shah, V.; et al. Source sector and fuel contributions to ambient PM_{2.5} and attributable mortality across multiple spatial scales. *Nature Communications* **2021**, *12*, 3594.
- Mu, Q.; Denby, B.R.; Wærsted, E.G.; Fagerli, H. Downscaling of air pollutants in Europe using uEMEP_v6. *Geoscientific Model Development* **2022**, *15*, 449–465.
- 595 Nuterman, R.; Mahura, A.; Baklanov, A.; Amstrup, B.; Zakey, A. Downscaling system for modeling of atmospheric composition on regional, urban and street scales. *Atmospheric Chemistry and Physics* **2021**, *21*, 11099–11112.



- OECD. *Redefining "Urban": A New Way to Measure Metropolitan Areas*; OECD Publishing: Paris, France, **2012**.
- Rey-Pommier, A.; Pisoni, E.; Beauchamp, M.; Bessagnet, B.; de Meij, A.; Zauli Sajani, S.; Thunis, P. Deep learning-based downscaling of PM_{2.5}, NO₂ and O₃ over Europe across emission-reduction scenarios. *Zenodo*, **2026**, <https://doi.org/10.5281/zenodo.23376831>.
- 600 Serifi, A.; Günther, T.; Ban, N. Spatio-temporal downscaling of climate data using convolutional and error-predicting neural networks. *Frontiers in Climate* **2021**, *3*, 656479.
- Simpson, D.; Benedictow, A.; Berge, H.; Bergström, R.; Emberson, L.D.; Fagerli, H.; Flechard, C.R.; Hayman, G.D.; Gauss, M.; Jonson, J.E.; et al. The EMEP MSC-W chemical transport model—technical description. *Atmospheric Chemistry and Physics* **2012**, *12*, 7825–7865.
- Soares, P.M.; Johannsen, F.; Lima, D.C.; Lemos, G.; Bento, V.A.; Bushenkova, A. High-resolution downscaling of CMIP6 Earth system and
605 global climate models using deep learning for Iberia. *Geoscientific Model Development* **2024**, *17*, 229–259.
- Song, Y.; Sohl-Dickstein, J.; Kingma, D.P.; Kumar, A.; Ermon, S.; Poole, B. Score-based generative modeling through stochastic differential equations, **2020**.
- Tarrasón, L.; Fagerli, H.; Jonson, J.E.; Klein, H.; van Loon, M.; Simpson, D.; Tsyro, S.; Vestreng, V.; Wind, P.; Posch, M.; Solberg, S.; Spranger, T.; Thunis, P.; White, L. *Transboundary Acidification, Eutrophication and Ground Level Ozone in Europe*; Norwegian Meteorological Institute: Oslo, Norway, **2004**.
610
- Thunis, P.; Clappier, A. Indicators to support the dynamic evaluation of air quality models. *Atmospheric Environment* **2014**, *98*, 402–409.
- Thunis, P.; Pisoni, E.; Zauli-Sajani, S.; de Meij, A.; Rey-Pommier, A.; Pekar, F. *Urban PM_{2.5} Atlas: Air Quality in European Cities*; **2025** Report.
- Turnock, S.T.; Butt, E.W.; Richardson, T.B.; Mann, G.W.; Reddington, C.L.; Forster, P.M.; Haywood, J.; Crippa, M.; Janssens-Maenhout, G.; Johnson, C.E.; et al. The impact of European legislative and technology measures to reduce air pollutants on air quality, human health
615 and climate. *Environmental Research Letters* **2016**, *11*, 024010.
- Economic Commission for Europe. *Guidelines for Reporting Emissions and Projections Data under the Convention on Long-Range Transboundary Air Pollution*, **2023**.
- Vardoulakis, S.; Fisher, B.E.A.; Pericleous, K.; Gonzalez-Flesca, N. Modelling air quality in street canyons: A review. *Atmospheric Environment* **2003**, *37*, 155–182.
620
- Wang, F.; Tian, D.; Lowe, L.; Kalin, L.; Lehrter, J. Deep learning for daily precipitation and temperature downscaling. *Water Resources Research* **2021**, *57*, e2020WR029308.
- World Health Organization. WHO Global Air Quality Guidelines: Particulate Matter (PM_{2.5} and PM₁₀), Ozone, Nitrogen Dioxide, Sulfur Dioxide and Carbon Monoxide: Executive Summary; World Health Organization: Geneva, Switzerland, **2021**.
- 625 World Health Organization. Sustainable Development Goal Indicator 3.9.1: Mortality Attributed to Air Pollution; World Health Organization: Geneva, Switzerland, **2024**.
- Zhang, Y.; Li, K.; Li, K.; Wang, L.; Zhong, B.; Fu, Y. Image super-resolution using very deep residual channel attention networks. In *European Conference on Computer Vision*; Springer International Publishing: Cham, Switzerland, **2018**; pp. 294–310.