Universidad de Chile
Facultad del Ciencias Físicas y Matemáticas
Departamento de Ingeniería Mecánica

Improving OMI-NO_2 spatial resolution using a Stochastic Convolutional Neural Network over central and southern Chile

Tesis para optar al grado de
Magíster en Ciencias de la Ingeniería, mención Mecánica

Memoria para optar al título de
Ingeniero Civil Mecánico

Santiago Nicolás Parraguez Cerda

Profesora guía
Viviana Meruane Naranjo

Profesora co-guía
Laura Gallardo Klenner

Comisión
Axel Osses Alvarado

Santiago de Chile
2021

Abstract of the thesis for the degree of Master of Engineering Sciences in Mechanics and the degree of Mechanical Engineer.
Santiago Parraguez Cerda
November, 2021
Advisor: Viviana Meruane Naranjo

Improving OMI-NO_2 spatial resolution using a Stochastic Convolutional Neural Network over central and southern Chile

Both precision and spatial resolution of atmospheric data are fundamental aspects when evaluating air quality and the effects of different factors on the distribution of air pollutants. Satellite-borne instruments have been very useful to collect this type of data, providing valuable information for process understanding and when deciding public policies of both regional and national relevance. However, as sensors and satellites retrievals have improved, one finds discontinuities in quality and resolution that require homogenisation for establishing long-term trends.

This work seeks to assess the feasibility of improving the spatial resolution of satellite measurements while, simultaneously estimating the expected error. A stochastic approach based on Convolutional Neural Networks is presented, which succeeds in increasing the spatial resolution of nitrogen dioxide (NO_2) columns collected with spectrometers onboard satellites. Furthermore, the approach allows estimating the aleatoric uncertainty generated by errors inherent in spectrometer measurements. Also, the methodology allows achieving the estimation precision of conventional deep learning models. The results show that the reconstructed fields are robust to added noise on the data, presenting slight decreases in the evaluated metrics above 5% noise. An advantage of the presented methodology is that the models can be trained with small-scale images, and then applied without domain restriction, if the resolution used during training is maintained.

The results indicate that the methodology is appropriate for the stated objectives. Also, it is subject to further improvement by considering state-of-the-art models (ResNet, GAN). An application to reconstruct NO_2 column data from the Ozone Monitoring Instrument (OMI) onboard the Aura satellite is shown, illustrating the potential of the methods. Thus, this work contributes to an improvement in the monitoring of air quality for the country, and it is expected that it can be applied to obtain better prediction results (both precision and error estimation) and cover a larger area of application.

Resumen de la tesis para optar al grado de Magíster en Ciencias de la Ingeniería, mención Mecánica y al título de Ingeniero Civil Mecánico.
Santiago Parraguez Cerda
Noviembre, 2021
Profesora guía: Viviana Meruane Naranjo

Mejora de la resolución espacial de datos OMI de NO_2 utilizando una Red Neuronal Convolucional Estocástica sobre Chile centro-sur

La precisión y la resolución espacial de datos atmosféricos son aspectos fundamentales a la hora de evaluar la calidad del aire y los efectos de distintos factores sobre la distribución de contaminantes. Los instrumentos de medición satelital han sido de gran utilidad para recopilar este tipo de datos, proveyendo información valiosa para entender procesos y decidir políticas públicas con relevancia tanto regional como nacional. Sin embargo, a medida que mejora la calidad de sensores y satélites, se presentan discontinuidades tanto en la calidad como en la resolución de los datos, lo que requiere métodos de homogenización para poder establecer tendencias a largo plazo.

Este trabajo busca evaluar la factibilidad de mejorar la resolución espacial de imágenes satelitales de contaminantes y, al mismo tiempo, estimar el error esperado. Se presenta una metodología estocástica basada en Redes Neuronales Convolucionales, que logra aumentar la resolución espacial de datos de dióxido de nitrógeno (NO_2) obtenidos a partir de espectrómetros satelitales. Además, los modelos presentados permiten estimar la incertidumbre aleatoria generada por errores inherentes a las mediciones de espectrómetros, logrando mantener la precisión de la estimación comparándolo con modelos convencionales basados en Redes Neuronales. Los resultados muestran que los campos reconstruidos son robustos a ruido añadido sobre los datos, presentando leves disminuciones en las métricas evaluadas para ruidos sobre 5%. Una ventaja de la metodología presentada es que los modelos pueden ser entrenados con imágenes en pequeña escala, para luego ser aplicados sin restricción de dominio, mientras se mantengan las resoluciónes utilizadas durante el entrenamiento.

Los resultados indican que la metodología es apropiada para los objetivos planteados, siendo viable aplicarla a modelos del estado del arte para obtener mejores resultados (ResNet, GAN). Se muestra una aplicación para reconstruir los datos de columna de NO_2 del Ozone Monitoring Instrument (OMI), a bordo del satélite Aura, mostrando el potencial de aplicación del método. Este trabajo contribuye a un mejoramiento en el monitoreo de la calidad del aire para el país, y se espera que pueda ser profundizado de manera de obtener mejores resultados (tanto de predicción como de estimación de error) y abarcar un área mayor de aplicación.

This work was funded by the Center for Climate and Resilience Research $(CR)^2$, Chile (ANID/FONDAP N°15110009), and supported by the supercomputer infrastructure of the National Laboratory for High Performance Computing (NLHPC, ECM-02).

Este trabajo fue financiado por el Centro de Ciencia del Clima y la Resiliencia $(CR)^2$, Chile (ANID/FONDAP N°15110009), y apoyado por la infraestructura de supercómputo del Laboratorio Nacional de Computación de Alto Rendimiento (NLHPC, ECM-02).

A la memoria de Rafael Cautivo Sánchez,
quién me enseñó a siempre querer aprender

Acknowledgements

En primer lugar, quiero mencionar a mis profesoras y profesor guía, a quienes les debo agradecer más que solamente el apoyo como docentes. Viviana Meruane, quién ha tenido la mejor disposición para apoyarme y resolver dudas más allá de lo académico; Axel Osses, quién entregó su valioso apoyo desde el campo de las Matemáticas y cuando necesité consejo; y Laura Gallardo, quién ha sido mucho más que una profesora guía durante este tiempo, abriéndome oportunidades y dándome herramientas tanto personales como profesionales. Muchas gracias por su tiempo y dedicación.

Quiero agradecer también a mi familia: a mi madre Javiera, mi padre Nelson y mi hermana Dominga. Ha sido una vida de apoyos, enseñanzas y risas, las cuales jamás terminaré de agradecer. Además agradezco a mis abuelos, abuelas, primos, primas, tíos y tías, con quienes en mayor o menor medida he compartido hermosos momentos durante tantos años.

A mis amigos y amigas del colegio, con quienes tengo un arsenal de recuerdos, viajes, risas, tallas malas y largas conversaciones. Atesoro mucho todas las alegrías que hemos pasado y me alegra que el tiempo nos entregue más. Además, me es importante mencionar a las profesoras y profesores que tuve. Quisiera detenerme especialmente en Cecilia, de Historia, quién me enseñó a perseverar y a amar la historia; y en Rafa, de Física, quién me hizo fascinarme con montones de cosas durante sus clases y fue un ejemplo de dedicación y amor por su profesión, le debo mucho reconocimiento.

Gracias a Vicente, por ayudarme con el trabajo gráfico, no solo de este documento, si no que también de múltiples presentaciones relacionadas que he tenido que realizar. También agradezco a aquellas personas que me dieron recomendaciones, me corrigieron cosas y dedicaron tiempo a ayudarme.

También debo mencionar a toda la gente que conocí en mi paso por la universidad, con quiénes he compartido mucho en estos 8 años. Muchísimas gracias Paulina y Yonjairo, llenaron de risas, conversaciones, bromas, cariño y abrazos esos edificios, no hay páginas para agradecerles, sin duda hubiese sido muy difícil sin ustedes. También todos aquellos con quienes compartí, durante los primeros años, durante mi paso por Mecánica y ahora último desde que salimos. Gracias por las tocatas, los paseos en bici, los almuerzos y los trabajos.

Muchas gracias a Daniela, quién ha sido un apoyo y una fuente de ánimo inagotable durante este tiempo. Me ha enseñado mucho y sin duda ha hecho este proceso mucho más sencillo. Te estaré siempre agradecido.

Finalmente quisiera agredecer a todas aquellas personas que no caben en las menciones anteriores, son demasiadas como para detenerme en todas, pero creo que cada una ha puesto su granito de arena en mi vida.

Gracias a todas y todos.

Table of Contents

List of Tables

List of Figures

1. Introduction

Tropospheric nitrogen dioxide (NO_2) is a primary pollutant that is harmful to human health (WHO, 2003). Accordingly, the World Health Organization has reduced the annual average guideline from 40 to 10 µg/m^3 in its most recent update (WHO, 2021). Nitrogen dioxide is the main source of nitrate aerosols, thus contributing to fully inhalable particles ($PM_{2.5}$), also NO_2 is, via photolysis ($\lambda < 424$ nm), the source of odd oxygen involved in the formation of ozone (O_3), and an air pollutant by itself (Crutzen, 1979; Seinfeld & Pandis, 2016). Furthermore, NO_2 has a chemistry-mediated climate impact through nitrate aerosols, O_3, and thus, methane (Szopa et al., 2021).

Nitrogen oxides are heterogeneously distributed in space and time due to their short atmospheric turn-over time (hours to days), and due to the variable distribution of sources and sinks. At the global scale, anthropogenic emissions of nitrogen oxides are estimated to *ca.* 42 TgN/yr, being fossil fuel combustion by the transportation sector, including aviation and shipping, roughly 50% of the total (Hoesly et al., 2018; Szopa et al., 2021). Natural sources of nitrogen oxides include lightning (*ca.* 5 TgN/yr within a factor of two) and soil processes (*ca.* 4.7 - 16.8 TgN/yr) according to Szopa et al. (2021). The main sink of NO_2 is its oxidation to nitric acid and nitrate both through gaseous-phase and multi-phase processes (Seinfeld & Pandis, 2016).

Satellite measurements have been very useful for monitoring not only NO_2, but also many trace gases and aerosols in the atmosphere. Over the past decades, many improvements have been made to satellite instruments, and the data retrieved with each of them has resulted in better environmental monitoring and modelling (Georgoulias et al., 2019). Over time, retrieval algorithms and instruments improve leading to higher spatial resolution and lower measurement errors (Judd et al., 2019). Thus, when studying long-term trends, such changes have to be taken into consideration. Different methods have been developed to handle this problem, trying to obtain a better resolution of older data by applying post-processing techniques for image up-sampling[1], usually known as super-resolution (SR) methods.

Over the last few years, different deep learning algorithms have been used to improve the resolution of images (Z. Wang et al., 2020), generally based on Convolutional Neural Networks (CNN), which are designed to process gridded data. Some applications have focused on satellite derived climatological data, with promising results (Cheng et al., 2020; Leinonen et al., 2021; Stengel et al., 2020; Vandal et al., 2017). These works focused on obtaining a good reconstruction in high resolution, leaving aside the uncertainty that this process inherently entails. Since environmental modelling problems generally handle noisy data (often

[1]In literature, both terms "up-sampling" and "down-scaling" can be found. Up-sampling often means an operation that increases the number of pixels in an image and, thus, reduces the physical size of each pixel. This definition is commonly used in image processing. On the other hand, the term down-scaling is used in climate science for an operation that reduces the physical size of a pixel and, therefore, increase its spatial resolution (Leinonen et al., 2021).

heteroscedastic, i.e. having an input dependent variance), it becomes relevant not only to try to predict the mean of these values, but also the entire conditional distribution. Hence, having an estimate of the uncertainty of a prediction can add valuable information when evaluating model results or assessing the impact of policy decisions (Cawley et al., 2007).

The objective of this work is to enable an improved spatial reconstruction of NO_2 column derived data from the Ozone Monitoring Instrument (OMI) onboard the Aura satellite (Levelt et al., 2006), as well as an estimation of the uncertainty of the reconstructed data. To this end, a Stochastic Super-Resolution Convolutional Neural Network model is proposed, which also estimates the error of the prediction by a stochastic approach of the problem. Instead of focusing on achieving high up-sampling factors, the research is centred on the error estimation method, assessing the feasibility of the proposed approach for these types of problems. Hence, two Convolutional Neural Networks (CNN) models and their performances are analysed, based on a simple configuration (Dong et al., 2016). In future work, once the feasibility of this methodology has been established, implementations with state-of-the-art architectures will be explored.

The following section details the satellite-borne column density data retrieval process, along with the data sources and the domain selected to carry out this work. Thereafter, the problem of super resolution (SR) is addressed and its approach with CNNs. Once the background is established, the research methodology is presented. This is divided into data processing, model definitions, and evaluation experiments. Subsequently, the results are presented and discussed, to end with the conclusions of the work.

2. Tropospheric NO_2 Column Density

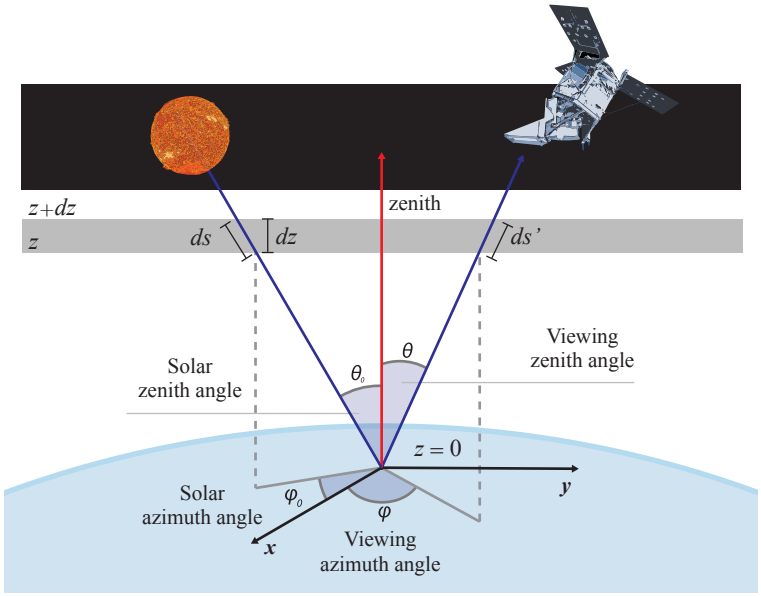
2.1. Satellite retrieval method

The important role of NO_2 in atmospheric composition requires monitoring its concentration with reliable, high-resolution observations. Satellite instruments have been very useful for measuring different components of both the troposphere and stratosphere since they enable us to estimate the concentrations of several molecules at the same time in a vertical column. Because higher pixel size decreases the spatial heterogeneity often associated with highly polluted features, spatial resolution of these measurements can bias the maximum values over urban regions (Duncan et al., 2016; Judd et al., 2019). Consequently, in the last decades, better instruments with higher spatial resolution have been continuously put into orbit, improving the data recovered with each of them and providing almost daily global coverage of valuable information of, among others, NO_2 trends (Georgoulias et al., 2019). This has driven an important advance both in modelling and in monitoring atmospheric behaviour, with benefits in multiple fields of science.

To retrieve tropospheric NO_2 column density data, satellite instruments use a Differential Optical Absorption Spectroscopy (DOAS) technique (Platt & Stutz, 2008), measuring the total amount of NO_2 between the satellite and the surface (see diagram in Figure 1). The basis of this method is Beer-Lambert's law (Swinehart, 1962), which relates the concentration of species along a path with the incident radiation at different frequencies. Each species has a unique absorption cross-section spectrum, that is, an amount of radiation which absorbs at different wavelengths (Platt & Stutz, 2008). Satellite instruments can measure the slant column density (SCD), defined by the concentration integrated over the light path in the atmosphere. Using knowledge of absorption cross-section of several trace gases (seen for NO_2 in Figure 2), the SCD can be estimated by obtaining the difference between the incident beam and a reference value. Afterwards, the SCD is separated into its stratospheric and tropospheric fraction on the basis of information coming from a data assimilation system (based on a transport model). Finally, an estimation of the air-mass factor is used to calculate the tropospheric vertical column density (VCD), which is the concentration of the trace gas vertically integrated over the troposphere.

2.2. Satellite data sources

For this study, data were collected by the Ozone Monitoring Instrument (OMI), onboard the Aura mission (Levelt et al., 2006); and the Tropospheric Monitoring Instrument (TROPOMI), onboard the Sentinel-5 Precursor mission (Veefkind et al., 2012). OMI is a spectrometer launched in July 2004, having an unprecedented spatial resolution for the time, which achieves ground pixels of 13 km along-track and 24 km across-track at nadir. This instrument has successfully provided daily global coverage during more than 15 years (Levelt



Source: Self-elaboration remastered from European Space Agency.

Figure 1: Scheme of satellite measurements principle. The instrument compares radiation received from Earth with a baseline measurement from the Sun to estimate vertical column burden.

et al., 2018), giving a high-value insight regarding the NO_2 spatial distribution and variability over the Earth (Duncan et al., 2016; Krotkov et al., 2016). TROPOMI, a spectrometer as well, was launched in October 2017, and has been providing almost global coverage each day since then. It has gone further with a spatial resolution of 7.2 km (5.6 km as of August 2019) along-track by 3.6 km across-track at nadir. Recent studies have validated and demonstrated the potential of having such detailed and precise measurements with broad coverage (Griffin et al., 2019; Lorente et al., 2019; Marais et al., 2021).

These two instruments were chosen due to the similarity between their orbits, so the recovered data at any coordinates correspond to similar local time. The spectral range of both sensors is similar for NO_2 detection, as can be seen in Figure 3. In addition, the TROPOMI NO_2 retrieval algorithm (Van Geffen et al., 2021) is based on the same approach introduced for the OMI retrieval (Boersma et al., 2007, 2011), and afterwards it has been used for reprocessing OMI NO_2 measurements (Boersma et al., 2018). Further, several studies have assessed differences between the two products, showing that TROPOMI data can be considered a more reliable higher resolution OMI data (Griffin et al., 2019; Judd et al., 2019; Lorente et al., 2019; Marais et al., 2021; Van Geffen et al., 2020).

Error sources should be considered and well estimated due to the importance of satellite data for air pollution monitoring, pollution trend studies and policy decisions. Despite numerous contributions to improve the retrieval algorithms of satellite-borne data, measurement uncertainties have been reduced but not altogether eliminated; hence it is important to assess them. Certain sources of uncertainty are inherent to the tropospheric column density recovery method. Three of them are especially relevant: instrument noise when slant column

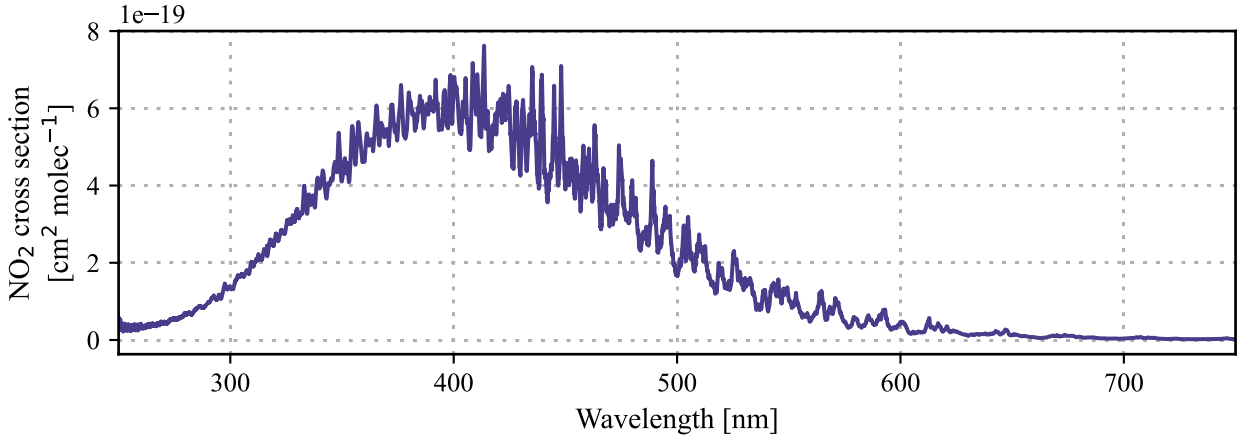
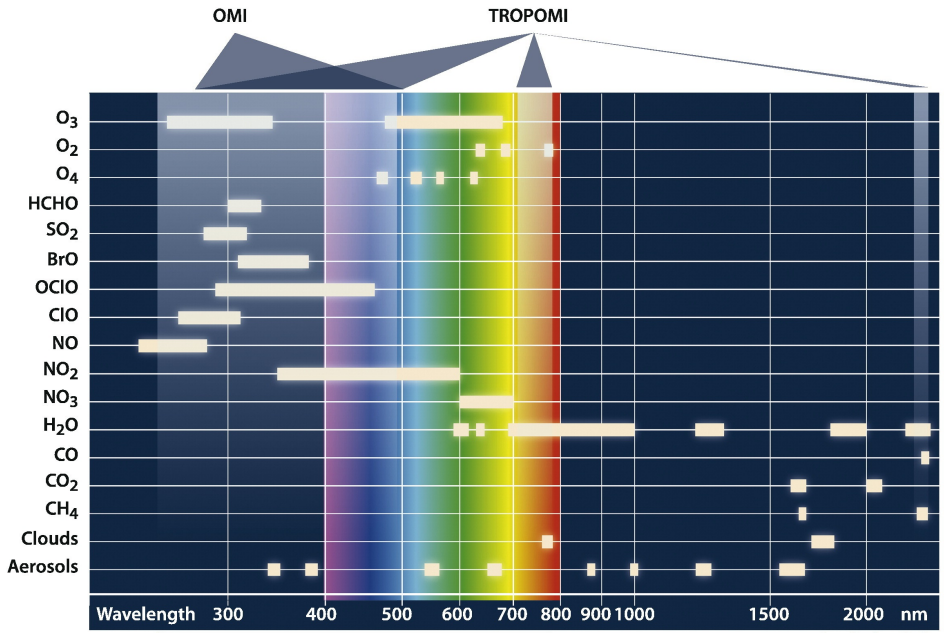


Figure 2: NO_2 absorption cross-section spectrum measured at 243°K. Data source: Institute of Environmental Physics (IUP), University of Bremen (Voigt et al., 2002).

density is retrieved; estimate of stratospheric fraction; and tropospheric air mass calculation (Boersma et al., 2004; Boersma et al., 2018; Lorente et al., 2017). For these reasons, as OMI and TROPOMI are no exceptions, uncertainties of both instruments have been assessed in previous studies (Boersma et al., 2018; Van Geffen et al., 2020; Zara et al., 2018). Likewise, this study seeks to handle these uncertainties in the data retrieved.

2.3. Study domain selection

For this work, the study area is restricted to central and southern Chile, which is shown in Figure 4a (Lon: 75°W – 69°W; Lat: 32°S – 40°S). Most of the country's population is concentrated here; hence, higher anthropic emissions. These emissions are mainly concentrated in the city of Santiago (seen in Figure 4b; Álamos et al., 2021), where several studies address their fate and effects (Gallardo et al., 2018; Gallardo et al., 2012; Menares et al., 2020; Seguel et al., 2020). Accordingly, a better understanding of NO_2 spatial distribution in the area is relevant.



Source: Veefkind et al. (2012).

Figure 3: Spectral ranges for TROPOMI and OMI, and high-absorption rate bands for various species, including NO_2.

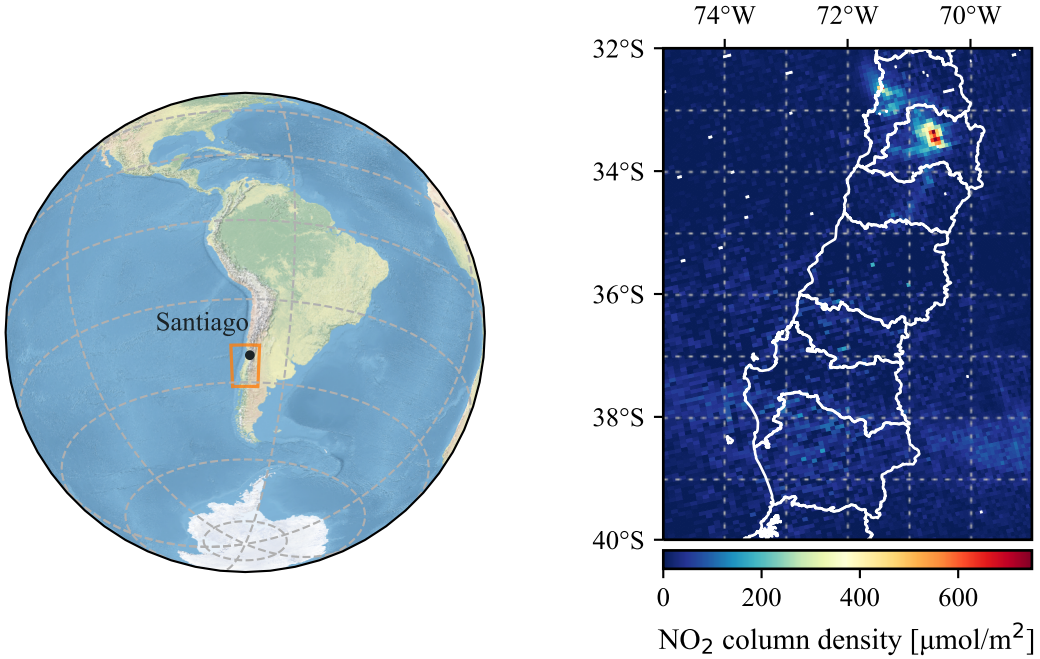


Figure 4: Study area. (a) Location of Santiago in South America; and (b) example of data retrieved with TROPOMI together with regional limits, showing higher concentration levels over Santiago.

3. Convolutional Neural Networks for Super Resolution

3.1. Single-Image Super Resolution problem

The single-image super resolution (SISR) problem consists in trying to infer a high-resolution (HR) image from a low-resolution (LR) one. Generic SISR algorithms are developed for general images, while domain-specific algorithms focus on specific types of images (*e.g.*, landscapes, faces, artwork, etc.). Improving resolution is an inherently ill-posed problem since there are always multiple HR reconstructions for a single LR image. Therefore, the problem has been approached with a variety of methodologies, obtaining different results depending on method and specific application (Nasrollahi & Moeslund, 2014; Yang et al., 2014).

Since image super-resolution aims to recover a unique HR image $\hat{I}_y$, from an input LR image I_x, the problem can be posed as

$$\hat{I}_y = \mathcal{F}(I_x, \theta)\,, \tag{1}$$

where $\mathcal{F}$ is the super-resolution model and θ denotes the parameters of $\mathcal{F}$. Usually, when modelling $\mathcal{F}$, approaches try to generate $\hat{I}_y$ as similarly as possible to the original I_y by optimising θ.

3.2. Convolutional Neural Networks

Convolutional Neural Networks (LeCun et al., 1989), known as CNNs, are specialised Neural Networks in grid-like topologies (for further read see Goodfellow et al., 2016, Chap. 9; Aggarwal, 2018, Chap. 8). These can be applied to data with different dimensionality, such as time series (1D) or images (2D). Due to their handling of correlation between neighbouring data, different Convolutional Network models have been used successfully in variety of applications such as image classification, image segmentation, features extraction, *i.a.* (Dong et al., 2016; Girshick et al., 2014; He et al., 2016; Huang et al., 2017; Krizhevsky et al., 2012; Long et al., 2015; Ren et al., 2016; Simonyan & Zisserman, 2015; Szegedy et al., 2017; Vincent et al., 2008).

3.2.1. Convolution operation

CNNs are based on the discrete convolution operation to process data, which considers the correlation between neighbouring values. In the case of a two-dimensional image, this is accomplished by a discrete weight function $w(i, j)$, also called the kernel of a convolutional

layer, which is applied iteratively to the entire image x. The convolution s is defined, for two-dimensional discrete functions, as follows:

$$s(i,j) = (x * w)(i,j) = \sum_m \sum_n x(m,n) \cdot w(i-m, j-n)\,. \tag{2}$$

Since it is not necessary to have the commutative property (obtained by flipping the kernel), the cross-correlation function is generally implemented, which is the same as convolution but without flipping the weight function:

$$s(i,j) = (w * x)(i,j) = \sum_m \sum_n x(i+m, j+n) \cdot w(m,n)\,. \tag{3}$$

This operation is graphically shown in Figure 5. Thus, each CNN layer consists of several filters, which are cross-correlated to the input image and optimised during training to obtain the desired result. After the filter, an activation function is applied (commonly non-linear, *e.g.*, ReLu), which leads to the application of the next layer.

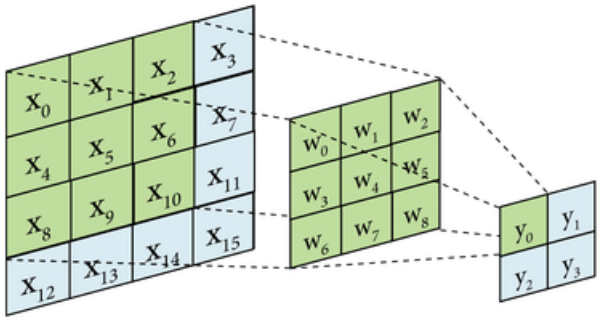


Figure 5: Convolutional operation in a CNN layer. The kernel w is applied to the input image x, obtaining an output image y.

3.2.2. Pooling

Sometimes it is desirable for a CNN layer to be invariant to small translations at the input. In these cases, a pooling operation is added after the activation function. The pooling operation is a statistical function that summarises the values between nearby outputs. For example, a max pooling operation returns the maximum value in a rectangular neighbourhood. As well as max pooling, average and min pooling are commonly used. Adding pooling in a convolutional layer helps to keep the same representation when the input is translated in a small amount, making the model more robust to input variations.

3.3. Application to SR problem

In recent years, deep learning (DL) techniques have developed rapidly, and newly super-resolution (SR) models based on CNNs now achieve state-of-the-art results (Z. Wang et al., 2020). These models range from the early Super-Resolution Convolutional Neural Network (SRCNN) model, proposed by Dong et al. (2016), to the recently exploited SR applications of Generative Adversarial Networks (GAN) models (Goodfellow et al., 2014) and Residual Network (ResNet) models (He et al., 2016). Later applications have shown outstanding results by combining these DL techniques for generic SR models (*e.g.*, Lai et al., 2017; Ledig et al., 2017; Lim et al., 2017; X. Wang et al., 2019).

When approached with CNN, the SISR problem consists of extracting latent information using successive convolutional layers (Z. Wang et al., 2020). At a certain point of the network a method of up-sampling is required, resulting in an output with higher resolution than the input. Considering the problem described in Equation 1, it can be approached using the intrinsic quality of neural networks optimisation methods, that is, training a model using a set $\mathcal{D} = \{I_x, I_y\}_{k=1}^{m}$ of images to get optimum parameters θ for the model. The optimisation of the model is achieved by solving the following optimisation problem:

$$\theta^* = \arg\min_{\theta} \; J(\theta); \quad J(\theta) = \frac{1}{m} \sum_{k=1}^{m} \mathcal{L}_k(\theta) + \Phi(\theta) \,, \tag{4}$$

where $\mathcal{L}(\theta)$ represents a loss function to compare the generated image and the ground truth image, and $\Phi(\theta)$ is a regularisation term over the parameters. Usually, the regularisation function can be of the L1 type (Lasso Regression) or of the L2 type (Ridge Regression), which fulfil the task of penalising the parameters with large values in the model. These functions are defined as:

$$\Phi_{L1}(\theta) = \lambda \sum_{j=1}^{p} |\theta_j| \,, \quad \Phi_{L2}(\theta) = \lambda \sum_{j=1}^{p} \theta_j^2 \,, \tag{5}$$

with λ the coefficient of regularisation.

Due to the impact of spatial resolution on atmospheric measurements (Judd et al., 2019), it is relevant to apply SR models to atmospheric data with state-of-the-art techniques. Previous studies have explored the application of CNN models to improve spatial resolution of data in different fields of atmospheric science. For instance the work on precipitation, irradiance and wind fields, referred in Cheng et al. (2020), Stengel et al. (2020), and Vandal et al. (2017).

3.4. Uncertainties in Neural Network models

There are two main sources of uncertainties that can be quantified in modelling: aleatoric (random) and epistemic (Kiureghian & Ditlevsen, 2009). An aleatoric uncertainty is presumed to be the intrinsic randomness of the measurement of a phenomenon, *i.e.*, the uncertainty which will be carried by the data utilised. Further, aleatoric uncertainty can be classified as homoscedastic uncertainty, which stays constant for different input values; and heteroscedastic, which varies depending on the input values, with some of them noisier than others. On the other hand, epistemic uncertainty refers to the lack of evidence, *i.e.*, the lack of information that a model has to make an accurate prediction, inherent in incomplete or erroneous models.

When applying DL models, these two uncertainties remain present, with aleatoric uncertainty being caused by inaccuracies presented in the data utilised, and epistemic uncertainty being caused by the limitations of the model and its lack of knowledge (Hüllermeier & Waegeman, 2021). To date, several studies have developed methods to estimate aleatoric and epistemic uncertainties, proposing different approaches to deal with the problem (Bishop, 1995; Khosravi et al., 2011; Papadopoulos et al., 2001). Recent studies have proposed methods to simultaneously estimate and separate both uncertainties (Amini et al., 2020; Huseljic et al., 2020; Kendall & Gal, 2017). This work focuses on addressing the aleatoric uncertainty inherent in satellite data, leaving aside the epistemic uncertainty the model may have.

4. Data and Methodology

In this work, we use Convolutional Neural Networks for SISR applied to satellite data, *i.e.*, in a specific domain. The aim is to improve OMI NO_2 data by training a model using TROPOMI data, given the need to have high-resolution images to carry out the process. Further, we present a stochastic approach to approximate the aleatoric uncertainty with a SRCNN, to consider the noise in satellite-borne tropospheric column data. Since our objective is to assess the feasibility of uncertainty estimation with CNN models in SR problems, we mainly focus on the estimation of the error, rather than on the up-sampling factor or the quality of the reconstruction. Hence, we present a simple SRCNN model which also estimates aleatoric uncertainty.

4.1. Data and pre-processing

Since our goal is to improve OMI's spatial resolution, a set $\mathcal{D}$ is needed for training the model, where I_x has OMI's resolution (25x25 km^2) and I_y has the target resolution. In this work, we define our HR images as 12.5x12.5 km^2, that is, an up-sampling by a factor of two. Data utilised in this work were retrieved by both OMI (Lamsal et al., 2021) and TROPOMI (Copernicus Sentinel-5P, 2021) (see Section 2.2). The latter, can be considered as a HR version of OMI. Therefore, TROPOMI data is utilised to train the model, sampling each image available to both the HR and the LR mentioned, so the optimised model can be later applied to improve the resolution of OMI images.

Datasets retrieved by satellite are generally characterised by a considerable amount of data gaps, which constitutes a source of error in their analysis. These data gaps can be generated due to the orbit of each satellite, cloud contamination, instrumental failure, or high ground reflectance (Van Geffen et al., 2021). An efficient method of filling data gaps is required, allowing the dataset to be used with the least possible missing values. In the case of the data retrieved for this work, both products have a quality assurance value (qa_value), which serves as a filter of the observations, indicating which pixels are useful and valid. Since a qa_value lower than 0.5 is not recommended for use in either case, all pixels with values lower than 0.5 are filtered out and considered invalid (Van Geffen et al., 2021). This process leaves the data with certain gaps between pixels, which makes it difficult to use it for CNN models because they do not inherently handle missing data.

To generate the dataset, all images are regularly gridded at the desired resolution (12.5x12.5 km^2), maintaining the amount of concentration according to the area. Additionally, missing values are gridded and saved in the new data for further filling. Thereafter, to overcome the missing data problem, two steps are performed on each image obtained. The first step consists in discarding those images that contain a high number of spatially close/grouped missing values (15 pixels or above), as this will make it difficult to fill them in with confidence. Second, a penalised least square method is used to fill missing values, which has been

successfully applied to geophysical datasets before (Garcia, 2010; G. Wang et al., 2012). This method is described below, along with its implementation. Once the gaps have been filled, it is possible to sample the images at low resolution (25x25 km^2) to be used for model optimisation.

4.1.1. Gap filling method

We use a penalised least square method based on two-dimensional discrete cosine transforms (DCT), which describes the data in terms of the sum of cosine functions oscillating at different frequencies (Garcia, 2010). This method has been used successfully to fill geophysical datasets when the smoothing factor s used is very small (≈ 0) (G. Wang et al., 2012). Let's consider a two-dimensional data z (*i.e.*, an image) with missing values, and $\hat{z}$ an approximation of the values with the missing values filled. Iteratively applying the DCT and its inverse (IDCT), it is possible to determine $\hat{z}$ by least squares solving:

$$\hat{z}_{k+1} = \mathrm{IDCT}\Big(\Gamma \circ \mathrm{DCT}\Big(W \circ (z - \hat{z}_k) + \hat{z}_k\Big)\Big)\,, \tag{6}$$

where $\circ$ stands for the Hadamard (elementwise) product. The inclusion of a binary W matrix indicates where the missing values are. Here, Γ represents a two-dimensional filtering tensor defined as

$$\Gamma_{i_1,i_2} = \left(1 + s\left(\sum_{j=1}^{2}\left(2 - 2\cos\frac{(i_j - 1)\,\pi}{n_j}\right)\right)^2\right)^{-1}\,, \tag{7}$$

where i_j denotes the ith element along the jth dimension, and n_j denotes the size of z along this dimension. Also, a robust iteration mode is defined, which helps to avoid overweight outlier values using a re-weighted process updating the residuals of the iteration[2].

Once the optimal $\hat{z}$ is found, the results can be used to fill in the missing data in the original images. To select which s value use, we proceed to randomly remove 0.5% of the pixels (real missing values in this dataset are around 0.3%) from the available data and then fill them in with both the normal and the robust implementation. Then, the reconstruction root mean squared error (RMSE) is measure to select the optimal s. The results obtained for different s, with and without robustness, are shown in Figure 6. From these, $s = 0.01$ is used to fill the data gaps as it is the value with the lowest reconstruction error.

4.2. Stochastic SRCNN model

In this work, besides the SR problem, we focus on the estimation of the aleatoric uncertainty, which is determined mainly by the satellite NO_2 data recovery errors. Thus, we

[2]The implementation of the algorithm can be found at github.com/SanParraguez/smoothn.

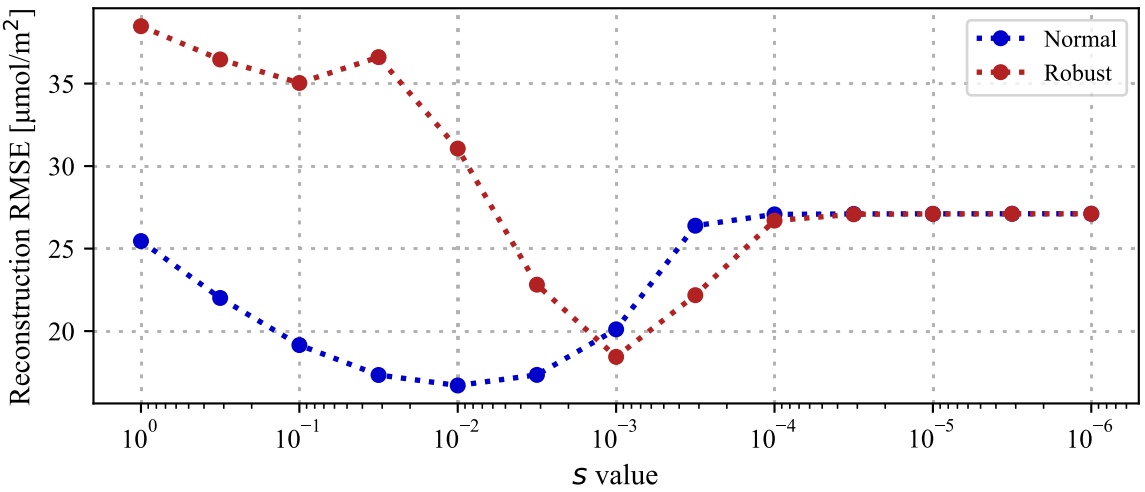


Figure 6: Reconstruction RMSE over the data with randomly removed pixels for given s values. Normal curve corresponds to the results of the basic method; Robust curve corresponds to the results applying the re-weighting method to avoid outlier data effects.

aim to assess the feasibility of using CNN not only to reconstruct satellite HR images, but also to use this model to estimate the error of the predictions, *i.e.*, the aleatoric uncertainty. The main novelty of the uncertainty estimation with CNN lies in the estimation of multiple correlated distributions in two dimensions, which is more complex than the estimation of a single distribution.

One of the most straightforward methods to estimate these errors is approximating a previously determined parametric probability distribution function (PDF) by Maximum Likelihood Estimation (MLE), proposed by Nix and Weigend (1994). This approach has been compared with Bayesian and Bootstrap methods with similar results and shorter training times (Khosravi et al., 2011; Papadopoulos et al., 2001) and used as the aleatoric uncertainty estimation method of some Neural Network based models (Amini et al., 2020; Huseljic et al., 2020; Kendall & Gal, 2017). Here, we attempt to approximate a function $f(x)$ by modelling our data as

$$y(x) \sim \mathcal{N}(\mu, \sigma^2); \quad \hat{\mu},\, \hat{\sigma} = f(x)\,, \tag{8}$$

where σ is the aleatoric uncertainty, which behaves as additive noise on the target values. For this, a network must be designed so that its outputs estimate $\hat{\mu}$ and $\hat{\sigma}$ as the mean and the standard deviation of a Gaussian distribution. For the purpose of this research, our data is assumed to have a Gaussian noise with 0 mean because this distribution is seen in data over the Pacific Ocean (Lon: 110°W; Lat: 25°S) far from coasts and ship tracks (Krotkov et al., 2016). However, our method does not depend on the assumed noise distribution, and other distributions might be used. Additionally, it is considered a heteroscedastic behaviour in the data, assuming a single distribution for each pixel.

Hence, a Stochastic Super Resolution Convolutional Neural Network (SSRCNN) is proposed, based on the approach of Dong et al. (2016), but adding a so-called Gaussian layer as the network's output. This layer aims to estimate $\hat{\mu}$ and $\hat{\sigma}$ parameters of the distribution by maximising the likelihood of $y(x)$. Due to spatial continuity of NO_2 concentrations, as well as μ, σ is dependant not only on the pixel value itself, but also on its neighbours, unlike problems where a single value is predicted. Therefore, the optimisation of θ can be achieved by minimising the negative log-likelihood loss function of many Gaussian distributions:

$$\mathcal{L}(\theta) = -\frac{1}{n}\sum_{i=1}^{n} \log\, p(\, y_i \,|\, \hat{\mu}_i, \hat{\sigma}_i^2) = \frac{1}{2n}\sum_{i=1}^{n} \log(2\pi\hat{\sigma}_i^2) + \frac{(y_i - \hat{\mu}_i)^2}{\hat{\sigma}_i^2}\,, \tag{9}$$

with n the number of pixels in the image. While learning θ, the likelihood function successfully models the aleatoric uncertainty. Notice that our model does estimate the epistemic uncertainty as it is an intrinsic model property and not of the data (Hüllermeier & Waegeman, 2021).

4.3. Training method

To train the CNN model, the complete set of TROPOMI NO_2 data available is used (from 01 May 2018 to 30 April 2021). The low-resolution images obtained by sampling TROPOMI ones are used as inputs of the model, aiming to reconstruct the original high-resolution ones. The application of the data is shown in Figure 7, differentiating a standard SRCNN model and our SSRCNN model outputs. To carry out the training, 15% of the data is set aside for subsequent evaluation; in addition, during the training a fraction of 20% of the data is used for validation.

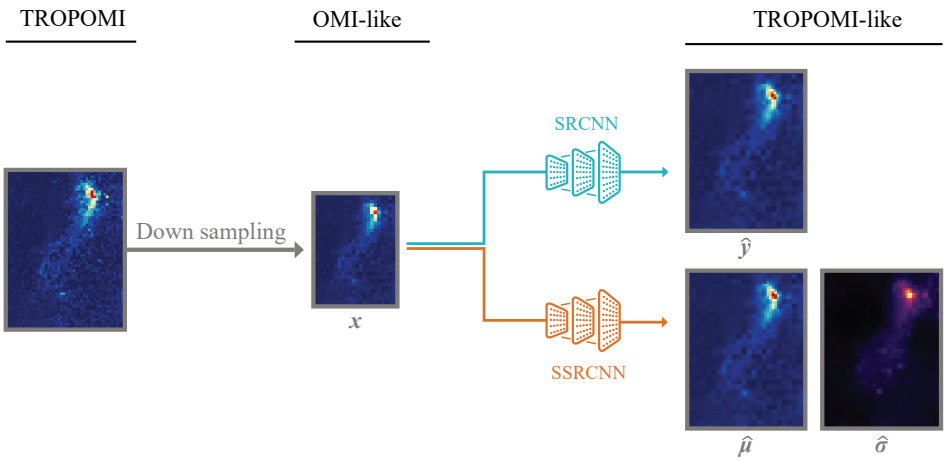


Figure 7: Data application process during training. Original images of high-resolution are down sampled and then the models seek to replicate these original images. In the case of SSRCNN the expected error is also estimated.

The optimisation process is carried out by applying the widely used Adam algorithm, using the *backpropagation* technique (Rumelhart et al., 1986) to estimate gradients (see Appendix A for loss functions gradients calculation). This method uses first-order gradients to compute individual adaptive learning rates for different parameters from estimates of first and second moments of the gradients (Kingma & Ba, 2015). For each iteration, the parameters θ of the model are updated by

$$\theta_t = \theta_{t-1} - \alpha \cdot \frac{\hat{m}_t}{\sqrt{\hat{v}_t} + \epsilon} \,, \tag{10}$$

where α is the stepsize, usually known as the learning rate, and $\hat{m}$ and $\hat{v}$ are the bias-corrected first moment estimate and the bias-corrected second raw moment estimate, respectively. The training process, similar for both SRCNN and SSRCNN cases, is detailed in Figure 8 and is implemented using Tensorflow framework (Abadi et al., 2016).

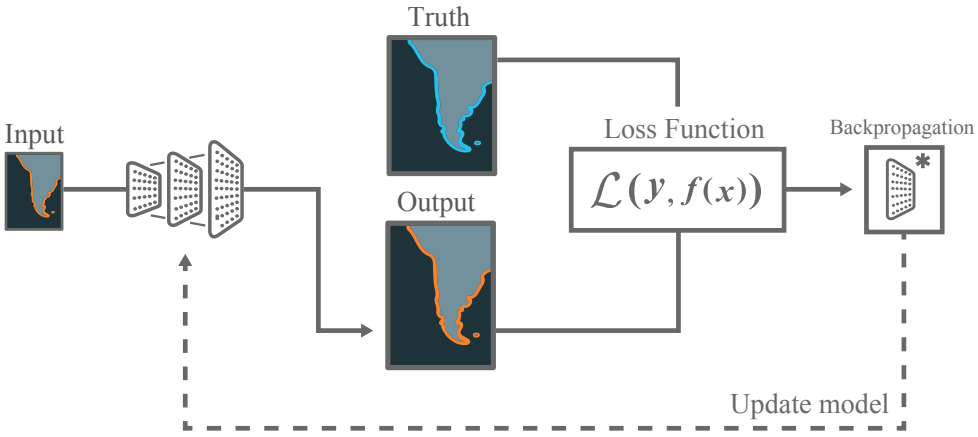


Figure 8: Training process of CNN models. The output of the model is compared against the ground truth by a loss function to afterwards propagate the errors and update the parameters. This process is repeated for the complete dataset in each iteration.

4.4. Evaluation experiments

There are two main characteristics that must be evaluated in the SSRCNN model: its capability to make accurate predictions, and its capability to estimate the error of each prediction. Therefore, both the $\hat{\mu}$ and the $\hat{\sigma}$ values of the output must be tested separately.

Prior to the evaluation of the model, two baselines are selected against which the results obtained in the experiments can be compared. First, a bicubic interpolation is used, acting as the base case; and second, a SRCNN model, based on the approach of Dong et al. (2016), is trained (following the same training methodology previously described and using mean squared error as loss function), which acts as a CNN base case. Hence, suitable architectures

must be sought for both the SRCNN and the SSRCNN models, under the assumption that the methods require different knowledge of the data. Having these baselines allows assessing differences in the precision of the predictions, *i.e.* evaluate $\hat{\mu}$ of the SSRCNN model.

To measure the model performance of the approach rather than that of a particular architecture, an appropriate network architecture must be selected for each model, which, when its parameters have been optimised, can be evaluated. This is carried out by parameterising the architecture of both CNN models, defining a series of variables that allow the construction of different networks. The function generates a bullet-shape architecture, determining parameters such as, *inter alia*, number of layers, kernel size and number of filters per layer (see Appendix B). Afterwards, a series of random searches are performed on a given domain of the parameterisation for both models. Once this stage is finished, a grid search is carried out to refine the initial results and obtain a better model in both cases.

Once the architectures are defined and both models are optimised, it is necessary to evaluate their performance on the test data, which has not been used previously during training. To perform this, different image quality assessment metrics are used. Which are selected to have a better insight of the reconstruction quality since objective computational metrics are not necessarily consistent between each other. These metrics are the Peak Signal-to-Noise Ratio (PSNR), the Structural Similarity Index Measurement (SSIM) and the Mean Absolute Percentage Error (MAPE). Additionally, the training time of the models is measured to better assess convenience of the methods.

PSNR is commonly used to quantify image reconstruction quality, comparing the original one and its reconstruction based in the mean squared error of each pixel. Given the ground truth image I with N pixels, and a reconstruction $\hat{I}$, the PSNR is defined as:

$$PSNR = 10 \cdot \log_{10} \left(\frac{L^2}{\frac{1}{n} \sum_{i=1}^{n} (I_i - \hat{I}_i)^2} \right) , \tag{11}$$

where L stands for the maximum pixel value. As a complementary measure, SSIM is used, which compares luminance, contrast, and image structure by sections (Horé & Ziou, 2010; Z. Wang et al., 2004). Here, the simplified form proposed by Z. Wang et al. (2004) is used, corresponding to:

$$SSIM = \frac{(2\mu_a\mu_b + c_1)(2\sigma_{ab} + k_2)}{(\mu_a^2 + \mu_b^2 + c_1)(\sigma_a^2 + \sigma_b^2 + k_2)} , \tag{12}$$

where a and b are the two images compared, and μ and σ are their mean intensity and its standard deviation (notice that this notation is used only at this equation to keep it as presented by Z. Wang et al.). Constants are used as proposed in the original article, with $k_1 = 0.01$ and $k_2 = 0.03$. Furthermore, MAPE is evaluated to have a more intuitive quantity of the error, which is relative to the value predicted. We calculated MAPE by following:

$$MAPE = \frac{100}{n} \sum_{i=1}^{n} \frac{| I_i - \hat{I}_i |}{\max(|I_i|, th)} , \tag{13}$$

where th is a threshold equal to 12 μmol/m^2 to avoid values too large when I_i is near zero. This value is used because it correspond to the mean value of NO_2 column over the Pacific Ocean where zero values are expected.

To evaluate our model, the $\hat{\mu}$ prediction of the SSRCNN is compared trough the said metrics against a bicubic interpolation and the SRCNN model. In addition to the calculated metrics, the dispersion of each of the three predictions is evaluated along with their correlations. This was also visually inspected and evaluated for single-image predictions, providing insight in the procedure. Afterwards, the same predictions are evaluated in detail for the SSRCNN model by comparing the prediction values and error values, as well as measuring the uncertainty through differential entropy (Cover & Thomas, 2006). The differential entropy h is calculated by its definition, for a Gaussian distribution with probability density function f, as following:

$$h(f) = -\int_S f(x) \log f(x) dx = \frac{1}{2} \log(2\pi\sigma^2 e) . \tag{14}$$

Further, we evaluated the robustness of the models when considering increasing amounts of noise in the incoming data. So, we performed a SSRCNN model sensitivity experiment. Thus, different noise levels are added to the data by multiplying the original value by a factor sampled from a Gaussian distribution with mean 1. The noise is applied to all pixels of the I_x images of the test data, while the original I_y are kept with their ground truth values. This is carried out with different variances for the distribution, assessing the metrics obtained by the model's predictions for each instance.

5. Results and Discussion

The results are presented below, separating them into four sections corresponding to the different experiments carried out to evaluate the proposed methodology. These experiments correspond to the architectures found and subsequently used, the model evaluations carried out, the noise sensitivity analysis, and a first application of the model to OMI data.

5.1. Architectures of models

First, architectures for the CNN models had to be chosen and trained. This procedure was carried out by means of a random search and then a refinement of the grid, finding above-average architectures for both CNN models by their loss functions (see Appendix B). Recall that the objective was to evaluate the feasibility of the approach, rather than trying to obtain the best architecture.

The SSRCNN model was designed to have two different outputs, which are the parameters of the given distribution. The selected architecture is detailed in Figure 9a, which specifies the layers of the model. The model has a bilinear up-sampling layer, three deep layers, and two independent output layers. Each layer has a dropout rate of 0.1 during training, and a regularisation for both L1 and L2 types of 1e-8. In contrast, the output layer $\hat{\mu}$ does not have regularisation, and the $\hat{\sigma}$ layer is penalised with L1 = 1e-5 and L2 = 1e-8.

Turning to the SRCNN model, a single output is required, which seeks to approximate y. The utilised architecture for this model is shown in Figure 9b. This model, as in the previous case, has three deep layers, but different number of channels and kernel sizes. The up-sampling method and the dropout ratio remain the same, bilinear and 0.1 respectively. Nevertheless, the regularisation weights are different; in this case both the deep layers and the output layer only have an L2 regularisation equal to 1e-7.

Once the architectures have been selected and trained, both the numbers of parameters to be adjusted and the time it takes to train each model are obtained. The models were trained for a maximum of 25,000 iterations, with a patience of 500 iterations before stopping the training in case there were no improvements. It should be noted that the same hardware was used in both cases to better compare these differences. These data are set out in Table 1. It is worth noticing that, despite the greater number of parameters that must be optimised, the SRCNN model takes considerably less time to be optimised. This may be attributable to the fact that more gradients must be propagated in the case of SSRCNN, leading to a greater number of calculations necessary to adjust the model parameters. In the following sections, the performances of these two models will be evaluated against a bicubic interpolation.

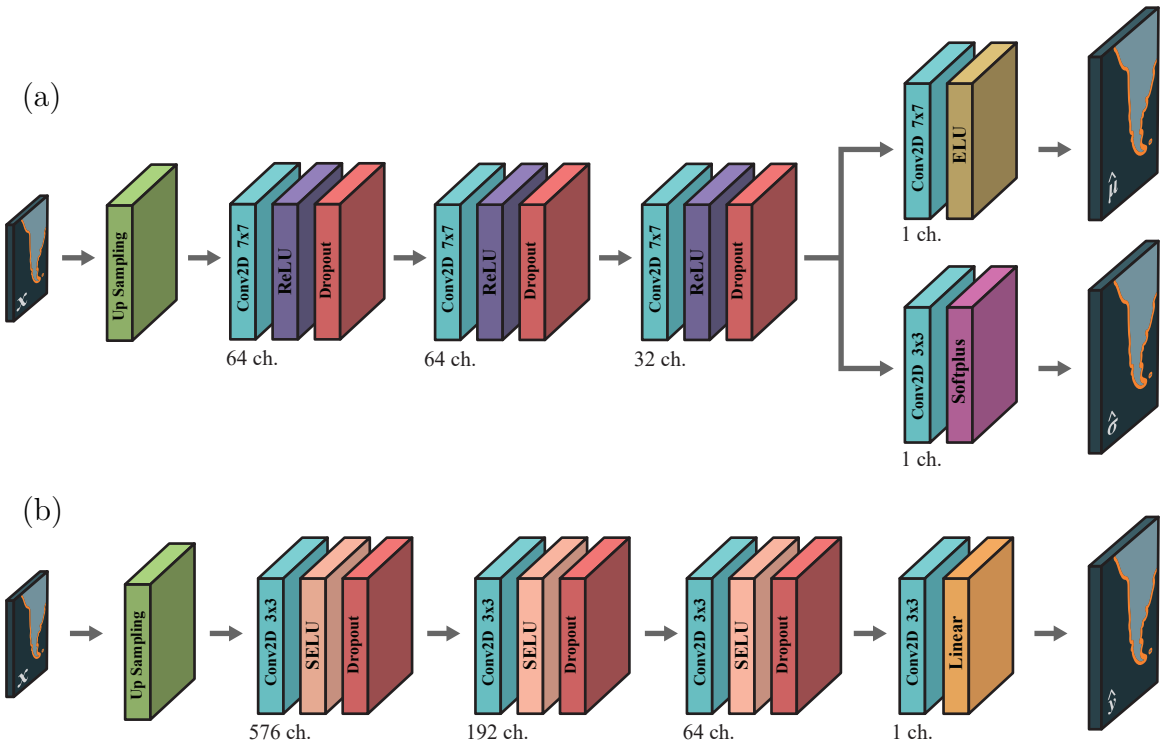


Figure 9: Architectures schemes of both applied CNN models. (a) SSRCNN model architecture; and (b) SRCNN model architecture.

5.2. Performance evaluation

During this section, the results of the experiments to evaluate the models are presented and discussed. As previously detailed, 15% of the database was used to assess the performance of the model and compare its results against the baselines. The mean results of the metrics (with improvement over bicubic interpolation) applied to the test data are set out in Table 2, where both CNN models perform better than the bicubic baseline in all images tested. Slight performance differences can be observed between the models. The clearest difference is found in MAPE, where the SSRCNN model obtains the best performance.

Table 1
Number of parameters and training time for both CNN models.

	SRCNN	SSRCNN
Number of parameters	1,112,513	306,210
Training time [min]*	23.2	69.8

*All models were trained in a GPU NVidia Tesla V100.

Table 2
Metrics comparison between bicubic interpolation and CNN models (best performance in bold). Percentage improvement over bicubic interpolation is also shown.

Metric	Bicubic	SRCNN		SSRCNN	
		Value	Diff. [%]	Value	Diff. [%]
PSNR [dB]	34.78	**35.32**	1.67 ± 0.34%	35.29	1.53 ± 0.23%
SSIM	0.820	0.833	1.69 ± 0.18%	**0.834**	1.76 ± 0.18%
MAPE [%]	37.0	36.6	0.99 ± 0.12%	**36.2**	2.23 ± 0.13%

The differences in metric improvement may be due to optimisation methods and the definition of the metric itself (Z. Wang et al., 2004). In this case, the better PSNR results of the SRCNN when compared to the SSRCNN model could be attributed to the loss function used, since PSNR is defined directly from the MSE between pixels. Besides, the SSRCNN model outperformed the baselines in both SSIM and MAPE. This means that, on the one hand, the SSRCNN model creates structures more similar to the ground truth; and on the other hand, it gets percentage-wise errors closer to the ground truth, where higher improvement is observed.

For gaining insight regarding the metrics' distributions, Figure 10 shows the cumulative empirical distributions obtained with each metric. A similar behaviour is seen between the three models evaluated, where, as in the case of the mean values, an improvement is observed in the CNN models with respect to the bicubic interpolation. Further, Figure 11 compares the improvement of both CNN models against the bicubic interpolation for each of the assessed metrics. The most significant finding is seen in MAPE, where the SSRCNN clearly outperforms the base CNN model, which also performs worse than bicubic interpolation on some images. Nevertheless, SRCNN presents considerably higher differences in PSNR for certain images, which can be attributed again to the loss function selected.

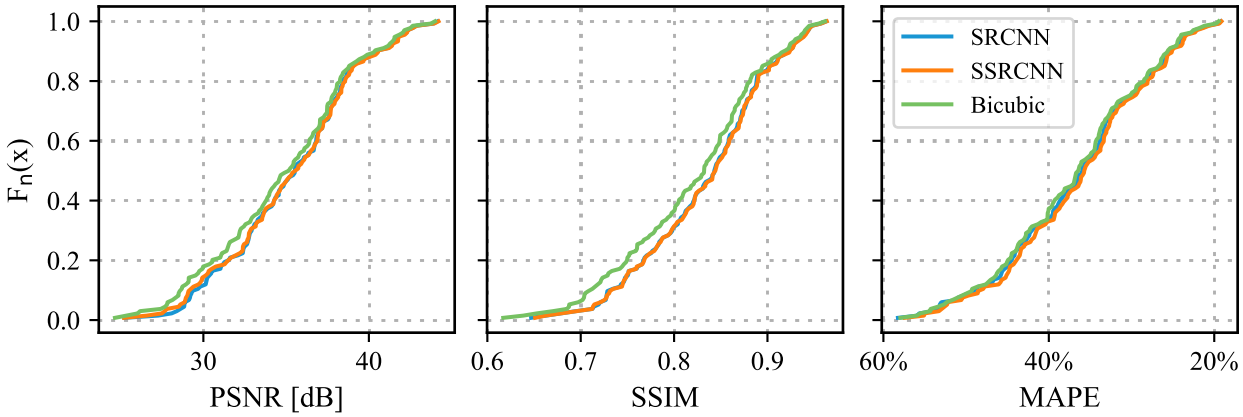


Figure 10: Empirical cumulative distributions of models' metrics for test images. Higher PSNR and SSIM is better, and lower MAPE is better.

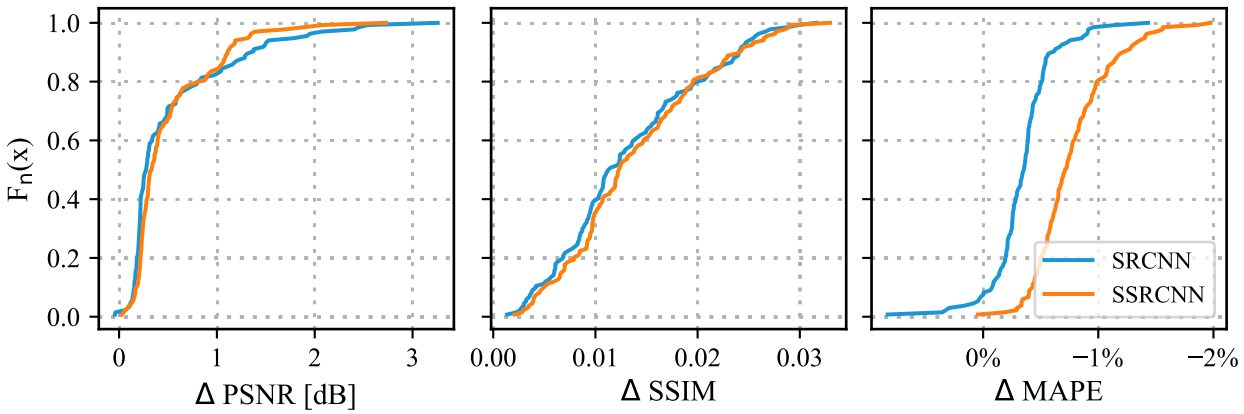


Figure 11: Empirical cumulative distributions of the improvement of models metrics over a bicubic interpolation for test images. Higher PSNR and SSIM is better, and lower MAPE is better.

Subsequently, the correlation between the different predictions and the expected values was assessed. The three panels in Figure 12 show the dispersion of the predictions for each pixel in the test images. Both CNN models outperform the bicubic interpolation, and the SRCNN model obtains a higher R^2, showing again a slightly better performance over the SSRCNN model.

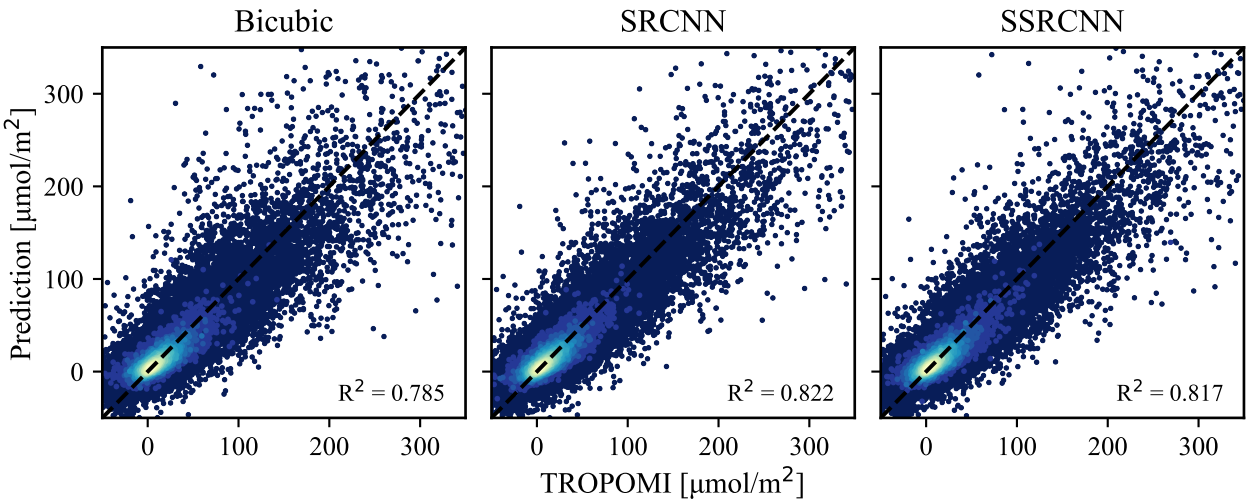


Figure 12: Scatter plot comparison of predicted values and TROPOMI observed values for the interpolation and models.

Dong et al. (2016) present greater improvements regarding bicubic interpolation than the ones obtained in this work. This difference is attributable to two main factors: the database used, which is much broader and more diverse in their case, and the type of images used (satellite data are characterised by being smooth and without such defined edges, which improves performance of a bicubic interpolation). On the other hand, the results presented are consistent with those of Dong et al. (2016), who show that SRCNN models outperform a bicubic interpolation. Moreover, the experiments suggest that the SSRCNN models can

demonstrate similar performances and may surpass SRCNN in some metrics such as SSIM or MAPE. Hence, the method seems capable of achieving similar results compared to a model which only predicts the desired value, instead of both the desired value and the expected error. Nonetheless, the training time of the model presented is approximately four times longer for the selected architectures, which must be considered when selecting an approach.

To have a visual perception of the reconstructions obtained by each model, an example from the test data was retrieved. As can be seen in Figure 13, both CNN models better replicate the areas with higher pixel values, which is expected for the SRCNN model. Because of this, a more detailed analysis was carried out to determine if the CNN models improve their performance in pixels with higher concentrations.

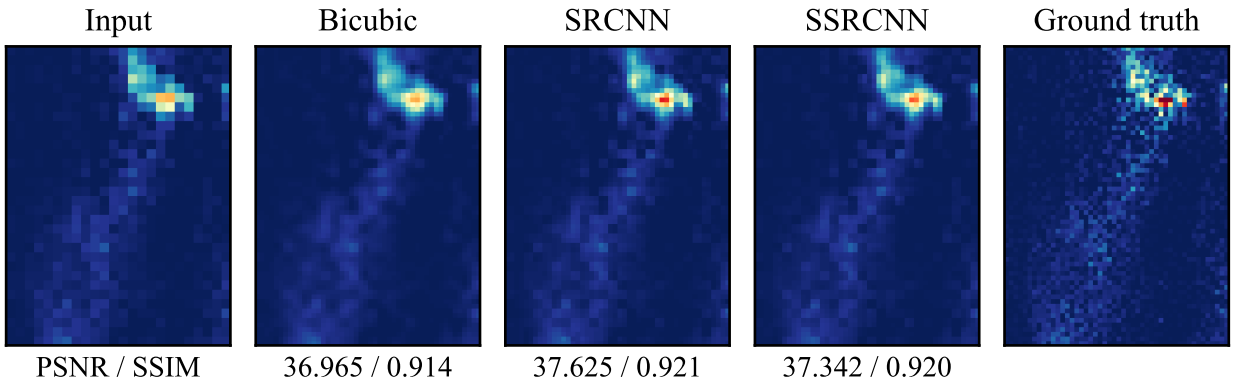


Figure 13: Comparison of a single prediction between interpolation and CNN models. (a) Input image for the models; (b) bicubic interpolation; (c) SRCNN model prediction; (d) SSRCNN model prediction; and (e) target image. Below each panel one finds the metrics obtained for this image (higher is better).

5.2.1. Evaluation of column density data maxima

This section presents a more in-depth evaluation of the behaviour of the models for areas with high NO_2 column values. To do this, some of the experiments already carried out were repeated, but this time only those pixels where the ground truth value is greater than 50 μmol/m^2 were considered. The evaluation metrics were calculated again, omitting the SSIM because it does not allow a pixel-by-pixel comparison, but rather by areas of the image. The results obtained are shown in Table 3, in which it could be seen an increase in the improvement of the CNN models over the bicubic interpolation, compared with the results of Table 2.

Similarly, the dispersions per pixel are obtained again, filtering those lower than the imposed threshold. The results of the correlations for each model are presented in Figure 14, from which two important results emerged. The first is that the bicubic interpolation obtains a lower R^2 with respect to the experiments with the complete images; the second is that the CNN models not only obtained greater differences with respect to interpolation, but also

Table 3
Metrics comparison between bicubic interpolation and CNN models for pixels above 50 μmol/m^2. Percentage improvement over bicubic interpolation is also shown.

Metric	Bicubic	SRCNN		SSRCNN	
		Value	Diff. [%]	Value	Diff. [%]
PSNR [dB]	29.19	30.06	3.69 ± 0.51%	30.11	3.56 ± 0.45%
MAPE [%]	30.31	27.43	8.80 ± 1.13%	27.68	8.91 ± 1.21%

increased their R^2 compared to the original experiment. These results are in line with what can be seen in Figure 13, which suggests a better reconstruction of CNN models of high-value pixels. Accordingly, if only values over 50 μmol/m^2 are considered, the differences between the models and the bicubic interpolation increase. This behaviour is useful when it is necessary to apply the models to study data over cities or other high-polluted areas.

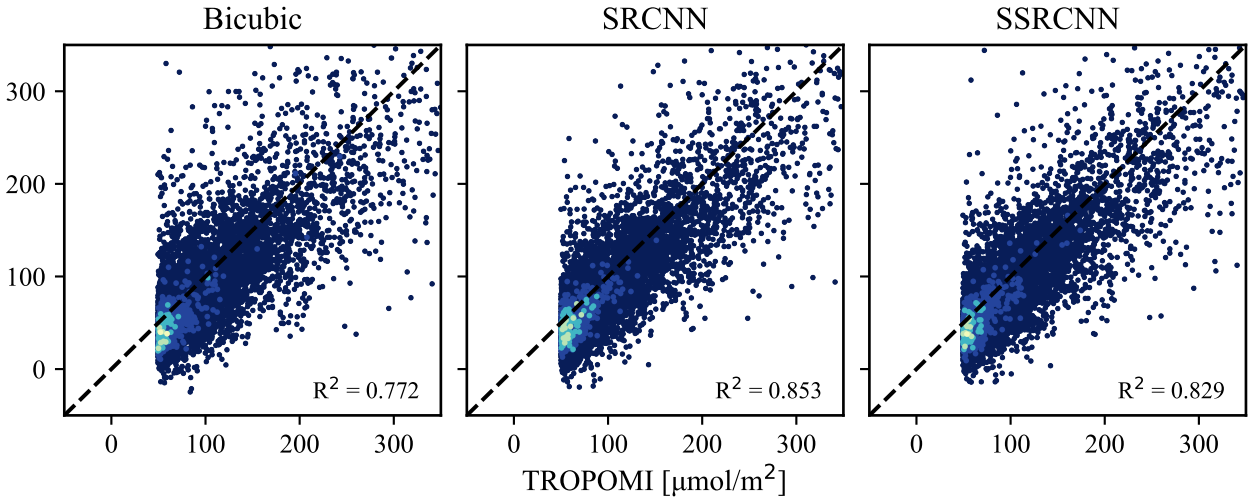


Figure 14: Scatter plot comparison of predicted values and TROPOMI observed values for the interpolation and models. Only TROPOMI values above 50 μmol/m^2 are considered.

These findings support the idea that CNN models outperform bicubic interpolation especially for high values. In addition, they further suggest that for smooth images the improvement is reduced due to the lack of edges. Although the results are not directly comparable due to differences in architectures and data, the increase in the metrics in high values is closer to the results obtained by Dong et al. (2016).

5.2.2. Uncertainty estimation

Up to now, the predicted values have been assessed comparing the performance of the trained models. The main improvement of the SSRCNN developed is its capability to estimate the error, *i.e.*, predict the uncertainty of its predictions, which must be evaluated as well. Therefore, Figure 15 shows the prediction of the SSRCNN model for the same image as in

Figure 13, comparing the error estimation against the actual prediction error. Due to the similarity between predicted σ and the error of the estimation, it can be assumed that the second one behaves as a sample of the approximated distributions. Further, the entropy is calculated to have a more intuitive measure of the uncertainty of the predictions.

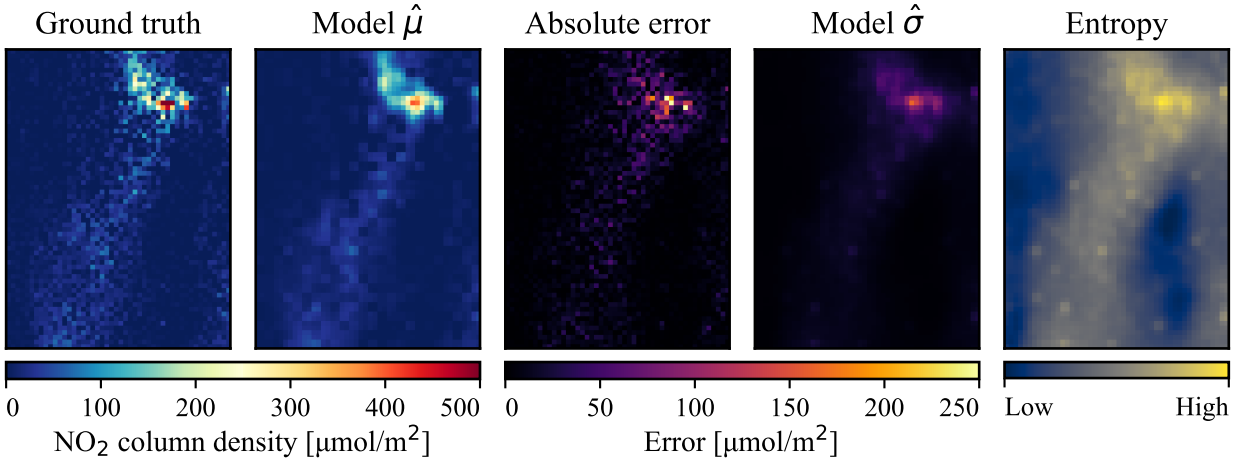


Figure 15: SSRCNN model single prediction result. (a) TROPOMI truth image; (b) model prediction; (c) absolute error of the prediction; (d) estimated error by the model; and (e) entropy as a measure of uncertainty.

By visually evaluating a larger number of images, it is possible to see that this behaviour is maintained for different distributions and concentrations in the data (see Appendix C). Future work will aim to better assess the quality of this error estimate; for the moment, the results presented here indicate that the model is capable of correctly estimating aleatoric uncertainty.

5.3. Robustness of the SSRCNN approach

To ensure the robustness of the proposed approach, a noise sensitivity analysis of the three metrics considered is carried out. For this analysis, random noise was added to the input images by multiplying each value by a noise factor, keeping the original output images. The amount of added noise ranged from 0.1 to 25% and applies to the entire evaluation dataset, comparing the metrics obtained in each step.

The results of the noise sensitivity analysis are summarised in Figure 16, in which the panels show the results of the three methods for each metric. The data obtained show that both CNN models maintain better results than the bicubic interpolation when adding different noise levels. From noise values of around 5%, a continuous decrease in performance can be seen in the three cases shown, and it is more pronounced after values of around 10%. In MAPE curves, the SRCNN model and the bicubic interpolation for noise ranges above 10% are closer, while the SSRCNN model remains slightly above. Hence, these findings provide

support to consider both CNN models robust against noise in the data, making them suitable for applications with the type of data used.

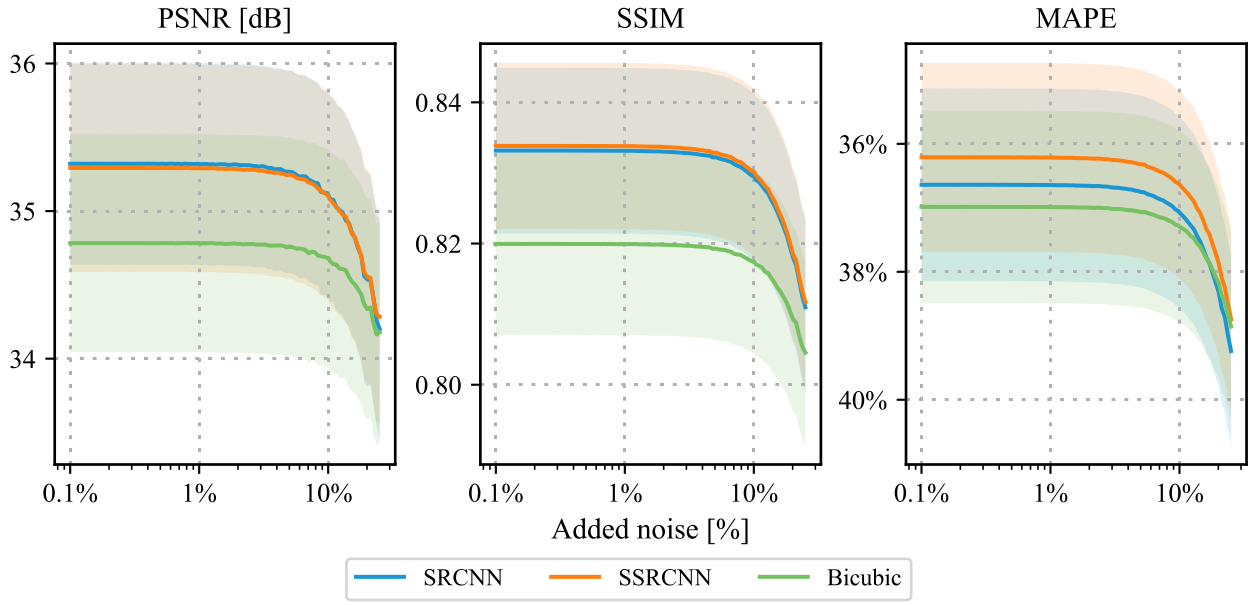


Figure 16: Noise sensitivity of models over test data. Evolution of metrics when different amount of random noise is added to the data. In transparent, 95% confidence intervals for the test images.

5.4. Application to OMI NO_2 data

The methodology proposed in this work focuses, but is not limited, to the application on OMI NO_2 data. Given the results presented, the trained model seems suitable to be used on OMI images successfully. A first application is made for the year 2016, when TROPOMI had not yet been launched, in which the images of the entire year are retrieved from OMI, and afterwards the SSRCNN model is applied, comparing annual average against an updated NO2 emissions inventory (Álamos et al., 2021). The original OMI yearly mean, the outputs of the model and the annual inventory can be compared in Figure 17, where an improvement in resolution can clearly be noticed.

One of the objectives of this research is to create a high-resolution OMI NO_2 data repository for future use. Therefore, in future work the model will be applied to the entire OMI database for the selected area, and both the training of better models and the application for larger areas will be evaluated. In addition, a dispersion model will be considered in order to be able to better compare the results obtained with the available emissions inventories.

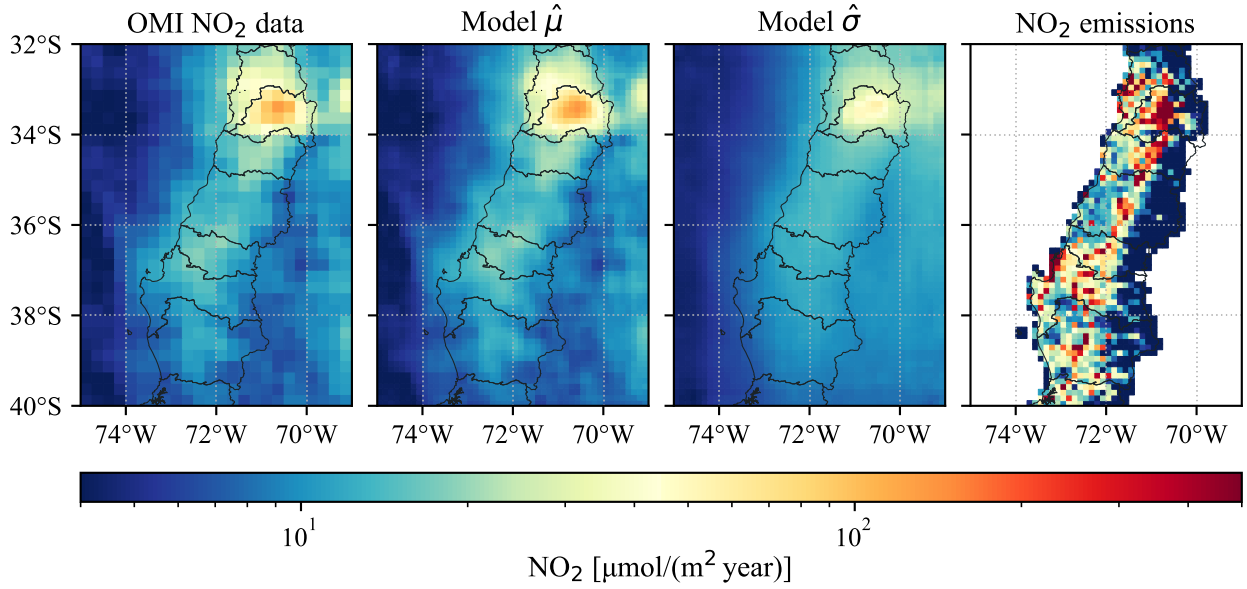


Figure 17: SSRCNN model application. Spatial distribution for 2016 of: (a) NO_2 as observed from OMI at 25x25 km^2 resolution (NASA); (b) NO_2 as estimated by the SSRCNN model at 12.5x12.5 km^2 resolution; (c) the model estimated error at 12.5x12.5 km^2; and (d) the emission fluxes interpolated to 12.5x12.5 km^2; all for year 2016.

6. Conclusions

The present work aimed to evaluate the feasibility of using CNN-based models to improve the spatial resolution of NO_2's satellite measurements, as well as to estimate the random uncertainty of the prediction, inherently provided by these data. It focused on improving the resolution of OMI NO_2, using the data provided by TROPOMI as a reference. To this end, a methodology was presented to address the problem based on a SRCNN model, which was modified to estimate the desired uncertainty.

To assess the advantages and limitations of the methodology, an SSRCNN model was trained and evaluated, comparing it with an SRCNN and a bicubic interpolation using appropriate reconstruction quality metrics. The experiments were carried out in a domain constrained to the south-central zone of Chile, using data at a specific resolution. The findings suggest that the proposed methodology can achieve performances like a CNN model without estimating uncertainty. In fact, some results showed a better performance for the SSRCNN model, as in the case of MAPE, so, in certain applications, these models could be an alternative with greater estimation precision. Moreover, the model is robust against adversarial noise of up to 10%, so the methodology could be applied to data from instruments with high noise levels. However, SSRCNN models apparently take much longer to converge, which is a limitation when time is a priority or not enough computing resources are available. It should be noted that this problem occurs only during the training of the model, not when applying it.

At last, a first application to OMI NO_2 data was presented. Since an important objective of the work carried out consists in generating a high-resolution OMI data repository, in future work, the optimised model will be applied to reconstruct the complete OMI time series available. This high-resolution database is expected to be useful for future studies in Chile that may benefit from improved resolution. Meanwhile, other architectures and models based in state-of-the-art methods will be tested, along with less constrained domains, seeking to obtain better results and extend the built repository.

There are some important limitations of the study, which must be acknowledged. First, more detailed searches for optimal architectures could lead to models having greater differences in performance. Unfortunately, it is difficult to cover a wide range of architectures due to the time required to train and evaluate each one. Second, despite the similarities between OMI and TROPOMI, there is always a level of error generated by moving from one domain to another. On the other hand, better results would be expected using a greater amount of data, exposing the model to a greater variety of patterns, which contributes to a better generalisation of the data, that is, better models. Finally, further assessment of the uncertainty estimation should be carried out, giving more certainty about the quality of these estimates.

In summary, the methodology appears adequate and robust to estimate high-resolution NO_2 fields, with performances at the level of models without estimation of uncertainty, but at the cost of an increasing in training time. This suggests that it would be worth applying

the methodology in state-of-the-art SR models, seeking to improve precision, up-sampling factors, and the uncertainty estimation. Given the relevance of considering uncertainties, the findings of this research provide valuable insights on the convenience of using deep learning techniques to estimate two-dimensional correlated uncertainties in SR problems, an area that has not yet been widely covered in the literature. Since this work focuses on NO_2 satellite data, it would be relevant for future work to evaluate the performance of SSRCNN models in other types of images.

Data availability

All the data used is publicly available at NASA GES DISC repository, including both TROPOMI and OMI products used. In the future, the data derived from the application of the model will be published altogether with the model at github.com/SanParraguez.

Bibliography

Abadi, M., Bahrman, P., Chen, J., Chen, Z., Davis, A., ... Zheng, X. (2016). TensorFlow: A System for Large-Scale Machine Learning. *USENIX Symposium on Operating Systems Design and Implementation*. https://www.usenix.org/system/files/conference/osdi16/osdi16-abadi.pdf

Aggarwal, C. C. (2018). *Neural Networks and Deep Learning*. Springer. https://doi.org/10.1007/978-3-319-94463-0

Álamos, N., Huneeus, N., Opazo, M., Osses, M., Puja, S., ... Calvo, R. (2021). High resolution inventory of atmospheric emissions from transport, industrial, energy, mining and residential sectors of Chile. *Earth System Science Data Discussions*, *2021*, 1–31. https://doi.org/10.5194/essd-2021-216

Amini, A., Schwarting, W., Soleimany, A., & Rus, D. (2020). Deep Evidential Regression. *Advances in Neural Information Processing Systems*, *33*, 14927–14937. https://proceedings.neurips.cc/paper/2020/file/aab085461de182608ee9f607f3f7d18f-Paper.pdf

Bishop, C. (1995). *Neural Networks for pattern recognition* (1st ed.). Oxford University Press.

Boersma, K. F., Eskes, H. J., Veefkind, J. P., Brinksma, E. J., van der A, R. J., ... Bucsela, E. J. (2007). Near-real time retrieval of tropospheric NO_2 from OMI. *Atmospheric Chemistry and Physics*, *7*(8), 2103–2118. https://doi.org/10.5194/acp-7-2103-2007

Boersma, K. F., Eskes, H. J., Dirksen, R. J., van der A, R. J., Veefkind, J. P., ... Brunner, D. (2011). An improved tropospheric NO_2 column retrieval algorithm for the Ozone Monitoring Instrument. *Atmospheric Measurement Techniques*, *4*(9), 1905–1928. https://doi.org/10.5194/amt-4-1905-2011

Boersma, K. F., Eskes, H. J., & Brinksma, E. J. (2004). Error analysis for tropospheric NO_2 retrieval from space. *Journal of Geophysical Research: Atmospheres*, *109*(4). https://doi.org/10.1029/2003jd003962

Boersma, K. F., Eskes, H. J., Richter, A., De Smedt, I., Lorente, A., ... Compernolle, S. (2018). Improving algorithms and uncertainty estimates for satellite NO_2 retrievals: Results from the Quality Assurance for Essential Climate Variables (QA4ECV) project. *Atmospheric Measurement Techniques Discussions*, 1–70. https://doi.org/10.5194/amt-2018-200

Cawley, G. C., Janacek, G. J., Haylock, M. R., & Dorling, S. R. (2007). Predictive uncertainty in environmental modelling. *Neural Networks*, *20*(4), 537–549. https://doi.org/10.1016/j.neunet.2007.04.024

Cheng, J., Kuang, Q., Shen, C., Liu, J., Tan, X., & Liu, W. (2020). ResLap: Generating High-Resolution Climate Prediction through Image Super-Resolution. *IEEE Access*, *8*, 39623–39634. https://doi.org/10.1109/access.2020.2974785

Copernicus Sentinel-5P (processed by ESA). (2021). TROPOMI Level 2 Nitrogen Dioxide total column products. Version 02. *European Space Agency*. https://doi.org/10.5270/s5p-9bnp8q8

Cover, T. M., & Thomas, J. A. (2006). *Elements of Information Theory*. Wiley-Interscience.

Crutzen, P. J. (1979). The role of NO and NO_2 in the chemistry of the troposphere and stratosphere. *Annual Review of Earth and Planetary Sciences*, *7*(1), 443–472. https://doi.org/10.1146/annurev.ea.07.050179.002303

Dong, C., Loy, C. C., He, K., & Tang, X. (2016). Image Super-Resolution Using Deep Convolutional Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *38*(2), 295–307. https://doi.org/10.1109/tpami.2015.2439281

Duncan, B. N., Lamsal, L. N., Thompson, A. M., Yoshida, Y., Lu, Z., ... Pickering, K. E. (2016). A space-based, high-resolution view of notable changes in urban NO_x pollution around the world (2005-2014). *Journal of Geophysical Research: Atmospheres*, *121*(2), 976–996. https://doi.org/10.1002/2015jd024121

Gallardo, L., Barraza, F., Ceballos, A., Galleguillos, M., Huneeus, N., ... Véliz, K. D. (2018). Evolution of air quality in Santiago: The role of mobility and lessons from the science-policy interface. *Elementa*, *6*. https://doi.org/10.1525/elementa.293

Gallardo, L., Escribano, J., Dawidowski, L., Rojas, N., de Fátima Andrade, M., & Osses, M. (2012). Evaluation of vehicle emission inventories for carbon monoxide and nitrogen oxides for Bogotá, Buenos Aires, Santiago, and São Paulo. *Atmospheric Environment*, *47*, 12–19. https://doi.org/10.1016/j.atmosenv.2011.11.051

Garcia, D. (2010). Robust smoothing of gridded data in one and higher dimensions with missing values. *Computational Statistics and Data Analysis*, *54*(4), 1167–1178. https://doi.org/10.1016/j.csda.2009.09.020

Georgoulias, A. K., van der A, R. J., Stammes, P., Boersma, K. F., & Eskes, H. J. (2019). Trends and trend reversal detection in 2 decades of tropospheric NO_2 satellite observations. *Atmospheric Chemistry and Physics*, *19*(9), 6269–6294. https://doi.org/10.5194/acp-19-6269-2019

Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. *IEEE Conference on Computer Vision and Pattern Recognition*, 580–587.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press. http://www.deeplearningbook.org

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., ... Bengio, Y. (2014). Generative Adversarial Nets. *Advances in Neural Information Processing Systems*, *27*. https://proceedings.neurips.cc/paper/2014/file/5ca3e9b122f61f8f06494c97b1afccf3-Paper.pdf

Griffin, D., Zhao, X., McLinden, C. A., Boersma, F., Bourassa, A., ... Wolde, M. (2019). High-Resolution Mapping of Nitrogen Dioxide With TROPOMI: First Results and Validation Over the Canadian Oil Sands. *Geophysical Research Letters*, *46*(2), 1049–1060. https://doi.org/10.1029/2018gl081095

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, *45*(8), 770–778. https://doi.org/10.1109/cvpr.2016.90

Hoesly, R. M., Smith, S. J., Feng, L., Klimont, Z., Janssens-Maenhout, G., ... Zhang, Q. (2018). Historical (1750–2014) anthropogenic emissions of reactive gases and aerosols from the Community Emissions Data System (CEDS). *Geoscientific Model Development*, *11*(1), 369–408. https://doi.org/10.5194/gmd-11-369-2018

Horé, A., & Ziou, D. (2010). Image quality metrics: PSNR vs. SSIM. *2010 20th International Conference on Pattern Recognition*, 2366–2369. https://doi.org/10.1109/icpr.2010.579

Huang, G., Liu, Z., van der Maaten, L., & Weinberger, K. Q. (2017). Densely Connected Convolutional Networks. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Hüllermeier, E., & Waegeman, W. (2021). Aleatoric and epistemic uncertainty in machine learning: an introduction to concepts and methods. *Machine Learning*, *110*(3), 457–506. https://doi.org/10.1007/s10994-021-05946-3

Huseljic, D., Sick, B., Herde, M., & Kottke, D. (2020). Separation of aleatoric and epistemic uncertainty in deterministic deep neural networks. *International Conference on Pattern Recognition*, 9172–9179. https://doi.org/10.1109/icpr48806.2021.9412616

Judd, L. M., Al-Saadi, J. A., Janz, S. J., Kowalewski, M. J., Bradley Pierce, R., ... Williams, D. (2019). Evaluating the impact of spatial resolution on tropospheric NO_2 column comparisons within urban areas using high-resolution airborne data. *Atmospheric Measurement Techniques*, *12*(11), 6091–6111. https://doi.org/10.5194/amt-12-6091-2019

Kendall, A., & Gal, Y. (2017). What Uncertainties Do We Need in Bayesian Deep Learning for Computer Vision? In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, ... R. Garnett (Eds.), *Advances in neural information processing systems*. Curran Associates, Inc. https://proceedings.neurips.cc/paper/2017/file/2650d6089a6d640c5e85b2b88265dc2b-Paper.pdf

Khosravi, A., Nahavandi, S., Creighton, D., & Atiya, A. F. (2011). Comprehensive review of Neural Network-based prediction intervals and new advances. *IEEE Transactions on Neural Networks*, *22*(9), 1341–1356. https://doi.org/10.1109/tnn.2011.2162110

Kingma, D. P., & Ba, J. (2015). Adam: A method for stochastic optimization. *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA*. http://arxiv.org/abs/1412.6980

Kiureghian, A. D., & Ditlevsen, O. (2009). Aleatory or epistemic? Does it matter? *Structural Safety*, *31*(2), 105–112. https://doi.org/10.1016/j.strusafe.2008.06.020

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep Convolutional Neural Networks. *Advances in Neural Information Processing Systems*, *25*, 1097–1105.

Krotkov, N. A., McLinden, C. A., Li, C., Lamsal, L. N., Celarier, E. A., ... Streets, D. G. (2016). Aura OMI observations of regional SO2 and NO2 pollution changes from 2005 to 2015. *Atmospheric Chemistry and Physics*, *16*(7), 4605–4629. https://doi.org/10.5194/acp-16-4605-2016

Lai, W.-S., Huang, J.-B., Ahuja, N., & Yang, M.-H. (2017). Deep Laplacian Pyramid Networks for Fast and Accurate Super-Resolution. *IEEE Conference on Computer Vision and Pattern Recognition*, 5835–5843. https://doi.org/10.1109/cvpr.2017.618

Lamsal, L. N., Krotkov, N. A., Vasilkov, A., Marchenko, S., Qin, W., ... Bucsela, E. (2021). Ozone Monitoring Instrument (OMI) Aura nitrogen dioxide standard product version 4.0 with improved surface and cloud treatments. *Atmospheric Measurement Techniques*, *14*(1), 455–479. https://doi.org/10.5194/amt-14-455-2021

LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., ... Jackel, L. D. (1989). Backpropagation applied to handwritten zip code recognition. *Neural Computation*, *1*(4), 541–551.

Ledig, C., Theis, L., Huszar, F., Caballero, J., Cunningham, A., ... Shi, W. (2017). Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 105–114. https://doi.org/10.1109/cvpr.2017.19

Leinonen, J., Nerini, D., & Berne, A. (2021). Stochastic Super-Resolution for Downscaling Time-Evolving Atmospheric Fields With a Generative Adversarial Network. *IEEE Transactions on Geoscience and Remote Sensing*, *59*(9), 7211–7223. https://doi.org/10.1109/tgrs.2020.3032790

Levelt, P. F., Joiner, J., Tamminen, J., Veefkind, J. P., Bhartia, P. K., ... Wargan, K. (2018). The Ozone Monitoring Instrument: Overview of 14 years in space. *Atmospheric Chemistry and Physics*, *18*(8), 5699–5745. https://doi.org/10.5194/acp-18-5699-2018

Levelt, P. F., van Den Oord, G. H., Dobber, M. R., Mälkki, A., Visser, H., ... Saari, H. (2006). The Ozone Monitoring Instrument. *IEEE Transactions on Geoscience and Remote Sensing*, *44*(5), 1093–1100. https://doi.org/10.1109/tgrs.2006.872333

Lim, B., Son, S., Kim, H., Nah, S., & Lee, K. M. (2017). Enhanced Deep Residual Networks for Single Image Super-Resolution. *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, *2017-July*, 1132–1140. https://doi.org/10.1109/cvprw.2017.151

Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. *IEEE Conference on Computer Vision and Pattern Recognition*, 3431–3440.

Lorente, A., Boersma, K. F., Eskes, H. J., Veefkind, J. P., van Geffen, J. H. G. M., ... Krol, M. C. (2019). Quantification of nitrogen oxides emissions from build-up of pollution over Paris with TROPOMI. *Scientific Reports*, *9*(1). https://doi.org/10.1038/s41598-019-56428-5

Lorente, A., Boersma, K. F., Yu, H., Dörner, S., Hilboll, A., ... Krol, M. (2017). Structural uncertainty in air mass factor calculation for NO_2 and HCHO satellite retrievals. *Atmospheric Measurement Techniques*, *10*(3), 759–782. https://doi.org/10.5194/amt-10-759-2017

Marais, E. A., Roberts, J. F., Ryan, R. G., Eskes, H., Boersma, K. F., ... Navarro-Comas, M. (2021). New observations of NO2 in the upper troposphere from TROPOMI. *Atmospheric Measurement Techniques*, *14*(3), 2389–2408. https://doi.org/10.5194/amt-14-2389-2021

Menares, C., Gallardo, L., Kanakidou, M., Seguel, R., & Huneeus, N. (2020). Increasing trends (2001–2018) in photochemical activity and secondary aerosols in Santiago, Chile. *Tellus B: Chemical and Physical Meteorology*, *72*(1), 1–18. https://doi.org/10.1080/16000889.2020.1821512

Nasrollahi, K., & Moeslund, T. B. (2014). Super-resolution: a comprehensive survey. *Machine Vision and Applications*, *25*(6), 1423–1468. https://doi.org/10.1007/s00138-014-0623-4

Nix, D. A., & Weigend, A. S. (1994). Estimating the mean and variance of the target probability distribution. *IEEE International Conference on Neural Networks - Conference Proceedings*, *1*, 55–60. https://doi.org/10.1109/icnn.1994.374138

Papadopoulos, G., Edwards, P. J., & Murray, A. F. (2001). Confidence estimation methods for neural networks: A practical comparison. *IEEE Transactions on Neural Networks*, *12*(6), 1278–1287. https://doi.org/10.1109/72.963764

Platt, U., & Stutz, J. (2008). *Differential Optical Absorption Spectroscopy* (1st ed.). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-75776-4

Ren, S., He, K., Girshick, R., & Sun, J. (2016). Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE transactions on pattern analysis and machine intelligence*, *39*(6), 1137–1149.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, *323*(6088), 533–536. https://doi.org/10.1038/323533a0

Seguel, R. J., Gallardo, L., Fleming, Z. L., & Landeros, S. (2020). Two decades of ozone standard exceedances in Santiago de Chile. *Air Quality, Atmosphere & Health*, *13*(5), 593–605. https://doi.org/10.1007/s11869-020-00822-w

Seinfeld, J. H., & Pandis, S. N. (2016). *Atmospheric Chemistry and Physics: From Air Pollution to Climate Change* (3rd ed.). Wiley-Blackwell.

Simonyan, K., & Zisserman, A. (2015). Very deep Convolutional Networks for large-scale image recognition. *3rd International Conference on Learning Representations*. https://arxiv.org/abs/1409.1556v6

Stengel, K., Glaws, A., Hettinger, D., & King, R. N. (2020). Adversarial super-resolution of climatological wind and solar data. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(29), 16805–16815. https://doi.org/10.1073/pnas.1918964117

Swinehart, D. F. (1962). The Beer-Lambert law. *Journal of Chemical Education*, *39*(7), 333–335. https://doi.org/10.1021/ed039p333

Szegedy, C., Ioffe, S., Vanhoucke, V., & Alemi, A. (2017). Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning. https://www.aaai.org/ocs/index.php/AAAI/AAAI17/paper/view/14806/14311

Szopa, S., Naik, V., Adhikary, B., Artaxo, P., Berntsen, T., ... Zanis, P. (2021). Short-Lived Climate Forcers. *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press. https://www.ipcc.ch/report/ar6/wg1

Van Geffen, J., Boersma, K. F., Eskes, H. J., Sneep, M., Ter Linden, M., ... Veefkind, J. P. (2020). S5P TROPOMI NO_2 slant column retrieval: Method, stability, uncertainties and comparisons with OMI. *Atmospheric Measurement Techniques*, *13*(3), 1315–1335. https://doi.org/10.5194/amt-13-1315-2020

Van Geffen, J., Eskes, H. J., Boersma, K. F., Maasakkers, J. D., & Veefkind, J. P. (2021). TROPOMI ATBD of the total and tropospheric NO_2 data products. (version 2.2.0). https://sentinel.esa.int/documents/247904/2476257/sentinel-5p-tropomi-atbd-no2-data-products

Vandal, T., Kodra, E., Ganguly, S., Michaelis, A., Nemani, R., & Ganguly, A. R. (2017). DeepSD: Generating high resolution climate change projections through single image super-resolution. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, *Part F1296*, 1663–1672. https://doi.org/10.1145/3097983.3098004

Veefkind, J. P., Aben, I., McMullan, K., Förster, H., de Vries, J., ... Levelt, P. F. (2012). TROPOMI on the ESA Sentinel-5 Precursor: A GMES mission for global observations of the atmospheric composition for climate, air quality and ozone layer applications. *Remote Sensing of Environment*, *120*(2012), 70–83. https://doi.org/10.1016/j.rse.2011.09.027

Vincent, P., Larochelle, H., Bengio, Y., & Manzagol, P.-A. (2008). Extracting and composing robust features with denoising autoencoders. *Proceedings of the 25th International Conference on Machine learning*, 1096–1103.

Voigt, S., Orphal, J., & Burrows, J. (2002). The temperature and pressure dependence of the absorption cross-sections of NO_2 in the 250–800 nm region measured by Fourier-transform spectroscopy. *Journal of Photochemistry and Photobiology A: Chemistry*, *149*(1), 1–7. https://doi.org/https://doi.org/10.1016/S1010-6030(01)00650-5

Wang, G., Garcia, D., Liu, Y., de Jeu, R., & Johannes Dolman, A. (2012). A three-dimensional gap filling method for large geophysical datasets: Application to global satellite soil moisture observations. *Environmental Modelling and Software*, *30*, 139–142. https://doi.org/10.1016/j.envsoft.2011.10.015

Wang, X., Yu, K., Wu, S., Gu, J., Liu, Y., ... Loy, C. C. (2019). ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks. *European Conference on Computer Vision 2018*. https://doi.org/10.1007/978-3-030-11021-5

Wang, Z., Chen, J., & Hoi, S. C. (2020). Deep learning for image super-resolution: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *8828*(100), 1–22. https://doi.org/10.1109/tpami.2020.2982166

Wang, Z., Bovik, A., Sheikh, H., & Simoncelli, E. (2004). Image Quality Assessment: From Error Visibility to Structural Similarity. *IEEE Transactions on Image Processing*, *13*(4), 600–612. https://doi.org/10.1109/tip.2003.819861

World Health Organization. (2021). *WHO global air quality guidelines: particulate matter ($PM_{2.5}$ and PM_{10}), ozone, nitrogen dioxide, sulfur dioxide and carbon monoxide*. World Health Organization. https://apps.who.int/iris/handle/10665/345329

World Health Organization. Regional Office for Europe. (2003). *Health aspects of air pollution with particulate matter, ozone and nitrogen dioxide*. Report on a WHO working group, Bonn, Germany 13-15 January 2003. Copenhagen: WHO Regional Office for Europe. https://apps.who.int/iris/handle/10665/107478

Yang, C.-Y., Ma, C., & Yang, M.-h. (2014). Single-Image Super-Resolution: A Benchmark. *Computer vision – eccv 2014* (pp. 372–386). Springer. https://doi.org/10.1007/978-3-319-10593-2_25

Zara, M., Boersma, K. F., De Smedt, I., Richter, A., Peters, E., ... Eskes, H. J. (2018). Improved slant column density retrieval of nitrogen dioxide and formaldehyde for OMI and GOME-2A from QA4ECV: intercomparison, uncertainty characterisation, and trends. *Atmospheric Measurement Techniques*, *11*(7), 4033–4058. https://doi.org/10.5194/amt-11-4033-2018

Appendix A. Loss function gradients

During training process, a loss function is used for assessing the model performance and then propagate the errors to the parameters. This is known as backpropagation, which allows to update the parameters at each iteration during model optimisation. The gradients for the cost functions used in the training of the two models are detailed below.

A.1. Maximum Likelihood for Gaussian distribution

The Maximum Likelihood Estimation (MLE) seeks to maximise the probability of finding the desired value given a distribution. To use this approach as a minimisation problem, the negative log-likelihood is taken, given the following loss function:

$$\mathcal{L}(\theta) = -\frac{1}{n}\sum_{i=1}^{n} \log\, p(y_i)\,, \tag{A1}$$

where p is the probability of finding y. For this case, this probability is given by the Gaussian distribution, which is set as:

$$p(\, y \,|\, \hat{\mu}, \hat{\sigma}^2\,) = \frac{1}{\sqrt{2\pi\hat{\sigma}^2}} \exp\left(-\frac{(y-\hat{\mu})^2}{2\hat{\sigma}^2}\right). \tag{A2}$$

Then, replacing on Equation A1, the loss function can bee simplified by:

$$\mathcal{L}(\theta) = -\frac{1}{n}\sum_{i=1}^{n} \log\left(\frac{1}{\sqrt{2\pi\hat{\sigma}_i^2}} \exp\left(-\frac{(y_i-\hat{\mu}_i)^2}{2\hat{\sigma}_i^2}\right)\right) = \frac{1}{2n}\sum_{i=1}^{n} \log(2\pi\hat{\sigma}_i^2) + \frac{(y_i-\hat{\mu}_i)^2}{\hat{\sigma}_i^2}\,. \tag{A3}$$

Subsequently, the derivatives of the loss function must be calculated to perform the backpropagation and update the parameters θ. Since μ and σ are calculated, both derivatives must be obtained separately as following:

$$\frac{\partial \mathcal{L}}{\partial \hat{\mu}} = -\frac{1}{n}\sum_{i=1}^{n} \frac{y_i-\hat{\mu}_i}{\hat{\sigma}_i^2}\,, \tag{A4}$$

$$\frac{\partial \mathcal{L}}{\partial \hat{\sigma}^2} = \frac{1}{2n}\sum_{i=1}^{n} \frac{1}{\hat{\sigma}_i^2} - \frac{(y_i-\hat{\mu}_i)^2}{\hat{\sigma}_i^4}\,. \tag{A5}$$

These values can be used to estimate the gradient of each parameter of the model and perform the optimisation process.

A.2. Mean Squared Error

The Mean Squared Error (MSE) is a commonly used loss function in Neural Networks, strongly penalising higher differences between the prediction values and the ground truth. This function is defined as:

$$\mathcal{L}(\theta) = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \,, \tag{A6}$$

where is only one prediction value to optimise. Therefore, the only gradient to calculate follows:

$$\frac{\partial \mathcal{L}}{\partial \hat{y}} = \frac{2}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i) \,. \tag{A7}$$

Hence, parameters of the model are updated by propagating only this loss.

Appendix B. Parameterisation of models' architectures

The parameterisation implemented to generate the architectures consists in 14 variables common to both models, 4 variables for the SRCNN model and 9 variables for the SSRCNN model. Each of these parameters are shown and explained in Table B1. The differences in the parameters lie exclusively in the output layer of each architecture. Both models are built with a bullet shape, where it maintains a stretch of layers of constant size and then descends, linearly, towards the output layer.

Both CNN models' architectures are optimised by random search and then grid search. The first step consists of generating a search domain in which parameters are randomly selected to generate an architecture to be trained. The probability of selecting each value within a parameter is uniform in each domain presented in Table B1. After this first search, a grid search is carried out to refine the initially obtained results. The same number of searches was carried out for both models; that is, 1000 architectures by random search followed by 200 architectures by grid search in each case.

Table B1
Input parameters to generate the architectures.

Parameter	Model	Possible values	Description
batch_size	–	{32, 64}	Size of the batch of data handle at the same time during an iteration.
learning_rate	–	{1e-3, 5e-4, 1e-4, 5e-5, 1e-5}	Step size given to the Adam optimiser.
n_layers	–	{2, 3, 4, ..., 7}	Number of layers in the network.
slope	–	{1.1, 1.2, 1.3, ..., 3.5}	Factor of reduction of the number of neurons in each layer.
last_layer	–	{8, 16, 32, 64}	Number of neurons in the last layer before the output.
const_layers	–	[0, 1]	Fraction of the layers with constant number of neurons.
dropout	–	{0.0, 0.1, 0.2}	Fraction of filters to be dropped during training.
up_sampling	–	[0, 1]	Fraction of the network where the up-sampling is performed.
ker_lr	–	{3, 5, 7, 9}	Size of the kernel utilised before the up-sampling.
ker_hr	–	{3, 5, 7, 9}	Size of the kernel utilised after the up-sampling.
l1_lay	–	$\{0\} \cup \{10^{-k} \mid k \in \mathbb{Z} \ \wedge \ 1 \leq k \leq 9\}$	L1 regularisation of the deep layers.
l2_lay	–	$\{0\} \cup \{10^{-k} \mid k \in \mathbb{Z} \ \wedge \ 1 \leq k \leq 9\}$	L2 regularisation of the deep layers.
layer_act	–	{linear, ReLU, ELU, SELU}	Activation function for the deep layers.
ker_out	Base	{3, 5, 7, 9}	Size of the kernel of the output convolution.
ker_mu	Stch	{3, 5, 7, 9}	Size of the kernel of the output μ convolution.
ker_si	Stch	{3, 5, 7, 9}	Size of the kernel of the output σ convolution.
l1_out	Base	$\{0\} \cup \{10^{-k} \mid k \in \mathbb{Z} \ \wedge \ 1 \leq k \leq 9\}$	L1 regularisation of the output layer.
l2_out	Base	$\{0\} \cup \{10^{-k} \mid k \in \mathbb{Z} \ \wedge \ 1 \leq k \leq 9\}$	L2 regularisation of the output layer.
l1_mu	Stch	$\{0\} \cup \{10^{-k} \mid k \in \mathbb{Z} \ \wedge \ 1 \leq k \leq 9\}$	L1 regularisation of the μ output layer.
l2_mu	Stch	$\{0\} \cup \{10^{-k} \mid k \in \mathbb{Z} \ \wedge \ 1 \leq k \leq 9\}$	L2 regularisation of the μ output layer.
l1_sig	Stch	$\{0\} \cup \{10^{-k} \mid k \in \mathbb{Z} \ \wedge \ 1 \leq k \leq 9\}$	L1 regularisation of the σ output layer.
l2_sig	Stch	$\{0\} \cup \{10^{-k} \mid k \in \mathbb{Z} \ \wedge \ 1 \leq k \leq 9\}$	L2 regularisation of the σ output layer.
output_act	Base	{linear, ReLU, ELU, SELU}	Activation function for the output layer
mu_act	Stch	{linear, ReLU, ELU, SELU}	Activation function for the μ output layer.
sig_act	Stch	{Softplus, Exponential}	Activation function for the σ output layer.

Base: SRCNN model, Stch: SSRCNN model

Appendix C. SSRCNN prediction examples

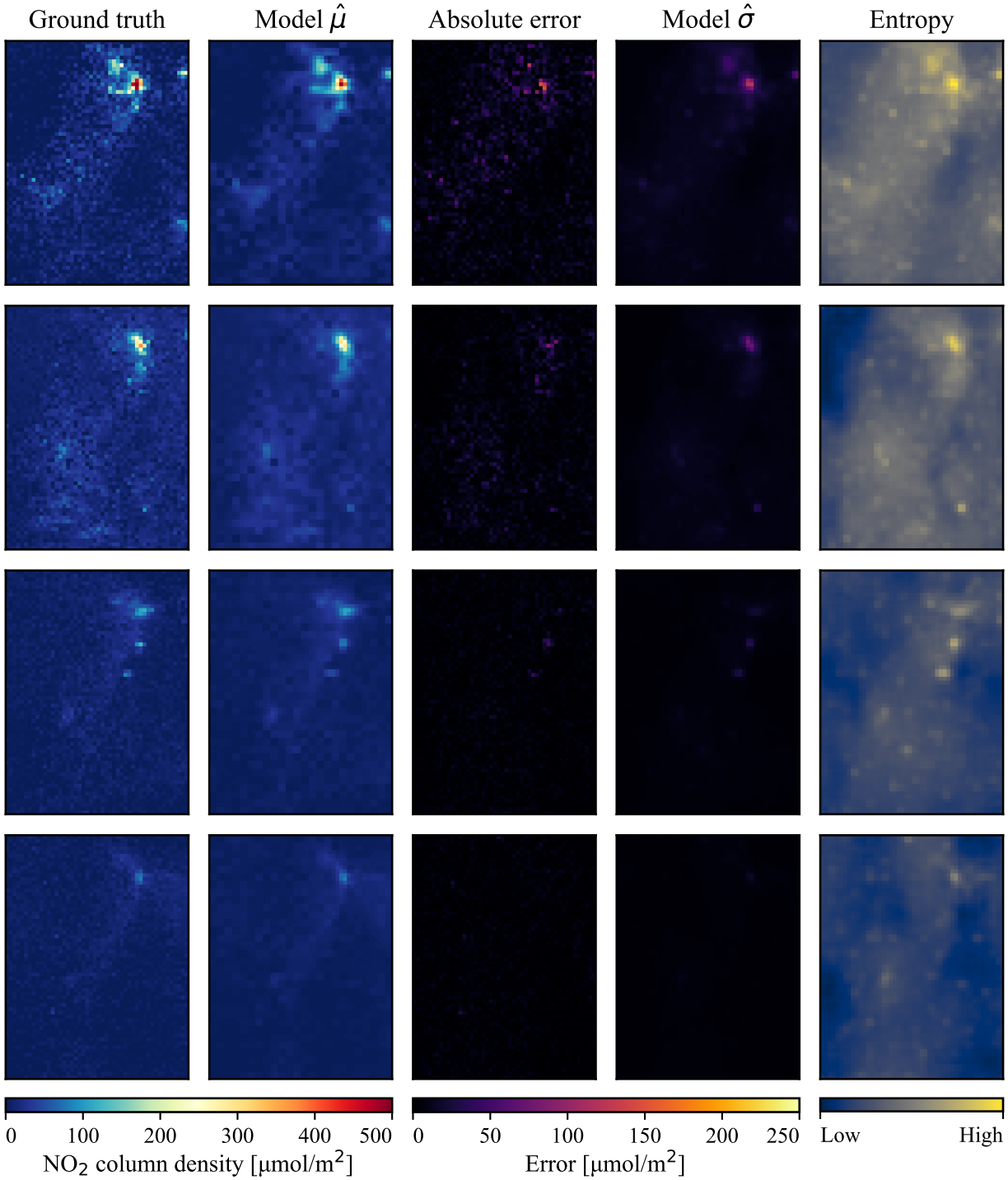


Figure C1: SSRCNN model prediction examples. By column: (a) TROPOMI truth image; (b) model prediction; (c) absolute error of the prediction; (d) estimated error by the model; and (e) entropy.