

Similar to Dumont Le Brazidec et al. (2024), the concentration map can be expressed as a linear combination of various tracers, given by

$$I = \alpha_{\text{bg}} I_{\text{bg}} + \sum_i^N \alpha_{\text{ps},i} s_i I_{\text{ps},i} + \alpha_{\text{anth}} I_{\text{anth}} + \epsilon. \quad (2)$$

Here  $I_{\text{bg}}$ ,  $I_{\text{ps},i}$  and  $I_{\text{anth}}$  denote the background CO<sub>2</sub> concentration, CO<sub>2</sub> from  $i$ th of total  $N$  point sources and CO<sub>2</sub> from other anthropogenic sources, respectively.  $\alpha_{\text{bg}}$ ,  $\alpha_{\text{ps},i}$  and  $\alpha_{\text{anth}}$  denote the corresponding scaling factors. The emission rates are scaled accordingly.  $s_i$  denotes a binary switch variable for the  $i$ th point source. The scaling factors and switching variables are derived from random distributions.  $\epsilon$  denotes the observation noise, which is simplified as Gaussian noise.

We introduce three scaling factors, given by

$$\alpha_{\text{bg}} = \frac{c_{\text{bg}}}{|I_{\text{bg}}|}, \alpha_{\text{ps},i} = \frac{e_{\text{ps},i}}{e_{\text{ps},i}^{\text{ref}}}, \alpha_{\text{anth}} = \frac{e_{\text{anth}}}{e_{\text{anth}}^{\text{ref}}}. \quad (3)$$

Here,  $|I_{\text{bg}}|$  denotes the mean concentration of the background CO<sub>2</sub> tracer from the WRF-GHG output;  $e_{\text{ps},i}^{\text{ref}}$  and  $e_{\text{anth}}^{\text{ref}}$  denote the reference emission rate for the  $i$ th point source and the non-point-source anthropogenic sources, respectively.  $c_{\text{bg}}$ ,  $e_{\text{ps},i}$  and  $e_{\text{anth}}$  are random scalars drawn from several distributions. To account for the observed correlation between emissions and background concentrations, as noted by such as Hakkarainen et al. (2016),  $c_{\text{bg}}$  and  $e_{\text{anth}}$  are sampled from their empirical joint distribution, estimated using EDGAR emission data (Crippa et al., 2024) and Carbon-Tracker XCO<sub>2</sub> data (Jacobson et al., 2023) in global CO<sub>2</sub> hotspot areas, specifically, 100 km × 100 km areas surrounding power plants emitting over 5 MtCO<sub>2</sub> yr<sup>-1</sup>.  $c_{\text{bg}}$  is then adjusted to track the predicted global mean XCO<sub>2</sub> under SSP1-2.6 (Meinshausen et al., 2020), i.e., the 2 °C scenario of the Paris Agreement, spanning the full lifetime of the TanSat-2 mission. Each  $e_{\text{ps},i}$  is sampled from the major power plants in the CARMA v3.0 inventory (Ummel, 2012). Finally, we introduce binary switch variables,  $s_i \sim \text{Bernoulli}(p)$ , where  $N \times p$  is the expectation of the quantities of power plants within the sampling area. The Supplement provides further details on the relevant distributions.

## 2.2 Deep neural network for GHG point source extraction (GHGPSE-Net)

A key goal of spaceborne GHG point source monitoring is the efficient and accurate detection, localization, and quantification of emission sources. This is essential for source attribution (Rafiq et al., 2020) and for coordinating with other observational missions (Irakulis-Loitxate et al., 2022; Chiba et al., 2019), particularly when the satellite spatial resolution is sparse. Traditional segmentation-based methods, however, cannot extract both source locations and emission rates automatically from a single concentration map.

Over the past decade, the remote-sensing community has widely adopted the deep learning based object-detection techniques, which focus on object localization and classification rather than pixel-wise segmentation (Zhang et al., 2023). There are two main paradigms, anchor-based neural network and anchor-free neural network. The anchor-based networks, such as Single Shot MultiBox Detector (SSD) (Liu et al., 2016) and Mask Region-based Convolutional Neural Network (Mask R-CNN) (He et al., 2017), predict bounding boxes directly. In contrast, anchor-free networks, such as CornerNet (Law and Deng, 2019) and CenterNet (Duan et al., 2019), extract object centers as key points. CenterNet, in particular, generates a heatmap with peaks corresponding to object centers whose intensities represent attributes such as length, width and orientation (Zhou et al., 2019).

Inspired by CenterNet (Duan et al., 2019), we propose GHGPSE-Net (shown in Fig. 4), a CNN-based model that converts an XCO<sub>2</sub> concentration map into an emission heatmap composed of Gaussian kernels, as to mitigate human interventions. Source locations and emission rates are then derived using Gaussian kernel fitting (GKF).

### 2.2.1 Deep learning model for heatmap prediction

We represent GHG point-source emissions as a heatmap generated by the summation of a series of two-dimensional Gaussian kernels, serving as the neural network's learning target. Each kernel's center and amplitude correspond to the source location and the emission rate. At pixel coordinate  $\mathbf{x} = [x, y]^T$ , the heatmap is given by

$$I(\mathbf{x}) = \sum_{i=1}^N a_i G_i(\mathbf{x}; \boldsymbol{\mu}_i), \quad (4)$$

where  $a_i \geq 0$  denotes the scale of the  $i$ th kernel centered at  $\boldsymbol{\mu}_i = [\mu_{x,i}, \mu_{y,i}]^T$ . The Gaussian function  $G_i$  is given by

$$G_i(\mathbf{x}; \boldsymbol{\mu}_i) = \exp\left(-\frac{1}{2} \frac{(\mathbf{x} - \boldsymbol{\mu}_i)^T (\mathbf{x} - \boldsymbol{\mu}_i)}{\sigma^2}\right), \quad (5)$$

where  $\sigma$  defines the spatial extent (hereafter intuitively referred to as the kernel size).

We train a CNN deep learning model using supervised learning to infer emission heatmaps from XCO<sub>2</sub> concentration maps. Deep neural networks are a class of universal machine learning algorithms, which have been mathematically proved to approximate any continuous real function within a hypercube (Cybenko, 1989). A CNN primarily consists of convolutional layers, activation functions, pooling layers and linear layers, allowing it to extract compact feature representations from complex, sparse inputs. Popular image-to-image CNN architectures, such as UNet (Ronneberger et al., 2015) and HourglassNet (Newell et al., 2016), preserve spatial resolution with output feature maps similar in size to the input images, making them suitable for emission heatmap inference.