On the theory of the divergence method for quantifying source emissions from satellite observations

Erik Franciscus Maria Koene[1], Dominik Brunner[2], and Gerrit Kuhlmann[3]

[1]Empa, Swiss Federal Laboratories for Materials Science and Technology
[2]EMPA
[3]Empa

May 08, 2024

On the theory of the divergence method for quantifying source emissions from satellite observations

E. F. M. Koene[1], **D. Brunner**[1] **& G. Kuhlmann**[1]

[1]Empa, Swiss Federal Laboratories for Materials Science and Technology, Überlandstrasse 129, Dübendorf, Switzerland

Key Points:

- The divergence method for estimating emissions from satellite images is derived from first principles.
- Assumptions made on the way are made explicit, highlighting discrepancies with existing literature.
- A comparison is made against the cross-sectional flux method which follows from the same theory.

0322 Constituent sources and sinks

0394 Instruments and techniques

3336 Numerical approximations and analyses

3360 Remote sensing

Corresponding author: Erik Koene, `erik.koene@empa.ch`

Abstract

The divergence method, a lightweight approach for estimating emission fluxes from satellite images, rests on a few implicit assumptions. This paper explicitly outlines these assumptions by deriving the method from first principles. The assumptions are: the enhanced mass flux is dominated by advection, normal fluxes vanish at the top and bottom of the atmosphere, steady-state conditions apply, sources are multiplications of temporal and spatial functions, sinks are described as first-order reactions, and effective wind fields are concentration-weighted wind fields. No such assumptions have to be made for the background field. A 'topography correction term' does not follow from the theory, but is rather shown to be a practical correction for topography-dependent effective wind speed errors. The cross-sectional flux method follows naturally from the derived theory, and the methods are compared. Effects of discrete pixels and finite-difference operations are explored, leading to recommendations, primarily the recommendation to integrate over small regions only to minimize the influence of noise. Numerical examples featuring Gaussian plumes and COSMO-GHG simulated plumes are provided. The Gaussian plume example suggests that the divergence method might underestimate emissions when assuming only advection in the presence of cross-wind diffusion. Conversely, the cross-sectional flux method remains unaffected, provided fluxes are integrated across the entire plume. The COSMO-GHG example reveals frequent violations of the steady-state assumption, although the assumption remains valid proximal to the source (<20 km in this example). It is the hope that this paper provides a solid theoretical foundation for the divergence and cross-sectional flux methods.

Plain Language Summary

Power plants, megacities, and other regions can be places where considerable amounts of gases are emitted into the atmosphere. Satellite data is capable of recording the enhanced concentrations in the atmosphere due to these sources. How does one estimate emissions of these sources based on this data? Typically by running an atmospheric inverse model; but one computationally lightweight method gaining popularity is the 'divergence method'. With this method, one applies a simple mathematical operation to the satellite image multiplied by the horizontal wind speeds – the result of which is related to the emissions. The method has been used before, but here we systematically state the assumptions involved; factors like how gases are transported horizontally and over time, and what happens when we apply the method on pixelated data (such as satellite images). We find that some other analyses in the literature may be partially incorrect. Another popular method (the 'cross-sectional flux') is easily derived from the same theory. Synthetic examples are used to illustrate that various assumptions are not met in realistic data; although the examples also show that one can work around these limitations by a careful application of the methods, and by acquiring more estimates over many days.

1 Introduction

Turning trace gas concentration satellite observations into quantified source contributions is a hard inverse problem. The relation between the source emissions on the one hand, and the satellite observations on the other, depends on atmospheric transport. Solving this inverse problem thus typically involves the use of an atmospheric transport model (see, e.g., Houweling et al., 2015; Jacob et al., 2016; Broquet et al., 2018; Ye et al., 2020; Kaminski et al., 2022). However, atmospheric transport simulations at plume-resolving scale, i.e., at a resolution of the order of 1 km or better (Brunner et al., 2023; Koene & Brunner, 2023), are computationally very expensive. It can be expected that the upcoming Copernicus CO_2 Monitoring (CO2M) satellite mission (Janssens-Maenhout et al., 2020; Meijer et al., 2020) will observe several tens of thousands of plumes globally every year for CO_2 alone; that is, 900 strong point sources (> 3.5 Mt/yr) and up to 300 megacities, imaged about 10–50 times annually, e.g., when conditions are cloud-free (Kuhlmann et al., 2021; Wang et al., 2020; Koene et al., 2021). Automatically carrying out tens of thousands of plume-resolving numerical inversions yearly using transport models is currently unsuitably expensive. Hence, the use of simplified inversion methods is of interest for implementing a relatively inexpensive monitoring and verification system (MVS) to track trace gas emissions over space and time (Pinty et al., 2017).

Several light-weight inversion approaches tailored to satellite or airborne total column observations have been presented in the literature, which may be grouped into three general classes:

1. **Fitting of simulated data to observed data using simplified transport models**. Replacing a complex atmospheric transport model by a simplified model can save orders of magnitudes of computational cost. Examples of simplified models that are able to represent the main features of plume transport and dispersion are Gaussian plume models (e.g., Stockie, 2011; Bovensmann et al., 2010; Nassar et al., 2017; Zheng et al., 2020) and Lagrangian particle dispersion models (e.g., Wu et al., 2018; Lauvaux et al., 2022), though the latter is already much more expensive than the former. The inverse problem is then reduced to fitting the simulated plumes to the observed plumes, which directly provides the likely source emission rates.
2. **Direct application of mass conservation**. Atmospheric transport takes place through advective, turbulent and convective transport. Whatever the transport mechanism, however, the principle of mass conservation holds: net outflow of mass plus net accumulation of mass in a volume corresponds (for a non-reactive gas) directly to the generation and/or depletion of mass within that region. Hence, computations of the mass flux based on the satellite image and known meteorological transport velocities can be used to estimate source emission rates. Examples of such mass balance methods are the cross-sectional flux (e.g., Conley et al., 2017; Varon et al., 2018; Kuhlmann et al., 2019) and the divergence methods (e.g., Beirle et al., 2019, 2021, 2023).
3. **Learning-based methods**. Whereas the first two classes are based on an inductive and a deductive step (i.e., choosing a physical model, then estimating emissions based on that model), the learning-based methods take a transductive step. They immediately estimate the emission rates from a set of known input-output pairs. For example, the integrated mass enhancement (IME) method relates the total integrated plume mass estimated from a satellite image directly to the source emission rate, after calibrating the relation between IME and effective wind for simulated cases in a range of meteorological settings (Varon et al., 2018). A machine learning approach can be similarly used to obtain such a direct inversion (e.g., Bréon et al., 2021; Dumont Le Brazidec et al., 2023).

In this paper, we take a close look at the theory for methods in the second identified class, i.e., those based on considerations of mass conservation, with a particular focus on the divergence method. That there is a link between the divergence and cross-sectional flux method is straightforward to appreciate through Gauss' divergence theorem (nicely illustrated in Conley et al., 2017), and we will explicitly show how this equality applies also for 2D integrated columns.

The divergence method has mainly been applied to TROPOMI data (primarily NO_2, but also CO and CH_4 data (e.g., Beirle et al., 2019, 2021, 2023; Liu et al., 2021; de Foy & Schauer, 2022; Sun, 2022; Rey-Pommier et al., 2022, 2023; Dix et al., 2022; Filonchyk & Peterson, 2023; Lonsdale & Sun, 2023). Recently, the method has also been applied to other satellites such as GEMS NO_2 data (Xu et al., 2023). A study of its application to synthetic CO_2 data for the upcoming CO2M mission can be found in Hakkarainen et al. (2022), which notes that a background and noise removal step must be applied to make the method robust for CO_2 data. Since its inception by Beirle et al. (2019), notable improvements to the method include a different order of operations (differentiate-then-average instead of average-then-differentiate) as proposed in de Foy and Schauer (2022), computing the results on the satellite swaths themselves before remapping to a uniform grid as proposed in de Foy and Schauer (2022), skipping computations proportional to the wind divergence and adding a correction term proportional to the gradient of the topography as proposed in Sun (2022), and approximating the divergence operation using finite-difference stencils that can account for non-pixel-aligned transport as proposed in Sun (2022) and Veefkind et al. (2023).

In this paper we will derive the divergence method from first principles, i.e., the continuity equation, and show all the assumptions made in the process, which have thus far not been discussed in detail in the literature. Furthermore, we show when and why it is advantageous (and physically sound) to remove the background component prior to computing divergence maps, and discuss the finding of Sun (2022) that a topography correction is required, showing that it does not follow from the continuity equation, but rather from not accounting for topography when computing the effective winds. The goal is to provide a solid theoretical grounding for the divergence method and cross-sectional flux method alike. Furthermore, we derive the method in Cartesian and spherical coordinates to aid processing of data in either coordinate systems. Finally, we show a number of numerical examples to demonstrate the theoretical and practical applicability of, and differences between, the divergence and cross-sectional flux methods.

2 Mathematical framework and assumptions

We will make a derivation of the divergence method in Cartesian coordinates. The Cartesian coordinate grid corresponds to a projected planar coordinate system, e.g., the transverse Mercator projection. An analogous treatment of the theory in spherical coordinates can be found in the Supplementary Information.

2.1 The continuity equation in Cartesian coordinates

The conservative form of the mass continuity equation for points (x, y, z) in Cartesian coordinates and at time t states that the net addition of sources and sinks (S) minus the time-varying accumulation of mass equals the net mass outflow through a volume V,

$$\iiint_V \left(S(x,y,z,t) - \frac{\partial \rho(x,y,z,t)}{\partial t} \right) \mathrm{d}V = \iiint_V \nabla \cdot \mathbf{F}(x,y,z,t;\rho,\mathbf{v})\, \mathrm{d}V, \tag{1}$$

where S [kg m^{-3} s^{-1}] describes the source (positive) and sink (negative) contribution within volume V, ρ [kg m^{-3}] here describes the density field of the considered trace gas (not the density of air, as common in meteorology), and $\mathbf{F}(x,y,z,t;\rho,\mathbf{v}) = \rho\mathbf{v}$ [kg m^{-2} s^{-1}] is the mass flux vector field for density ρ multiplied with transport vector $\mathbf{v}$, which describes the transport of the trace gas by the wind field. In Cartesian space, we use the volume element $\mathrm{d}V = \mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z$. We take x to be the coordinate in the easterly direction, y the coordinate in the northerly direction, and z represents the vertical coordinate. It will be helpful to consider that our density field can be decomposed into two parts, $\rho = \rho_{\mathrm{BG}} + \rho_{\mathrm{E}}$, where ρ_{BG} is a background component and ρ_{E} is an enhancement over the background. Because the divergence operator and the mass flux vector are linear functions, use of this density decomposition in eq. (1) yields two simultaneous equations whose sum corresponds to the total continuity equation,

$$\iiint_V \left(-\frac{\partial \rho_{\mathrm{BG}}(x,y,z,t)}{\partial t}\right) \mathrm{d}V = \iiint_V \nabla \cdot \mathbf{F}(x,y,z,t;\rho_{\mathrm{BG}},\mathbf{v})\,\mathrm{d}V, \tag{2}$$

$$\iiint_V \left(S(x,y,z,t) - \frac{\partial \rho_{\mathrm{E}}(x,y,z,t)}{\partial t}\right) \mathrm{d}V = \iiint_V \nabla \cdot \mathbf{F}(x,y,z,t;\rho_{\mathrm{E}},\mathbf{v})\,\mathrm{d}V. \tag{3}$$

The background field is that part of the density field which passively enters our domain via a mass flux through the boundaries (i.e., devoid of any internal source or sink component, which is strictly only true for a non-reactive trace gas – for a prominent reacting trace gas like NO_2, that means none of the field can be considered to be 'background'), while we define the enhanced field as the density field which was generated (or depleted) within the domain of interest. We only require eq. (3) to obtain estimates of sources and sinks, and it is thus the equation that we study further. Note that in the rest of the text, a reader can always take $\rho_{\mathrm{BG}} = 0$ for $\rho_{\mathrm{E}} = \rho$, in the sense that making the split between enhancement and background is not a strict requirement of the theory and we can instead work with the full concentration field – however, this comes with the penalty that all assumptions made in the theory apply to both the enhanced field as well as the background, rather than only the enhanced field. We see two particular reasons to recommend removing the background and working only with the enhanced field, which will become clear later:

1. We only require the wind field around the plume (usually within the planetary boundary layer) rather than over the entire vertical column.
2. We need to make no assumptions (such as the stringent assumption of steady-state conditions) on this background field.

Such a background removal has been shown to be important to improve the results of the divergence method with synthetic CO_2 satellite observations (Hakkarainen et al., 2022). Alternatively, an approximation of the divergence method as proposed by (Sun, 2022) has the implicit effect of removing the background, which is discussed in Section 6.

2.2 The definition of the mass flux vector

The mass flux field **F** describes the flow of density per unit time through a unit area. In Cartesian coordinates it is given as

$$\mathbf{F}(x,y,z,t;\rho_{\mathrm{E}},\mathbf{v}) = \rho_{\mathrm{E}}(x,y,z,t) \begin{bmatrix} v_x(x,y,z,t) \\ v_y(x,y,z,t) \\ v_z(x,y,z,t) \end{bmatrix}, \tag{4}$$

where the vector $(v_x, v_y, v_z)^T$ describes the *transport velocities* in the respective x, y and z directions. This transport can be described by a variety of mechanisms, e.g., advection, turbulent mixing, and moist convection. While advective transport is explicitly resolved by atmospheric transport models, turbulent mixing needs to be parameterized as a sub-grid scale process. The simplified representation of turbulence often represents a major source of uncertainty in atmospheric transport models, especially in terms of vertical mixing in the atmospheric boundary layer (Karion et al., 2019; Brunner et al., 2023; Katharopoulos et al., 2022).

We will now make our first assumption, which allows us to circumvent the complexities introduced by turbulent mixing. Potential limitations of this assumption will be discussed in Section 6.

Assumption 1. We assume the mass flux is dominated by advection. (5)

Thus, we assume $(v_x, v_y, v_z)^T = (u, v, w)^T$ where u, v and w are the grid-resolved easterly, northerly, and vertical wind speeds, respectively. Under Assumption 1, we assume the following definition for the flux vector,

$$\mathbf{F}(x,y,z,t) \equiv \rho_{\mathrm{E}}(x,y,z,t) \begin{bmatrix} u(x,y,z,t) \\ v(x,y,z,t) \\ w(x,y,z,t) \end{bmatrix} = \begin{bmatrix} F_x(x,y,z,t) \\ F_y(x,y,z,t) \\ F_z(x,y,z,t) \end{bmatrix}. \tag{6}$$

The Cartesian mass flux field may be written as $\mathbf{F} = F_x\mathbf{i} + F_y\mathbf{j} + F_z\mathbf{k}$ for unit basis vectors $\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$ in the easterly, northerly and vertical directions, respectively.

2.3 Integrating over columns

We will now simplify the expression for the integrated flux divergence in Eq. (1), by focusing on a column stretching between the earth surface and the top of the atmosphere as typically measured from a satellite. We will make an assumption concerning the fluxes at these limits:

Assumption 2. The normal fluxes at the top and bottom of the atmosphere vanish, $\mathbf{F}\cdot\hat{\mathbf{n}} = 0$. (7)

That is: we assume the species will neither leak to space, nor penetrate into the ground (as the wind speed normal to the surface goes to 0).

In Cartesian coordinates, we define the volume of such a column as the set of points satisfying $V_{\mathrm{A}} = \{\,(x,y,z) \mid (x,y) \in S_{\mathrm{A}}, z_0 \leq z \leq z_T\,\}$, where S_{A} describes a horizontal slice or surface through the column, and where $z = z_0(x,y)$ describes the (possibly spatially varying) ground surface and $z = z_T$ describes the top of the atmosphere, see Figure 1. Under this assumption, the volume integration over volume V may be carried out as a vertical integration from $z = z_0$ to $z = z_T$, followed by the horizontal integration. We define $\hat{\mathbf{n}}(z_0) = \nabla(z - z_0(x,y))$ as the normal vector pointing out of the surface defined by the plane $z - z_0(x,y) = 0$, and $\hat{\mathbf{n}}(z_T) = \nabla(z - z_T(x,y))$ as the normal vector pointing out of the surface defined by the plane $z - z_T(x,y) = 0$. Using the fundamental theorem of calculus ($\int_a^b f'(x)\,\mathrm{d}x = f(b) - f(a)$ for a continuous function $f(x)$) and Leibniz' integral rule for interchanging integration and differentiation operators ($\int_{a(x)}^{b(x)} \partial f(x,z)/\partial x\,\mathrm{d}z = [\int_{a(x)}^{b(x)} f(x,z)\,\mathrm{d}z]' + a'(x)f(x,a(x)) - b'(x)f(x,b(x))$), we then obtain

$$\iiint_{V_{\mathrm{A}}} \nabla\cdot\mathbf{F}(x,y,z,t)\,\mathrm{d}V_{\mathrm{A}} = \iint_{S_{\mathrm{A}}} \int_{z_0}^{z_T} \left(\frac{\partial F_x(x,y,z,t)}{\partial x} + \frac{\partial F_y(x,y,z,t)}{\partial y} + \frac{\partial F_z(x,y,z,t)}{\partial z} \right) \mathrm{d}z\,\mathrm{d}S_{\mathrm{A}}, \tag{8}$$

$$\begin{aligned} = \iint_{S_{\mathrm{A}}} \Bigg(& \frac{\partial \int_{z_0}^{z_T} F_x(x,y,z,t)\,\mathrm{d}z}{\partial x} + \frac{\partial \int_{z_0}^{z_T} F_y(x,y,z,t)\,\mathrm{d}z}{\partial y} + \\ & \underbrace{-\frac{\partial z_T}{\partial x} F_x(x,y,z_T,t) - \frac{\partial z_T}{\partial y} F_y(x,y,z_T,t) + F_z(x,y,z_T,t)}_{\mathbf{F}(z_T)\cdot\hat{\mathbf{n}}(z_T)=0} + \\ & \underbrace{+\frac{\partial z_0}{\partial x} F_x(x,y,z_0,t) + \frac{\partial z_0}{\partial y} F_y(x,y,z_0,t) - F_z(x,y,z_0,t)}_{-\mathbf{F}(z_0)\cdot\hat{\mathbf{n}}(z_0)=0} \Bigg) \mathrm{d}S_{\mathrm{A}}, \end{aligned} \tag{9}$$

$$= \iint_{S_{\mathrm{A}}} \nabla_{x,y}\cdot\mathbf{F}^Z(x,y,t)\,\mathrm{d}x\,\mathrm{d}y, \tag{10}$$

as we assumed $\mathbf{F}(z_T)\cdot\hat{\mathbf{n}}(z_T) = \mathbf{F}(z_0)\cdot\hat{\mathbf{n}}(z_0) = 0$ following Assumption 2, and where we defined the horizontal divergence operator $\nabla_{x,y} = (\partial/\partial x, \partial/\partial y)^T$ applied to a vertically integrated vector field $\mathbf{F}^Z = \int_{z_0}^{z_T}(F_x\mathbf{i} + F_y\mathbf{j})\,\mathrm{d}z = F_x^Z\mathbf{i} + F_y^Z\mathbf{j}$.

Our derivation closely follows that presented by Sun (2022), with two notable distinctions. The first is that the derivation by Sun (2022) starts from the continuity equation without a source term present. Second, but related to the first distinction, their derivation then redefines $F_z(x, y, z_0, t)$ as the emissions term E. The consequence of this latter step is that their equation does not satisfy $\mathbf{F}(z_0)\cdot\hat{\mathbf{n}}(z_0) = 0$ in regions of topography (i.e., our Assumption 2). However, our assumption appears valid, as the flux into the ground describes the transport of the trace gas by the wind field, and there is no wind blowing into the surface. Removing the term related to the vertical flux at the surface as in Sun (2022) leaves a term proportional to the gradient of the surface elevation, which was demonstrated to improve results with the divergence method (e.g., Sun, 2022; Beirle et al., 2023). In Section 6 we show that this topography-related correction term (which, again, does not follow from the theory) becomes necessary when the effective wind is replaced by wind at a (few) level(s) above the surface. In that case, it turns out that the method suggested by Sun (2022) provides a good approximation of the divergence of the effective wind field.

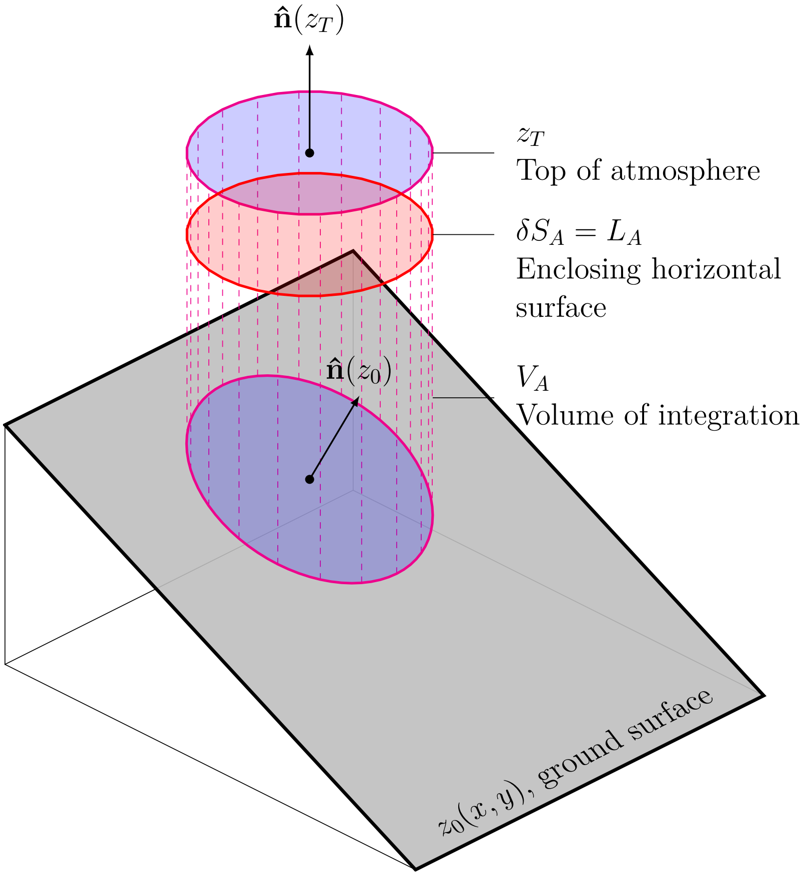

Figure 1. The integration domain V_A consists of a column ranging from the ground ($z = z_0$) to the top ($z = z_T$) of the atmosphere, enclosed by the two blue ellipsoids. A horizontal slice of the column is given by the area S_A (in red), enclosed by the line $\partial S_\mathrm{A} = L_\mathrm{A}$.

2.4 Computing fluxes for total vertical column observations

Satellite data will not provide the integrated fluxes needed in Eq. (10). Instead, the satellite product will be an estimate of the total vertical column density (TVCD), which is an integration of the chemical species' density over a column, divided by the column area. In Cartesian coordinates this corresponds to

$$\mathrm{TVCD}_{S_\mathrm{A}}(t) = \frac{\iiint_{V_\mathrm{A}} \rho(x,y,z,t)\,\mathrm{d}V_\mathrm{A}}{\iint_{S_\mathrm{A}} \mathrm{d}S_\mathrm{A}}. \tag{11}$$

For the decomposed density function ($\rho = \rho_\mathrm{BG} + \rho_\mathrm{E}$) that we are considering, we furthermore define the background column (BG),

$$\mathrm{BG}_{S_\mathrm{A}}(t) = \frac{\iiint_{V_\mathrm{A}} \rho_\mathrm{BG}(x,y,z,t)\,\mathrm{d}V_\mathrm{A}}{\iint_{S_\mathrm{A}} \mathrm{d}S_\mathrm{A}} \tag{12}$$

and the enhanced vertical column density (EVCD),

$$\mathrm{EVCD}_{S_\mathrm{A}}(t) = \frac{\iiint_{V_\mathrm{A}} \rho_\mathrm{E}(x,y,z,t)\,\mathrm{d}V_\mathrm{A}}{\iint_{S_\mathrm{A}} \mathrm{d}S_\mathrm{A}} = \mathrm{TVCD}_{S_\mathrm{A}}(t) - \mathrm{BG}_{S_\mathrm{A}}(t). \tag{13}$$

We may further pose that we can formulate a simple density model that inverts relation (13),

$$\rho_\mathrm{E}(x,y,z,t) = \mathrm{EVCD}_{S_\mathrm{A}}(t) C_0(x,y,z,t), \tag{14}$$

with a concentration function for which it holds that $\int_z C_0\,\mathrm{d}z = 1$ and $\iiint_{V_\mathrm{A}} C_0\,\mathrm{d}V_\mathrm{A} = \iint_{S_\mathrm{A}} \mathrm{d}S_\mathrm{A}$. To provide a simple example with a Dirac delta distribution, the concentration function $C_0(x,y,z,t) = \delta(z_e - z)$ would simply imply that a species' enhancement is fully present at a uniform altitude of z_e. Using this expression for the density, we may obtain an expression for $\mathbf{F}^Z$ in Eq. (10) as

$$\mathbf{F}^Z(x,y,t) = \int_{z_0(x,y)}^{z_T} \rho_\mathrm{E}(x,y,z,t) \begin{bmatrix} u(x,y,z,t) \\ v(x,y,z,t) \end{bmatrix} \mathrm{d}z, \tag{15}$$

$$= \mathrm{EVCD}_{S_\mathrm{A}}(t) \int_{z_0(x,y)}^{z_T} C_0(x,y,z,t) \begin{bmatrix} u(x,y,z,t) \\ v(x,y,z,t) \end{bmatrix} \mathrm{d}z, \tag{16}$$

$$= \mathrm{EVCD}_{S_\mathrm{A}}(t) \begin{bmatrix} U_Z(x,y,t) \\ V_Z(x,y,t) \end{bmatrix}, \tag{17}$$

where U_Z and V_Z are the wind speeds weighted by the function C_0 as per Eq. (16) (aka "effective wind speeds"). As mentioned before, here lies one of the advantages of working with the enhanced concentrations rather than the full concentration field, as the wind fields are only needed around the plume enhancement, which will generally be the wind field close to the Earth's surface. Conversely, if we work with the total vertical column density (e.g., when using CO_2 columns), we would need the wind field and concentration profile accurately known for the entire atmospheric column. Conversely, for a plume within a well-mixed planetary boundary layer, a density-weighted average wind speed within the layer would be a good approximation.

For the further discussion it is vital to turn the assertions made here into an explicit assumption:

Assumption 3. We can compute the effective wind speeds U_Z and V_Z accurately. (18)

As is clear, as soon as we violate this assumption and use wind speeds that are not the correct effective wind speeds, we do not compute the correct $\mathbf{F}^Z$ anymore, and correspondingly we do not compute the correct solutions to the continuity equation. This mistake is easily made in practice, e.g., when taking a wind speed at a single altitude above ground.

In the Supplementary Information we show a few examples (S1-S6 in particular) of how the effective wind speed may be a function of topography.

At this stage we managed to write equivalent systems for both the Cartesian and the spherical coordinate systems as found in the Supplementary Information. That is, the vectors $\mathbf{F}^Z$ and $\mathbf{F}^R$ are simply the enhanced vertical column density multiplied with a weighted wind field; the divergence operators are simply differentiations applied along the respective coordinate axes (x and y for the Cartesian system, or θ and ϕ for the spherical coordinate system); which may be used to obtain the right-hand sides of the continuity equations (3) and S3.

2.5 The left-hand side of the continuity equation

Our analysis thus far was mostly concerned with rewriting the right-hand side of the continuity equation (1) in terms of the enhanced vertical column density. The left-hand side corresponds to the sum of the sources and sinks within V and its ground surface, and the temporal accumulation of mass.

Mass can accumulate in a transient situation with temporally changing wind fields or changing sources and sinks. However, as we generally just have a single overpass image, we state the next assumption,

Assumption 4. We assume steady-state conditions during the time of overpass,

$$\iiint_{V_\mathrm{A}} \frac{\partial \rho_\mathrm{E}(x,y,z,t)}{\partial t}\,\mathrm{d}V_\mathrm{A} = \iint_{S_\mathrm{A}} \frac{\partial \mathrm{EVCD}(x,y,t)}{\partial t}\,\mathrm{d}S_\mathrm{A} \approx 0. \tag{19}$$

If we had access to the instantaneous evolution of the enhanced vertical column density (say, we had access to the EVCD at two time instances a few minutes apart), we could estimate the expression in Eq. (19). This could be possible for a geostationary but not for a polar orbiting satellite. Instead, one can hope that the error introduced by the steady state assumption follows a normal Gaussian distribution in time and space, such that this error decreases when multiple independent estimates from different overpass days or different integrating areas can be obtained. Note that we do *not* have to assume steady state conditions for the background field ρ_BG, which does not appear in Eq. (19), apart from the assumption that we could successfully compute the enhanced field $\rho_\mathrm{E} = \rho - \rho_\mathrm{BG}$. Here lies the second big advantage of removing the background field and only working with the enhanced columns – we require no assumptions on the field we do not take into account. And as mentioned before, such a background-correction step was shown to considerably improve CO_2 divergence results (Hakkarainen et al., 2022).

Two final assumptions remain. The first considers the source term within volume V_A and its bottom boundary. We will assume that we can write the source and sink field as the product of Q_i (the emission rate of source i in units of kg/s) with a Dirac delta distribution for each source.

Assumption 5. Sources are given as temporal distributions multiplied by spatial distributions,

$$\iiint_{V_\mathrm{A}} \sum_i Q_i(t)\delta(x-x_i, y-y_i, z-z_i)\,\mathrm{d}V_\mathrm{A} = \sum_{i \text{ if } (x_i,y_i)\in S_a} Q_i(t). \tag{20}$$

(Note that other assumed horizontal distributions, like 2D Gaussians or rectangular functions, are possible too, but then this must be accounted for in the further theory. We merely provide a nice way to discretize our source fields).

The final assumption concerns the sink term within the volume. A zero-th order assumption is to assume no presence of sinks – but if this assumption is incorrect we will generally underestimate the emission rates. An alternative assumption made in Beirle et al. (2019) is that of a first-order reaction, following the partial differential equation

$\partial\rho_\mathrm{E}/\partial t = -k\rho_\mathrm{E}$, i.e., where the concentration decays proportional to the concentration field itself, with a rate k describing the decay constant (for half-life $t_{1/2} = \ln(2)/k$). Rewriting the expression for a sink as $\iiint_{V_\mathrm{A}} \partial\rho_\mathrm{E}(x,y,z,t)/\partial t \,\mathrm{d}V_\mathrm{A} = \iiint_{V_\mathrm{A}} -k(x,y,z)\rho_\mathrm{E}(x,y,z,t)\,\mathrm{d}V_\mathrm{A}$, and using our density model of eq. (14) we can find

Assumption 6. Sinks are described by first-order reactions: $\iint_{S_\mathrm{A}} -k_\mathrm{A}(x,y)\mathrm{EVCD}(x,y,t)\,\mathrm{d}S_\mathrm{A}$,

$$\tag{21}$$

where $k_\mathrm{A} = \int_z C_0 k \,\mathrm{d}z$ is the vertically weighted reaction rate, similar to how the windfield is weighted vertically. This allows us to account for a vertically varying rate (e.g., due to temperature-dependent chemical reactions). It is typical to make the assumption of a constant rate, i.e., $k_\mathrm{A}(x,y) \equiv k$.

Gathering all six assumptions made up to this point, we obtain the final form for the continuity equation,

$$\sum_{i \text{ if } (x_i,y_i)\in S_a} Q_i(t) = \iint_{S_\mathrm{A}} \left(\nabla_{x,y}\cdot\mathbf{F}^Z(x,y,t) + k_\mathrm{A}(x,y)\mathrm{EVCD}(x,y,t)\right)\mathrm{d}S_\mathrm{A}. \tag{22}$$

The expression in spherical coordinates is entirely analogous.

3 The divergence method and cross-sectional flux method

We are now at a point where we can define the divergence method to estimate source emission rates Q_i [kg/s].

3.1 The divergence method

Equation (22) contains the divergence method as described by Beirle et al. (2019). Following their notation, a divergence field $D(x,y,t) = \nabla_{x,y}\cdot\mathbf{F}^Z(x,y,t)$ is defined, which is averaged over multiple satellite images. After adding the chemical sink term (if applicable) as per Eq. (22) and integrating over many small horizontal slices S_A, this provides a map of emission rates within each element.

We can describe the full process by defining a horizontal slice S_A as a small rectangular patch, formally defined by $S_\mathrm{A} = \{\,(x,y) \mid x_0 \le x \le x_1, y_0 \le y \le y_1\,\}$. Then, Eq. (10) yields (following the fundamental theorem of calculus),

$$\iint_{S_A} \nabla_{x,y}\cdot\mathbf{F}^Z(x,y,t)\,\mathrm{d}S_A = \int_{x_0}^{x_1}\int_{y_0}^{y_1} \nabla_{x,y}\cdot\mathbf{F}^Z(x,y,t)\,\mathrm{d}x\,\mathrm{d}y, \tag{23}$$

$$= \int_{y_0}^{y_1} \left(F_x^Z(x_1,y,t) - F_x^Z(x_0,y,t)\right)\mathrm{d}y + \int_{x_0}^{x_1} \left(F_y^Z(x,y_1,t) - F_y^Z(x,y_0,t)\right)\mathrm{d}x, \tag{24}$$

$$\approx \left(F_x^Z(x_1,y,t) - F_x^Z(x_0,y,t)\right)\Delta y + \left(F_y^Z(x,y_1,t) - F_y^Z(x,y_0,t)\right)\Delta x, \tag{25}$$

for $\Delta x = x_1 - x_0$ and $\Delta y = y_1 - y_0$. In this discretized approximation of the integrated divergence field, we have assumed that F_x is constant in the y direction, while F_y is constant in the x direction. Repeating the computation of Eq. (25) for a grid of small patches in the x and y directions then gives an approximate map of emission rates per patch. An example of this process is shown in Figure 2(c). Then, following Eq. (22) we can relate the obtained values to emission rates per cell.

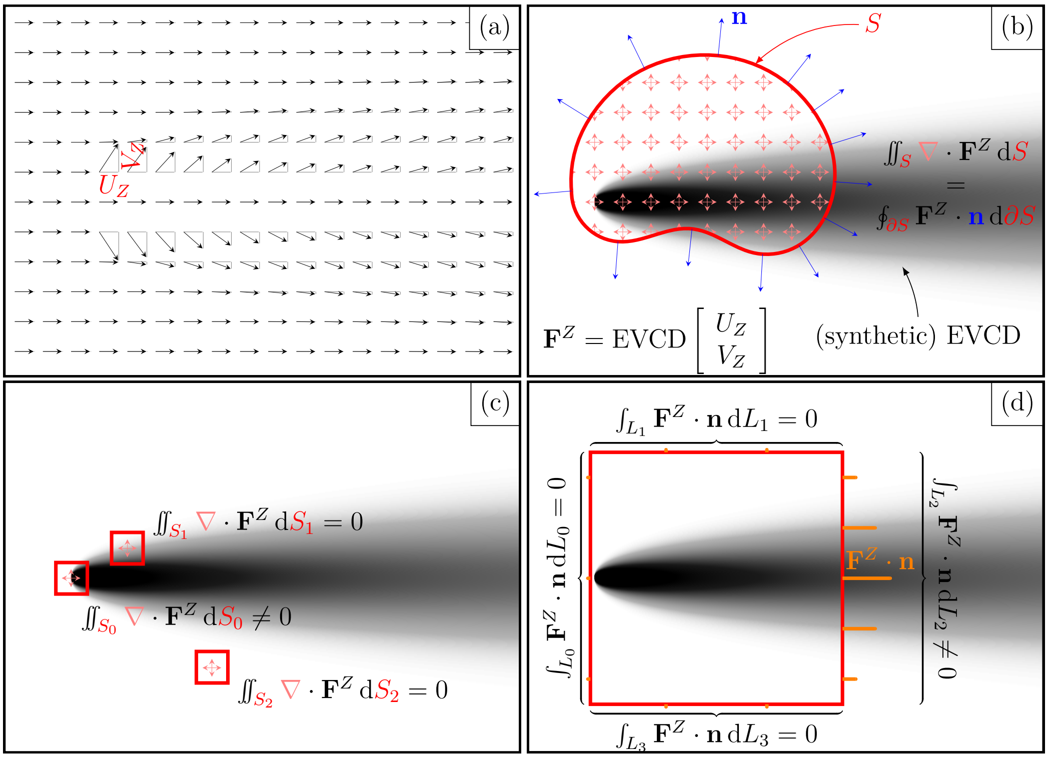

Figure 2. Schematic example of the divergence method and the cross-sectional flux method applied to a source producing a Gaussian plume. (a) The effective transport (explicitly not just wind-)speeds in the easterly and northerly directions, U_0 and V_0 respectively, form a transport vector at every position. Their explicit form is given in eq. (33). (b) The total vertical column density (TVCD) can be used to compute the vertically integrated flux vector $\mathbf{F}_\mathrm{V}$. Then, defining an arbitrary area S enclosing the source, we have that the area integral of the divergence of the flux equals the closed line integral of the normal flux along the area, ∂S. (c) The divergence method consists of evaluating the divergence for small patches. Under the steady-state assumption, without presence of sinks, and assuming the transport vectors U_0 and V_0 are known accurately, this will only yield a non-zero value at the location of the emission source. (d) The cross-sectional flux method consists of computing the closed line integral. The curve part in the 'upstream' direction (L_0), and two sections parellel but far away from the plume center line (L_1 and L_3) can be assumed to not contribute to the total line integral. As a result, only a cross-sectional line section (L_2) remains.

3.2 The cross-sectional flux method

We point out that the cross-sectional flux method follows from an application of Gauss' divergence theorem to Eq. (10),

$$\iiint_{V_\mathrm{A}} \nabla \cdot \mathbf{F}(x,y,z,t)\,\mathrm{d}V_\mathrm{A} = \iint_{S_\mathrm{A}} \nabla_{x,y} \cdot \mathbf{F}^Z(x,y,t)\,\mathrm{d}S_\mathrm{A} = \oint_{\partial S_\mathrm{A}} \mathbf{F}^Z(x,y,t) \cdot \mathbf{n}\,\mathrm{d}L_\mathrm{A}, \qquad (26)$$

where $\mathbf{n}\,\mathrm{d}L_\mathrm{A} = \mathbf{i}\,\mathrm{d}y - \mathbf{j}\,\mathrm{d}x$, using $\mathbf{n}$ as the outward pointing unit vector on every point of the closed curve $\partial S_\mathrm{A} = L_\mathrm{A}$ (see Figure 1). A reader may confirm that for an identical rectangular choice of the horizontal slice S_A, equations (24) and (26) are indeed identical, hence, the cross-sectional flux method and divergence method are mathematically identical. A judicious choice for the coordinates of the closed contour L_A enclosing the known source can simplify the equations. An example is shown in Figure 2(d), where $L_\mathrm{A} = L_0 \cup L_1 \cup L_2 \cup L_3$ is the union of four straight lines, of which three line integrals are (approximately) zero because they lie upstream of the source or parallel to the source at large offsets. Equation (26) can then thus be computed as the line integral of the flux field along the only line that crosses through the source enhancement field. Then, following Eq. (22) we can relate the obtained values to emission rates within the area under consideration.

3.3 The difference between the two methods

Although the divergence method and cross-sectional flux methods are identical from a mathematical point of view, it is worth pointing out that they differ in practice. We note the following six important differences:

- In the divergence method, one first produces a map of emission rates; this map is then used to quantify the emissions for specific sources of interest (Beirle et al., 2019, 2021; Hakkarainen et al., 2022). Conversely, in the cross-sectional flux method, one first picks a source of interest, and then emissions are quantified by choosing an appropriate integrating contour surrounding this source. The order of operations is thus opposite ('compute-emissions-then-pick' in the divergence method, versus 'pick-then-compute-emissions' for the cross-sectional flux method).
- The divergence method 'ignores' additional information from the plume downwind, whereas the cross-sectional flux method can use data from multiple offsets along the plume (see, e.g., Kuhlmann et al., 2020, 2021). On te other hand, this means that the divergence method is not influenced by additional sources downwind. The cross-sectional flux method, in contrast, is more likely to introduce biases when plumes of various sources mix in the downwind direction.
- Both the divergence and cross-sectional flux methods require that multiple emission estimates are made, to overcome issues with noisy and uncertain data and processing. For the divergence method, the only means of getting more estimates is to repeat the estimation for multiple satellite overpasses. The cross-sectional flux method (per the previous point) can compute multiple estimates from a single image alone. The obtained signal from the cross-sectional flux method can thus, potentially, be superior, as more estimates can be obtained from the same data.
- By averaging over multiple instances in time, the divergence method may pick up on small sources that in a single image produce only a faintly discernible plume enhanced over the background, visible in the pixel overlapping with the source and a few next to it. The cross-sectional flux method applies primarily to large plumes that are clearly visible in instantaneous images, not to plumes that are only faintly present in a handful of pixels.
- The divergence method requires little input of the user to run. Conversely, the cross-sectional flux method requires multiple steps to be carried out (e.g., selecting the point source of interest, selecting corresponding pixel enhancements over the background, selecting suitable integrating contours, making sure plumes are not over-

lapping). Therefore, the divergence method is easily automated, while the cross-sectional flux method requires a very tuned algorithm to carry out the required steps for the chosen product (satellite and trace gas) in an automated fashion.

- The divergence method is 'receptive' to the first-order derivative of the enhanced field (i.e., requires sharp discontinuities to find sources), while the cross-sectional flux method is 'receptive' to just the amplitude of the enhanced field (i.e., requires large enhancements to find sources). The divergence method may thus struggle to find sources that are not well-described as point-sources (e.g., cities or other area sources), while the cross-sectional flux method is not hindered by this restriction (e.g., see Hakkarainen et al., 2022; Kuhlmann et al., 2021). Conversely, the cross-sectional flux method is only dependent on large enhancements over the background, not their first-order derivatives.

4 Considerations of discretization for the divergence method

The enhanced vertical column density will only be known at discrete pixels rather than as a continuous field. Calculus operations like line integrals and divergence operations must thus be replaced with their discrete counterparts. In this section we shortly describe the effect of working with discretized data when using the divergence method.

4.1 Finite-difference and finite-volume differentiation

Given the discrete nature of the enhanced vertical column density field, we will introduce an additional subscript notation of the form $f_{i,j}(t) = f(i\Delta x, j\Delta y, t)$ where f is any function and i and j are integers. Then the mass flux $\mathbf{F}^Z$ given in Eq. (17) can be written as

$$\begin{bmatrix} F^Z_{x|i,j}(t) \\ F^Z_{y|i,j}(t) \end{bmatrix} = \begin{bmatrix} U_{Z|i,j}(t) \\ V_{Z|i,j}(t) \end{bmatrix} \mathrm{EVCD}_{i,j}(t). \tag{27}$$

We consider this formulation a 'coincident' definition of the flux, as the wind and density fields are defined at coincident locations. In practice, this will require interpolation of the satellite observations and the wind fields to a common Cartesian or spherical grid. From this definition we may compute a discrete version of the divergence using centered finite-differences (FD) using the following smallest possible stencil,

$$(\nabla_{x,y} \cdot \mathbf{F}^Z)_{i,j}(t) \approx D^C_{ij}(t) = \frac{F_{x|i+1,j}(t) - F_{x|i-1,j}(t)}{2\Delta x} + \frac{F_{y|i,j+1}(t) - F_{y|i,j-1}(t)}{2\Delta y}. \tag{28}$$

Note that the computation at position $(i\Delta x, j\Delta y)$ is independent of $\mathrm{EVCD}_{i,j}$.

An alternative is to use a finite-volume approximation of the divergence method, which could already be discerned in Eq. (25): instead of a centered FD approach, we could use the flux at the cell boundaries. In this case, the wind field would be interpolated to the cell boundaries, whereas the satellite observations would be interpolated to the cell centers. We can use the concept of an 'upwind' flux to compute the fluxes at cell edges as

$$F^Z_{x|i+\frac{1}{2},j}(t) = U_{Z|i+\frac{1}{2},j}(t) \times \begin{cases} \mathrm{EVCD}_{i,j}(t) & \text{if } U_{Z|i+\frac{1}{2},j}(t) \geq 0, \\ \mathrm{EVCD}_{i+1,j}(t) & \text{otherwise,} \end{cases} \tag{29}$$

$$F^Z_{y|i,j+\frac{1}{2}}(t) = V_{Z|i,j+\frac{1}{2}}(t) \times \begin{cases} \mathrm{EVCD}_{i,j}(t) & \text{if } V_{Z|i,j+\frac{1}{2}}(t) \geq 0, \\ \mathrm{EVCD}_{i,j+1}(t) & \text{otherwise} \end{cases} \tag{30}$$

In other words, we transport the mass of the cell upstream from the wind direction into the downstream direction. Then, the divergence can be approximated using a more compact stencil,

$$(\nabla_{x,y} \cdot \mathbf{F}^Z)_{i,j}(t) \approx D^U_{ij}(t) = \frac{F_{x,i+\frac{1}{2},j}(t) - F_{x,i-\frac{1}{2},j}(t)}{\Delta x} + \frac{F_{y,i,j+\frac{1}{2}}(t) - F_{y,i,j-\frac{1}{2}}(t)}{\Delta y}. \tag{31}$$

The divergence approximation with eq. (31) uses spatial offsets of Δx and Δy rather than $2\Delta x$ and $2\Delta y$. Furthermore, unlike eq. (28), the computation at position $(i\Delta x, j\Delta y)$ now depends on $\mathrm{EVCD}_{i,j}$ at the same location. Hence, one can expect that the divergence as computed with the upwind fluxes results in a more 'focused' emissions estimation; at the cost of shifting the estimated fluxes (and thus the enhancements) to the *border* of a pixel containing a source rather than the center of a pixel. Hence, the estimated emissions would be slightly offset with respect to the true source location.

4.1.1 Gaussian plume example

Consider an example of the two divergence computations for a vertically integrated Gaussian plume for a point-source at $(x, y) = (0, 0)$ with source rate Q,

$$G(x,y) = H(x)\frac{Q}{U}\frac{1}{\sqrt{2\pi}\sigma(x)}e^{\frac{-y^2}{2\sigma^2(x)}}, \tag{32}$$

where $H(x)$ is the (Heaviside) step function, U is the wind speed in the x direction to describe the advective process, and $\sigma(x) = H(x)\sqrt{2K_y x/U}$ describes the crosswind diffusion through the eddy diffusion coefficient K_y. The transport speed in the y direction due to turbulence can be found to be $V = Uy/(2x)$, following, e.g., Stockie (2011). We horizontally integrate eq. (32) to form pixels of Δx by Δy using a numerical quadrature, to compute $\mathrm{EVCD}_{i,j}$. For the effective wind, we use

$$U_{Z|i,j} = U, \qquad V_{Z|i,j} = \frac{\int_{(i-\frac{1}{2})\Delta x}^{(i+\frac{1}{2})\Delta x}\int_{(j-\frac{1}{2})\Delta y}^{(j+\frac{1}{2})\Delta y} G(x,y)\frac{Uy}{2x}\,\mathrm{d}y\,\mathrm{d}x}{\int_{(i-\frac{1}{2})\Delta x}^{(i+\frac{1}{2})\Delta x}\int_{(j-\frac{1}{2})\Delta y}^{(j+\frac{1}{2})\Delta y} G(x,y)\,\mathrm{d}y\,\mathrm{d}x}. \tag{33}$$

The expression for V_Z implies that we weigh the local cross-wind *transport* velocity $Uy/(2x)$ with the horizontal concentration field, as indicated in eq. (16). Note that this definition of the effective wind field is analogous to what is proposed in Roberts et al. (2023), who point out that using the full transport speed of a Gaussian plume corrects for underestimated divergence estimates. The integrals may be either computed using calculus methods or by the same numerical quadrature used to compute the $\mathrm{EVCD}_{i,j}$ pixels.

Now, when comparing the divergence as computed using the coincident and the upwind flux definitions in Fig. 3, we note that the latter method gives a figure that is considerably more 'focused'. The reason for this is that the divergence using the upwind flux definition is a more localized operation compared to the coincident flux version. Note, furthermore, in Fig. 3 that even in this simple analytical example, the divergence map is not actually restricted to just the pixels at, or surrounding, the source position. The reason for this is the fact that we replaced the continuous divergence operation with an FD approximation. The FD approximation is erroneous for higher wavenumbers (i.e., sharp features, such as around a point source), with as a result a blurring (or, more precisely, a wavenumber-dependent phase shift) of the originally expected point source result.

4.2 Numerical integration

Just as differentiation is replaced with finite-difference approximations, we must replace integration with numerical summations over discrete values. Assuming we want to integrate over a rectangular area of $N_x\Delta x$ by $N_y\Delta y$, the numerical integration can be carried out as

$$\iint_{S_\mathrm{A}} (\nabla_{x,y}\cdot\mathbf{F}^Z(t))\,\mathrm{d}S_\mathrm{A} \approx \sum_{i=0}^{N_x-1}\sum_{j=0}^{N_y-1} D_{i,j}(t)\Delta x\Delta y, \tag{34}$$

where $D_{i,j}$ represents the discretely computed divergence field. If we use this formulation on the divergence results of Fig. 3, we find $Q = 3$ for both the finite difference and finite volume versions, which was also used to model the Gaussian plume.

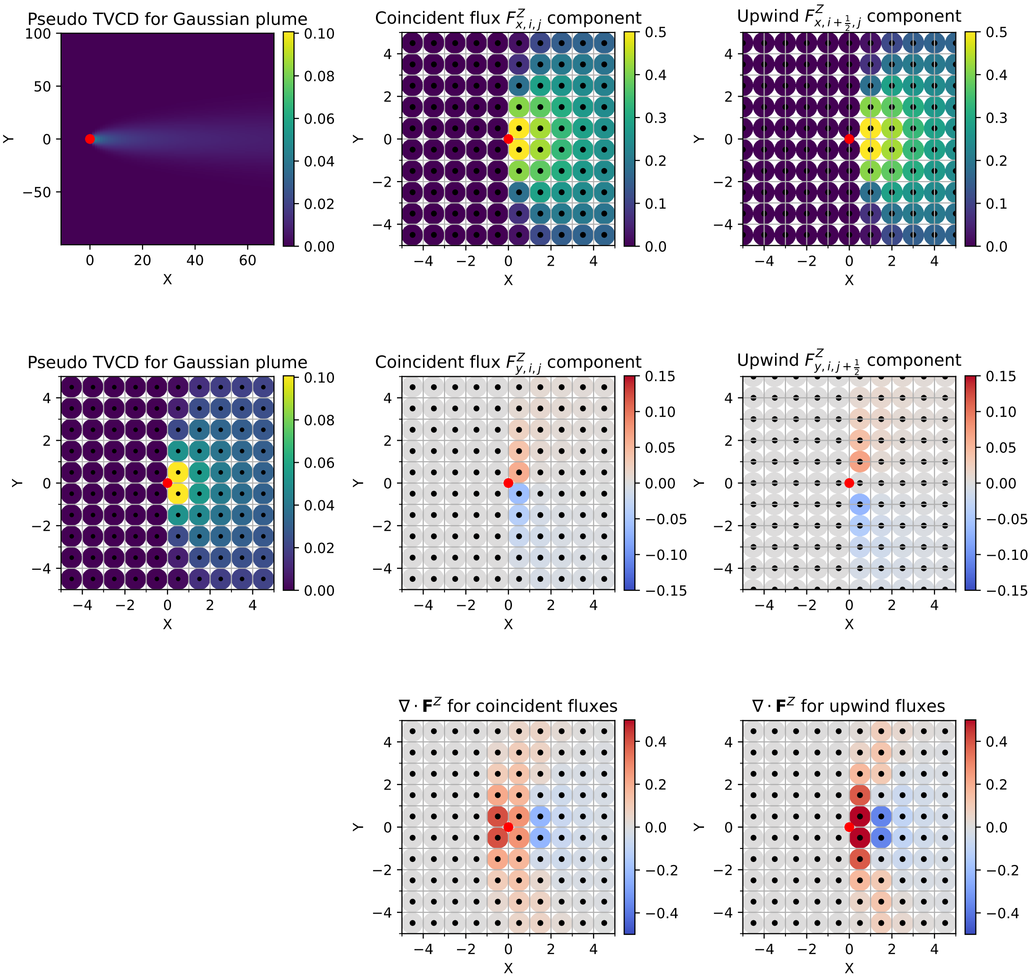

Figure 3. Coincident flux versus upwind flux definitions lead to different results when creating the divergence map $\nabla \cdot \mathbf{F}^Z$. The source position is indicated with a red dot, and pixel values are indicated with circles. The upwind fluxes are staggered with respect to the input data. Note that the divergence map using the upwind flux definition is more focused, less smeared, than the divergence map created using the coincident flux definition. However, it is not centered on the source location but shifted towards the right, following the wind direction.

The structure of the discrete divergence operator $D_{i,j}$ can be treated as a telescoping series of internally cancelling contributions, such that only the terms at the borders of the domain remain (which is a numerical demonstration of Gauss' divergence theorem). That is, for the coincident flux definition of the divergence, we could also compute

$$\begin{aligned}\sum_{i=0}^{N_x-1}\sum_{j=0}^{N_y-1} D^C_{i,j}(t)\Delta x\Delta y = \Delta y &\left(\sum_{j=0}^{N_y-1}\left[\frac{-F_{x|-1,j}-F_{x|0,j}}{2}\right]+\sum_{j=0}^{N_y-1}\left[\frac{F_{x|N_x-1,j}+F_{x|N_x,j}}{2}\right]\right)\\ &+\Delta x\left(\sum_{i=0}^{N_x-1}\left[\frac{-F_{y|i,-1}-F_{y|i,0}}{2}\right]+\sum_{i=0}^{N_x-1}\left[\frac{F_{y|i,N_y-1}+F_{y|i,N_y}}{2}\right]\right),\end{aligned}\tag{35}$$

and for the upwind flux definition of the divergence, we have that

$$\begin{aligned}\sum_{i=0}^{N_x-1}\sum_{j=0}^{N_y-1} D^U_{i,j}(t)\Delta x\Delta y = \Delta y &\left(\sum_{j=0}^{N_y-1}\left[\frac{-F_{x|-\frac{1}{2},j}}{1}\right]+\sum_{j=0}^{N_y-1}\left[\frac{F_{x|N_x-\frac{1}{2},j}}{1}\right]\right)\\ &+\Delta x\left(\sum_{i=0}^{N_x-1}\left[\frac{-F_{y|i,-\frac{1}{2}}}{1}\right]+\sum_{i=0}^{N_x-1}\left[\frac{F_{y|i,N_y-\frac{1}{2}}}{1}\right]\right).\end{aligned}\tag{36}$$

Note that eq. (35) contains the spatial average between two cells to the locations explicitly indicated in (36). This formulation is one possible basis upon which a cross-sectional flux implementation rests. It will become relevant in the next section that the numerical integral for the coincident flux depends on 8 EVCD values, while the integral using the upwind fluxes depends only on 4 EVCD values.

4.3 Noise

Assume that independent and identically distributed (i.i.d.) Gaussian additive noise is present in the computed mass flux vector, such that we can only work with the noisy flux, $F^{\text{noisy}}_{x|i,j} = F_{x|i,j} + N_{i,j}$, with noise component $N_{i,j} \sim \mathcal{N}(0,\sigma^2)$ for a zero-mean Gaussian distribution $\mathcal{N}$ with variance σ^2. In that case, the coincident divergence computation in eq. (28) becomes

$$D^{C,\text{noisy}}_{i,j}(t) = \frac{F^{\text{noisy}}_{x|i+1,j}-F^{\text{noisy}}_{x|i-1,j}}{2\Delta x}+\frac{F^{\text{noisy}}_{y|i,j+1}-F^{\text{noisy}}_{y|i,j-1}}{2\Delta y}, \tag{37}$$

$$= D^C_{i,j}(t)+\frac{N_{i+1,j}-N_{i-1,j}}{2\Delta x}+\frac{N_{i,j+1}-N_{i,j-1}}{2\Delta y}, \tag{38}$$

$$\sim \mathcal{N}\left(D^C_{i,j}(t),\sigma^2\left[\frac{1}{2\Delta x^2}+\frac{1}{2\Delta y^2}\right]\right). \tag{39}$$

Hence, the finite-difference approximation of the coincident divergence operator would be a random variable centered around the noise-free divergence approximation, with a variance inversely proportional to the cell dimensions. Substituting the same noise model in the numerical integral of the divergence, eq. (35), yields

$$\sum_{i=0}^{N_x-1}\sum_{j=0}^{N_y-1} D^{C,\text{noisy}}_{i,j}(t)\Delta x\Delta y \sim \mathcal{N}\left(\sum_{i=0}^{N_x-1}\sum_{j=0}^{N_y-1} D^C_{i,j}(t)\Delta x\Delta y,[N_x\Delta x^2+N_y\Delta y^2]\sigma^2\right). \tag{40}$$

We can repeat the computation for the upwind definition of the fluxes, eq. (31), to find

$$D^{U,\text{noisy}}_{i,j}(t) = \frac{F^{\text{noisy}}_{x|i+\frac{1}{2},j}-F^{\text{noisy}}_{x|i-\frac{1}{2},j}}{\Delta x}+\frac{F^{\text{noisy}}_{y|i,j+\frac{1}{2}}-F^{\text{noisy}}_{y|i,j-\frac{1}{2}}}{\Delta y}, \tag{41}$$

$$\sim \mathcal{N}\left(D^U_{i,j}(t),\sigma^2\left[\frac{2}{\Delta x^2}+\frac{2}{\Delta y^2}\right]\right), \tag{42}$$

and (for eq. (36))

$$\sum_{i=0}^{N_x-1}\sum_{j=0}^{N_y-1} D_{i,j}^{U,\text{noisy}}(t)\Delta x\Delta y \sim \mathcal{N}\left(\sum_{i=0}^{N_x-1}\sum_{j=0}^{N_y-1} D_{i,j}^{C}(t)\Delta x\Delta y, 2[N_x\Delta x^2 + N_y^2\Delta y]\sigma^2\right), \quad (43)$$

This means that the noise variance for integrating the divergence map with the upwind flux definition is expected to be exactly twice as large as with the coincident flux; even though the divergence map computed with eq. (42) has a noise variance four times as high. The reason for this was highlighted in the previous section, which showed that twice as many cells are summed in a numerical integration of the coincident flux integration, eq. (35), than for the upwind flux integration, eq. (36).

4.3.1 Gaussian plume example with noise

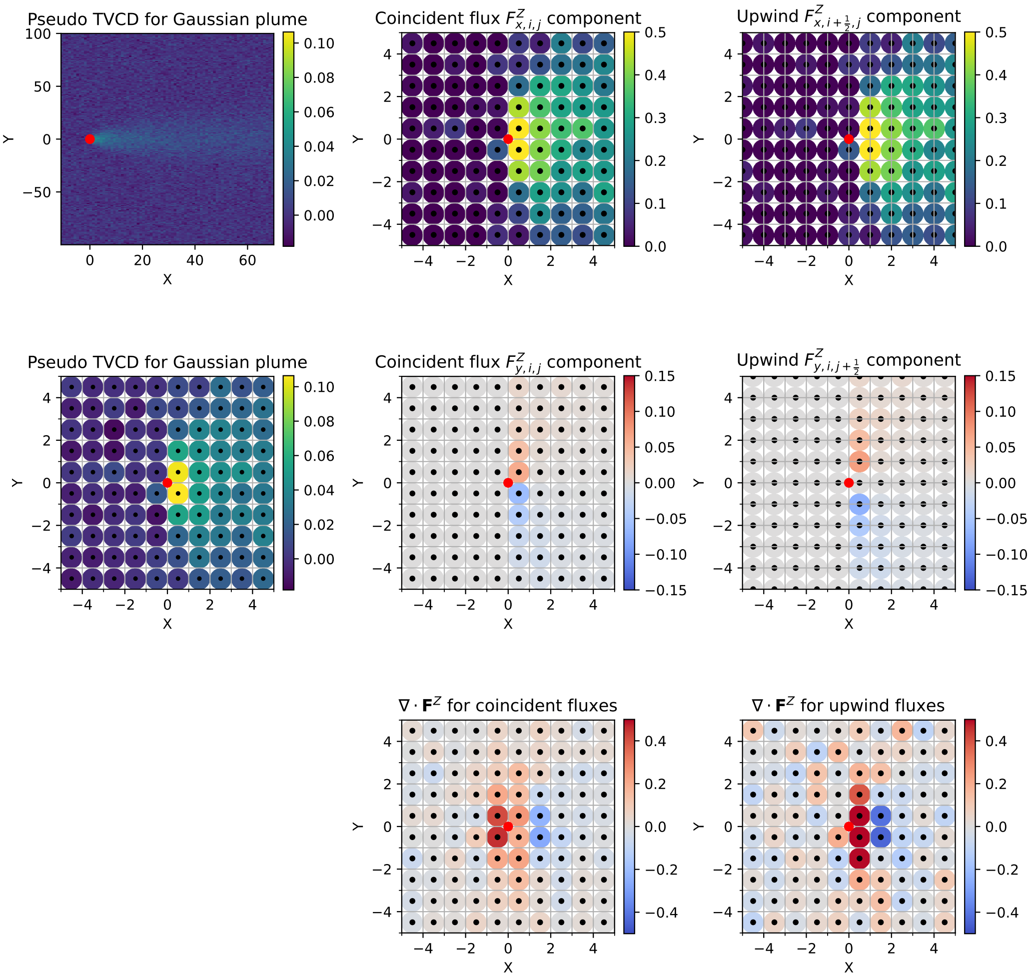

Figure 4. Coincident flux versus upwind flux definitions lead to different results when creating the divergence map $\nabla \cdot \mathbf{F}^{Z,\text{noisy}}$ with a 0.1 SNR. Compare to Fig. 3. The divergence map using the upwind flux definition remains more strongly elevated over the background noise and less smeared than the divergence map created using the coincident flux definition.

Analogous to the previous noise-free Gaussian plume example, we now provide an example of the coincident and upwind flux definitions in Fig. 4 with a pseudo 'noisy' satellite image with signal-to-noise ratio $\mathrm{SNR} = \mathrm{EVCD}^2_{\mathrm{average}}/\sigma^2_{\mathrm{im}} = 0.1$. Compared to Fig. 3, we see that the upwind divergence map remains more 'focused' around the true source location (that is: on a single image; the result would start to approach the results of the divergence method if we would average over many images with various wind directions). Hence, if the sole purpose of computing the divergence map is to *locate* sources, it is clear that the upwind flux map may continue to yield a superior result. However, if the purpose is to compute source emissions accurately, the coincident flux definition is more robust to noise. Repeating the source estimation step 1 000 000 times, we can show a kernel density plot like Fig. 5 to show the distribution in the estimated emissions for this Gaussian plume example. As derived, the coincident flux source estimate follows a Gaussian distribution with $\sigma^2_{\mathrm{est}} \cong (N_x\Delta x^2 + N_y\Delta y^2)U^2\sigma^2_{\mathrm{im}}$, while the upwind flux estimates follow a distribution with $2\sigma^2_{\mathrm{est}}$.

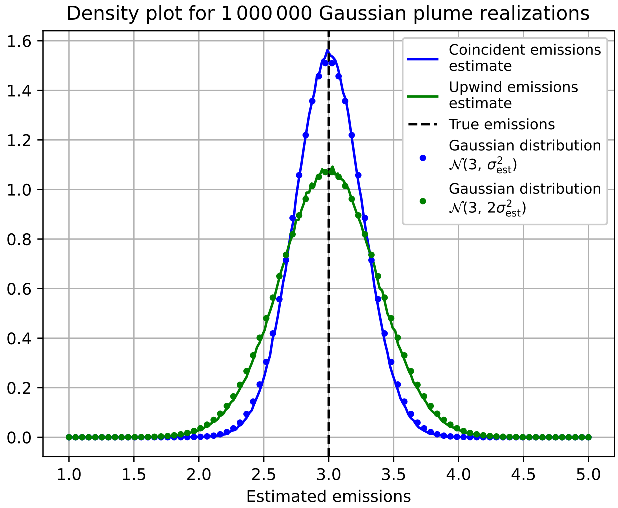

Figure 5. Kernel density plot for source estimations based on 1 000 000 realizations of a noisy Gaussian plume image with a SNR=0.1. One such realization is shown in Fig. 4. The coincident flux divergence method is more precise in estimating the emissions than the upwind divergence method under presence of noise.

4.3.2 Recommendations

The results from eqs. (39)–(43) indicate five general recommendations for keeping the noise level of the source estimations in check:

1. **Obtain more estimates**. If we can make M estimates of the emissions $Q_i(t)$, then their arithmetic mean will have its variance decreased by a factor $1/M$. For example, with the cross-sectional flux method, one might be able to estimate the emissions from a single source at multiple offsets, and apply an appropriate averaging operation to decrease the uncertainty of emissions from a single image (e.g., see Kuhlmann et al., 2019, 2021).
2. **Denoise the image**. If we can work with an image that has its noise variance σ^2 reduced through some denoising procedure, less noise can leak into the final result. This was for example explored by Hakkarainen et al. (2022), where it was

shown that application of a mean filter increased the possibility of obtaining accurate source estimates for a number of point sources using a simulated CO2M satellite dataset (Kuhlmann et al., 2019). Alternatively, as Beirle et al. (2019) explored, averaging over many temporal instances of the flux field also achieves a denoising of the flux map by bringing down the sample error per cell.

3. **Include a robust fitting process before integrating**. Following eqs. (39)–(43), the divergence estimate is inversely proportional to the cell dimensions while the source emission estimate is directly proportional to the integration size. Hence, to reduce the error we may attempt to perform a robust fitting procedure prior to the integration step. For example, Beirle et al. (2019) proposes to fit 2-D Gaussian distributions to the divergence map for possible source locations rather than integrating the divergence map directly.
4. **Compute the divergence first, prior to further processing**. Temporal averaging and divergence computations (and area integrals) are linear operators, thus they commute from a mathematical point of view. Computing area integrals over temporally averaged divergence maps (de Foy & Schauer, 2022; Beirle et al., 2023), or over divergence computations of time-averaged flux maps (Beirle et al., 2019), thus seem mathematically identical. However, when working with finite data (e.g., a finite swath length, patchy data due to cloud cover, . . .) commutativity breaks down. For instance (to give a simple example), if each satellite overpass yields just one usable pixel, divergence computations are not possible with the individual overpass images. Conversely, if one first bins and time-averages these pixels, one can suddenly compute the divergence. However, this latter divergence map would likely be erroneous as each pixel is drawn from a different distribution, invalidating a steady-state assumption (e.g., if each overpass featured different emission rates, the divergence computation is not giving a meaningful result). Clearly, the order of operations matters. Thus, as is also done in de Foy and Schauer (2022), it is imperative to take the average of the divergences rather than the other way around, as this is the desired result. With regards to the question of whether one should bin the pixels first and then compute the divergence, or compute the divergence first and then bin the results, de Foy and Schauer (2022) and Beirle et al. (2023) shows that the latter version produces superior results. This makes sense as the swath is the best approximation of the continuous field, and binning (as a nonlinear operator) weakens this continuity. Hence, the best results can be obtained by computing the divergence result prior to the binning, averaging and integration steps, both in theory and in practice.
5. **Estimate the source rate over the smallest region possible to avoid noise**. Equations (40)–(43) with Gaussian i.d.d. noise suggest that using a large integration region around a source negatively affects the results of the divergence method. The primary reason is that, when integrating over a larger region, we are simply adding random noise of pixels that are not related to the source, which increases the variance of the estimate without improving upon its expected value.

5 Examples

5.1 Numerical example for a Gaussian plume

We already considered two examples for a Gaussian plume model in Figures 3 and 4. However, we will show one additional Gaussian plume example, where we explore the effect of making the assumption of pure advection while knowing that the Gaussian plume transport speeds are described as both advective (downwind) and diffusive (cross-wind). Furthermore, we will consider the effect of different integration areas on the results. Figure 6 shows results for both the divergence method (top) and the cross-sectional-flux method (bottom). The left panels work from the assumption that only an advective flux is present, while the right panels make the assumption that both the advective and diffusive trans-

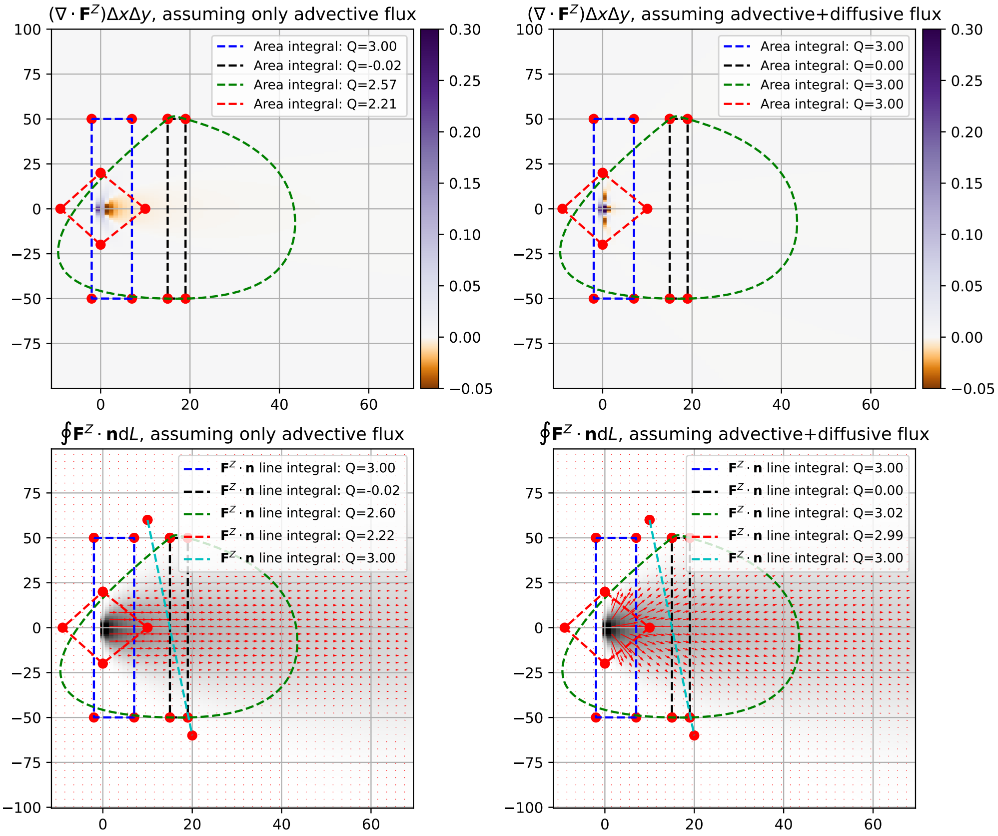

Figure 6. An example of a Gaussian plume to which we applied the divergence method (top panels) and the cross-sectional flux method (bottom panels). In the left panels we assume that only an advective flux is present, while in the right panels we assume the full 2-D flux vector for the Gaussian plume model. The emissions for this example are $Q = 3$ (in arbitrary but consistent units). Shown in color in the top panels are the corresponding divergence maps. Shown in grayscale in the bottom panels is the TVCD of the Gaussian plume.

port speeds are used to compute the flux fields. The example shows that the assumption of an advective flux yields no errors, *provided that* the area of integration is chosen with straight lines crossing the entire plume (as in the 'blue' dotted domain). Conversely, when we choose arbitrary different domains, missing the diffusive component of the flux underestimates the obtained emissions (as in the 'green' and 'red' dotted domains). Note, furthermore, that the divergence map is more 'focused' when the full flux vector is used, even when we are using the coincident flux method here. This result is in line with findings of Roberts et al. (2023) who recommend modifying the flux map using the full transport vector rather than just the wind. In Figure 6 we can moreover see that we can compute the line integral for a single cross-sectional line (shown in 'cyan') downwind of the source, to obtain a good estimate of the emission rate. Note that this line does not have to be perpendicular to the wind direction, it merely has to be straight and cross the whole plume; as then the cross-wind fluxes cancel. We show this mathematically in the Supplementary Information. This follows from the assumptions shown in Figure 2(d), namely that the upwind and far-away parallel fluxes along the plume do not contribute to the line integral (when the background field is removed appropriately). There is no analogous single-line method for the divergence method.

The conclusion we can draw from this Gaussian plume example is that the assumption of only an advective flux is appropriate *provided that* we can sample full plume cross-sections along straight lines (as in the blue, black and cyan line sections of Fig. 6). Conversely, when we do not sample the entire plume cross-section along a straight line, the assumption leads to underestimations (as in the green and red line sections). This conclusion has a negative consequence for the divergence method *[It is relevant to remark here that taking the area integral of the divergence map is yet another commuting linear operator; so whether one first time-averages all divergence maps and then integrates to estimate the source strength, or one integrates all divergence maps to get individual source strength estimates and then time-averages the results, one obtains the same results. In this example we see that the latter approach may underestimate source emissions on single images, and thus the total average over many Gaussian plumes in arbitrary directions will also be negatively biased.]*: one can generally expect that the integration domain on which the divergence is computed is not cutting through the plume in a straight line. The best solution for this problem is thus likely the one pointed out by Roberts et al. (2023): include the diffusive flux prior to taking the divergence. For the cross-sectional flux method, we see that as long as we can take a cross-section along a straight line (which should generally be possible), we obtain accurate results, regardless of whether we do or do not include the effects of diffusion.

Of course, we remark that these conclusions are based on a Gaussian plume model. In the next section we take a closer look at more realistically modeled plumes.

5.2 Example using a COSMO-GHG modeled enhanced plume

As an alternative to the theoretical and analytical model of the Gaussian plumes, we run a full-fledged atmospheric transport model, COSMO-GHG, to generate a more realistic and time-varying plume (Jähn et al., 2020). COSMO-GHG is run with a rotated latitude-longitude grid at approximately $1\,\mathrm{km} \times 1\,\mathrm{km}$ resolution and 60 vertical levels extending to 23 km altitude. As meteorological initial and boundary conditions we use the ECMWF operational analysis data. We simulate a point source with constant emissions of 1241 kg CO_2/s at the location of the Bełchatów power station in Poland. This power plant has been quantified various times with space-based capabilities, e.g., Nassar et al. (2022) and Fuentes Andrade et al. (2023). Here we show a purely synthetic example. We prescribe a vertical emission profile to account for buoyant plume rise, with the majority of the emissions centered around 600 m altitude (Brunner et al., 2019, 2023). Output is generated in time steps of 15 minutes, on June 5, 2018. By vertically integrating the simulated CO_2 field we compute pseudo satellite TVCD images of CO_2 [kg/sr].

5.2.1 Single time – varying distances

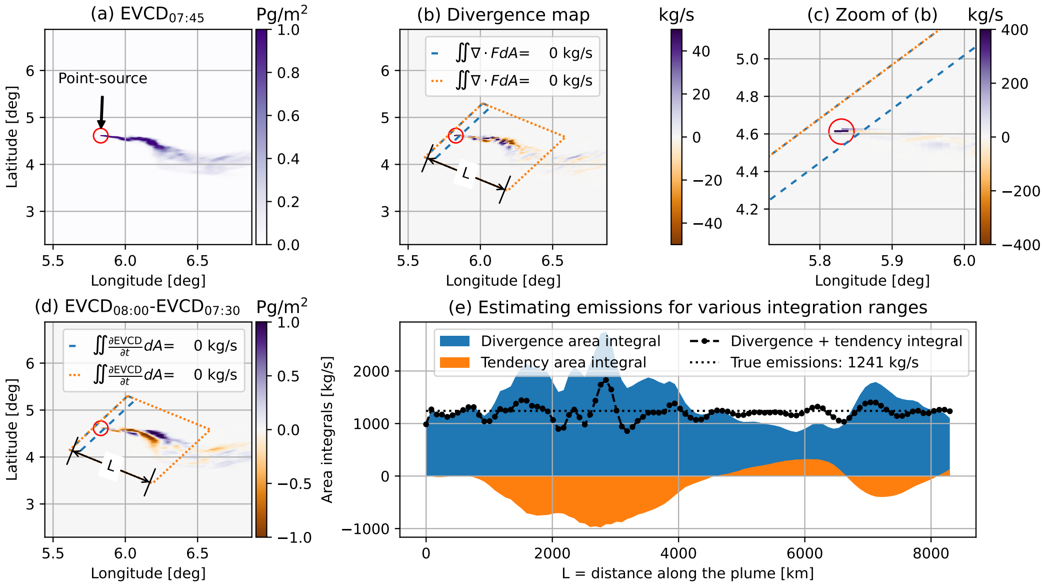

Figure 7. The source emission estimates for a synthetic TVCD at various integration ranges at 07:45 UTC. Panel (a) shows the TVCD at 07:45 UTC, panel (b) shows the divergence map (multiplied with the cell areas) for this region, panel (c) a zoom of (b) close to the source, panel (d) shows the difference between TVCDs at 08:00 UTC and 07:30 UTC. In the legend of (b) and (d) we report the integral of the divergence and tendency area integrals for two distances denoted with L, respectively. Panel (e) shows the estimation of the source strength for a large range of L values. The divergence is estimated for a varying distance along the plume (see Fig. 7). We can see that the temporal area integral compensates for over-estimations otherwise resulting from the divergence area integral and vice versa.

We begin our examination by considering a morning-time plume at 07:45 UTC (Figure 7(a)), for which we compute the (coincident flux) divergence field; the result is shown in Fig. 7(b)–(c). We can see that the divergence map has a clear maximum around the location of the point-source (clear to see in Figure 7(c)); though some positive and negative fluctuations also exist at downstream locations (clear to see in Figure 7(b)). The reason for this is (1) that tracer transport in COSMO-GHG is not just (wind-based) advective transport but also contains a diffusive component, (2) the situation is not a steady-state situation and is thus not fully described by the divergence computation alone, and (3) we are only approximating the divergence operation with a finite-difference operator.

Further drawn in Fig. 7(b)–(d) are two integration areas denoted in dashed blue (at about 2 km distance to the source) and dotted orange (at about 93 km from the source). Within these integration ranges, we compute the area integrals $\iint \mathrm{dEVCD}/\,\mathrm{d}t\,\mathrm{d}A$ (the integral of the tendency) and $\iint \nabla{\cdot}\mathbf{F}^Z\,\mathrm{d}A$ (the integral of the divergence). The tendency term is what is typically assumed to be 0 following Assumption 4, i.e., it corresponds to the deviation from the steady-state situation. We calculate this term using simulated EVCD values at time steps immediately before and after 07:45, namely at 07:30 and 08:00, via a central finite-difference approximation.

In Fig. 7(e), emissions estimates are presented for 100 integration ranges with varying distances L along the plume. We can see that merely integrating the area up to a distance of about 50 km along the plume would over-estimate the true emissions by a factor 2; while integration to around a distance of 100 km along the plume would under-estimate the emissions by about 25%. Conversely, when taking the area integral of the temporal difference, we see that it plays a compensating role – with negative values where the divergence computation yields overly large values, and positive values where the divergence computation yields overly small values. If we take the average over the 100 estimates along the plume, we would estimate emissions of about 1465 kg/s when just looking at the divergence map, while we obtain an average of about 1215 kg/s when taking the temporal component into account, which is considerably closer to the true simulated value of 1240 kg/s.

5.2.2 *Multiple times – one distance*

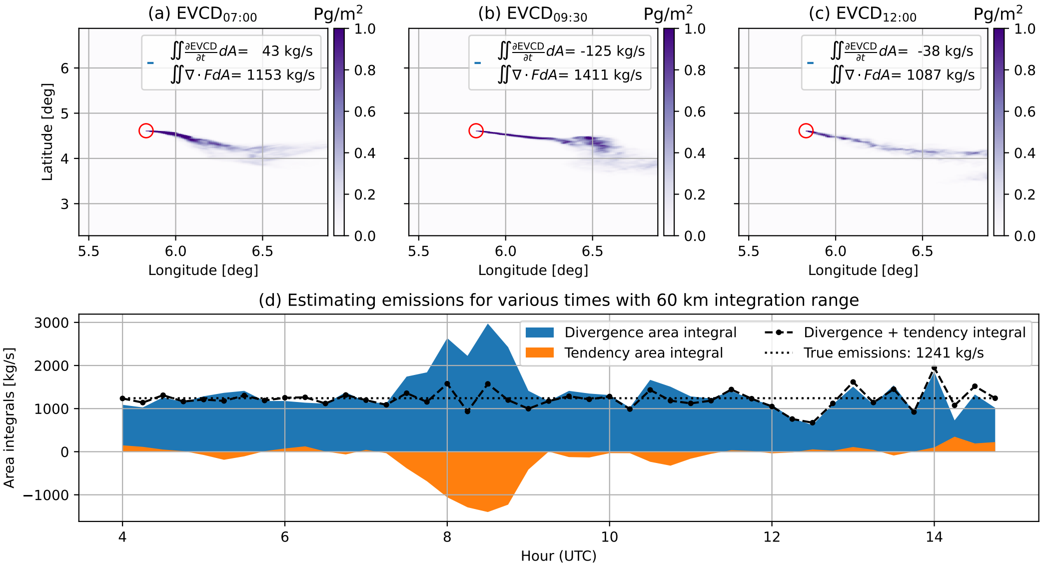

Figure 8. The divergence is estimated for a synthetic TVCD with an area reaching 60 km distance along the plume, at every 15 minutes between 04:00–14:45 UTC on June 5, 2018. Panels (a)–(c) show four different TVCDs corresponding to their subscripted times. Panel (d) shows the estimated emissions. We can see that the temporal area integral compensates for over-estimations otherwise resulting from the divergence area integral and vice versa. The average for just the divergence area integral is 1369 kg/s, while the average for both components lies at 1223 kg/s.

Alternatively, we can fix the spatial integration domain and slice the dataset at different times between 04:00 UTC up to 14:45 UTC. The result of this procedure is shown in Figure 8. We can see that Fig. 8(d) is remarkably similar to Fig. 7(e), despite the different horizontal axis to display the data. Again, the tendency area integral (i.e., the integral of the $\partial \mathrm{EVCD}/\partial t$ term) can compensate for instances where the plume is considerably over- or under-estimated by the divergence area integral. In other words, using the full continuity equation without assuming steady-state conditions leads to better estimates. Unlike in the previous figure, we observe that between 12:00 and 13:00, the source emissions are under-estimated even when taking both the divergence and temporal component into account. The reason for this is likely (1) that tracer transport takes place through more mechanisms than wind alone in COSMO-GHG, and (2) that the temporal finite-difference approximation using instantaneous model outputs is a too coarse approximation of the true tendency, missing small scale turbulent wind fluctuations.

5.2.3 Multiple times – multiple distances

In Figure 9 we illustrate the case for multiple times and multiple distances, i.e., summarize the over-all results from the previous subsections. We observe the same trends as already described before, which are that neglecting the temporal variations may lead to large deviations from the true fluxes; though that smaller errors still remain even after taking the tendency area integral into account.

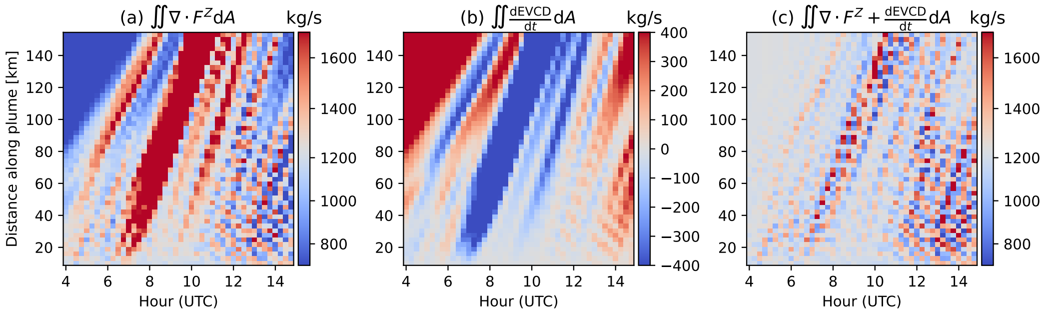

Figure 9. The divergence is estimated for a synthetic EVCD at varying distances and varying times on June 6 2018 (see Figs. 7–8). The true source emission rate is $Q = 1241$ kg/s, around which the colormap of (a) and (c) is centered. Panel (a) corresponds to the divergence estimate for various times and different along-plume integration ranges; panel (b) corresponds to the tendency area integral for the same data; panel (c) corresponds to the sum of panels (a) and (b). We can see that the tendency area integral largely compensates for over-estimations otherwise resulting from the divergence area integral and vice versa.

It is interesting to see that the errors have a spatio-temporal consistency. For example, an overly large divergence estimate (panel Fig. 9(a)) is visible in red from 07:00 UTC at a distance along the plume between 20–50 km, and this overly large estimate appears to propagate along the plume for later times. The tendency area integral (panel Fig. 9(b)) exhibits the same spatio-temporal trends, but with a reversed sign. When taking the sum of Fig. 9(a) and Fig. 9(b), as seen in Fig. 9(c), the two integrals largely cancel out, getting much closer to the desired value of Q=1241 kg/s. Thus, a deviation from the steady state can propagate through the domain with some consistency. This is further demonstrated in Fig. 10, where it can indeed be seen that the difference between EVCDs shows a signal that appears to propagate downstream (e.g., in Fig. 10(a) through Fig. 10(c)).

An important observation from Fig. 9(b) is that the tendency integral is smallest, close to the source. Hence, at a distance close to the source (i.e., less than 20 km in this case), the steady-state assumption appears to be most valid. This further motivates the use of a small integration domain when using the divergence method, as Assumption 4 (steady-state) is more valid for small domains.

6 Discussion

In the derivation of the divergence and mass balance methods, we made a number of assumptions. We will now shortly discuss their validity.

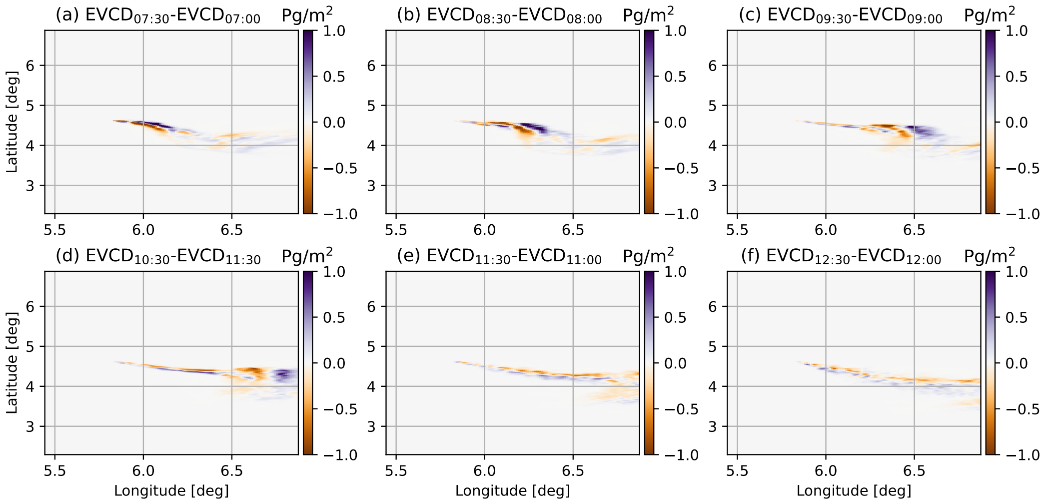

Figure 10. The difference between EVCDs half an hour apart, for 6 consecutive hours of the COSMO simulation. The degree to which the values deviate from zero is indicative of deviations from the steady state.

1. **Mass flux is dominated by linear advection.** This is generally a good assumption, but it can be problematic in situations with low wind speeds and high turbulence. Probably the assumption is also better at larger distances downstream where plumes become more homogeneously mixed. In the initial phase, plumes spread out rapidly (like in the Gaussian case) in the across-wind direction. The Gaussian plume example (Fig. 6) shows that the divergence map is not as focused as it can be when we assume purely advective wind and disregard the diffusive mixing. In the theory and examples we show that the cross-sectional flux method is relatively robust under the assumption of linear advection in the downstream direction (see Fig. 6): if we can truly take cross-sections cutting straight through the plume, the result is not affected by cross-wind mixing.
   For the divergence method, which can't easily be rotated around a specific source, an underestimation of the source strength may occur due to underestimating the full transport vector. The observation that diffusion (following Fick's law of diffusion) must be added into the divergence method was explicitly proposed by Roberts et al. (2023). One proposed fix by them is to integrate over a large enough area to account for this problem, which for both synthetic and real data essentially removes the underestimation problem. However, as shown in this paper, using a larger integration range increases noise receptivity and makes the result more susceptible to non-steady-state features. An alternative is to follow Roberts et al. (2023) and after going through the full calculus for topographically varying boundaries, eq. (22) may be written including advection and diffusion as

$$\sum_{i \text{ if } (x_i,y_i)\in S_a} Q_i(t) = \iint_{S_\mathrm{A}} \Big(\nabla_{x,y}\cdot \mathbf{F}^Z(x,y,t) + k\mathrm{EVCD}(x,y,t) \\ -K\nabla^2_{x,y}\left(\mathrm{EVCD}(x,y,t)+\rho_E(x,y,z_0,t)z_0(x,y)\right)\Big)\,\mathrm{d}S_\mathrm{A}, \qquad (44)$$

   with K a constant assumed eddy covariance term that must be estimated for the data. Note that there is an added term proportional to the second derivative of the topography multiplied with ground concentrations. A potential problem is that this procedure of computing second-order derivatives of real satellite images will be very sensitive to noise. Alternatively, if one manages to fit a local coordinate

system to individual point sources prior to taking the divergence, one can follow the Gaussian plume example given in this paper and include diffusion by adopting a cross-wind speed $V = Uy/(2x)$ together with along-wind speed U.
An additional factor that must be considered for very large plumes are Coriolis effects that add an additional (virtual) transport velocity (see, e.g., Potts et al., 2023, for more details).

2. **The normal fluxes at the bottom and top of the atmosphere vanish.** The assumption regarding a vanishing flux at the top of the atmosphere is certainly valid for space-based remote sensing if one considers the total columns. If one works with remote sensing at lower altitudes, one needs to be more careful and cannot simply ignore such a term if there is considerable vertical wind at the instrument altitude. Regarding the vanishing transport flux at the bottom of the atmosphere, this assumption is also physically sound as there is no transport into the ground. We remark that any source or sink at or through the surface (e.g., seepage, dry deposition) will still be quantified if it is captured in the satellite image (as the flux $\mathbf{F}$ describes the advective transport, not the additive emission flux).
3. **Effective wind fields can be computed correctly.** The effective wind speed as derived in the theory section should be a column-averaged wind speed (explicitly given as Assumption 3), but in practice one is likely to take the wind on (a few) fixed model levels. This deviation from the theory may cause systematic errors. A method to mitigate such errors may be found in Sun (2022), who proposes an alternative divergence operator D_2, which is based on the 'standard' divergence method D_0 with two deviations:

$$\begin{aligned} D_0 &= \nabla_{x,y} \cdot \left(\begin{bmatrix} U_Z \\ V_Z \end{bmatrix} \mathrm{EVCD} \right) \\ &= \begin{bmatrix} U_Z \\ V_Z \end{bmatrix} \cdot \nabla_{x,y} \mathrm{EVCD} + \mathrm{EVCD} \nabla_{x,y} \cdot \begin{bmatrix} U_Z \\ V_Z \end{bmatrix}, \end{aligned} \tag{45}$$

$$D_1 = \begin{bmatrix} U_Z \\ V_Z \end{bmatrix} \cdot \nabla_{x,y}(\mathrm{EVCD}), \tag{46}$$

$$D_2 = D_1 + X \begin{bmatrix} U_Z \\ V_Z \end{bmatrix} \mathrm{EVCD} \cdot \nabla_{x,y}(z_0). \tag{47}$$

D_1 computes only the first term of the full divergence D_0. D_2 adds a term proportional to the gradient of the surface topography, z_0. The operator D_2 that we present here is slightly different from the one derived by Sun (2022), as they suggest to replace the effective wind fields by the surface wind fields in the topographical correction term, but we could not reproduce that derivation. As we shall see, however, it fortuitously holds that

$$\mathrm{EVCD} \nabla_{x,y} \cdot \begin{bmatrix} U_Z \\ V_Z \end{bmatrix} \approx X \begin{bmatrix} U_Z \\ V_Z \end{bmatrix} \mathrm{EVCD} \cdot \nabla_{x,y}(z_0). \tag{48}$$

Thus, the topography correction term in D_2 actually *adds back* the wind divergence term that went ignored in D_1; thus, $D_2 \approx D_0$. We demonstrate this observations for our simulated COSMO-GHG CO_2 data in Supplementary Information Figure S3. However, and perhaps most importantly, it also appears that

$$\mathrm{EVCD} \nabla_{x,y} \cdot \begin{bmatrix} U_Z \\ V_Z \end{bmatrix} \approx X \begin{bmatrix} U_{Z,\text{fixed levels}} \\ V_{Z,\text{fixed levels}} \end{bmatrix} \mathrm{EVCD} \cdot \nabla_{x,y}(z_0). \tag{49}$$

In other words, the topography correction term using not the effective wind speeds but rather the wind from a fixed (few) level(s) is a good approximation of the true effective wind divergence term. This is also shown in Supplementary Information Figure S3. Thus, when using the wrong effective wind (which will generally be the case), more accurate results are obtained when using D_2 rather than D_0.
The reason we give for this improvement is because the effective wind should vary with the topography (terrain-forced flows, for example, force wind speeds to be

higher over a hill than its surroundings; while valleys are protected from high wind speeds) but our chosen effective wind speeds will usually not capture these features appropriately. As shown in Supplementary Information S1–S8, we can derive that whereas one would like to compute $D_0 = \nabla_{x,y}\cdot\mathbf{F}^Z$, one in practice computes $\nabla_{x,y}\cdot(\mathbf{F}^Z + \mathbf{E}^Z)$ instead, where $\mathbf{E}^Z$ represents the erroneous part of the horizontal mass flux, or represents the ignored wind divergence term. As sketched in the Supplementary Information examples, if the error is proportional to the surface due to ignoring terrain-forced flows then $\nabla_{x,y}\cdot\left(\mathbf{E}^Z\right) \approx -X\mathbf{F}^Z\cdot\nabla_{x,y}(z_0)$. Hence, by adding $X\mathbf{F}^Z\cdot\nabla_{x,y}(z_0)$ we (partially) correct for this type of error in the effective wind speed. As remarked above, we thus argue that the topography correction term as devised by Sun (2022) is in fact a less 'noisy' approximation of the wind divergence term than EVCD$\nabla_{x,y}\cdot[U_{Z,\text{fixed levels}}\,V_{Z,\text{fixed levels}}]^T$.

In the supplementary information Figures S1 and S2, we provide two examples of computing the divergence method for a full year of synthetic data using the different divergence operators, and show that, indeed operator D_2 gives more artefact-free results than operator D_0 if erroneous effective wind fields are used.

It is clear that more research into an effective wind field determination would be useful, as it is such a crucial factor in the determination of emissions that small errors in its determination have large influences on the flux image and the determined (or determinable) emission fluxes.

4. **Conditions are steady-state during the time of overpass.** This assumption is strictly incorrect, and the COSMO-GHG example shows that the assumption can lead to significant over- and under-estimations by ignoring temporal variability. The example (Fig. 9) additionally shows that the effect is smallest close to the source. The use of two satellites spaced only a few minutes apart, and with a good SNR, would help to estimate the temporal variability. The deviations due to the non-steady-stateness of real weather situations potentially cancel out if observations are made multiple times on different days. However, since the satellite typically overpasses at the same local time, systematic diurnal variability in wind patterns could induce systematic errors that do not cancel out. We also refer back to the previous discussion point, where it was shown that ignoring the term proportional to the wind divergence is a method to partially avoid violations of the steady-state assumption.
5. **Sources are given as emission rates multiplied with delta distributions.** This is a relatively minor assumption, and merely a way to express that the source consists of a spatial distribution multiplied with a temporal distributions. Regardless, the examples show that our use of finite-difference operations does not *recover* the source as a delta distribution (which would occupy just a single cell). Instead, the use of the FD operations introduces a blurring of the result. Hence, the integration area to determine the emissions for a point source must encompass more cells than just the cell containing a source.
6. **Sinks are described by first-order reactions.** The assumption of first-order reactions is attractive because it gives a way to estimate chemical activity straight from the observations. However, it is likely that this is a severe simplification of the actual chemical activity in the atmosphere. An example is the conversion of NO to NO_2 which only happens in the presence of OH and O_3 which is depleted faster at the edges of a plume than at the center – indicating that the chemistry is likely far from linear. Adapting the method for higher-order chemistry schemes applied to vertical column density images is, however, far from straightforward.

A topic not mentioned yet in this paper is the effect of missing pixels (e.g., due to cloud cover, or being located at the edge of the domain). We cannot make particular general remarks about what to do in case of cloud cover, this will be implementation-specific and up to the implementing party. For example, in Kuhlmann et al. (2019, 2021) Gaussian distributions were fitted through cross-sectional slices, which could effectively account for missing pixels, by assuming a plume cross-section is roughly Gaussian. For the divergence method, we would generally recommend to use the central finite-difference stencil away from cloudy pixels, but use an upwind finite-difference approximation at the edge of missing pixels (as the stencil is smaller, it is capable of recovering more data). Finally, we refer to the various tips and tricks given in Beirle et al. (2023) for further recommendations of using the divergence method in practice.

7 Conclusions

We have derived the divergence and cross-sectional flux method based on first principles, and shown the five assumptions made in that process. The divergence and cross-sectional flux method are, mathematically speaking, identical methods. However, in practice they differ in various aspects. For the purposes of estimating emissions from point sources, the primary difference is that the divergence method is most sensitive around the source location itself (i.e., related to the derivative of the observation field), whereas the cross-sectional flux method is sensitive also in the downstream directions (i.e., related to the amplitude of the observation field). Our derivation differs from that of Sun (2022), as we do not find *theoretical* grounds for adding a topography-dependent correction term in addition to removing the wind divergence term. However, we show how such modifications are advantageous on *practical* grounds, as the topography-dependent correction term turns out to be a more artefact-free approximation of the wind divergence term in case the effective wind is not known perfectly.

The transport speed of a species is not necessarily the same as the wind speed, which can cause errors in the divergence and cross-sectional flux methods. However, if we position our integration area perpendicular to the along-plume axis, this effect can be minimized (see Fig. 6). We show this for a Gaussian plume case, and demonstrate that for the cross-sectional flux method, as long as we take the cross-section truly perpendicular to the plume, we don't suffer from unaccounted for cross-wind transport effects. For the divergence method it's advised to integrate over a large-enough area to minimize this error (see Roberts et al., 2023), or to modify the system into a form like eq. (44). However, this conflicts with the recommendation to keep the integration area small to avoid issues with the assumption of steady-state conditions and minimize noise. It will therefore be necessary to find a suitable compromise to minimize uncertainties.

We furthermore show that the effective wind in the divergence and cross-sectional flux methods should correspond to our *a priori* estimate of the concentration distribution in space (horizontally as well as vertically), weighted over each pixel. Use of plume-rise calculations for specific point-sources may be of relevance in this context, while at distances further downstream one should likely average over the entire planetary boundary layer (assuming an overpass in the late morning and afternoon).

We show that the divergence method can be made more receptive to the finding of point-sources by using the *upstream* flux around a source, which results in a more compact finite-difference stencil. Conversely, we show that in presence of Gaussian additive white noise, source emission estimations are more robust against noise by using a central finite-difference approximation. Still, in presence of many missing pixels (e.g., due to patchy cloud cover) the upwind method may have some use.

Our derivation shows that removing the background eliminates the steady-state assumption for the background component, so removing the background can already fix a lot of the error introduced through this assumption, and is thus recommended (as previously also observed in Hakkarainen et al., 2022).

A number of examples are given to demonstrate the performance of the method on synthetic data. The primary effect observed is that the steady-state assumption is faulty and may lead to big (i.e, a factor 2) over- and underestimations of the true source strength even *without* presence of noise. If the error made in this context is roughly Gaussian distributed in space and time, it will reduce when we can make many emission estimates. In particular, with the cross-sectional flux method we can make more estimates with a single image (for different downstream slices) and thus potentially suffer less from such effects.

Appendix A Open Research

The Jupyter Notebooks to execute the analysis in the paper can be found at `https://mybinder.org/v2/gh/efmkoene/divergence_method_figures/HEAD`, are hosted at `https://github.com/efmkoene/divergence_method_figures` and are preserved at [this will be uploaded to Zenodo once accepted].

Acknowledgments

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 958927 (CoCO2). We wish to thank Janne Hakkarainen, Johanna Tamminen and Grégoire Broquet and three anonymous reviewers for insightful discussions and suggestions. The COSMO-GHG simulations were conducted at the Swiss National Supercomputing Centre (CSCS) under grant No. s1152 and were supported by the Center for Climate Systems Modeling (C2SM).

References

Beirle, S., Borger, C., Dörner, S., Eskes, H., Kumar, V., de Laat, A., & Wagner, T. (2021). Catalog of NO_x emissions from point sources as derived from the divergence of the NO_2 flux for tropomi. *Earth System Science Data*, *13*(6), 2995–3012.

Beirle, S., Borger, C., Dörner, S., Li, A., Hu, Z., Liu, F., … Wagner, T. (2019). Pinpointing nitrogen oxide emissions from space. *Science advances*, *5*(11), eaax9800.

Beirle, S., Borger, C., Jost, A., & Wagner, T. (2023). Improved catalog of NO_x point source emissions (version 2). *Earth System Science Data Discussions*, 1–37.

Bovensmann, H., Buchwitz, M., Burrows, J., Reuter, M., Krings, T., Gerilowski, K., … Erzinger, J. (2010). A remote sensing technique for global monitoring of power plant CO_2 emissions from space and related applications. *Atmospheric Measurement Techniques*, *3*(4), 781–811.

Bréon, F.-M., David, L., Chatelanaz, P., & Chevallier, F. (2021). On the potential of a neural network-based approach for estimating XCO_2 from OCO-2 measurements. *Atmospheric Measurement Techniques Discussions*, 1–21.

Broquet, G., Bréon, F.-M., Renault, E., Buchwitz, M., Reuter, M., Bovensmann, H., … Ciais, P. (2018). The potential of satellite spectro-imagery for monitoring CO_2 emissions from large cities. *Atmospheric Measurement Techniques*, *11*(2), 681–708.

Brunner, D., Kuhlmann, G., Henne, S., Koene, E., Kern, B., Wolff, S., … Fix, A. (2023). Evaluation of simulated CO_2 power plant plumes from six high-resolution atmospheric transport models. *Atmospheric Chemistry and Physics*, *23*(4), 2699–2728.

Brunner, D., Kuhlmann, G., Marshall, J., Clément, V., Fuhrer, O., Broquet, G., … Meijer, Y. (2019). Accounting for the vertical distribution of emissions in atmospheric CO_2 simulations. *Atmospheric Chemistry and Physics*, *19*(7), 4541–4559.

Conley, S., Faloona, I., Mehrotra, S., Suard, M., Lenschow, D. H., Sweeney, C., … Pifer, J. (2017). Application of gauss's theorem to quantify localized surface emissions from airborne measurements of wind and trace gases. *Atmospheric*

Measurement Techniques, *10*(9), 3345–3358.

de Foy, B., & Schauer, J. J. (2022). An improved understanding of NOx emissions in South Asian megacities using TROPOMI NO2 retrievals. *Environmental Research Letters*, *17*(2), 024006.

Dix, B., Francoeur, C., Li, M., Serrano-Calvo, R., Levelt, P. F., Veefkind, J. P., . . . De Gouw, J. (2022). Quantifying NOx Emissions from US Oil and Gas Production Regions Using TROPOMI NO2. *ACS Earth and Space Chemistry*, *6*(2), 403–414.

Dumont Le Brazidec, J., Vanderbecken, P., Farchi, A., Bocquet, M., Lian, J., Broquet, G., . . . Lauvaux, T. (2023). Segmentation of XCO$_2$ images with deep learning: application to synthetic plumes from cities and power plants. *Geoscientific Model Development*, *16*(13), 3997–4016. Retrieved from `https://gmd.copernicus.org/articles/16/3997/2023/` doi: 10.5194/gmd-16-3997-2023

Filonchyk, M., & Peterson, M. P. (2023). NO$_2$ emissions from oil refineries in the Mississippi Delta. *Science of The Total Environment*, *898*, 165569.

Fuentes Andrade, B., Buchwitz, M., Reuter, M., Bovensmann, H., Richter, A., Boesch, H., & Burrows, J. P. (2023). A method for estimating localized CO$_2$ emissions from co-located satellite XCO$_2$ and NO$_2$ images. *EGUsphere*, *2023*, 1–41.

Hakkarainen, J., Ialongo, I., Koene, E., Szeląg, M. E., Tamminen, J., Kuhlmann, G., & Brunner, D. (2022). Analyzing local carbon dioxide and nitrogen oxide emissions from space using the divergence method: An application to the synthetic smartcarb dataset. *Frontiers in Remote Sensing*, 64.

Houweling, S., Baker, D., Basu, S., Boesch, H., Butz, A., Chevallier, F., . . . others (2015). An intercomparison of inverse models for estimating sources and sinks of co2 using gosat measurements. *Journal of Geophysical Research: Atmospheres*, *120*(10), 5253–5266.

Jacob, D. J., Turner, A. J., Maasakkers, J. D., Sheng, J., Sun, K., Liu, X., . . . Frankenberg, C. (2016). Satellite observations of atmospheric methane and their value for quantifying methane emissions. *Atmospheric Chemistry and Physics*, *16*(22), 14371–14396.

Jähn, M., Kuhlmann, G., Mu, Q., Haussaire, J.-M., Ochsner, D., Osterried, K., . . . Brunner, D. (2020). An online emission module for atmospheric chemistry transport models: implementation in cosmo-ghg v5. 6a and cosmo-art v5. 1-3.1. *Geoscientific Model Development*, *13*(5), 2379–2392.

Janssens-Maenhout, G., Pinty, B., Dowell, M., Zunker, H., Andersson, E., Balsamo, G., . . . Veefkind, P. (2020). Toward an operational anthropogenic CO$_2$ emissions monitoring and verification support capacity. *Bulletin of the American Meteorological Society*, *101*(8), E1439 - E1451. Retrieved from `https://journals.ametsoc.org/view/journals/bams/101/8/bamsD190017.xml` doi: https://doi.org/10.1175/BAMS-D-19-0017.1

Kaminski, T., Scholze, M., Rayner, P., Houweling, S., Voßbeck, M., Silver, J., . . . Knorr, W. (2022). Assessing the impact of atmospheric co2 and no2 measurements from space on estimating city-scale fossil fuel co2 emissions in a data assimilation system. *Frontiers in Remote Sensing*, *3*, 887456.

Karion, A., Lauvaux, T., Lopez Coto, I., Sweeney, C., Mueller, K., Gourdji, S., . . . Whetstone, J. (2019). Intercomparison of atmospheric trace gas dispersion models: Barnett shale case study. *Atmospheric Chemistry and Physics*, *19*(4), 2561–2576.

Katharopoulos, I., Brunner, D., Emmenegger, L., Leuenberger, M., & Henne, S. (2022). Lagrangian particle dispersion models in the grey zone of turbulence: Adaptations to FLEXPART-COSMO for simulations at 1 km grid resolution. *Boundary-Layer Meteorology*, *185*(1), 129–160.

Koene, E., & Brunner, D. (2023). *Assessment of plume model performance* (Tech.

Rep.). ECMWF. (CoCO2 report D4.4)

Koene, E., Brunner, D., Kuhlmann, G., Hakkkarainen, J., Le Brazidec, J., & Broquet, G. (2021). *Documentation of plume detection and quantification methods* (Tech. Rep.). ECMWF. (CoCO2 report D4.3)

Kuhlmann, G., Broquet, G., Marshall, J., Clément, V., Löscher, A., Meijer, Y., & Brunner, D. (2019). Detectability of CO_2 emission plumes of cities and power plants with the Copernicus anthropogenic CO_2 monitoring (CO2M) mission. *Atmospheric Measurement Techniques*, *12*(12), 6695–6719.

Kuhlmann, G., Brunner, D., Broquet, G., & Meijer, Y. (2020). Quantifying CO_2 emissions of a city with the copernicus anthropogenic CO_2 monitoring satellite mission. *Atmospheric Measurement Techniques*, *13*(12), 6733–6754.

Kuhlmann, G., Henne, S., Meijer, Y., & Brunner, D. (2021). Quantifying CO_2 emissions of power plants with CO_2 and NO_2 imaging satellites. *Frontiers in Remote Sensing*, 14.

Lauvaux, T., Giron, C., Mazzolini, M., d'Aspremont, A., Duren, R., Cusworth, D., … Ciais, P. (2022). Global assessment of oil and gas methane ultra-emitters. *Science*, *375*(6580), 557–561.

Liu, M., Van Der A, R., Van Weele, M., Eskes, H., Lu, X., Veefkind, P., … Sun, J. (2021). A new divergence method to quantify methane emissions using observations of Sentinel-5P TROPOMI. *Geophysical Research Letters*, *48*(18), e2021GL094151.

Lonsdale, C. R., & Sun, K. (2023). Nitrogen oxides emissions from selected cities in North America, Europe, and East Asia observed by the TROPOspheric Monitoring Instrument (TROPOMI) before and after the COVID-19 pandemic. *Atmospheric Chemistry and Physics*, *23*(15), 8727–8748.

Meijer, Y., et al. (2020). *Copernicus CO_2 Monitoring Mission Requirements Document* (Tech. Rep.). European Space Agency. Retrieved from `https://esamultimedia.esa.int/docs/EarthObservation/CO2M_MRD_v3.0_20201001_Issued.pdf` (Issue 3.0, EOP-SM/3088/YM-ym)

Nassar, R., Hill, T. G., McLinden, C. A., Wunch, D., Jones, D. B., & Crisp, D. (2017). Quantifying CO2 emissions from individual power plants from space. *Geophysical Research Letters*, *44*(19), 10–045.

Nassar, R., Moeini, O., Mastrogiacomo, J.-P., O'Dell, C. W., Nelson, R. R., Kiel, M., … Crisp, D. (2022). Tracking CO2 emission reductions from space: A case study at Europe's largest fossil fuel power plant. *Frontiers in Remote Sensing*, *3*, 98.

Pinty, B., Janssens-Maenhout, G., Dowell, M., Zunker, H., Brunhes, T., Ciais, P., … Scholze, M. (2017). *An Operational Anthropogenic CO_2 Emissions Monitoring & Verification Support capacity - Baseline Requirements, Model Components and Functional Architecture* (Tech. Rep.). European Commission Joint Research Centre. Retrieved from `https://www.copernicus.eu/sites/default/files/2019-09/CO2_Red_Report_2017.pdf` (EUR 28736 EN)

Potts, D. A., Timmis, R., Ferranti, E. J., & Vande Hey, J. D. (2023). Identifying and accounting for the Coriolis effect in satellite NO_2 observations and emission estimates. *Atmospheric Chemistry and Physics*, *23*(7), 4577–4593.

Rey-Pommier, A., Chevallier, F., Ciais, P., Broquet, G., Christoudias, T., Kushta, J., … Sciare, J. (2022). Quantifying NOx emissions in Egypt using TROPOMI observations. *Atmospheric Chemistry and Physics*, *22*(17), 11505–11527.

Rey-Pommier, A., Chevallier, F., Ciais, P., Kushta, J., Christoudias, T., Bayram, I. S., & Sciare, J. (2023). Detecting nitrogen oxide emissions in Qatar and quantifying emission factors of gas-fired power plants–a 4-year study. *Atmospheric Chemistry and Physics*, *23*(21), 13565–13583.

Roberts, C., IJzermans, R., Randell, D., Jones, M., Jonathan, P., Mandel, K., … Shorttle, O. (2023). Avoiding methane emission rate underestimates when using the divergence method. *arXiv preprint arXiv:2304.10303*.

Stockie, J. M. (2011). The mathematics of atmospheric dispersion modeling. *Siam Review*, *53*(2), 349–372.

Sun, K. (2022). Derivation of emissions from satellite-observed column amounts and its application to TROPOMI NO_2 and CO observations. *Geophysical Research Letters*, *49*(23), e2022GL101102.

Varon, D. J., Jacob, D. J., McKeever, J., Jervis, D., Durak, B. O., Xia, Y., & Huang, Y. (2018). Quantifying methane point sources from fine-scale satellite observations of atmospheric methane plumes. *Atmospheric Measurement Techniques*, *11*(10), 5673–5686.

Veefkind, J., Serrano-Calvo, R., de Gouw, J., Dix, B., Schneising, O., Buchwitz, M., … Levelt, P. (2023). Widespread frequent methane emissions from the oil and gas industry in the Permian basin. *Journal of Geophysical Research: Atmospheres*, *128*(3), e2022JD037479.

Wang, Y., Broquet, G., Bréon, F.-M., Lespinas, F., Buchwitz, M., Reuter, M., … Zheng, B. (2020). PMIF v1.0: assessing the potential of satellite observations to constrain CO_2 emissions from large cities and point sources over the globe using synthetic data. *Geoscientific Model Development*, *13*(11), 5813–5831.

Wu, D., Lin, J. C., Fasoli, B., Oda, T., Ye, X., Lauvaux, T., … Kort, E. A. (2018). A Lagrangian approach towards extracting signals of urban CO_2 emissions from satellite observations of atmospheric column CO_2 (XCO_2): X-Stochastic Time-Inverted Lagrangian Transport model ("X-STILT v1"). *Geoscientific Model Development*, *11*(12), 4843–4871.

Xu, T., Zhang, C., Xue, J., Hu, Q., Xing, C., & Liu, C. (2023). Estimating Hourly Nitrogen Oxide Emissions over East Asia from Geostationary Satellite Measurements. *Environmental Science & Technology Letters*.

Ye, X., Lauvaux, T., Kort, E. A., Oda, T., Feng, S., Lin, J. C., … Wu, D. (2020). Constraining fossil fuel co2 emissions from urban area using oco-2 observations of total column co2. *Journal of Geophysical Research: Atmospheres*, *125*(8), e2019JD030528.

Zheng, B., Chevallier, F., Ciais, P., Broquet, G., Wang, Y., Lian, J., & Zhao, Y. (2020). Observing carbon dioxide emissions over China's cities and industrial areas with the Orbiting Carbon Observatory-2. *Atmospheric Chemistry and Physics*, *20*(14), 8501–8510.