

Bayesian denoising of satellite images using co-registered NO₂ images – Supplementary material

Erik Franciscus Maria Koene¹, Gerrit Kuhlmann¹, and Dominik Brunner¹

¹Empa, Laboratory for Air Pollution / Environmental Technology, Dübendorf, Switzerland

Correspondence: Erik Koene (erik.koene@empa.ch)

S1 Alternative derivation of jMMSE estimator

The jMMSE estimator derived in the main text provides a straightforward route to obtaining an estimator of the noise-free CO₂ data based on the joint observational model of eq. (2). That is, we first derive the maximum a posteriori solution, and rewrite this to the linear minimum mean square estimator. The original derivation, however, was done along the following lines – it may be considered to be an alternative derivation of the method, for which the generalization to other setups might be obtained more easily.

The linear minimum mean square error problem can be formulated as

$$\arg \min_{\mathbf{h}, b} \mathbb{E} \left[\left(\mathbf{h}^T \tilde{\mathbf{M}} + b - c \right)^2 \right], \quad (\text{S1})$$

i.e., we try to estimate c by $\mathbf{h}^T \tilde{\mathbf{M}} + b$. Note that \mathbf{h} acts on the *noisy data*, while before \mathbf{H} acted on the *noise-free column*, thus these are entirely different vectors. In the ideal noise-free case we have that $\mathbf{h} = \mathbf{w} = [1 \ 0]^T$. We take a derivative with respect to b and equate the result to zero,

$$\frac{\partial \mathbb{E} \left[\left(\mathbf{h}^T \tilde{\mathbf{M}} + b - c \right)^2 \right]}{\partial b} = 2 \mathbb{E} \left[\mathbf{h}^T \tilde{\mathbf{M}} + b - c \right] \equiv 0, \quad (\text{S2})$$

$$\iff \hat{b} = \mathbb{E}[c] - \mathbf{h}^T \mathbb{E}[\tilde{\mathbf{M}}] \quad (\text{S3})$$

Substituting this in eq. (S1) we get

$$\mathbb{E} \left[\left(\mathbf{h}^T \tilde{\mathbf{M}} + b - c \right)^2 \right] = \mathbb{E} \left[\left(\mathbf{h}^T (\tilde{\mathbf{M}} - \mathbb{E}[\tilde{\mathbf{M}}]) \right)^2 + (c - \mathbb{E}[c])^2 - 2 \mathbf{h}^T (\tilde{\mathbf{M}} - \mathbb{E}[\tilde{\mathbf{M}}]) (c - \mathbb{E}[c]) \right], \quad (\text{S4})$$

$$\begin{aligned} &= \underbrace{\mathbf{h}^T \left(\mathbb{E}[(\tilde{\mathbf{M}} - \mathbb{E}[\tilde{\mathbf{M}}])(\tilde{\mathbf{M}} - \mathbb{E}[\tilde{\mathbf{M}}])^T] \right) \mathbf{h}}_{\mathbf{h}^T \mathbf{C}_{dd} \mathbf{h}} + \underbrace{\mathbb{E}[(c - \mathbb{E}[c])^2]}_{\sigma_c^2} \\ &\quad - \underbrace{2 \mathbf{h}^T \mathbb{E}[(\tilde{\mathbf{M}} - \mathbb{E}[\tilde{\mathbf{M}}])(c - \mathbb{E}[c])]}_{2 \mathbf{h}^T \mathbf{C}_{dc}}, \end{aligned} \quad (\text{S5})$$

with \mathbf{C}_{dd} the data covariance matrix as given in eq. (5), σ_c^2 the variance of the prior, and \mathbf{C}_{dc} is given as following, assuming signal-independent noise $\mathbb{E}[c\mathbf{n}] = \mathbb{E}[c]\mathbb{E}[\mathbf{n}]$, and recalling the covariance rule $\mathbb{E}[(a - \mathbb{E}[a])(b - \mathbb{E}[b])] = \mathbb{E}[ab] - \mathbb{E}[a]\mathbb{E}[b]$,

$$20 \quad \mathbf{C}_{dc} = \mathbb{E}[(\tilde{\mathbf{M}} - \mathbb{E}[\tilde{\mathbf{M}}])(c - \mathbb{E}[c])], \quad (\text{S6})$$

$$= \mathbb{E}[\tilde{\mathbf{M}}c] - \mathbb{E}[\tilde{\mathbf{M}}]\mathbb{E}[c], \quad (\text{S7})$$

$$= \mathbb{E}[(\mathbf{H}c + \mathbf{n})c] - \mathbb{E}[\mathbf{H}c + \mathbf{n}]\mathbb{E}[c], \quad (\text{S8})$$

$$= \left(\mathbf{H} \underbrace{(\mathbb{E}[c^2] - \mathbb{E}[c]^2)}_{\sigma_c^2} + \underbrace{\mathbb{E}[\mathbf{n}c] - \mathbb{E}[\mathbf{n}]\mathbb{E}[c]}_{=0} \right) \underbrace{\mathbf{H}^T \mathbf{w}}_{=1} \quad (\text{S9})$$

$$= \mathbf{H}\mathbf{H}^T \sigma_c^2 \mathbf{w} \quad (\text{S10})$$

$$25 \quad = (\mathbf{C}_{dd} - \mathbf{C}_{nn}) \mathbf{w}, \quad (\text{S11})$$

where we used eq. (8) to find a convenient model-free expression.

Differentiating eq. (S5) with respect to \mathbf{h} and equating the result to zero yields

$$\frac{\partial \mathbb{E} \left[\left(\mathbf{h}^T \tilde{\mathbf{M}} + b - c \right)^2 \right]}{\partial \mathbf{h}} = 2\mathbf{C}_{dd}\mathbf{h} - 2\mathbf{C}_{dc} \equiv 0, \quad (\text{S12})$$

$$\iff \hat{\mathbf{h}} = \mathbf{C}_{dd}^{-1} (\mathbf{C}_{dd} - \mathbf{C}_{nn}) \mathbf{w} = (\mathbf{I} - \mathbf{C}_{dd}^{-1} \mathbf{C}_{nn}) \mathbf{w}. \quad (\text{S13})$$

30 The obtained least-squares optimal values for $\hat{\mathbf{h}}$ and \hat{b} yield the jMMSE estimate for the denoised CO₂ column,

$$\hat{c} = \hat{\mathbf{h}}^T \tilde{\mathbf{M}} + \hat{b}, \quad (\text{S14})$$

$$= \hat{\mathbf{h}}^T (\tilde{\mathbf{M}} - \mathbb{E}[\tilde{\mathbf{M}}]) + \mathbb{E}[c], \quad (\text{S15})$$

$$= \mathbf{w}^T (\mathbf{I} - \mathbf{C}_{nn} \mathbf{C}_{dd}^{-1}) (\tilde{\mathbf{M}} - \mathbb{E}[\tilde{\mathbf{M}}]) + \mathbb{E}[c], \quad (\text{S16})$$

which we recognize is the same as eq. (11).

35 S2 Explicit form of the Joint MMSE model

We can simplify eq. (11) by using the fact that $\mathbf{w}^T = [1 \quad 0]$,

$$\hat{c} = \tilde{C}\tilde{O}_2 - [\text{Cov}(n_{CO_2}, n_{CO_2}) \quad \text{Cov}(n_{CO_2}, n_{NO_2})] \mathbf{C}_{dd}^{-1} (\mathbf{M} - \overline{\mathbf{M}}), \quad (\text{S17})$$

and we may furthermore invert the data covariance matrix to write eq. (S17) as

$$\begin{aligned}
\hat{c} = & \tilde{C}\tilde{O}_2 - \frac{1}{1 - \frac{\text{Cov}(\tilde{C}\tilde{O}_2, \tilde{N}\tilde{O}_2)^2}{\text{Cov}(\tilde{C}\tilde{O}_2, \tilde{C}\tilde{O}_2)\text{Cov}(\tilde{N}\tilde{O}_2, \tilde{N}\tilde{O}_2)}} \frac{\text{Cov}(n_{CO_2}, n_{CO_2})}{\text{Cov}(\tilde{C}\tilde{O}_2, \tilde{C}\tilde{O}_2)} (\tilde{C}\tilde{O}_2 - \overline{C\tilde{O}_2}) \\
& + \frac{1}{\frac{\text{Cov}(\tilde{C}\tilde{O}_2, \tilde{C}\tilde{O}_2)\text{Cov}(\tilde{N}\tilde{O}_2, \tilde{N}\tilde{O}_2)}{\text{Cov}(\tilde{C}\tilde{O}_2, \tilde{N}\tilde{O}_2)^2} - 1} \frac{\text{Cov}(n_{CO_2}, n_{CO_2})}{\text{Cov}(\tilde{C}\tilde{O}_2, \tilde{N}\tilde{O}_2)} (\tilde{N}\tilde{O}_2 - \overline{N\tilde{O}_2}), \\
& - \frac{1}{\frac{\text{Cov}(\tilde{C}\tilde{O}_2, \tilde{C}\tilde{O}_2)\text{Cov}(\tilde{N}\tilde{O}_2, \tilde{N}\tilde{O}_2)}{\text{Cov}(\tilde{C}\tilde{O}_2, \tilde{N}\tilde{O}_2)^2} - 1} \frac{\text{Cov}(n_{CO_2}, n_{NO_2})}{\text{Cov}(\tilde{C}\tilde{O}_2, \tilde{N}\tilde{O}_2)} (\tilde{C}\tilde{O}_2 - \overline{C\tilde{O}_2}), \\
& + \frac{1}{1 - \frac{\text{Cov}(\tilde{C}\tilde{O}_2, \tilde{N}\tilde{O}_2)^2}{\text{Cov}(\tilde{C}\tilde{O}_2, \tilde{C}\tilde{O}_2)\text{Cov}(\tilde{N}\tilde{O}_2, \tilde{N}\tilde{O}_2)}} \frac{\text{Cov}(n_{CO_2}, n_{NO_2})}{\text{Cov}(\tilde{N}\tilde{O}_2, \tilde{N}\tilde{O}_2)} (\tilde{N}\tilde{O}_2 - \overline{N\tilde{O}_2}).
\end{aligned} \tag{S18}$$

Assuming no noise correlation between the CO_2 and NO_2 data, $\text{Cov}(n_{CO_2}, n_{NO_2}) = 0$, that simplifies to

$$\hat{c} = \tilde{C}\tilde{O}_2 - \frac{\text{Cov}(n_{CO_2}, n_{CO_2})}{\text{Cov}(\tilde{C}\tilde{O}_2, \tilde{C}\tilde{O}_2) - \frac{\text{Cov}(\tilde{C}\tilde{O}_2, \tilde{N}\tilde{O}_2)^2}{\text{Cov}(\tilde{N}\tilde{O}_2, \tilde{N}\tilde{O}_2)}} \left((\tilde{C}\tilde{O}_2 - \overline{C\tilde{O}_2}) - \frac{\text{Cov}(\tilde{C}\tilde{O}_2, \tilde{N}\tilde{O}_2)}{\text{Cov}(\tilde{N}\tilde{O}_2, \tilde{N}\tilde{O}_2)} (\tilde{N}\tilde{O}_2 - \overline{N\tilde{O}_2}) \right) \tag{S19}$$

S3 Details for the derivation

We provide some more detail to some equations in the main body of the text, to aid a reader in reproducing the derivation of the (joint) MMSE model.

S3.1 Sherman-Morrison-like matrix inversion identity

The Sherman-Morrison formula is typically given as

$$(\mathbf{A} + \mathbf{u}\mathbf{v}^T)^{-1} = \mathbf{A}^{-1} - \frac{\mathbf{A}^{-1}\mathbf{u}\mathbf{v}^T\mathbf{A}^{-1}}{1 + \mathbf{v}^T\mathbf{A}^{-1}\mathbf{u}}. \tag{S20}$$

By pre-multiplying with \mathbf{A} we obtain

$$\mathbf{A}(\mathbf{A} + \mathbf{u}\mathbf{v}^T)^{-1} = \mathbf{I} - \frac{\mathbf{u}\mathbf{v}^T\mathbf{A}^{-1}}{1 + \mathbf{v}^T\mathbf{A}^{-1}\mathbf{u}}. \tag{S21}$$

Furthermore, if we substitute $\mathbf{u} = \mathbf{v} = f\mathbf{w}$ for some constant factor f , we can rewrite the left- and right-hand side into

$$\mathbf{A}(\mathbf{A} + \mathbf{w}f^2\mathbf{w}^T)^{-1} = \mathbf{I} - \frac{\mathbf{w}f^2\mathbf{w}^T\mathbf{A}^{-1}}{1 + f^2\mathbf{w}^T\mathbf{A}^{-1}\mathbf{w}}. \tag{S22}$$

Dividing the numerator and denominator of the fraction by f^2 finally yields

$$\mathbf{A}(\mathbf{A} + \mathbf{w}f^2\mathbf{w}^T)^{-1} = \mathbf{I} - \frac{\mathbf{w}\mathbf{w}^T\mathbf{A}^{-1}}{f^{-2} + \mathbf{w}^T\mathbf{A}^{-1}\mathbf{w}}. \tag{S23}$$

In the main text we used $\mathbf{A} = \mathbf{C}_{nn}$, $\mathbf{w} = \mathbf{H}$, and $f^2 = \sigma_c^2$, i.e.,

$$\mathbf{C}_{nn} \underbrace{(\mathbf{C}_{nn} + \mathbf{H}\mathbf{H}^T\sigma_c^2)^{-1}}_{\mathbf{C}_{dd}^{-1}} = \mathbf{I} - \frac{\mathbf{H}\mathbf{H}^T\mathbf{C}_{nn}^{-1}}{\sigma_c^{-2} + \mathbf{H}^T\mathbf{C}_{nn}^{-1}\mathbf{H}}. \tag{S24}$$

S3.2 Moving from eq. (9) to eq. (10)

We started with eq. (9), also stated above,

$$60 \quad \mathbf{C}_{nn} \mathbf{C}_{dd}^{-1} = \mathbf{I} - \frac{\mathbf{H} \mathbf{H}^T \mathbf{C}_{nn}^{-1}}{\sigma_c^{-2} + \mathbf{H}^T \mathbf{C}_{nn}^{-1} \mathbf{H}}, \quad (\text{S25})$$

where $\mathbf{C}_{dd} = \mathbf{C}_{nn} + \mathbf{H} \mathbf{H}^T \sigma_c^2$. We have as our sole goal to rewrite this expression into a form that equals eq. (4).

We start by bringing the identity matrix to the left-hand side and multiplying with -1 ,

$$\mathbf{I} - \mathbf{C}_{nn} \mathbf{C}_{dd}^{-1} = \frac{\mathbf{H} \mathbf{H}^T \mathbf{C}_{nn}^{-1}}{\sigma_c^{-2} + \mathbf{H}^T \mathbf{C}_{nn}^{-1} \mathbf{H}}. \quad (\text{S26})$$

We then pre-multiply with $\mathbf{w}^T = [1 \quad 0]$ which satisfies $\mathbf{w}^T \mathbf{H} = 1$ to find

$$65 \quad \mathbf{w}^T (\mathbf{I} - \mathbf{C}_{nn} \mathbf{C}_{dd}^{-1}) = \frac{\mathbf{H}^T \mathbf{C}_{nn}^{-1}}{\sigma_c^{-2} + \mathbf{H}^T \mathbf{C}_{nn}^{-1} \mathbf{H}}, \quad (\text{S27})$$

and we post-multiply with $\tilde{\mathbf{M}}$ to get

$$\mathbf{w}^T (\mathbf{I} - \mathbf{C}_{nn} \mathbf{C}_{dd}^{-1}) \tilde{\mathbf{M}} = \frac{\mathbf{H}^T \mathbf{C}_{nn}^{-1} \tilde{\mathbf{M}}}{\sigma_c^{-2} + \mathbf{H}^T \mathbf{C}_{nn}^{-1} \mathbf{H}}. \quad (\text{S28})$$

Compared to eq. (4) we now only lack a factor $\sigma_c^{-1} \mathbb{E}[c] / (\sigma_c^{-2} + \mathbf{H}^T \mathbf{C}_{nn}^{-1} \mathbf{H})$. The simplest way to gain this factor is to simply add $\mathbb{E}[c]$ to both sides,

$$70 \quad \mathbf{w}^T (\mathbf{I} - \mathbf{C}_{nn} \mathbf{C}_{dd}^{-1}) \tilde{\mathbf{M}} + \mathbb{E}[c] = \frac{\mathbf{H}^T \mathbf{C}_{nn}^{-1} \tilde{\mathbf{M}}}{\sigma_c^{-2} + \mathbf{H}^T \mathbf{C}_{nn}^{-1} \mathbf{H}} + \mathbb{E}[c], \quad (\text{S29})$$

and realizing we can make $\mathbb{E}[c]$ part of the fraction by multiplying it with the denominator,

$$\mathbf{w}^T (\mathbf{I} - \mathbf{C}_{nn} \mathbf{C}_{dd}^{-1}) \tilde{\mathbf{M}} + \mathbb{E}[c] = \frac{\mathbf{H}^T \mathbf{C}_{nn}^{-1} \tilde{\mathbf{M}} + (\sigma_c^{-2} + \mathbf{H}^T \mathbf{C}_{nn}^{-1} \mathbf{H}) \mathbb{E}[c]}{\sigma_c^{-2} + \mathbf{H}^T \mathbf{C}_{nn}^{-1} \mathbf{H}}, \quad (\text{S30})$$

$$= \frac{\mathbf{H}^T \mathbf{C}_{nn}^{-1} \tilde{\mathbf{M}} + \sigma_c^{-2} \mathbb{E}[c]}{\sigma_c^{-2} + \mathbf{H}^T \mathbf{C}_{nn}^{-1} \mathbf{H}} + \frac{\mathbf{H}^T \mathbf{C}_{nn}^{-1} \mathbf{H} \mathbb{E}[c]}{\sigma_c^{-2} + \mathbf{H}^T \mathbf{C}_{nn}^{-1} \mathbf{H}}. \quad (\text{S31})$$

We can see on the right-hand side that we got our desired term with $\sigma_c^{-2} \mathbb{E}[c]$ and an extra, unwanted term. There is a convenient
75 expression for this extra unwanted term which we can deduce from eq. (S27), namely,

$$\mathbf{w}^T (\mathbf{I} - \mathbf{C}_{nn} \mathbf{C}_{dd}^{-1}) \mathbf{H} \mathbb{E}[c] = \frac{\mathbf{H}^T \mathbf{C}_{nn}^{-1} \mathbf{H} \mathbb{E}[c]}{\sigma_c^{-2} + \mathbf{H}^T \mathbf{C}_{nn}^{-1} \mathbf{H}}. \quad (\text{S32})$$

Subtracting (S32) from (S31) gives

$$\mathbf{w}^T (\mathbf{I} - \mathbf{C}_{nn} \mathbf{C}_{dd}^{-1}) (\tilde{\mathbf{M}} - \mathbf{H} \mathbb{E}[c]) + \mathbb{E}[c] = \frac{\mathbf{H}^T \mathbf{C}_{nn}^{-1} \tilde{\mathbf{M}} + \sigma_c^{-2} \mathbb{E}[c]}{\sigma_c^{-2} + \mathbf{H}^T \mathbf{C}_{nn}^{-1} \mathbf{H}}. \quad (\text{S33})$$

Finally, we can see that $\mathbf{H} \mathbb{E}[c] = \mathbb{E}[\mathbf{H}c] = \mathbb{E}[\mathbf{M} - \mathbf{n}]$, which we can use as a substitution on the left-hand side of the
80 equation,

$$\mathbf{w}^T (\mathbf{I} - \mathbf{C}_{nn} \mathbf{C}_{dd}^{-1}) (\tilde{\mathbf{M}} - \mathbb{E}[\mathbf{M}] + \mathbb{E}[\mathbf{n}]) + \mathbb{E}[c] = \frac{\mathbf{H}^T \mathbf{C}_{nn}^{-1} \tilde{\mathbf{M}} + \sigma_c^{-2} \mathbb{E}[c]}{\sigma_c^{-2} + \mathbf{H}^T \mathbf{C}_{nn}^{-1} \mathbf{H}}. \quad (\text{S34})$$

Under the assumption of zero mean noise $\mathbb{E}[\mathbf{n}] = 0$ which underpins the Bayesian solution of eq. (4), we obtain the final expression,

$$\mathbf{w}^T (\mathbf{I} - \mathbf{C}_{nn} \mathbf{C}_{dd}^{-1}) (\tilde{\mathbf{M}} - \mathbb{E}[\tilde{\mathbf{M}}]) + \mathbb{E}[c] = \frac{\mathbf{H}^T \mathbf{C}_{nn}^{-1} \tilde{\mathbf{M}} + \sigma_c^{-2} \mathbb{E}[c]}{\sigma_c^{-2} + \mathbf{H}^T \mathbf{C}_{nn}^{-1} \mathbf{H}}. \quad (\text{S35})$$

85 As a sidenote, we remark that we can obtain the posterior covariance also from eq. (S27), namely,

$$\mathbf{w}^T (\mathbf{I} - \mathbf{C}_{nn} \mathbf{C}_{dd}^{-1}) \mathbf{C}_{nn} \mathbf{w} = \frac{1}{\sigma_c^{-2} + \mathbf{H}^T \mathbf{C}_{nn}^{-1} \mathbf{H}}, \quad (\text{S36})$$

see, e.g., eq. 6.9 in Fichtner (2021). The left-hand portion here allows for a simplification, e.g.,

$$\mathbf{w}^T (\mathbf{I} - \mathbf{C}_{nn} \mathbf{C}_{dd}^{-1}) \mathbf{C}_{nn} \mathbf{w} = \sigma_{CO_2}^2 - \mathbf{w}^T \mathbf{C}_{nn} \mathbf{C}_{dd}^{-1} \mathbf{C}_{nn} \mathbf{w}, \quad (\text{S37})$$

explicitly corresponding to the uncertainty reduction.

90 S3.3 Moving from (10) to (11)

Finally, we can simplify eq. (10), also present in the previous subsection, to a simplified expression,

$$\hat{c} = \mathbf{w}^T (\mathbf{I} - \mathbf{C}_{nn} \mathbf{C}_{dd}^{-1}) (\tilde{\mathbf{M}} - \mathbb{E}[\mathbf{M}]) + \mathbb{E}[c], \quad (\text{S38})$$

$$= \mathbf{w}^T \mathbf{I} (\tilde{\mathbf{M}} - \mathbb{E}[\mathbf{M}]) - \mathbf{w}^T (\mathbf{C}_{nn} \mathbf{C}_{dd}^{-1}) (\tilde{\mathbf{M}} - \mathbb{E}[\mathbf{M}]) + \mathbb{E}[c], \quad (\text{S39})$$

$$= \underbrace{\mathbf{w}^T \tilde{\mathbf{M}}}_{=CO_2} - \underbrace{\mathbf{w}^T \mathbb{E}[\mathbf{M}] + \mathbb{E}[c]}_{=0} - \mathbf{w}^T (\mathbf{C}_{nn} \mathbf{C}_{dd}^{-1}) (\tilde{\mathbf{M}} - \mathbb{E}[\mathbf{M}]), \quad (\text{S40})$$

$$95 = \tilde{CO}_2 - \mathbf{w}^T (\mathbf{C}_{nn} \mathbf{C}_{dd}^{-1}) (\tilde{\mathbf{M}} - \mathbb{E}[\mathbf{M}]). \quad (\text{S41})$$

S4 Divergence results for SMARTCARB

In the main paper we present divergence results for a full year of TROPOMI. However, we can equally present divergence results for a full year of SMARTCARB. The advantage of the SMARTCARB dataset is that we have access to the noise free simulated data. Therefore, we can compare the divergence results between the various noisy and denoised estimates to the true noise free result. As also mentioned in the main body of the paper, emission estimation depends on a good signal to noise ratio, but also on other factors such as a good estimate of the effective wind speed and on steady-state conditions.

The code to produce the divergence images is identical to what was used for TROPOMI. SMARTCARB consists of data simulated for up to 6 synthetic satellites, allowing a user to make constellations of 6 or less than 6 satellites. However, for this study we used all 6 satellites, meaning we have access to 1562 images.

105 We can first look at the resulting divergence images as in Figure S1.

Then we integrate around each source to a radius of 15 km. This produces a bar plot as shown in Figure S2. We can notice that the various methods do *not* necessarily get close to the true (known) emissions, but they do get relatively close to the noise-free emissions. What is of particular importance for us here is to note that there is no obvious bias present for the various

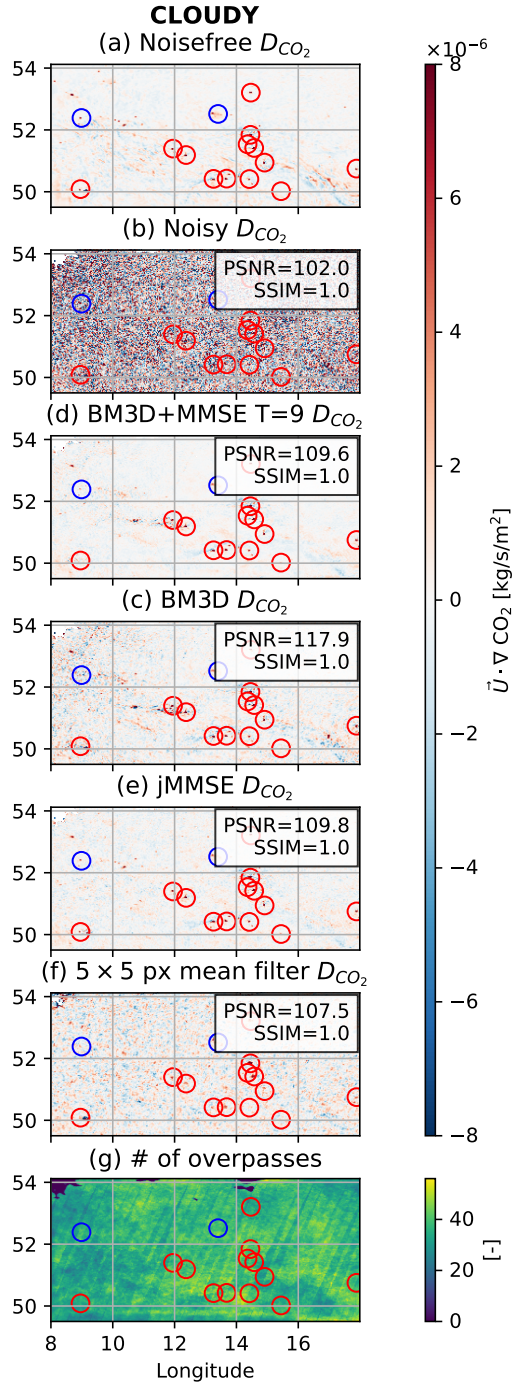


Figure S1. An overview of the various SMARTCARB divergence images as obtained using the various methods. In panel (f) we show the number of overpasses where the CO_2 signal had a cloud cover of less than 1%.

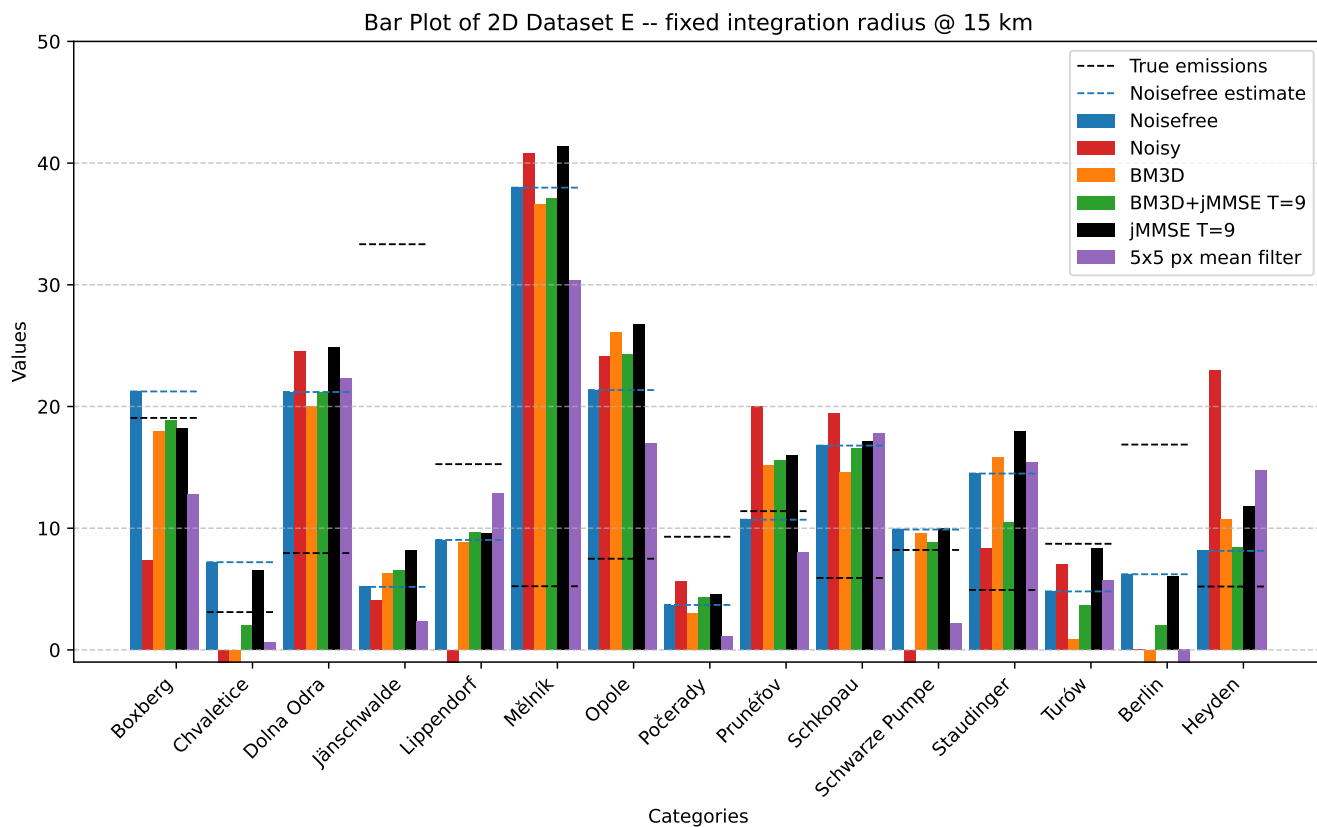


Figure S2. A bar plot of the estimated emissions for the SMARTCARB dataset, using the various denoising methods

denoising methods, other than the 5×5 pixel mean filter which has a tendency to underestimate the emissions marginally. We show this also in the scatter plot of Figure S3.

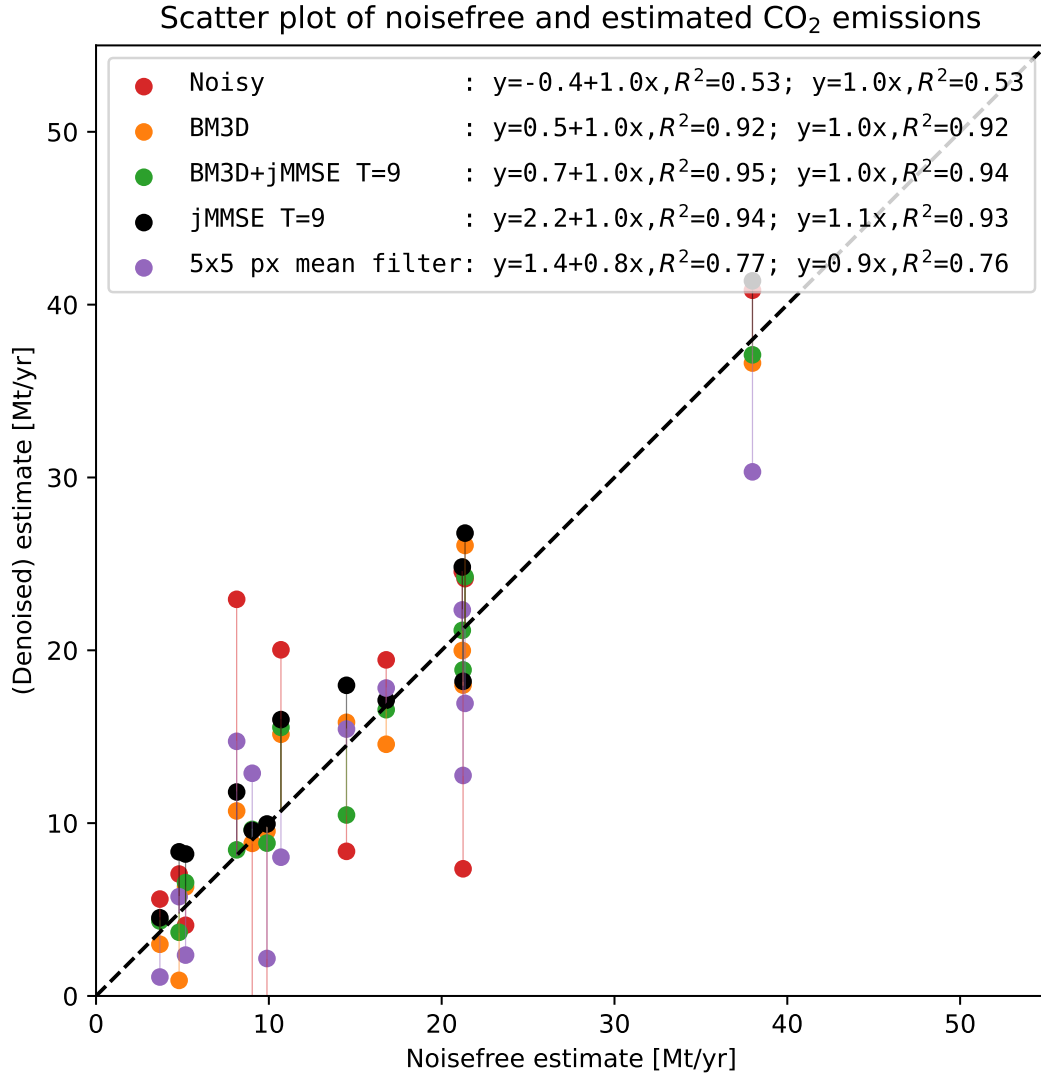


Figure S3. A scatter plot of the estimated emissions for the SMARTCARB dataset, using the various denoising methods. Added to the plot legend are least squares fits comparing the noise-free to the denoised estimates – once for a model of $y(x) = a + bx$ and once for a model $y(x) = bx$, and associated R^2 values corresponding to goodness-of-fit are given. It is here clear that the BM3D+jMMSE T=9 approach yields the best results for both fitting models in terms of R^2 .

S5 Additional figures

S5.1 Synthetic case – jMMSE for different window sizes

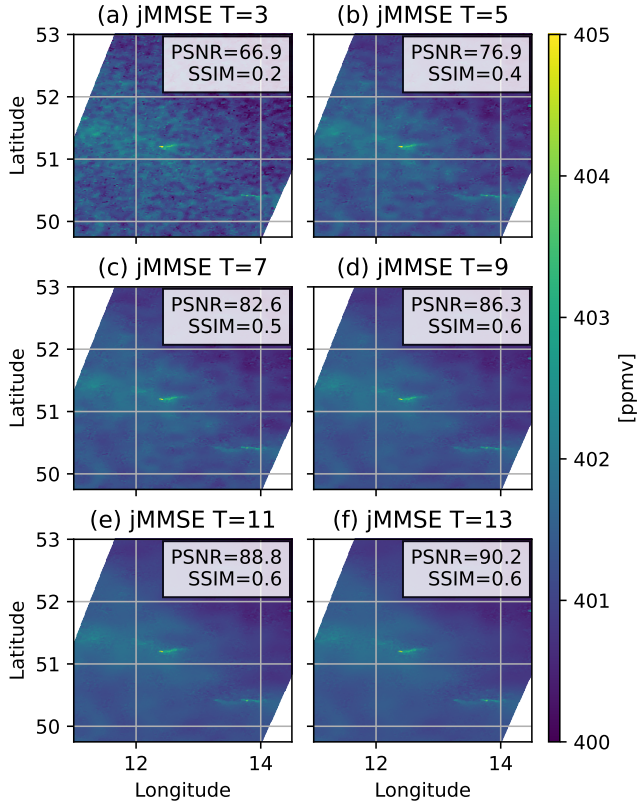


Figure S4. The noisy data from which Figure 3(b) was drawn, though we now consider a larger area, denoised using the jMMSE with $T = 3$ to $T = 13$ leads to increasing PSNR and SIM scores. This indicates enhanced performance with larger window sizes.

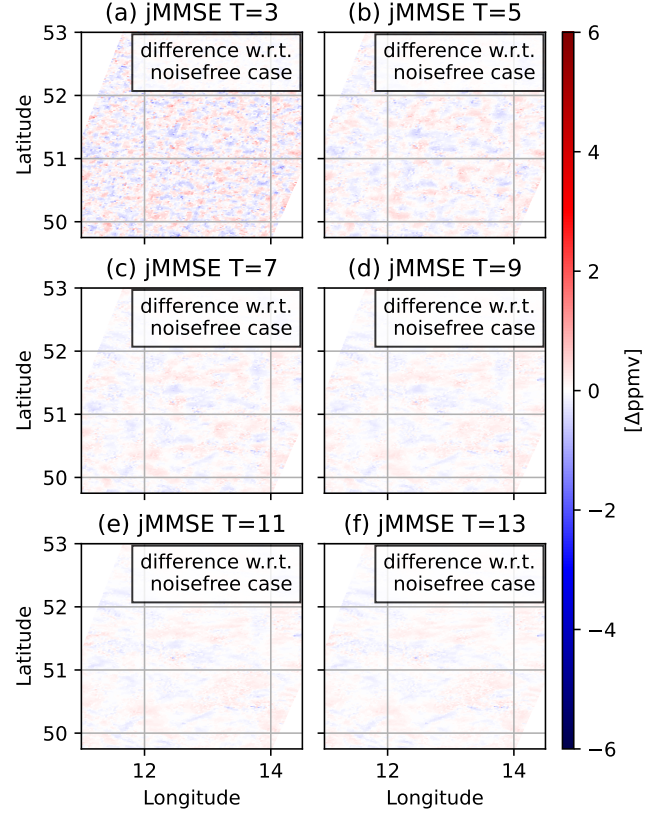


Figure S5. Same as Figure S4, but now we show the difference with respect to the noise-free case. We see that we remove increasingly more noise, but do not improve our recovery of some plumes (visible as ‘blue’ features, i.e., where our recovery underestimates the true source strength).

S5.2 TROPOMI SO₂ case – jMMSE for different window sizes

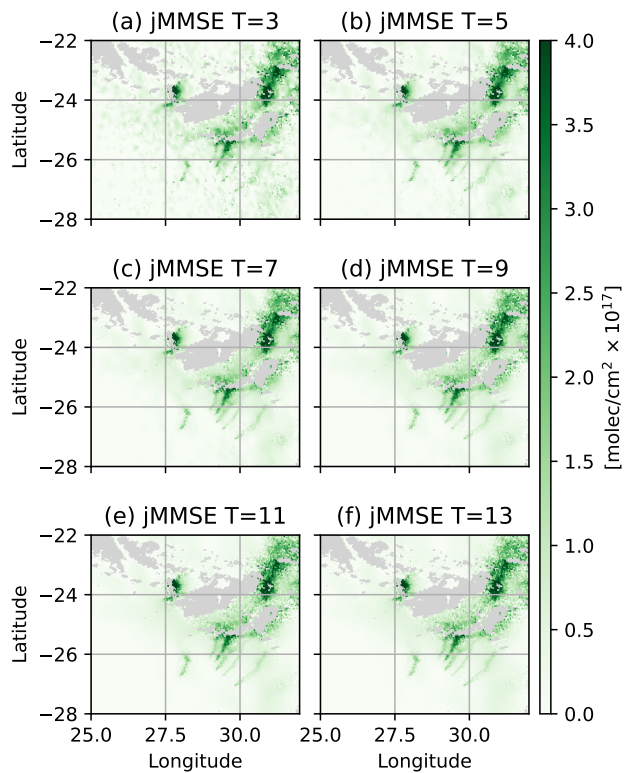


Figure S6. The noisy data from Figure 5(c) denoised using the jMMSE with $T = 3$ to $T = 13$ (without an application of BM3D). We can see that the data becomes increasingly less noisy.

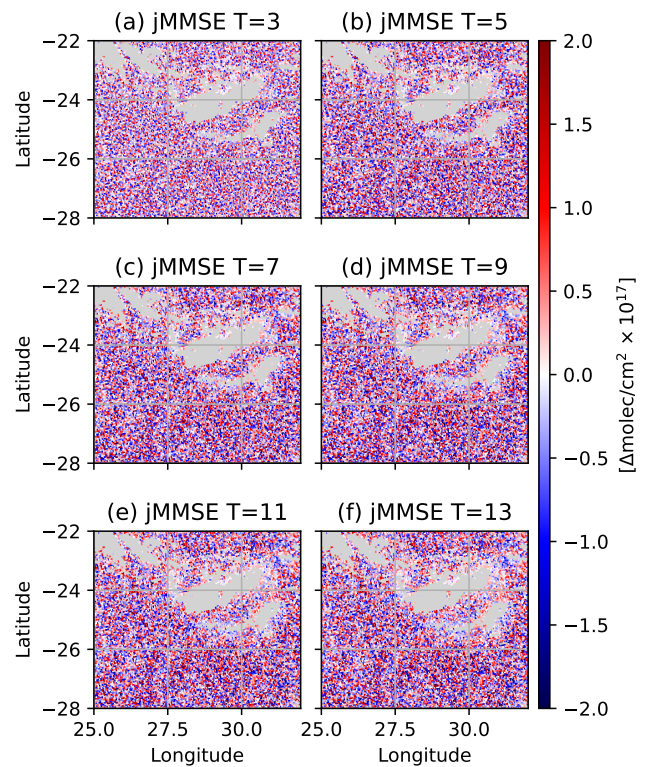


Figure S7. Same as Figure S6, but now we show the difference with respect to the original SO₂ input data. We notice diminishing returns regarding noise removal as T grows.

S5.3 SO₂ emission estimates

115 In the main body of the paper, we display the emission estimate ratios between the ‘noisy’ divergence and denoised SO₂
divergence estimates. Here, we provide the actual estimates we obtained in the form of a heatmap. Some emission estimates
are negative, specifically for Newcastle steel works, Camden power station, and Kelvin power station. This indicates that the
integration radius may be inadequate or that an important nearby sink was overlooked. As noted in the main body, a more
thorough study, with improved AMF corrections and carefully chosen integration ranges for each source, would likely yield
120 more reliable numbers. However, this is beyond the scope of the current paper.

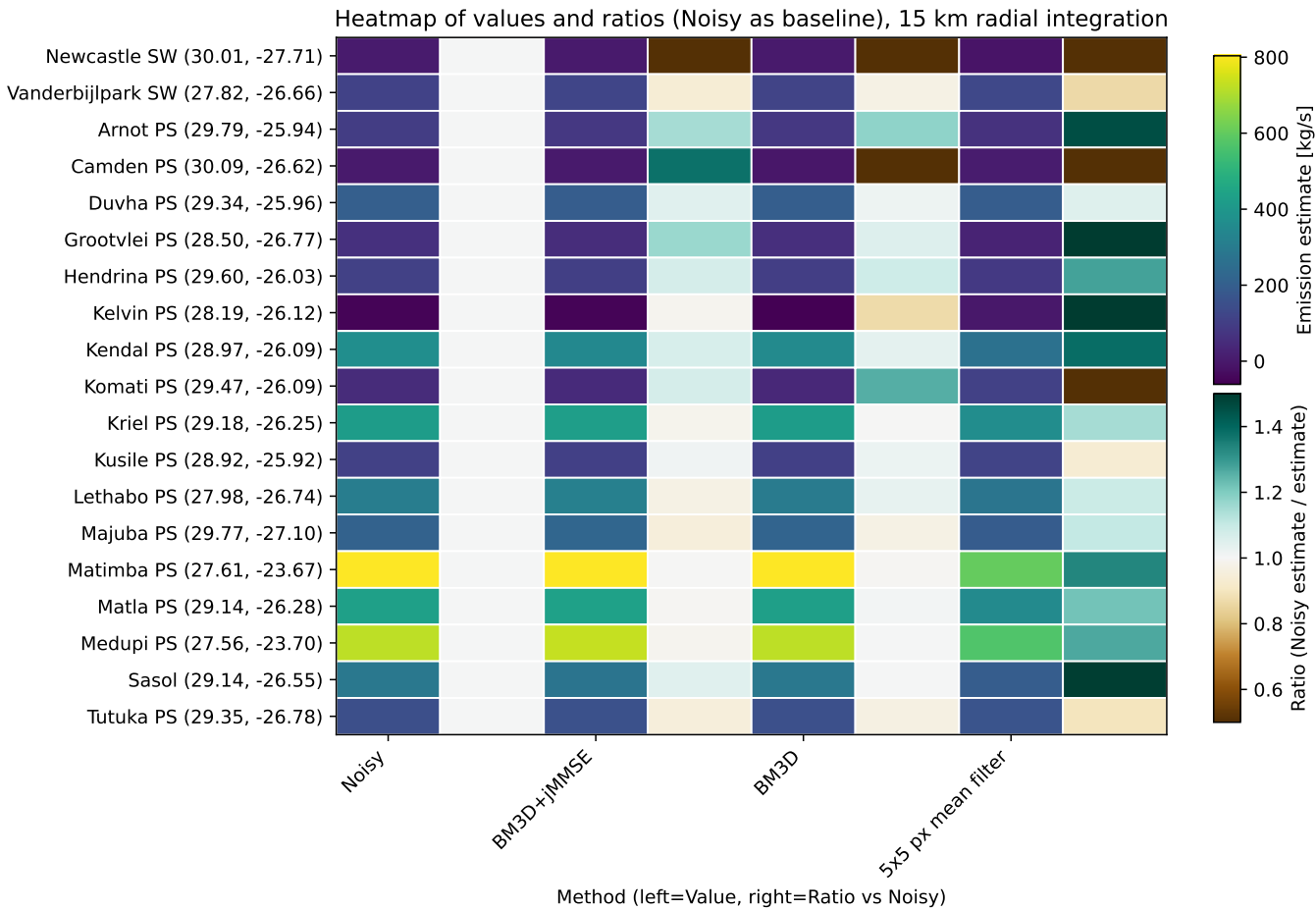


Figure S8. Heatmap of SO₂ emission estimates. The abbreviation SW stands for ‘steel works’ and ‘PS’ stands for power station.

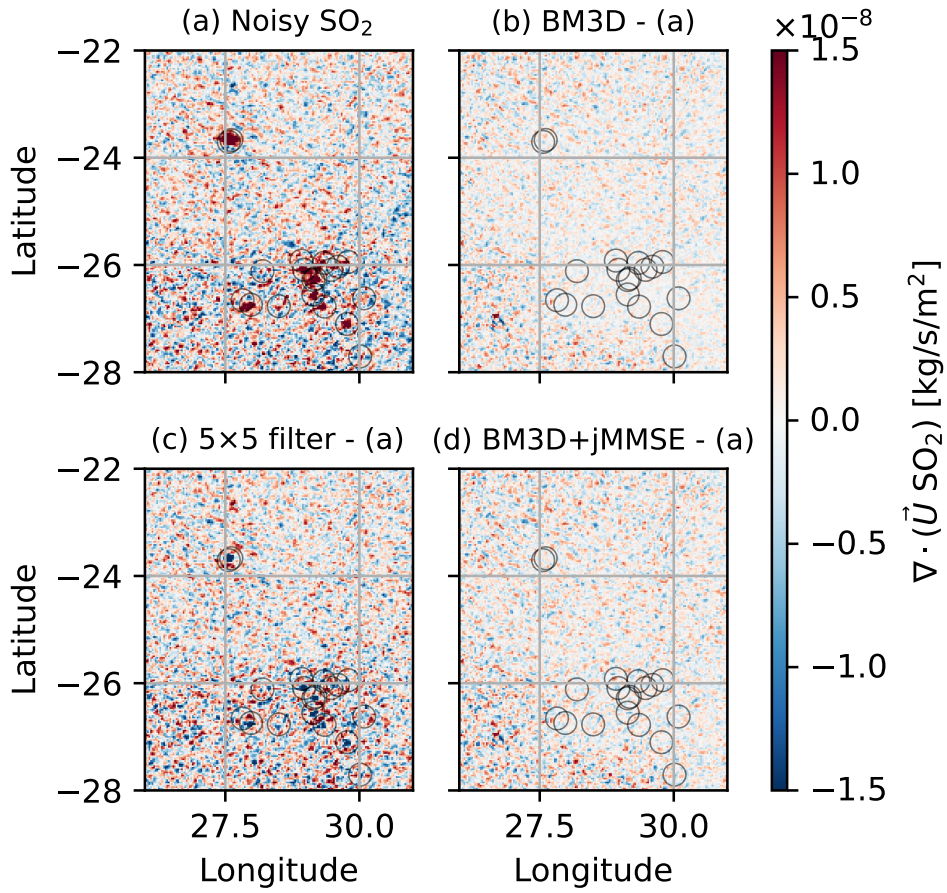


Figure S9. Difference plot of the divergence, by subtracting the SO_2 divergence map, reproduced in panel (a), from the denoised estimates shown in Figure 6. We can see that the 5×5 filter shows considerable loss of signal, while the BM3D and BM3D+jMMSE methods, in their difference plots, essentially just show a noise reduction without affecting the sources (apart from some residual ‘red’ spots, which indicate that the signal has been *boosted* by the denoising process).

References

Fichtner, A.: Lecture Notes on Inverse Theory, Tech. rep., ETH Zurich, Switzerland, <https://doi.org/10.33774/coe-2021-qpq2j>, 2021.

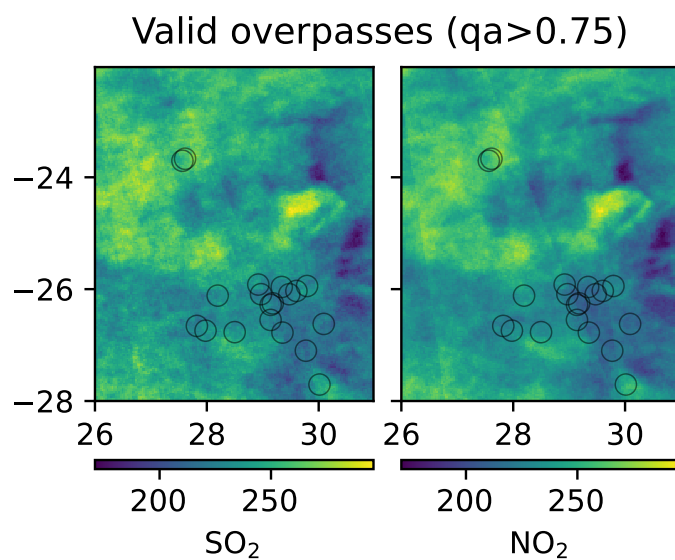


Figure S10. The number of valid overpasses for SO_2 (left) and NO_2 (right) for the selected year 2021 over the South African region. Both products yield around 200 to 250 valid overpasses around most of the sources of interest.