well with reference data, but we find indications of a weak positive bias). The continuous GAW CH_4 reference data cover seasonal cycle signals and have a larger amplitude than the AirCore data. We demonstrate that the lower tropospheric partial column averaged mixing ratio generated from the combined data product is able to capture these signals much better than the respective IASI product or the TROPOMI total column averaged product.

There might be a chance to further improve the quality of the combined data product by performing detailed investigations on the inconsistency between the TROPOMI and the MUSICA IASI XCH_4 data. The availability of additional CH_4 profile reference data for low latitudes (e.g. obtained by the AirCore system) would be very beneficial for such purpose.

The proposed method takes benefit from the outputs generated by the dedicated individual TROPOMI and IASI retrievals, it needs no extra retrievals, and is thus computationally very efficient. This makes it ideal for an application at large scale, and allows the combination of operational IASI and TROPOMI products in an efficient and sustained manner. This has a particular attraction, because IASI and TROPOMI successor instruments will be jointly aboard the upcoming Metop (Meteorological operational) Second Generation satellites (guaranteeing observations from the 2020s to the 2040s). IASI and TROPOMI successor instruments will have globally-distributed and perfectly-collocated observations (over land) in the order of several hundred thousand per day, for which a combined product can be generated in a computationally very efficient way.

Data availability. Access to the MUSICA IASI data is provided via http://www.imk-asf.kit.edu/english/musica-data.php. The TROPOMI XCH_4 data used in this study are available for download at ftp://ftp.sron.nl/open-access-data-2/TROPOMI/tropomi/ch4/14_14_Lorente_et_al_2020_AMTD/. TCCON data are made available via the TCCON Data Archive, hosted by CaltechDATA, California Institute of Technology, California (USA), http://tccondata.org. For Trainou AirCore data please contact Michel Ramonet (michel.ramonet@lsce.ipsl.fr) and for Sodankylä AirCore data please contact Huilin Chen (huilin.chen@rug.nl). The GAW surface in-situ data are available via the World Data Centre for Greenhouse Gases (WDCGG), https://gaw.kishou.go.jp/search/.

Appendix A: Theoretical considerations

In this appendix, we give a brief overview on the theory of optimal estimation remote sensing methods and follow the notation as recommended by the TUNER activity (von Clarmann et al., 2020), which is closely in line with the notation used by Rodgers (2000). The overview focuses on the equations that are important for our work, i.e. the a posteriori combination of two independently retrieved optimal estimation remote sensing products. For a more detailed and general insight into the theory of optimal estimation remote sensing methods we refer to Rodgers (2000) and for a general introduction on vector and matrix algebra dedicated textbooks are recommended.

A1 Basics on retrieval theory

Atmospheric remote sensing instruments measure radiance spectra (written as state vector $\boldsymbol{y}$), which can be well simulated by models ($\boldsymbol{F}$) whenever the actual atmospheric state (the vector $\boldsymbol{x}$) is known. Using the a priori atmospheric state vector $\boldsymbol{x_a}$ we can linearise and write:

$$\boldsymbol{F}(\boldsymbol{x}) - \boldsymbol{F}(\boldsymbol{x_a}) = \Delta \boldsymbol{y} = \mathbf{K}(\boldsymbol{x} - \boldsymbol{x_a}). \tag{A1}$$

Here, $\mathbf{K}$ is the Jacobian matrix, i.e. derivatives that capture how the measurement vector (the measured radiances) will change for changes of the atmospheric state (the atmospheric state vector $\boldsymbol{x}$). A remote sensing retrieval inverts Eq. (A1) and provides an estimation of the difference between the atmospheric state and the a priori atmospheric state. For a moderately non-linear problem (according to Chapter 5 of Rodgers, 2000), the retrieved optimal estimation product ($\hat{\boldsymbol{x}}$) can be written as:

$$\hat{\boldsymbol{x}} - \boldsymbol{x_a} = \mathbf{G}\Delta \boldsymbol{y} = \mathbf{G}[\mathbf{K}(\boldsymbol{x} - \boldsymbol{x_a})]. \tag{A2}$$

$\mathbf{G}$ is the gain matrix and realises the inversion from the measurement domain (radiances) to the domain of the atmospheric states. It consists of derivatives that capture how the retrieved atmospheric state vector will change for changes in the measurement vector:

$$\begin{aligned} \mathbf{G} &= (\mathbf{K}^T {\mathbf{S_{y,n}}}^{-1}\mathbf{K} + {\mathbf{S_a}}^{-1})^{-1}\mathbf{K}^T{\mathbf{S_{y,n}}}^{-1} \\ &= \mathbf{S_a}\mathbf{K}^T(\mathbf{K}\mathbf{S_a}\mathbf{K}^T + \mathbf{S_{y,n}})^{-1}, \end{aligned} \tag{A3}$$

with $\mathbf{S_{y,n}}$ and $\mathbf{R} = {\mathbf{S_a}}^{-1}$ being the retrieval's noise covariance and the constraint matrices (in a strict optimal estimation sense, the constraint matrix is the inverse of the a priori covariance matrix $\mathbf{S_a}$), respectively. The equivalence of both lines in Eq. (A3) is demonstarted in Chapter 4.1 of Rodgers (2000), where the first line is called the n-form and the second line the m-form.

The averaging kernel

$$\mathbf{A} = \mathbf{G}\mathbf{K}, \tag{A4}$$

is an important component of a remote sensing retrieval, because according to Eq. (A2) it reveals how changes of the actual atmospheric state vector $\boldsymbol{x}$ affect the retrieved atmospheric state vector $\hat{\boldsymbol{x}}$.

A valuable diagnostic quantity is the a posteriori covariance matrix, which can be calculated as follows:

$$\mathbf{S}_{\hat{\boldsymbol{x}}} = (\mathbf{K}^T{\mathbf{S_{y,n}}}^{-1}\mathbf{K} + {\mathbf{S_a}}^{-1})^{-1}. \tag{A5}$$

The linearised formulation of the retrieval solution according to (A2) is very useful for the analytic characterisation of the product. The retrieval state's noise error covariance matrix for noise can be analytically calculated as:

$$\mathbf{S_{\hat{x},n}} = \mathbf{G}\mathbf{S_{y,n}}\mathbf{G}^T, \tag{A6}$$

where $\mathbf{S_{y,n}}$ is the covariance matrix for noise on the measured radiances $\boldsymbol{y}$.

Further very helpful equations are the relations between the a posteriori covariance, the averaging kernel, the constraint (or the a priori covariance), and the retrieval's state noise error covariance matrices:

$$\mathbf{S}_{\hat{\boldsymbol{x}}} = (\mathbf{I} - \mathbf{A})\mathbf{S_a}, \tag{A7}$$

and

$$\mathbf{S}_{\hat{\boldsymbol{x}},\mathbf{n}} = \mathbf{S}_{\hat{\boldsymbol{x}}}\mathbf{K}^T{\mathbf{S_{y,n}}}^{-1}\mathbf{K}\mathbf{S}_{\hat{\boldsymbol{x}}} = \mathbf{A}\mathbf{S}_{\hat{\boldsymbol{x}}}, \tag{A8}$$

with $\mathbf{I}$ being the identity matrix. Equations (A7) and (A8) follow from Eqs. (A3) - (A6).

A2 Optimal combination of retrieval data products

In this section, we discuss an optimal estimation retrieval that uses a combined measurement vector (two measurements from different instruments). First we show that the retrieval output of two profile retrievals performed on the same vertical grid can be used in way that yields to the same results as performing a retrieval with the combined measurement vector. Then we present an approach for combining the outputs of a retrieval that provides profiles and another retrieval that provides column data. We show that the combination of profile and column data can be realised in a computationally efficient manner via a Kalman filter. Finally, we discuss the validity of the methods and the requirements on the individual retrieval products.

A2.1 Inversion of a combined measurement vector

According to Eqs. (A2), (A3), and (A5) the retrieval product obtained from measurement $\boldsymbol{y}$ can be written as:

$$\hat{\boldsymbol{x}} - \boldsymbol{x_a} = (\mathbf{K}^T \mathbf{S_{y,n}}^{-1} \mathbf{K} + \mathbf{S_a}^{-1})^{-1} \mathbf{K}^T \mathbf{S_{y,n}}^{-1} \mathbf{K} (\boldsymbol{x} - \boldsymbol{x_a}). \tag{A9}$$

In the case of two individual measurements (measurement 1 and 2), we obtain from using a combined measurement vector $\{\boldsymbol{y_1}, \boldsymbol{y_2}\}$:

$$\begin{aligned} \hat{\boldsymbol{x}} - \boldsymbol{x_a} &= (\mathbf{K_1}^T \mathbf{S_{y_1,n}}^{-1} \mathbf{K_1} + \mathbf{K_2}^T \mathbf{S_{y_2,n}}^{-1} \mathbf{K_2} + \mathbf{S_a}^{-1})^{-1} (\mathbf{K_1}^T \mathbf{S_{y_1,n}}^{-1} \mathbf{K_1} + \mathbf{K_2}^T \mathbf{S_{y_2,n}}^{-1} \mathbf{K_2})(\boldsymbol{x} - \boldsymbol{x_a}) \\ &= (\mathbf{S_{\hat{x}_1}}^{-1} + \mathbf{S_{\hat{x}_2}}^{-1} - \mathbf{S_a}^{-1})^{-1} (\mathbf{K_1}^T \mathbf{S_{y_1,n}}^{-1} \mathbf{K_1} + \mathbf{K_2}^T \mathbf{S_{y_2,n}}^{-1} \mathbf{K_2})(\boldsymbol{x} - \boldsymbol{x_a}), \end{aligned} \tag{A10}$$

where $\mathbf{S_{y_1,n}}$ and $\mathbf{S_{y_2,n}}$ are the respective measurement noise covariances, $\mathbf{K_1}$ and $\mathbf{K_2}$ the respective Jacobians and $\mathbf{S_{\hat{x}_1}}$ and $\mathbf{S_{\hat{x}_2}}$ the respective a posteriori covariances. The second line follows from Eq. (A5). According to Eqs. (A3) - (A5) we can substitute $\mathbf{K}^T \mathbf{S_{y,n}}^{-1} \mathbf{K} (\boldsymbol{x} - \boldsymbol{x_a})$ by $\mathbf{S_{\hat{x}}}^{-1} (\hat{\boldsymbol{x}} - \boldsymbol{x_a})$ and write Eq. (A10) as

$$\hat{\boldsymbol{x}} - \boldsymbol{x_a} = (\mathbf{S_{\hat{x}_1}}^{-1} + \mathbf{S_{\hat{x}_2}}^{-1} - \mathbf{S_a}^{-1})^{-1} [\mathbf{S_{\hat{x}_1}}^{-1} (\hat{\boldsymbol{x}}_1 - \boldsymbol{x_a}) + \mathbf{S_{\hat{x}_2}}^{-1} (\hat{\boldsymbol{x}}_2 - \boldsymbol{x_a})]. \tag{A11}$$

Using Eq. (A11) we can realise an optimal combination of the two retrieval products that only needs the a priori covariance, the a posteriori covariances, and the two retrieval products. The Jacobians are not needed. This combination is mathematically equivalent to using the Jacobians of a combined measurement vector $\{\boldsymbol{y_1}, \boldsymbol{y_2}\}$.

A2.2 Combining profile and column data products

Equation (A11) requires two retrieval results on the same vertical grid and can be used to combine two profile products. Here we will develop a method for combining a profile and a column data product. For a column retrieval we can write in analogy to Eq. (A1)

$$\Delta \boldsymbol{x}^* = \boldsymbol{a}^{*T} (\boldsymbol{x} - \boldsymbol{x_a}), \tag{A12}$$

where $\boldsymbol{a}^{*T}$ is the column averaged mixing ratio according to Appendix C2. Equation (A12) poses an inverse problem of the same kind as Eq. (A1) and in order to optimally estimate a profile from an available column product we can apply the same solution approach as in Eqs. (A2) and (A3). A similar application of this approach is also presented in Sect. 4.2 of Rodgers and Connor (2003). For the application here we substitute in Eq. (A3) $\mathbf{K}$ by $\boldsymbol{a}^{*T}$ and $\mathbf{S_{y,n}}$ by the scalar $S^*_{\hat{\boldsymbol{x}},n}$ (the noise error variance of the column data product) and get the profile

$$\begin{aligned} \hat{\boldsymbol{x}} - \boldsymbol{x_a} &= (\boldsymbol{a}^* {S^*_{\hat{\boldsymbol{x}},n}}^{-1} \boldsymbol{a}^{*T} + \mathbf{S_a}^{-1})^{-1} \boldsymbol{a}^* {S^*_{\hat{\boldsymbol{x}},n}}^{-1} \boldsymbol{a}^{*T} (\boldsymbol{x} - \boldsymbol{x_a}) \\ &= (\boldsymbol{a}^* {S^*_{\hat{\boldsymbol{x}},n}}^{-1} \boldsymbol{a}^{*T} + \mathbf{S_a}^{-1})^{-1} \boldsymbol{a}^* {S^*_{\hat{\boldsymbol{x}},n}}^{-1} \boldsymbol{w}^{*T} \mathbf{S}_{\hat{\boldsymbol{x}}} \mathbf{K}^T {\mathbf{S}_{\mathbf{y},\mathbf{n}}}^{-1} \mathbf{K} (\boldsymbol{x} - \boldsymbol{x_a}). \end{aligned} \tag{A13}$$

We write the second line of Eq. (A13) to discuss similarities with Eq. (A9). The comparison of both reveals that for a retrieval providing only a column product, the Jacobian information provided by $\mathbf{K}$ is vertically aggregated according to the operator $\boldsymbol{a}^* {S^*_{\hat{\boldsymbol{x}},n}}^{-1} \boldsymbol{w}^{*T} \mathbf{S}_{\hat{\boldsymbol{x}}}$. The term $\mathbf{S}_{\hat{\boldsymbol{x}}}$ is the vertically resolved a posteriori covariance, which exist for a retrieval that internally inverts profiles, but only distributes the column products; however, it is only an internal measure of the retrieval, and actually not available.

Instead of the term of Eq. (A12) we now invert the term $\Delta \boldsymbol{x}^* = \boldsymbol{a_2^{*T}} (\boldsymbol{x} - \hat{\boldsymbol{x}}_\mathbf{1})$, i.e. we replace $\boldsymbol{x_a}$ by the profile product $\hat{\boldsymbol{x}}_\mathbf{1}$ of a first retrieval (retrieval 1) on the right side of (A12) and use $\boldsymbol{a_2^{*T}}$ and $S^*_{\hat{\boldsymbol{x}}_\mathbf{2},n}$ for the column averaging kernel and the noise error variance of a second retrieval (retrieval 2), respectively. Here and in the following, retrieval 1 is the profile retrieval and retrieval 2 the retrieval that provides only column products. The solution can easily be achieved by substituting in (A13) $\mathbf{S_a}$ by $\mathbf{S}_{\hat{\boldsymbol{x}}_\mathbf{1}}$, which is the a posteriori covariance of retrieval 1:

$$\hat{\boldsymbol{x}} - \hat{\boldsymbol{x}}_\mathbf{1} = (\boldsymbol{a_2^*} {S^*_{\hat{\boldsymbol{x}}_\mathbf{2},n}}^{-1} \boldsymbol{a_2^{*T}} + {\mathbf{S}_{\hat{\boldsymbol{x}}_\mathbf{1}}}^{-1})^{-1} \boldsymbol{a_2^*} {S^*_{\hat{\boldsymbol{x}}_\mathbf{2},n}}^{-1} \boldsymbol{a_2^{*T}} (\boldsymbol{x} - \hat{\boldsymbol{x}}_\mathbf{1}). \tag{A14}$$

We modify Eq. (A14) by using $\hat{\boldsymbol{x}}_\mathbf{1} = \mathbf{A_1}(\boldsymbol{x} - \boldsymbol{x_a}) + \boldsymbol{x_a}$:

$$\begin{aligned} \hat{\boldsymbol{x}} - \boldsymbol{x_a} &= \mathbf{A_1}(\boldsymbol{x} - \boldsymbol{x_a}) + ({\mathbf{S}_{\hat{\boldsymbol{x}}_\mathbf{1}}}^{-1} + \boldsymbol{a_2^*} {S^*_{\hat{\boldsymbol{x}},n}}^{-1} \boldsymbol{a_2^{*T}})^{-1} \boldsymbol{a_2^*} {S_{\hat{\boldsymbol{x}}_\mathbf{2},n}}^{-1} \boldsymbol{a_2^{*T}} (\mathbf{I} - \mathbf{A_1})(\boldsymbol{x} - \boldsymbol{x_a}) \\ &= ({\mathbf{S}_{\hat{\boldsymbol{x}}_\mathbf{1}}}^{-1} + \boldsymbol{a_2^*} {S^*_{\hat{\boldsymbol{x}},n}}^{-1} \boldsymbol{a_2^{*T}})^{-1} [{\mathbf{S}_{\hat{\boldsymbol{x}}_\mathbf{1}}}^{-1} \mathbf{A_1} + \boldsymbol{a_2^*} {S^*_{\hat{\boldsymbol{x}},n}}^{-1} \boldsymbol{a_2^{*T}}](\boldsymbol{x} - \boldsymbol{x_a}) \\ &= ({\mathbf{S}_{\hat{\boldsymbol{x}}_\mathbf{1}}}^{-1} + \boldsymbol{a_2^*} {S^*_{\hat{\boldsymbol{x}},n}}^{-1} \boldsymbol{a_2^{*T}})^{-1} [{\mathbf{S}_{\hat{\boldsymbol{x}}_\mathbf{1}}}^{-1} (\hat{\boldsymbol{x}}_\mathbf{1} - \boldsymbol{x_a}) + \boldsymbol{a_2^*} {S^*_{\hat{\boldsymbol{x}},n}}^{-1} (\hat{x}_2^* - \boldsymbol{w}^{*T} \boldsymbol{x_a})]. \end{aligned} \tag{A15}$$

In the third line of Eq. (A15) we use the column product $\hat{x}_2^* = \boldsymbol{a_2^{*T}}(\boldsymbol{x} - \boldsymbol{x_a}) + \boldsymbol{w}^{*T} \boldsymbol{x_a}$. Similarly to Eq. (A11) we can generate a combined product without the need of the Jacobian matrices. The combination is possible by using the profile and the column product ($\hat{\boldsymbol{x}}_\mathbf{1}$ and $\hat{x}_2^*$, respectively) together with the a posteriori covariance of the profile product and the noise error and averaging kernel of the column product.

A2.3 Linear Kalman filter

Here we show that the approach developed in Appendix A2.2 is equivalent to a Kalman filter. An important application of a Kalman filter (Kalman, 1960; Rodgers, 2000) is data assimilation in the context of atmospheric modelling. There, the filter operates sequentially in different time steps. Kalman filter data assimilation methods determine the analysis state ($\hat{\boldsymbol{x}}^{\boldsymbol{a}}$) by optimally combining the background (or forecast) state ($\hat{\boldsymbol{x}}^{\boldsymbol{b}}$) with the information as provided by a new observation ($\hat{\boldsymbol{x}}^{\boldsymbol{o}}$):

$$\hat{\boldsymbol{x}}^{\boldsymbol{a}} = \hat{\boldsymbol{x}}^{\boldsymbol{b}} + \mathbf{M}[\hat{\boldsymbol{x}}^{\boldsymbol{o}} - \mathbf{H}\hat{\boldsymbol{x}}^{\boldsymbol{b}}] \tag{A16}$$

Optimal means here that the uncertainties of both, the background state and the observation, are correctly taken into account by the Kalman gain matrix ($\mathbf{M}$):

$$\mathbf{M} \quad = \quad \mathbf{S}_{\hat{\boldsymbol{x}}^b}\mathbf{H}^T(\mathbf{H}\mathbf{S}_{\hat{\boldsymbol{x}}^b}\mathbf{H}^T + \mathbf{S}_{\hat{\boldsymbol{x}}^o,\mathbf{n}})^{-1}, \tag{A17}$$

with $\mathbf{S}_{\hat{\boldsymbol{x}}^b}$ and $\mathbf{S}_{\hat{\boldsymbol{x}}^o,\mathbf{n}}$ being the uncertainty covariances of background state and the new measurement, respectively. The matrix $\mathbf{H}$ is the measurement forward operator, which maps the background domain into the measurement domain.

By rearranging the n-form of (A14) as the m-form – in analogy to to Eq. (A3) – and by using again $\hat{\boldsymbol{x}}_{\mathbf{1}} = \mathbf{A_1}(\boldsymbol{x} - \boldsymbol{x_a}) + \boldsymbol{x_a}$ and $\hat{x}_2^* = \boldsymbol{a}^{*T}(\boldsymbol{x} - \boldsymbol{x_a}) + \boldsymbol{w}^{*T}\boldsymbol{x_a}$ we get

$$\begin{aligned}
\hat{\boldsymbol{x}} &= \hat{\boldsymbol{x}}_{\mathbf{1}} + \mathbf{S}_{\hat{\boldsymbol{x}}_1}\boldsymbol{a}_{\mathbf{2}}^*(\boldsymbol{a}_{\mathbf{2}}^{*T}\mathbf{S}_{\hat{\boldsymbol{x}}_1}\boldsymbol{a}_{\mathbf{2}}^* + S_{\hat{\boldsymbol{x}}_{\mathbf{2}},n}^*)^{-1}\boldsymbol{a}_{\mathbf{2}}^{*T}(\boldsymbol{x} - \hat{\boldsymbol{x}}_{\mathbf{1}}) \\
&= \hat{\boldsymbol{x}}_{\mathbf{1}} + \mathbf{S}_{\hat{\boldsymbol{x}}_1}\boldsymbol{a}_{\mathbf{2}}^*(\boldsymbol{a}_{\mathbf{2}}^{*T}\mathbf{S}_{\hat{\boldsymbol{x}}_1}\boldsymbol{a}_{\mathbf{2}}^* + S_{\hat{\boldsymbol{x}}_{\mathbf{2}},n}^*)^{-1}[\hat{x}_2^* - \boldsymbol{a}_{\mathbf{2}}^{*T}\hat{\boldsymbol{x}}_{\mathbf{1}} - (\boldsymbol{w}_{\mathbf{2}}^{*T}\boldsymbol{x_a} - \boldsymbol{a}_{\mathbf{2}}^{*T}\boldsymbol{x_a})] \\
&= \hat{\boldsymbol{x}}_{\mathbf{1}} + \boldsymbol{m}(\hat{x}_2^* - \boldsymbol{a}_{\mathbf{2}}^{*T}\hat{\boldsymbol{x}}_{\mathbf{1}}) - \boldsymbol{m}(\boldsymbol{w}_{\mathbf{2}}^{*T}\boldsymbol{x_a} - \boldsymbol{a}_{\mathbf{2}}^{*T}\boldsymbol{x_a})
\end{aligned} \tag{A18}$$

with

$$\boldsymbol{m} = \mathbf{S}_{\hat{\boldsymbol{x}}_1}\boldsymbol{a}_{\mathbf{2}}^*(\boldsymbol{a}_{\mathbf{2}}^{*T}\mathbf{S}_{\hat{\boldsymbol{x}}_1}\boldsymbol{a}_{\mathbf{2}}^* + S_{\hat{\boldsymbol{x}}_{\mathbf{2}},n}^*)^{-1}. \tag{A19}$$

Disregarding the term that accounts for the a priori information ($\boldsymbol{m}(\boldsymbol{w}_{\mathbf{2}}^{*T}\boldsymbol{x_a} - \boldsymbol{a}_{\mathbf{2}}^{*T}\boldsymbol{x_a})$), the Eqs. (A18) and (A19) are the same as Kalman filter Eqs. (A16) and (A17): retrieval 1 provides the background state and retrieval 2 the new observation. Compared to Eq. (A15) the form of Eq. (A18) has the advantage that no matrices have to be inverted only the scalar ($\boldsymbol{a}_{\mathbf{2}}^{*T}\mathbf{S}_{\hat{\boldsymbol{x}}_1}\boldsymbol{a}_{\mathbf{2}}^* + S_{\hat{\boldsymbol{x}}_{\mathbf{2}},n}^*$).

A2.4 Discussion and requirements

In the Appendices A2.2 and A2.3, we assume the usage of the same a priori for the two individual retrievals. Since generally two individually performed retrievals use two different a priori settings we have to perform an a priori adjustment. Using the a priori of retrieval 2 as the reference ($\boldsymbol{x}_{\mathbf{2},\boldsymbol{a}} = \boldsymbol{x_a}$), we can adjust the output of retrieval 1 by (see Eq. (10) of Rodgers and Connor, 2003):

$$\hat{\boldsymbol{x}}_{\mathbf{1}}' = \hat{\boldsymbol{x}}_{\mathbf{1}} + (\mathbf{A_1} - \mathbf{I})(\boldsymbol{x}_{\mathbf{1},\boldsymbol{a}} - \boldsymbol{x}_{\mathbf{2},\boldsymbol{a}}), \tag{A20}$$

where $\boldsymbol{x}_{\mathbf{1},\boldsymbol{a}}$ is the a priori used by retrieval 1.

For a combination according to Eq. (A11) we need retrieval 1 and 2 outputs obtained by using the same constraint (the inverse of the a priori covariance $\mathbf{S_a}$). This has to be accounted for before applying Eq. (A11), by adusting the contraint according to the formalism as presented in Chapter 10.4 of Rodgers (2000) or Sect. 4.2 of Rodgers and Connor (2003). By applying Eq. (A15) or the Kalman filter according to Eq. (A18) the common constraint is automatically set to the constraint of the retrieval 1 product and no extra modification is necesarry.

The synergetic combination of remote sensing profile and column products according to Eq. (A15) or (A18) is possible, whenever: (1) the two remote sensing observations are made at the same time and detect the same location, (2) the problems is moderately non-linear (according to Chapter 5 of Rodgers, 2000), and (3) the individual retrieval output as listed by Eq. (A15) or (A18) is made available. This is for the profile retrieval the a posteriori covariances ($\mathbf{S}_{\hat{\boldsymbol{x}}}$, which might also be reconstracted from $\mathbf{A}$ and $\mathbf{R} = \mathbf{S_a}^{-1}$ according to Eq. (A7)), the averaging kernels ($\mathbf{A}$), and the retrieved and a priori state vectors ($\hat{\boldsymbol{x}}$ and $\boldsymbol{x_a}$, respectively). For the column retrieval we need the noise variances (the scalar $S_{\hat{\boldsymbol{x}},n}$), the column averaging kernels (the row vector $\boldsymbol{a}^{*T}$), the column product ($\hat{x}_2^*$), and the a priori column data ($\boldsymbol{w}^{*T}\boldsymbol{x_a}$), respectively.

Appendix B: Operator for transformation between linear and logarithmic scales

Linear scale differentials and logarithmic scale differentials are related by $\partial x = x \partial \ln x$. For transforming differentials or covariances of a state vector with dimension nal (nal: number of atmospheric levels) from logarithmic to linear scale we define the $nal \times nal$ diagonal matrix $\mathbf{L}$:

$$\mathbf{L} = \begin{pmatrix} \hat{x}_1 & 0 & \cdots & 0 \\ 0 & \hat{x}_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \hat{x}_{nal} \end{pmatrix}. \tag{B1}$$

Here $\hat{x}_i$ is the value of the ith element of the retrieved state vector (i.e. in case of an atmospheric CH_4 state vector the CH_4 mixing ratios retrieved at the ith model level).

Approximatively, a logarithmic scale averaging kernel matrix $\mathbf{A^l}$ can then be expressed in the linear scale as:

$$\mathbf{A} \approx \mathbf{L}\mathbf{A^l}\mathbf{L}^{-1}. \tag{B2}$$

This is here an approximation, because on the right side the operator $\mathbf{L}$ should contain the actualy instead of the retrieved mixing ratios. It is a valid approximation as long as the a priori is reasonable and there is no large bias in the retrieval data.

Similarly a logarithmnic scale covariance matrix $\mathbf{S^l}$ can be approximately expressed in the linear scale as:

$$\mathbf{S} \approx \mathbf{L}\mathbf{S^l}\mathbf{L}^T. \tag{B3}$$

Here the approximation is because $\Delta x \approx x \Delta \ln x$.

Appendix C: Operators for column data

This appendix explains the calculation of operators for partial (and total) column data. Although some sections are similar to Appendix C of Schneider et al. (2021) we think it is a very useful reference here, because it facilitates the reproducibility of our results.

For converting mixing ratio profiles into amount profiles we set up a pressure weighting operator $\mathbf{Z}$, as a diagonal matrix with the following entries:

$$Z_{i,i} = \frac{\Delta p_i}{g_i m_{\mathrm{air}} (1 + \frac{m_{\mathrm{H_2O}}}{m_{\mathrm{air}}} \hat{x}_i^{\mathrm{H_2O}})}. \tag{C1}$$

Using the pressure p_i at atmospheric grid level i we set $\Delta p_1 = \frac{p_2 - p_1}{2} - p_1$, $\Delta p_{nal} = p_{nal} - \frac{p_nal - p_{nal-1}}{2}$, and $\Delta p_i = \frac{p_{i+1} - p_i}{2} - \frac{p_i - p_{i-1}}{2}$ for $1 < i < nal$. Furthermore, g_i is the gravitational acceleration at level i, m_{air} and $m_{\mathrm{H_2O}}$ the molecular mass of dry air and water vapour, respectively, and $\hat{x}_i^{\mathrm{H_2O}}$ the retrieved or modelled water vapour mixing ratio at level i.

We define an operator $\mathbf{W}^T$ for resampling fine gridded atmospheric amount profiles into coarse gridded atmospheric partial column amount profiles. It has the dimension $c \times nal$, where c is the number of the resampled coarse atmospheric grid levels and nal, the number of atmospheric levels of the original fine atmospheric grid. Each line of the operator has the value '1' for the levels that are resampled and '0' for all other levels:

$$\mathbf{W}^T = \begin{pmatrix} 1 & \cdots & 1 & 0 & \cdots & \cdots & \cdots & \cdots & 0 \\ 0 & \cdots & 0 & 1 & \cdots & 1 & 0 & \cdots & 0 \\ 0 & \cdots & \cdots & \cdots & \cdots & 0 & 1 & \cdots & 1 \end{pmatrix}. \tag{C2}$$

In analogy we can define a row vector $\boldsymbol{w}^T$ (with the dimension $1 \times nal$) with all elements having the value '1', which allows the resampling for the total column amounts.

C1 Column amounts

The kernel that discribes how a change in the amount at a certain altitude affects the retrieved partial (or total) colunm amount can be calculated as:

$$\mathbf{A}' = \mathbf{W}^T \mathbf{Z} \mathbf{A} \mathbf{Z}^{-1}. \tag{C3}$$

For the total column, we replace $\mathbf{W}^T$ by $\boldsymbol{w}^T$ and get the row vector $\boldsymbol{a'}^T$ (dimension $1 \times nal$). This is the total column kernel provided by the TROPOMI data and it is typically written as $\boldsymbol{a}^T$. Figure 3 shows examples of such total and partial column amount kernels. The total column amount kernel can be interpolated to different altitude grids. For the applications in Sects. 2 and 3 we interpolate the TROPOMI total column amount kernel to the vertical grid used by the MUSICA IASI retrieval.

C2 Column averaged mixing ratios

We can also combine the operators $\mathbf{Z}$ and $\mathbf{W}^T$ for the calculation of a pressure weighted resampling operator by:

$$\mathbf{W^*}^T = (\mathbf{W}^T \mathbf{Z} \mathbf{W})^{-1} \mathbf{W}^T \mathbf{Z}. \tag{C4}$$

This operator resamples linear scale mixing ratio profiles into linear scale partial column averaged mixing ratio profiles. The respective total column operator $\boldsymbol{w^*}^T$ can be calculated in analogy to Eq. (C4) by replacing $\mathbf{W}^T$ by $\boldsymbol{w}^T$

With operator $\mathbf{W^*}^T$ we can calculate a coarse gridded partial column averaged state $\hat{\boldsymbol{x}}^*$ from the fine gridded linear mixing ratio state $\hat{\boldsymbol{x}}$ by:

$$\hat{\boldsymbol{x}}^* = \mathbf{W^*}^T \hat{\boldsymbol{x}}. \tag{C5}$$

The kernels matrix of the partial column averaged mixing ratio state can then be calculated from the fine gridded linear scale kernel matrix ($\mathbf{A}$) by:

$$\mathbf{A}^* = \mathbf{W}^{*T}\mathbf{A}. \tag{C6}$$

This kernel discribes how a change in the mixing ratio at a certain altitude affects the retrieved partial colunm averaged mixing ratio. Covariances of the partial column averaged mixing ratio state can be calculated from the corresponding covariance matrices of the fine gridded linear scale ($\mathbf{S}$) by:

$$\mathbf{S}^* = \mathbf{W}^{*T}\mathbf{S}\mathbf{W}^*. \tag{C7}$$

The respective calculations for total column averaged mixing ratios can be made by replacing $\mathbf{W}^{*T}$ by $\boldsymbol{w}^{*T}$. For the total column avereraged mixing ratios the covariance is a simple variance (the scalar S^*) and the kernel has the dimension $1 \times nol$, i.e. it is a row vector $\boldsymbol{a}^{*T}$.

The total column amount kernel ($\boldsymbol{a_T}^T$) provided with the TROPOMI data set can be converted into a total column averaged mixing ratio kernel $\boldsymbol{a_T^*}^T$ by the following calculation (using Eqs. (C3), (C4), and (C6)):

$$\boldsymbol{a_T^*}^T = \boldsymbol{w}^{*T}\mathbf{A_T} = (\boldsymbol{w}^T\mathbf{Z}\boldsymbol{w})^{-1}\boldsymbol{a_T}^T\mathbf{Z}. \tag{C8}$$

The total column averaged mixing ratio kernel $\boldsymbol{a_T^*}^T$ used in Sects. 2 and 3 is valid for the vertical grid used by the MUSICA IASI retrieval. It is calculated from the TROPOMI total column amount kernel ($\boldsymbol{a_T}^T$) provided in the TROPOMI output files according to Eq. (C8), after its interpolation onto the MUSICA IASI grid (see also Appendix C1).

Author contributions. Matthias Schneider developed the idea for the optimal a posteriori combination of two remote sensing products and he prepared the figures and the manuscript. Benjamin Ertl developed and performed the continuous MUSICA IASI data processing, where he was supported by Matthias Schneider, Christopher J. Diekmann, Farahnaz Khosrawi, Amelie N. Röhling, Omaira E. García, and Eliezer Sepúlveda. Frank Hase developed the PROFFIT-nadir retrieval code used for the MUSICA IASI processing. Tobias Borsdorff, Jochen Landgraf, and Alba Lorente are responsible for the TROPOMI processing and made TROPOMI data available. Huilin Chen and Rigel Kivi are responsible for the AirCore profile measurements over Sodankylä. Thomas Laemmel, Michel Ramonet, Cyril Crevoisier, and Jérome Pernin are responsible for the AirCore profile measurements over Trainou. Martin Steinbacher and Frank Meinhardt are responsible for the GAW data of Jungfraujoch and Schauinsland, respectively. Rigel Kivi, Darko Dubravica, Frank Hase, Voltaire A. Velazco, David W. T. Griffith, Nicholas M. Deutscher, and David F. Pollard are responsible for the TCCON data from Sodankylä, Karlsruhe, Burgos, Darwin, Wollongong, and Lauder. All authors supported the generation of the final version of this manuscript.

Competing interests. The authors declare that they have no conflict of interest