

The degree of freedom for signal assessment of measurement networks for joint chemical state and emission analysis

Xueran Wu^{1,2,3}, Hendrik Elbern^{1,2}, and Birgit Jacob³

¹IEK-8, Forschungszentrum Jülich, Wilhelm-Johnen-Straße, 52428, Jülich, Germany.

²Rhenish Institute for Environmental Research at University of Cologne, Aachener Straße 209, 50931, Cologne, Germany.

³Mathematics Department, University of Wuppertal, Gaußstraße 20, 42119, Wuppertal, Germany.

Correspondence to: Xueran Wu (x.wu@fz-juelich.de)

Abstract. The degree of freedom for signal (DFS) is generalized and applied to estimate the potential observability of observation networks for augmented model state and parameter optimisation problems. Control parameters may include prognostic state variables, mostly the initial values and insufficiently known model parameters. As for chemistry-transport models, emission rates are at least as important as initial values for model evolution control. Extending the optimization parameter set must be met by appropriate observation networks, which only allow for controlling the full optimization task, if the observability is satisfied. In this paper, we introduce an approach to quantify the degrees of freedom for signal jointly but individually for initial trace gas state and emission rates for transport-diffusion models forced by emissions. By applying a Kalman smoother, the quantitative assessment method on the efficiency of observation configurations to optimize the initial value and emission rates is developed based on the singular value decomposition. For practical applications, we derive an ensemble based version of the approach and give several elementary experiments.

1 Introduction

Air quality and climate change are influenced by the fluxes of green house gases, reactive gas emissions and aerosols. The temporal evolution of chemistry in the atmosphere is usually modelled by atmospheric chemistry transport models. poorly known initial values, sources and sinks are a serious problem for the quality of simulation, which can be addressed by data assimilation and inverse modelling. Parameter mis-specifications in a model can only be identified within data assimilation intervals of space-time methods, if the simulation is sufficiently sensitive and the error related observability of the measurement network is given. This poses the observability problem. Otherwise, the forecast degrades beyond the observation controlled period.

In practice data assimilation problems are typically solved in circumstances, where the number of observations are markedly lower than the model degree of freedom (Daley 1991). Consequently, when aiming to improve the quality of analysis by observation configurations, several aspects can be considered. These include (i) to optimize the observation network, subject to given constraints, (ii) to evaluate the value of individual or types of observations for the analyses, and (iii) to quantify the degree of which the analysis can be influenced by the observations, which is related to the sensitivity.

The observation network optimization problem (i) has been addressed traditionally by Observation System Simulation Experiments (OSSEs, e.g Daley 1991). The advanced concept of targeted observations has been popularized with the FASTEX campaign (e.g Szunyogh et al. 1999; Langland et al. 1999). Theoretical studies are presented, for example by Bishop et al. (1999); Berliner et al. (1998), or recently by Belsky et al. (2014) for highly nonlinear dynamics, and Wu et al. (2016) for the optimal locations of observations for time-varying system within a finite-time interval.

The benefit assessment (ii) of individual observations or types of measurements or remote sensing data within a given network seeks to identify the value ranking of information sources, accounting for analysis achieved. This problem has been investigated by Cardinali et al. (2004), Cardinali (2009), and a sequence of related papers, or Liu and Kalnay (2008), the latter without use of an adjoint model. A related approach was described by Baker and Daley (2000), who exploited sensitivities to observations to identify their spatial extensions of impact.

Finally, the need to quantify the information content provided by the observations (iii) can be satisfied by suitable and calculable measures, such as entropy reduction or DFS. The concept of DFS has been applied to satellite retrieval problems, typically of lower dimensions as compared to data assimilation (see for example Eyre 1990; Rodgers 2000; Rabier et al. 2002; Fourrié et al. 2003; Martynenko et al. 2010; Fisher 2003). However, these studies focused on the classical data assimilation problem with initial values or prognostic state variables as the only parameters to be optimized. Yet for chemistry transport or greenhouse gas models, which highly depend on the emissions in the troposphere, the optimization of the initial state is no longer the only issue. Rather, the optimization of emissions play an equally important role as initial values. In order to get better analysis from combining the model with observations, efforts of joint optimization by adding the emission rates to concentrations have been made (Elbern et al. 2000, 2007; Bocquet 2012; Bocquet and Sakov 2013; Miyazaki et al. 2012; Tang et al. 2011, 2013; Winiarek et al. 2014). Yet the lack of ability to observe and estimate surface emission fluxes directly is a major roadblock, hampering the progress in predictive skills of climate and atmospheric chemistry models. Therefore, the capacity to distinguish between the degree of freedom for signal of both emission rates and concentrations is crucial to assess the value of a measurement network. These assessment results are dependent not only on observation network and its deployment with respect to emission sources, but also influenced by the assimilation window lengths and the meteorological transport conditions.

In this context a meanwhile classical task is greenhouse gas inversion, aiming at the estimation of carbon dioxide, methane, and nitrous oxide sources, from which a rich set of literature emerged (e.g Peter et al. 2005). The sensitivity of the model evolution with respect to the model errors and the observation network as its detector is a key quantity to be analysed. Several methodologies have been proposed to account for model errors in both variational and ensemble data assimilation (e.g. Belsky et al. 2014; Gillijns and De Moor 2007; Li et al. 2009; Smith et al. 2013; Tremolet 2007; Daescu 2004; Zupanski 2007). Navon (1997) outlined the perceptibility and stability in optimal parameter estimation in meteorology and oceanography. Elbern et al. (2000, 2007) took the strong constraint by a given diurnal profile shape of emission rates such that their amplitudes in addition to initial values are the only parameters to be optimized by 4D-variational inversion. A general framework to optimize a set of parameters controlling the 4D-var data assimilation system was introduced by Cioaca and Sandu (2014) and applied to shallow model state and other parameters in a related paper (Cioaca and Sandu 2014).

Singular value decomposition (SVD) is a well-known tool applied to identifying the priorities of observations by detecting the fastest growing uncertainties in meteorological models (e.g Lorenz 1965; Buizza and Palmer 1995; Khattatov et al. 1999; Johnson 2003; Liao et al. 2006; Daescu 2008; Abida and Bocquet 2009; Kang and Xu 2012; Singh et al. 2012, 2013; Sandu et al. 2013). However, singular vector analysis and other methods for atmospheric chemistry with emissions are different since emissions play a similarly important role in forecast accuracy with initial values. Goris and Elbern (2013) used the singular vector decomposition to determine the sensitivity of the chemical composition to emissions and initial values for a variety of typical chemical scenarios and integration length. This methodology has been generalized for the 3-dimensional EURAD-IM (European Air pollution Dispersion-Inverse Model) and applied to a field campaign with airship borne measurements by Goris and Elbern (2015).

In this paper we introduce the practical implementation of this approach to identify the impacts of the observation networks for controlling transport diffusion models with the Kalman smoother as the appropriate data assimilation method. The focus is placed on presenting an approach to identify the spatially resolved potential and limits of measurement networks to optimize chemical states and emission rates jointly, by comparing their relative sensitivities. In section 2, we describe an atmospheric transport model extended with emission rates by establishing the dynamic model for emission rates in a novel way. In section 3, based on the Kalman smoother, we derive the theoretical approach to determine the degree of freedom for signal for both initial values and emission rates. In section 4, we develop the ensemble approach to evaluate the degree of freedom for signal of the model. In section 5, we present the approach to identify the sensitivity of observations by determining the directions of maximum perturbation growth to the initial perturbation. In section 6, we extend a 3D advection-diffusion equation with the dynamic model of the emission rate and give several elementary experiments to verify the approaches. In section 7, we conclude the main contributions of this paper and discuss possible extensions.

2 Atmospheric inverse modelling extended by emission rates

We usually describe the concentration change rate by the following prognostic atmospheric transport model

$$\frac{dc(t)}{dt} = \mathcal{A}(c) + e(t), \quad (1)$$

where \mathcal{A} is a nonlinear model operator, $c(t)$ and $e(t)$ are the state vector of chemical constituents and emission rates at time t , respectively .

The prior estimate of the state vector of concentrations $c(t)$ is given and denoted by $c_b(t)$, termed as the background state. The prior estimate of emission rates, usually taken from emission inventories, is denoted by $e_b(t)$.

Let \mathbf{A} be the tangent linear operator of \mathcal{A} , $\delta c(t_0) = c(t_0) - c_b(t_0)$ and $\delta e(t) = e(t) - e_b(t)$. The linear evolution of the perturbation of $c(t)$ follows the tangent linear model as

$$\frac{d\delta c}{dt} = \mathbf{A}\delta c + \delta e(t). \quad (2)$$

By the discretization of the tangent linear model in space, it is straightforward to obtain the linear solution of (2) discretized in space and continuous in time as

$$\delta c(t) = M(t, t_0)\delta c(t_0) + \int_{t_0}^t M(t, s)\delta e(s)ds, \quad (3)$$

where $M(\cdot, \cdot)$ is the resolvent obtained from the spacial discretization of \mathbf{A} . Without loss of generality, we assume $\delta c(t) \in \mathbf{R}^n$, $\delta e(t) \in \mathbf{R}^n$, where n is the dimension of the partial phase space of concentrations and emission rates. Obviously, $M(\cdot, \cdot) \in \mathbf{R}^{n \times n}$.

In addition, let $y(t)$ be the observation vector of $c(t)$ and define

$$\delta y(t) = y(t) - \mathcal{H}(c_b)(t), \quad (4)$$

where $\delta y(t) \in \mathbf{R}^{m(t)}$, $m(t)$ is the dimension of the phase space of observation configurations at time t . $\mathcal{H}(t)$ is a nonlinear forward observation operator mapping the model space to the observation space. Then by linearizing the nonlinear operator \mathcal{H} as H , we present the observation system as

$$\delta y(t) = H(t)\delta c(t) + \nu(t), \quad (5)$$

where the observation error $\nu(t)$ of the Gaussian distribution has zero mean and variance $R(t) \in \mathbf{R}^{m(t) \times m(t)}$.

The Kalman smoother is a recursive estimator to provide the best linear unbiased estimates (BLUE) of the unknown variables with error estimates, using a sequence of observations (e.g Gelb 1974). In addition to 4D-Var approaches, Kalman smoothers not only can provide the best linear unbiased estimate by a series of observations over time for the state vector, but also update the forecasting error covariances of that estimate. In this paper our problems will be treated by the Kalman smoother from a theoretical viewpoint.

If the initial state of concentration is the only parameter to be optimized, it is feasible to apply the Kalman filter and smoother by the tangent linear model (2) with observations (5) within the time interval $[t_0, t_N]$. However, as mentioned before, in most cases the exact values of emission rates are poorly known. It has been shown by Elbern et al. (2007) that the diurnal profiles of emission rates are better known and hence can be considered as constraints, such that the amplitudes of the diurnal emission cycle can be taken as optimization parameters. Thus, we first reformulate the background evolution of emission rates from time s to t into a dynamic form

$$e_b(t) = M_e(t, s)e_b(s), \quad s \leq t, \quad (6)$$

where $e_b(\cdot)$ is a n -dimensional vector, of which the i^{th} element is denoted by $e_b^i(\cdot)$ and $M_e(t, s)$ is the scaling diagonal matrix defined as

$$M_e(t, s) = \begin{pmatrix} \frac{e_b^1(t)}{e_b^1(s)} & & & & \\ & \frac{e_b^2(t)}{e_b^2(s)} & & & \\ & & \ddots & & \\ & & & \ddots & \\ & & & & \frac{e_b^n(t)}{e_b^n(s)} \end{pmatrix}. \quad (7)$$

In this works we establish the dynamic model of emission rates subject to the constraint

$$\delta e(t) = M_e(t, s)\delta e(s), \quad s \leq t. \quad (8)$$

It was stated in several studies (e.g. Catlin 1989; Gelb 1974) that the estimation of the variable x by the fix-interval Kalman smoother generally equals to the conditional expectation based on the observations within the entire time interval, denoted by

5 $\mathbf{E}[x|\{y(t_{obs}), t_{obs} \in [t_0, t_N]\}]$. With the emission model (8), the estimate of $e(t)$ by Kalman smoother on $[t_0, t_N]$ follows the linear property of the conditional expectation,

$$\begin{aligned} & \mathbf{E}[e(t)|\{y(t_{obs}), t_{obs} \in [t_0, t_N]\}] \\ &= \mathbf{E}[M_e(t, s)e(s)|\{y(t_{obs}), t_{obs} \in [t_0, t_N]\}] = M_e(t, s)\mathbf{E}[e(s)|\{y(t_{obs}), t_{obs} \in [t_0, t_N]\}]. \end{aligned} \quad (9)$$

It implies that BLUEs of emission rates with the dynamic model (8) by Kalman smoother preserve the proportioned diurnal profiles of their backgrounds.

By rewriting (3) as $\delta c(t) = M(t, t_0)\delta c(t_0) + \int_{t_0}^t M(t, s)M_e(s, t_0)\delta e(t_0)ds$, we obtain the transport model with the state vector extended by emission rates

$$\begin{pmatrix} \delta c(t) \\ \delta e(t) \end{pmatrix} = \begin{pmatrix} M(t, t_0) & \int_{t_0}^t M(t, s)M_e(s, t_0)ds \\ 0 & M_e(t, t_0) \end{pmatrix} \begin{pmatrix} \delta c(t_0) \\ \delta e(t_0) \end{pmatrix}. \quad (10)$$

Typically, there is no direct observation for emissions, apart from the flux tower observations used for carbon dioxide, which are not considered here. Therefore, we can reformulate the observation mapping as

$$\delta y(t) = (H(t), 0_{n \times n}) \begin{pmatrix} \delta c(t) \\ \delta e(t) \end{pmatrix} + \nu(t), \quad (11)$$

where $0_{n \times n}$ is a $n \times n$ matrix with zero elements.

It is now clear that both concentrations and emission rates are included in the state vector of the homogeneous model (10). It allows us to apply the Kalman smoother in a fixed time interval $[t_0, t_N]$ in order to optimize both parameters. Besides, in Appendix A, a more general case of the transport model extended by emission is shown.

3 The degree of freedom for signal of concentrations and emissions

In this section we will introduce the theoretical approach to determine the DFS of concentrations and emissions, resting on the extended model in Section 2. This approach gives us access to determine the efficiency of observations to optimize each variable based on the Kalman smoother within a finite-time interval.

25 For convenience, we generalize the atmospheric transport model (10) by the following discrete-time linear system on the time interval $[t_0, t_1, \dots, t_N]$:

$$x(t_{k+1}) = M(t_{k+1}, t_k)x(t_k) + \varepsilon(t_k), \quad (12)$$

$$y(t_k) = H(t_k)x(t_k) + \nu(t_k), \quad (13)$$

where $x(\cdot) \in \mathbf{R}^n$ is the state variable and $y(t_k) \in \mathbf{R}^{m(t_k)}$ is the observation vector at time t_k . The model error $\varepsilon(t_k)$ and the observation error $\nu(t_k)$, $k = 1, \dots, N$ of Gaussian distributions have zero means. The model error covariance matrix is denoted by $Q(t_k)$, while the observation error covariance matrix is denoted by $R(t_k)$.

- 5 We denote the BLUE of $x(t_i)$ based on $\{y(t_0), \dots, y(t_k)\}$ by $\hat{x}(t_i|t_k)$, $t_i, t_k \in [t_0, \dots, t_N]$. Especially, the prior estimation of $x(t_0)$ is denoted by $\hat{x}(t_0|t_{-1})$. Correspondingly, $P(t_i|t_k)$ is defined as the error covariance of $\hat{x}(t_i|t_k)$, $t_i \in [t_0, \dots, t_N]$ and $t_k \in [t_{-1}, \dots, t_N]$. It is known that the inverse of the analysis error covariance matrix at initial time, $P^{-1}(t_0|t_N)$ of a fixed-interval Kalman smoother is the optimal Hessian of the underlying cost function of 4D-Var (Li and Navon, 2001). Thus, we have

$$10 \quad P^{-1}(t_0|t_N) = P^{-1}(t_0|t_{-1}) + \sum_{i=0}^N M^\top(t_i, t_0) H^\top(t_i) R^{-1}(t_i) H(t_i) M(t_i, t_0). \quad (14)$$

It is clear that (14) comprises the information of the initial condition, model evolution, observation configurations and errors over the entire time interval $[t_0, \dots, t_N]$. At the same time, it is independent of any specific data and state vector, apart from the reference model evolution $M(\cdot, \cdot)$ needed for the linearization, as well as the observation operator $H(\cdot)$. Actually, if we define

$$\mathcal{G} = \begin{pmatrix} H(t_0)M(t_0, t_0) \\ H(t_1)M(t_1, t_0) \\ \vdots \\ H(t_N)M(t_N, t_0) \end{pmatrix}, \quad \mathcal{R}^{-1} = \begin{pmatrix} R^{-1}(t_0) & & & \\ & R^{-1}(t_1) & & \\ & & \ddots & \\ & & & R^{-1}(t_N) \end{pmatrix}, \quad (15)$$

- 15 we can rewrite (14) as

$$P^{-1}(t_0|t_N) = P^{-1}(t_0|t_{-1}) + \mathcal{G}^\top \mathcal{R}^{-1} \mathcal{G}, \quad (16)$$

where $\mathcal{G}^\top \mathcal{R}^{-1} \mathcal{G}$ is the observability Gramian with respect to \mathcal{R}^{-1} in control theory (Brockett, 1994). It represents the observation capacity of the observation networks with respect to the model.

- Though (16) meets the demand to represent the estimate covariance by all available information before starting the data assimilation procedure, it cannot be applied directly to evaluate the potential improvement of the estimate by the Kalman smoother, due to the lack of clear statistical significance of the inverse of a covariance matrix. We aspire a matrix, which allows us for a direct and normalized comparison between sensitivities to initial values and emission rates. To this end, we consider matrix \tilde{P} with the following form:

$$\tilde{P} = P^{-\frac{1}{2}}(t_0|t_{-1})(P(t_0|t_{-1}) - P(t_0|t_N))P^{-\frac{1}{2}}(t_0|t_{-1}) = I - P^{-\frac{1}{2}}(t_0|t_{-1})P(t_0|t_N)P^{-\frac{1}{2}}(t_0|t_{-1}), \quad (17)$$

- 25 where I is the identity matrix and $P^{\frac{1}{2}}(t_0|t_{-1})$ satisfies $P^{\frac{1}{2}}(t_0|t_{-1})P^{\frac{1}{2}}(t_0|t_{-1}) = P(t_0|t_{-1})$.

The matrix \tilde{P} is a normalized matrix of the difference between the background forecast error covariance matrix $P(t_0|t_{-1})$ and the analysis error covariance matrix $P(t_0|t_N)$, as inferred the Kalman smoother. It shows how much the observation

networks improve the estimation of model states and is the foundation matrix to study the DFS of models(Fisher 2003; Rodgers 2000; Singh et al. 2013).

Since $P(t_0|t_N)$ is unknown prior to the data assimilation procedure, we use (16) to rewrite \tilde{P} as

$$\begin{aligned}
\tilde{P} &= P^{-\frac{1}{2}}(t_0|t_{-1})(P(t_0|t_{-1}) - P(t_0|t_N))P^{-\frac{1}{2}}(t_0|t_{-1}) \\
&= P^{-\frac{1}{2}}(t_0|t_{-1})(P(t_0|t_{-1}) - (P^{-1}(t_0|t_{-1}) + \mathcal{G}^\top \mathcal{R}^{-1} \mathcal{G})^{-1})P^{-\frac{1}{2}}(t_0|t_{-1}) \\
&= I - P^{-\frac{1}{2}}(t_0|t_{-1})(P^{-1}(t_0|t_{-1}) + \mathcal{G}^\top \mathcal{R}^{-1} \mathcal{G})^{-1}P^{-\frac{1}{2}}(t_0|t_{-1}) \\
&= I - (I + P^{\frac{1}{2}}(t_0|t_{-1})\mathcal{G}^\top \mathcal{R}^{-1} \mathcal{G}P^{\frac{1}{2}}(t_0|t_{-1}))^{-1}.
\end{aligned} \tag{18}$$

It is worth noting that in (18)

$$I + P^{\frac{1}{2}}(t_0|t_{-1})\mathcal{G}^\top \mathcal{R}^{-1} \mathcal{G}P^{\frac{1}{2}}(t_0|t_{-1}) \tag{19}$$

is always invertible even if the observation Gramian $\mathcal{G}^\top \mathcal{G}$ is not full-rank. Thus, \tilde{P} is well-defined for all models with invertible initial covariance and observation systems with invertible error covariances within assimilation window t_0 to t_N . Then, we apply the singular value decomposition to simplify (18)

$$P^{\frac{1}{2}}(t_0|t_{-1})\mathcal{G}^\top \mathcal{R}^{-\frac{1}{2}} = VSU^\top, \tag{20}$$

where V and U are unitary matrices consisting of the left and right singular vectors, respectively, while S is the rectangular diagonal matrix consisting of the singular values.

Then, (18) can be simplified as

$$\begin{aligned}
\tilde{P} &= I - (I + P^{\frac{1}{2}}(t_0|t_{-1})\mathcal{G}^\top \mathcal{R}^{-1} \mathcal{G}P^{\frac{1}{2}}(t_0|t_{-1}))^{-1} \\
&= I - (I + VSS^\top V^\top)^{-1} \\
&= VV^\top - (VV^\top + VSS^\top V^\top)^{-1} \\
&= VV^\top - (V(I + SS^\top)V^\top)^{-1} \\
&= V(I - (I + SS^\top)^{-1})V^\top \\
&= \sum_{i=1}^r \frac{s_i^2}{1 + s_i^2} v_i v_i^\top,
\end{aligned} \tag{21}$$

where r is the rank of (18) and v_i is the i^{th} left singular vector in V related to the singular value s_i , which is the i^{th} element on the diagonal of S .

It is clear that the sum of the diagonal entries of \tilde{P} can be used to evaluate the total improvements of model states. Thus, the nuclear norm is appropriately taken as the metric, which is defined as

$$\|A\|_1 = \text{tr}(\sqrt{A^\top A}), \tag{22}$$

where A is any matrix and $\text{tr}(\cdot)$ denotes the trace of the matrix.

From (21), we obtain

$$\|\tilde{P}\|_1 = \text{tr}(\tilde{P}) = \sum_{i=1}^r \frac{s_i^2}{1+s_i^2}. \quad (23)$$

This is well-known as the degree of freedom for signal (DFS) of the model (e.g Rodgers 2000).

It is obvious that $\|\tilde{P}\|_1 < \|I\|_1 = n$. Here n can be considered as the total relative improvement if the system is definitely
5 observed. Thus, if we consider the ratio

$$\tilde{p} = \frac{\|\tilde{P}\|_1}{\|I\|_1} = \frac{\|\tilde{P}\|_1}{n} \in [0, 1), \quad (24)$$

the percentage of the total improvement of the model is obtained, which is called the *relative degree of freedom for signal*.

In order to get a deeper insight into the capacity of the observation networks to improve the estimation of all model states,
we consider the corresponding value in the diagonal of \tilde{P} as *the contribution of the degree of freedom for signal*. Denote the
10 j^{th} element on the diagonal of \tilde{P} by \tilde{P}_j , from (21), the contribution of the j^{th} element of $x(t_0)$ to the degree of freedom for
signal can be expressed as

$$\tilde{P}_j = \sum_{i=1}^r \frac{s_i^2}{1+s_i^2} (v_{ij})^2, \quad (25)$$

where v_{ij} is the j^{th} element of v_i .

Besides, we can see that Eqn. (21) enables us to discriminate the DFS contributed to different optimization parameters, that
15 is emission rates and initial values. Without loss of generality, we divide (21) into the following block matrix according to the
dimension of c and e

$$\tilde{P} = \begin{pmatrix} \tilde{P}^c & \tilde{P}^{ce} \\ \tilde{P}^{ec} & \tilde{P}^e \end{pmatrix} = \sum_{i=1}^{2n} \frac{s_i^2}{1+s_i^2} \begin{pmatrix} v_i^c \\ v_i^e \end{pmatrix} (v_i^{c\top}, v_i^{e\top}) \in \mathbf{R}^{2n \times 2n}, \quad (26)$$

where $(v_i^{c\top}, v_i^{e\top})^\top = v_i$.

It is easy to see that

$$20 \quad \tilde{P}^c = \sum_{i=1}^{2n} \frac{s_i^2}{1+s_i^2} v_i^c v_i^{c\top}, \quad \tilde{P}^e = \sum_{i=1}^{2n} \frac{s_i^2}{1+s_i^2} v_i^e v_i^{e\top}. \quad (27)$$

Further, the degree of freedom for signal of j^{th} element in $c(t_0)$ and $e(t_0)$ are given by

$$\tilde{P}_j^c = \sum_{i=1}^{2n} \frac{s_i^2}{1+s_i^2} (v_{ij}^c)^2, \quad \tilde{P}_j^e = \sum_{i=1}^{2n} \frac{s_i^2}{1+s_i^2} (v_{ij}^e)^2, \quad (28)$$

where v_{ij}^c and v_{ij}^e are the j^{th} elements of v_i^c and v_i^e respectively.

Moreover, the degree of freedom for signal of concentration $\|\tilde{P}^c\|_1$ and emission rates $\|\tilde{P}^e\|_1$ are calculated by

$$25 \quad \|\tilde{P}^c\|_1 = \sum_{i=1}^{2n} \frac{s_i^2}{1+s_i^2} \text{tr}(v_i^c v_i^{c\top}), \quad \|\tilde{P}^e\|_1 = \sum_{i=1}^{2n} \frac{s_i^2}{1+s_i^2} \text{tr}(v_i^e v_i^{e\top}). \quad (29)$$

It is worth noticing that

$$\tilde{P}^c = (P^c(t_0|t_{-1}))^{-\frac{1}{2}}(P^c(t_0|t_{-1}) - P^c(t_0|t_N))(P^c(t_0|t_{-1}))^{-\frac{1}{2}} \quad (30)$$

$$\tilde{P}^e = (P^e(t_0|t_{-1}))^{-\frac{1}{2}}(P^e(t_0|t_{-1}) - P^e(t_0|t_N))(P^e(t_0|t_{-1}))^{-\frac{1}{2}} \quad (31)$$

if and only if there is no prior correlation between the initial concentration and emission rates. In this case $P^{ce}(t_0|t_{-1}) = 0_{n \times n}$,

5 the corresponding relative degrees of freedom for signal of concentration and emission rates are defined as

$$\tilde{p}^c = \frac{\|\tilde{P}^c\|_1}{n}, \quad \tilde{p}^e = \frac{\|\tilde{P}^e\|_1}{n}. \quad (32)$$

From (24), it is obvious that $\tilde{p}^c \in [0, 1)$ and $\tilde{p}^e \in [0, 1)$ can be considered as the percentages of the relative improvements of concentration and emission rates, respectively. However, efficient observation networks probably lead to both of them are close to 1 such that

$$10 \quad \frac{\|\tilde{P}^c\|_1}{n} + \frac{\|\tilde{P}^e\|_1}{n} > 1. \quad (33)$$

It indicates the normalization of \tilde{P} is only with respect to the extended covariance matrix rather than specified to the state c and emission rates e . The relative degree of freedom for signal cannot serve our objective to distinguish the observability of concentration and emission rates. However, by observing the block form of \tilde{P} , we have

$$\|\tilde{P}^c\|_1 + \|\tilde{P}^e\|_1 = \|\tilde{P}\|_1. \quad (34)$$

15 Thus, in order to compare the improvements of the concentration and emission rates, we define *relative ratio of the degree of freedom for signal* for concentrations or emission rates as

$$\tilde{p}^c = \frac{\|\tilde{P}^c\|_1}{\|\tilde{P}\|_1}, \quad \tilde{p}^e = \frac{\|\tilde{P}^e\|_1}{\|\tilde{P}\|_1}, \quad \tilde{p}^e + \tilde{p}^c \equiv 1. \quad (35)$$

If the degree or relative degree of freedom for signal of the observation network and assimilation window is almost zero, an improvement cannot be expected. In contrast, $\{\tilde{P}_j^c\}_{j=1}^n$ and $\{\tilde{P}_j^e\}_{j=1}^n$, which show the improvement of each parameter j of concentrations and emission rates respectively, can help us determining which parameters can be optimized by the existing observation configurations. Furthermore, comparing \tilde{p}^c with \tilde{p}^e , we can conclude that the estimate of the one with the larger relative ratio of freedom for signal can be improved more efficiently by the existing observation configurations than the other. In other words, if $\tilde{p}^c > \tilde{p}^e$, the existing observation configurations are more efficient to the initial values of concentrations. Conversely, if $\tilde{p}^c < \tilde{p}^e$, the observation configurations can improve the estimate of emission rates more. According to \tilde{p}^c and 20 \tilde{p}^e , the “weights” between the concentrations and emission rates can be identified quantitatively. In a data assimilation context, where observations are in a weighted relation to the background, the BLUE favors those parameters with higher observation efficiency.

The special case that \tilde{p}^e is very close to zero implies that observation network is nearly “blind” for emission rate optimization.

4 The ensemble approach to determine the DFS

The ensemble Kalman smoother (EnKS), as a Monte Carlo implementation derived from the Kalman smoother, is suitable for problems with a large number of control variables and is a frequently applied tool in the field of data assimilation (Evensen, 2009). In this section we will introduce the ensemble-based case of the approach in Section 3.

- 5 For the discrete-time system (12), we denote the ensemble samples of $\hat{x}(t_i|t_j)$ by $\hat{x}_k(t_i|t_j)$, $i, j = 1, \dots, N$, $k = 1, \dots, q$, where q is the number of ensemble members.

Correspondingly, the ensemble means of $\hat{x}(t_i|t_j)$ is given by

$$\bar{x}(t_i|t_j) = \frac{1}{q} X(t_i|t_j) \mathbf{1}_{q \times 1}, \quad (36)$$

- where $X(t_i|t_j) = (\hat{x}_1(t_i|t_j), \hat{x}_2(t_i|t_j), \dots, \hat{x}_q(t_i|t_j))$ is the $n \times q$ ensemble matrix, $\mathbf{1}_{i \times j}$ is a $i \times j$ matrix of which each element is equal to 1.

We calculate the ensemble forecast and analysis covariances as

$$\bar{P}(t_i|t_j) = \frac{1}{q-1} \tilde{X}(t_i|t_j) \tilde{X}^\top(t_i|t_j), \quad (37)$$

where $\tilde{X}(t_i|t_j) = X(t_i|t_j) - \frac{1}{q} X(t_i|t_j) \mathbf{1}_{q \times q}$ is the related perturbation matrix. We define the ensemble observation configurations in the entire assimilation window as

$$15 \quad y_k^f = \mathcal{G} \hat{x}_k(t_0|t_{-1}), \quad k = 1, \dots, q. \quad (38)$$

Further, the ensemble mean and the forecasting error covariance matrix of the ensemble observation configurations are given by

$$\bar{y}^f = \frac{1}{q} \sum_{k=1}^q y_k^f, \quad \bar{P}_{yy}^f = \frac{1}{q-1} \sum_{k=1}^q (\hat{y}_k^f - \bar{y}^f)(\hat{y}_k^f - \bar{y}^f)^\top = \mathcal{G} \bar{P}(t_0|t_{-1}) \mathcal{G}^\top. \quad (39)$$

Similarly, we denote the ensemble covariance between the initial states and the forecasting observations by

$$20 \quad \bar{P}_{xy}^f = \frac{1}{q-1} \sum_{k=1}^q (\hat{x}_k(t_0|t_{-1}) - \bar{x}(t_0|t_{-1}))(\hat{y}_k^f - \bar{y}^f)^\top = \bar{P}(t_0|t_{-1}) \mathcal{G}^\top. \quad (40)$$

Furthermore, defining the ensemble observations as

$$\hat{y}_k(t_i) = y(t_i) + \nu_k(t_i), \quad k = 1, \dots, q, \quad i = 1, \dots, N, \quad (41)$$

we assume $\bar{\nu}(t_i) = \frac{1}{q} \sum_{k=1}^q \nu_k(t_i) = 0$, $\bar{R}(t_i) = \frac{1}{q-1} \sum_{k=1}^q \nu_k(t_i) \nu_k^\top(t_i)$ and $\bar{\mathcal{R}}^{-1}$ is the block diagonal matrix with the diagonal $(\bar{R}^{-1}(t_0), \dots, \bar{R}^{-1}(t_N))$.

- 25 It is shown by Evensen (2009) that the ensemble forecast and analysis covariances have the same form with the covariances in the standard Kalman filter. However, the ensemble size q is significantly less than the dimension of the model n in practical applications. It causes that the initial ensemble covariance $\bar{P}(t_0|t_{-1})$ is not invertible. In this case, the pseudo inverse is a

widely used alternative of the inverse of a matrix, due to its best fitness and uniqueness. We denote the pseudo inverse of a matrix A by A^\dagger . Then for the initial ensemble covariance

$$\bar{P}(t_0|t_{-1}) = \frac{1}{q-1} \tilde{X}(t_0|t_{-1}) \tilde{X}^\top(t_0|t_{-1}), \quad (42)$$

we apply the singular value decomposition to

$$5 \quad \frac{1}{\sqrt{q-1}} \tilde{X}(t_0|t_{-1}) = V_0 S_0 U_0^\top, \quad (43)$$

where $V_0 \in \mathbf{R}^{n \times n}$ and $U_0 \in \mathbf{R}^{q \times q}$ consist of the left and right singular vectors respectively, and $S_0 \in \mathbf{R}^{n \times q}$ is a rectangular diagonal matrix with singular values $\{s_{0i} | s_{0i} \geq 0\}_{i=1}^q$ on its diagonal. Thus,

$$\bar{P}(t_0|t_{-1}) = V_0 S_0 U_0^\top U_0 S_0^\top V_0^\top = V_0 S_0 S_0^\top V_0^\top = V_0 \hat{S}_0^2 V_0^\top, \quad (44)$$

where $\hat{S}_0^2 = S_0 S_0^\top \in \mathbf{R}^{n \times n}$ is a block diagonal matrix with the diagonal $(s_{01}^2, \dots, s_{0r_0}^2, \mathbf{0}_{1 \times (n-r_0)})$, r_0 is the rank of S_0 . Hence,

10 we find a pseudo inverse

$$\bar{P}^{\dagger \frac{1}{2}}(t_0|t_{-1}) = V_0 \hat{S}_0^\dagger V_0^\top, \quad (45)$$

where \hat{S}_0^\dagger is the pseudo inverse of \hat{S}_0 with the diagonal $(1/s_{01}, \dots, 1/s_{0r_0}, \mathbf{0}_{1 \times (n-r_0)})$. Analog to (17), we define \bar{P} as

$$\bar{P} = \bar{P}^{\dagger \frac{1}{2}}(t_0|t_{-1}) (\bar{P}(t_0|t_{-1}) - \bar{P}(t_0|t_N)) \bar{P}^{\dagger \frac{1}{2}}(t_0|t_{-1}). \quad (46)$$

Likewise, corresponding to (13), we present the observation system in the entire time interval as

$$15 \quad y = \mathcal{G}x(t_0) + \nu, \quad (47)$$

where $y = (y^\top(t_0), \dots, y^\top(t_N))^\top$, $\nu = (\nu^\top(t_0), \dots, \nu^\top(t_N))^\top$ and \mathcal{G} as the observation configuration for $x(t_0)$. Then, for the analysis error covariance matrix, we obtain

$$\begin{aligned} & \bar{P}(t_0|t_N) \\ &= \bar{P}(t_0|t_{-1}) - \bar{P}(t_0|t_{-1}) \mathcal{G}^\top (\mathcal{G} \bar{P}(t_0|t_{-1}) \mathcal{G}^\top + \mathcal{R})^{-1} \mathcal{G} \bar{P}(t_0|t_{-1}) \\ 20 \quad &= \bar{P}(t_0|t_{-1}) - \bar{P}(t_0|t_{-1}) \mathcal{G}^\top \mathcal{R}^{-\frac{1}{2}} (I + \mathcal{R}^{-\frac{1}{2}} \mathcal{G} \bar{P}(t_0|t_{-1}) \mathcal{G}^\top \mathcal{R}^{-\frac{1}{2}})^{-1} \mathcal{R}^{-\frac{1}{2}} \mathcal{G} \bar{P}(t_0|t_{-1}) \\ &= \bar{P}(t_0|t_{-1}) - \bar{P}_{xy}^f \mathcal{R}^{-\frac{1}{2}} (I + \mathcal{R}^{-\frac{1}{2}} \bar{P}_{yy}^f \mathcal{R}^{-\frac{1}{2}})^{-1} \mathcal{R}^{-\frac{1}{2}} (\bar{P}_{xy}^f)^\top. \end{aligned} \quad (48)$$

Further, analog to (21), we obtain

$$\begin{aligned} \bar{P} &= \bar{P}^{\dagger \frac{1}{2}}(t_0|t_{-1}) (\bar{P}(t_0|t_{-1}) - \bar{P}(t_0|t_N)) \bar{P}^{\dagger \frac{1}{2}}(t_0|t_{-1}) \\ &= \bar{P}^{\dagger \frac{1}{2}}(t_0|t_{-1}) \bar{P}_{xy}^f \mathcal{R}^{-\frac{1}{2}} (I + \mathcal{R}^{-\frac{1}{2}} \bar{P}_{yy}^f \mathcal{R}^{-\frac{1}{2}})^{-1} \mathcal{R}^{-\frac{1}{2}} (\bar{P}_{xy}^f)^\top \bar{P}^{\dagger \frac{1}{2}}(t_0|t_{-1}). \end{aligned} \quad (49)$$

25 Let $\sum_{i=1}^N m(t_i) = m$ be the number of observations available within the assimilation window. To proceed with (49), we apply again the singular value decomposition into the following matrix

$$\bar{P}^{\dagger \frac{1}{2}}(t_0|t_{-1}) \bar{P}_{xy}^f \mathcal{R}^{-\frac{1}{2}} = \bar{V} \bar{S} \bar{U}^\top \in \mathbf{R}^{n \times m}, \quad (50)$$

where $\bar{U} \in \mathbf{R}^{m \times m}$ and $\bar{V} \in \mathbf{R}^{n \times n}$ consists of the eigenvectors of $\mathcal{R}^{-\frac{1}{2}} \mathcal{G} \bar{P}(t_0|t_{-1}) \mathcal{G}^\top \mathcal{R}^{-\frac{1}{2}}$ and $\bar{P}^{\frac{1}{2}}(t_0|t_{-1}) \mathcal{G}^\top \mathcal{R}^{-1} \mathcal{G} \bar{P}^{\frac{1}{2}}(t_0|t_{-1})$, respectively. $S \in \mathbf{R}^{n \times m}$ consists of the singular values on its diagonal.

We denote the rank of (50) by r . Then, we rewrite \bar{P} as

$$\begin{aligned}
\bar{P} &= \bar{P}^{\dagger \frac{1}{2}}(t_0|t_{-1})(\bar{P}(t_0|t_{-1}) - \bar{P}(t_0|t_N))\bar{P}^{\dagger \frac{1}{2}}(t_0|t_{-1}) \\
5 \quad &= \bar{V} \bar{S}^\top \bar{U}^\top (\bar{U} \bar{U}^\top + \bar{U} (\bar{S} \bar{S}^\top) \bar{U}^\top)^{-1} \bar{U} \bar{S} \bar{V}^\top \\
&= \bar{V} \bar{S}^\top (I + \bar{S}^\top \bar{S})^{-1} \bar{S} \bar{V}^\top \\
&= \sum_{i=1}^r \frac{\bar{s}_i^2}{1 + \bar{s}_i^2} \bar{v}_i \bar{v}_i^\top.
\end{aligned} \tag{51}$$

where r is the rank of \bar{P} and \bar{v}_i is the i^{th} left singular vector in V related to the singular value \bar{s}_i , which is the i^{th} element on the diagonal of S .

10 We observe that (51) and (21) have a similar form. By virtue of

$$\bar{P}^{\dagger \frac{1}{2}}(t_0|t_{-1}) \bar{P}_{xy}^f \bar{\mathcal{R}}^{-\frac{1}{2}} = \bar{P}^{\frac{1}{2}}(t_0|t_{-1}) \mathcal{G}^\top \bar{\mathcal{R}}^{-\frac{1}{2}}, \tag{52}$$

the final results of (21) and (51) are equivalent. However, compared with $P^{\frac{1}{2}}(t_0|t_{-1}) \mathcal{G}^\top \mathcal{R}^{-\frac{1}{2}}$, the ensemble expression $\bar{P}^{\dagger \frac{1}{2}}(t_0|t_{-1}) \bar{P}_{xy}^f \bar{\mathcal{R}}^{-\frac{1}{2}}$ processes the absolute advantage that in the calculation of \bar{P}_{xy}^f since we do not need the explicit form of \mathcal{G} . It allows us to code it line by line such that our approach is computationally more efficient.

15 Analog to the standard case, we can similarly define *the ensemble degree of freedom for signal* (EnDFS) as $\|\bar{P}\|_1$ and consider each element on the diagonal of \bar{P} as *the contribution to EnDFS* of the corresponding model state.

Since $\bar{P}(t_0|t_{-1})$ is typically not full rank,

$$\begin{aligned}
&\bar{P}^{\dagger \frac{1}{2}}(t_0|t_{-1})(\bar{P}(t_0|t_{-1}) - \bar{P}(t_0|t_N))\bar{P}^{\dagger \frac{1}{2}}(t_0|t_{-1}) \\
&= \bar{P}^{\dagger \frac{1}{2}}(t_0|t_{-1})\bar{P}(t_0|t_{-1})\bar{P}^{\dagger \frac{1}{2}}(t_0|t_{-1}) - \bar{P}^{\dagger \frac{1}{2}}(t_0|t_{-1})\bar{P}(t_0|t_N)\bar{P}^{\dagger \frac{1}{2}}(t_0|t_{-1}) \\
20 \quad &= V_0 \hat{S}_0^\dagger V_0^\top (V_0 \hat{S}_0^2 V_0^\top) V_0 \hat{S}_0^\dagger V_0^\top - \bar{P}^{\dagger \frac{1}{2}}(t_0|t_{-1})\bar{P}(t_0|t_N)\bar{P}^{\dagger \frac{1}{2}}(t_0|t_{-1}) \\
&= V_0 I_{r_0} V_0^\top - \bar{P}^{\dagger \frac{1}{2}}(t_0|t_{-1})\bar{P}(t_0|t_N)\bar{P}^{\dagger \frac{1}{2}}(t_0|t_{-1}),
\end{aligned} \tag{53}$$

where I_{r_0} is the diagonal matrix with the diagonal $(\mathbf{1}_{1 \times r_0}, \mathbf{0}_{1 \times (n-r_0)})$. It is clear from (48) that $\bar{P}^{\dagger \frac{1}{2}}(t_0|t_{-1})\bar{P}(t_0|t_N)\bar{P}^{\dagger \frac{1}{2}}(t_0|t_{-1})$ is still a nonnegative definite matrix.

Thus, the *ensemble relative degree of freedom for signal* (EnRDFS) is defined by

$$25 \quad \bar{p} = \frac{\|\bar{P}\|_1}{\|I_{r_0}\|_1} = \frac{\|\bar{P}\|_1}{r_0} \in [0, 1). \tag{54}$$

In order to distinguish the improvements for concentrations and emission rates, the *ensemble relative ratios of DFS* remain

$$\bar{p}^c = \frac{\|\bar{P}^c\|_1}{\|\bar{P}\|_1}, \quad \bar{p}^e = \frac{\|\bar{P}^e\|_1}{\|\bar{P}\|_1}. \tag{55}$$

If we further consider the nonlinear dynamic model, we can renew the definition of the forecasting observation configurations as

$$30 \quad y_k^f = \mathcal{G}(\hat{x}_k(t_0|t_{-1})), \quad k = 1, \dots, q, \tag{56}$$

such that it can follow the nonlinear model, where \mathcal{G} is again a combined model-observation nonlinear operator.

Correspondingly, the ensemble mean of \bar{y}_k^f and \bar{P}_{xy}^f can be calculated based on (56) with the nonlinear \mathcal{G} . Thus, the above ensemble-based approach is available for nonlinear models.

5 Sensitivity of observation networks

5 The above discussion about DFS aims to evaluate a predefined measurement network on its potential to analyze initial values and emission rates simultaneously. In Appendix B, independent of any concrete data assimilation method, we use the singular vector approach to identify the sensitive directions of observation networks to initial values and emission rates and show the association between the efficiency and sensitivity of observation networks.

From Appendix B, we can see that the singular value s_k shows the amplification of the impact of the initial state to the observation configurations during the entire time interval. The associated singular vector in the state space v_k is the direction of k^{th} growth of the perturbation of observations evolved from the initial perturbation. With the special choice $W_0 = P^{-1}(t_0|t_{-1})$ and $\mathcal{W} = \mathcal{R}^{-1}$ we compare the sensitivity analysis with the analysis in Section 3. Since $\frac{s_k^2}{1+s_k^2}$ are decreasing with the decrease of s_k , $k = 1, \dots, n$, it is clear that v_k is also the k^{th} direction which maximizes the relative improvement of estimates based on Kalman smoother. It indicates that the states contributing to DFS more are the same with the states more sensitive to the observation networks. Besides, the leading singular value s_1 is related to the operator norm of \tilde{P} as

$$\|\tilde{P}\| = \max_{\|x\|=1} \|\tilde{P}x\| = \frac{s_1^2}{1+s_1^2}, \quad (57)$$

which implies the upper boundedness of \tilde{P} . It gives us an access to approximate and target the sensitive parameters or areas with the metric of the leading singular vectors weighted by the corresponding singular values.

Moreover, due to the homogeneity of the atmospheric transport model state vector extended with emissions, the above sensitivity analysis can be easily applied by dividing singular vectors into the block form according to the dimensions of the initial state and emissions. The corresponding block parts of different singular vectors indicate the different sensitive directions of the initial state and emissions and allow for this relative quantification. Correspondingly, we can approximate and target the parameters sensitive to the existing observation networks for both initial values and emission rates.

6 Example

25 We consider a linear advection-diffusion model with Dirichlet horizontal boundary condition and Neumann boundary condition in the vertical direction on the domain $[0, 14] \times [0, 14] \times [0, 4]$,

$$\frac{\partial \delta c}{\partial t} = -v_x \frac{\partial \delta c}{\partial x} - v_y \frac{\partial \delta c}{\partial y} + \frac{\partial}{\partial z} \left(K(z) \frac{\partial \delta c}{\partial z} \right) + \delta e, \quad (58)$$

where δc , δe are the perturbations of the concentration, the emission rate of a species respectively. For vertical diffusion, $K(z)$ is a differentiable function of height z .

For velocity $v_x = v_y = 0.5$ and the time step $\Delta t = 0.5$, the numerical solution is based on the symmetric operator splitting technique (Yanenko, 1971) with the following operator sequence

$$\delta c(t + \Delta t) = T_x T_y D_z A D_z T_y T_x \delta c(t), \quad (59)$$

where T_x and T_y are transport operators in horizontal directions x and y , D_z is the diffusion operator in vertical direction z .

5 The parameters of emission and deposition rates are included in A . The Lax-Wendroff algorithm is chosen as the discretization method for horizontal advection with $\Delta x = \Delta y = 1$. The vertical diffusion is discretized with $\Delta z = 1$ by Crank-Nicolson scheme with the Thomas algorithm (Higham, 2002) as solver. The number of the grid points is $N_g = 1125$.

With the same temporal and spacial discretization of the concentration, the background knowledge of the emission rate is given by $e_b(t_n, i, j, l)$, where $n = 1, \dots, N$. We establish the discrete dynamic model of the emission rate according to (8)

$$10 \quad \delta e(t_{n+1}) = M_e(t_{n+1}, t_n) \delta e(t_n), \quad n = 1, \dots, N, \quad (60)$$

where $M_e(t_{n+1}, t_n) = e_b(t_{n+1})/e_b(t_n)$.

For expository reasons we assume δd is a constant over time and the only one fixed observation configuration is time-invariant. It indicates that the observation operator mapping the state space to the observation space is a $1 \times 2N_g$ time-invariant matrix.

15 In our simulations, we produce $q = 500$ (the ensemble number) samples for the initial concentration and emission rate respectively by pseudo independent random numbers and make the states correlated by the moving average technique. It has been tested that the computation cost of our approach is linearly increasing with the number of ensembles. In the following, we present three different tests, aiming to demonstrate the roles of variable winds, emissions, and vertical diffusion.

Advection test: The following part demonstrates the application of the DFS analysis tool by basic examples, designed to 20 show the expected elementary outcomes of the following situations, which exhibit the effects of assimilation window length in relation to emission location: these include (i) an assimilation window is too short to capture emission impacts at the observation site, (ii) an extended assimilation window with balanced signal of impacts of concentrations and emissions at the observation site, and (iii) a further increased assimilation window features a declining impact of initial values and growing emission impact. The first elementary advection test (Fig. 1 to Fig. 7) identifies the sensitivities of parameters subject to different wind direction 25 and data assimilation window (DAW) through the DFS. Focusing on the advection effects, we apply the model with a weak diffusion process ($K(z) = 0.5e^{-z^2}$).

In Fig. 1 to 3 we assume southwesterly winds and the assigned data assimilation windows are $10\Delta t$, $35\Delta t$ and $48\Delta t$ respectively. The computation times are approximately 8.1s, 28.5s and 39.4s in our tests with the above three different assimilation windows, from which we can verify that the computation cost is nearly linearly increasing with the length of data assimilation 30 window. The contributions to EnDFS of the initial states are shown in the left panels of Fig. 1 to 3. We can find that the horizontal fields at lowest layer ($z = 0$) where the estimates of the concentration are probably improved is enlarged with the extension of data assimilation windows. It is because that more and more grid points of the concentration are correlated with longer data assimilation windows.

The right panels of Fig. 1 to 3 show the EnDFS of the emission rate at each grid point with $z = 0$. From Fig. 1, we can observe that the EnDFS of the emission rate are smaller than the case of initial value in the influenced area. It indicates that the observations cannot detect the emission rate within $10\Delta t$ data assimilation window. Thus, in this case initial values alone can be optimized. It is shown in the right panels of Fig. 2 and Fig. 3 that the emission rate plays a more and more important role on the impact of observations. In this two cases, we consider both the concentration and emission rate as optimized parameters. The quantitative balance between the concentration and emission rate is provided in Table 1.

Fig. 4 exhibits in its upper row panels the singular values underlying the results shown in Fig. 1 to 3. We approximate the sensitivities of the initial concentrations by the first five leading singular vectors weighted by the associated singular values in the nuclear norm and show the results in the three panels of Fig. 4, lower row. It is clearly visible that the sensitive area can be well targeted by only few singular vectors, although the sensitivity analysis cannot provide the quantitative solutions with a clear statistical significance as the the degree of freedom for signal of the model. The areas of influence to the measurement site in dependence of wind direction and assimilation window lengths is clearly visualized, corresponding to expectations.

As counter example, Fig. 5 to 7 also show the EnDFS of the concentration and emission rate under the same assumptions as Fig. 1 to 3 respectively, except that northeasterly wind is assumed. Clearly, with the northeasterly wind, whatever the duration of the assimilation window is, the emission is not detectable and improvable by that particular observation configuration. This hypothesis is demonstrated by our method. The quantitative balances are exposed in Table 1 for the related figures, where the insensitivity to emission rate optimisation remains equally low and induced by numerical noise. .

Emission signal test: The purpose of emission signal test (Fig. 8 and Fig. 9) is to assess the impact of observation configurations to the emission rates evolved with different diurnal profiles. We have the same assumptions as Fig. 3 except the wind speed in Fig. 8 and Fig. 9 is accelerated such that the profiles of the emission rate is better detectable as to the observation. The only distinction between the situations in Fig. 8 and Fig. 9 is the pronounced diurnal cycle background profile of the emission rate during the assimilation window $48\Delta t$, schematically simulating a rush hour induced source. Since the profiles of emission rates are correlated with the emitted amount of that species during the data assimilation window, it is clearly shown in Table 1 that the distinct variation of the emission rate during the data assimilation window acts to level \bar{p}^c and \bar{p}^e , and thus helps to improve the estimates of source.

Diffusion test: The diffusion test (Fig. 10 to Fig. 12) aims to test the approach via comparing the EnDFS of the concentration and the emission rate at the layer $z = 0$ with a weak diffusion process and a strong diffusion process. We assume the observation configuration at each time step is located at $(12, 10, 4)$ in the diffusion test , with $K(z) = 0.5e^{-z^2}$ in Fig. 10 and $K(z) = 0.5e^{-z^2} + 1$ in Fig. 11. Besides, Fig. 10 and Fig. 12 preserve the same assumptions with Fig. 3.

It is obviously seen from Fig. 3 and Fig. 10 that the different observation locations strongly influence on the distribution of the concentration. Table 2 shows that with the same diffusion coefficient the degree ensemble degree of freedom for signal of the concentration in the lowest layer in Fig. 3 is definitely larger than the one for Fig. 10. Moreover, it can be seen from Table 1 that the observation configuration at the top layer is not efficient to the emission rate with such weak diffusion within $48\Delta t$ data assimilation window .

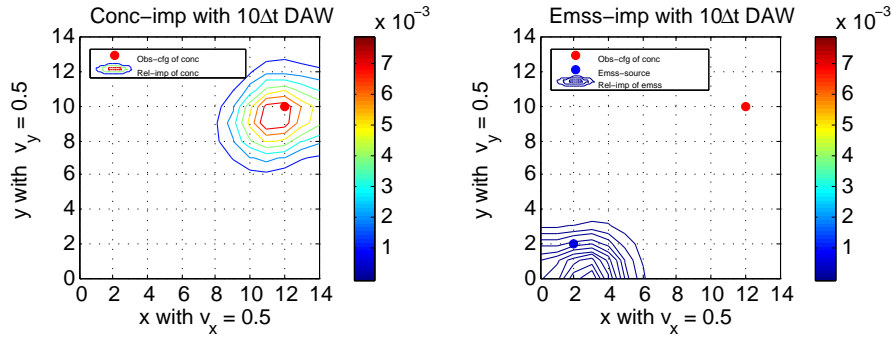


Figure 1. Advection test with $10\Delta t$ DAW and southwesterly wind. Isoleths of ensemble relative improvements of the concentration and emission rate are shown in the left and right figure panels respectively. The point located at (12, 10, 0) named as ‘Obs-cfg of conc’ shows the invariant observation configuration. The point located at (2, 2, 0) named as ‘Emss-source’ is the source of the emission rate.

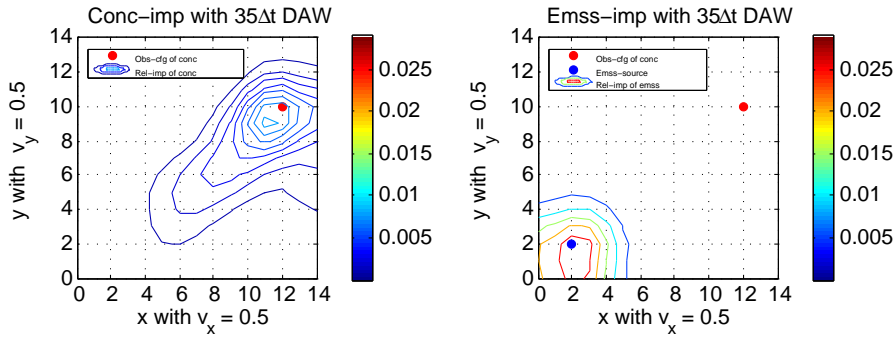


Figure 2. Advection test with $35\Delta t$ DAW and southwesterly wind. Plotting conventions are as in Fig. 1.

Comparing Fig. 10 with Fig. 11, we can find that the EnDFS of concentration and emission rate increase with the stronger diffusion process. The increasing impact of the observation configuration with the stronger diffusion is also verified by the EnDFS and ensemble relative ratios of DFS of the concentration and emission rate for Fig. 10 and 11 in Table 2. The balances between the concentration and emission rate for Fig. 10 and 11 are shown in Table 1. The significant difference of the ‘weight’ of emission rate in Table 1 implies that the observation configuration cannot detect the emission at the lowest layer with such a weak diffusion in Fig. 10 and with the stronger in Fig. 11 both the concentration and emission rate should be considered as optimized parameters with the corresponding ‘weights’.

Finally, similar to Fig. 4, the singular values of Fig. 10, Fig. 11 and the approximating targeting results of sensitive parameters are shown in Fig. 12.

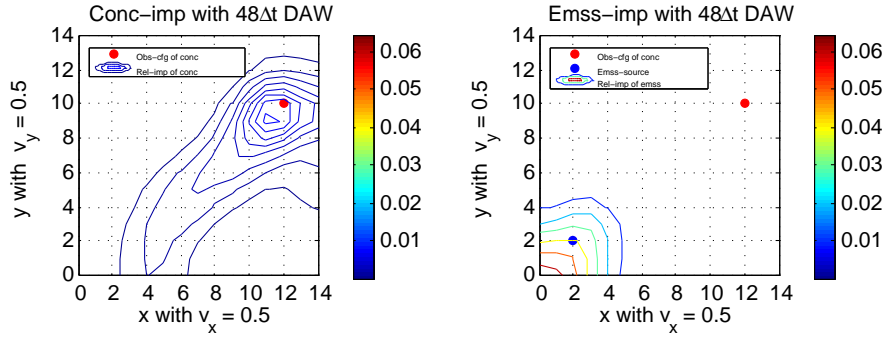


Figure 3. Advection test with $48\Delta t$ DAW and southwesterly wind. Plotting conventions are as in Fig. 1.

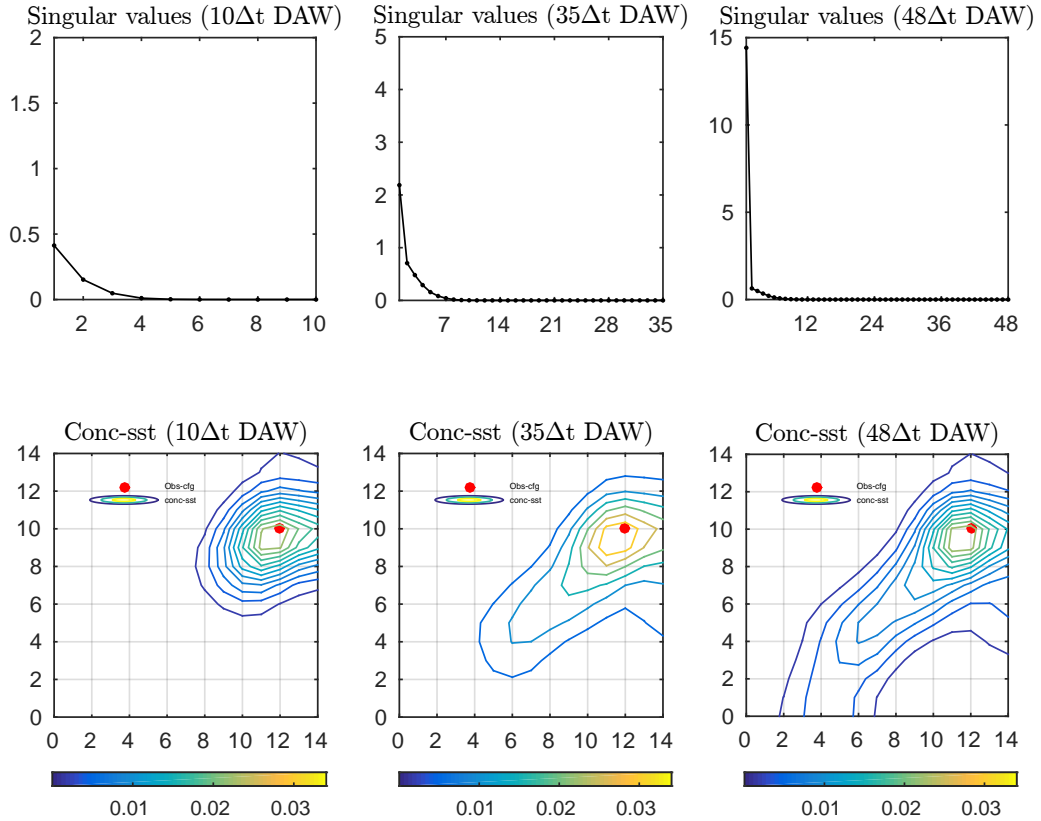


Figure 4. Singular values of Fig. 1 ~ 3 and sensitivities of initial states approximated by 5 leading singular values.

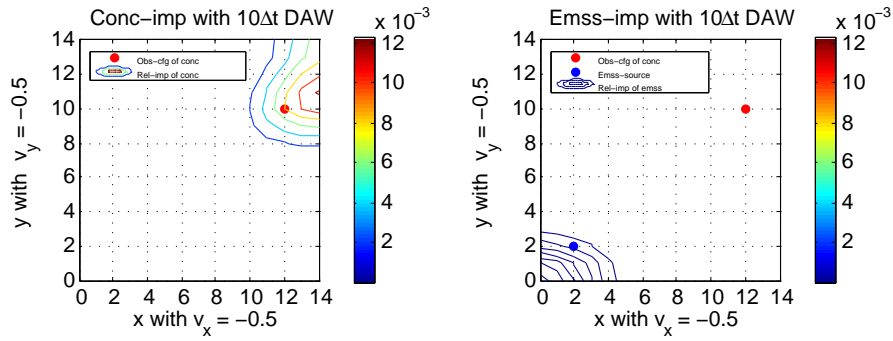


Figure 5. Advection test with $10\Delta t$ DAW and northeasterly wind. Plotting conventions are as in Fig. 1.

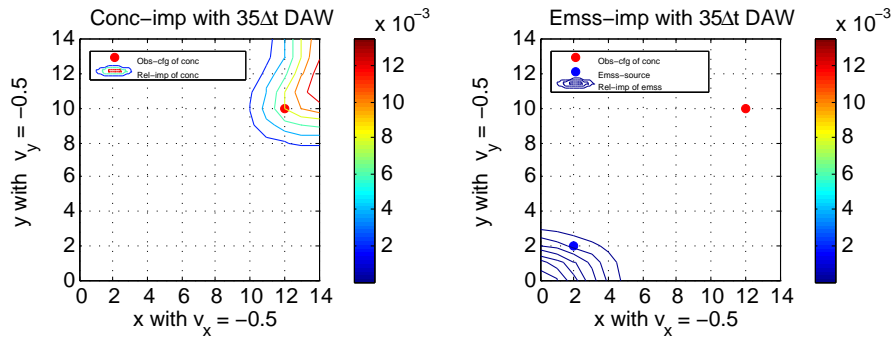


Figure 6. Advection test with $35\Delta t$ DAW and northeasterly wind. Plotting conventions are as in Fig. 1.

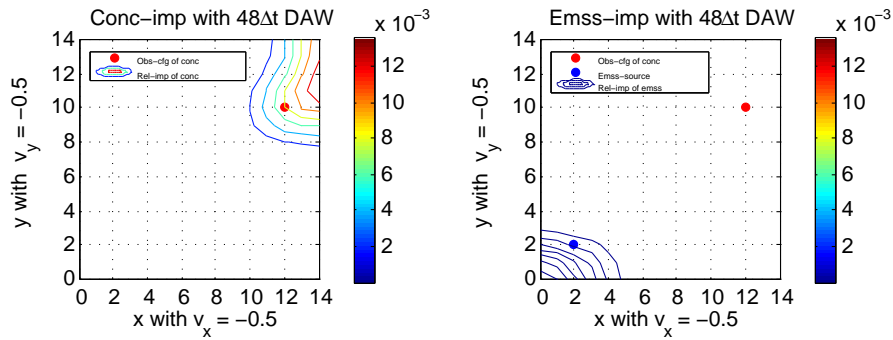


Figure 7. Advection test with $48\Delta t$ DAW and northeasterly wind. Plotting conventions are as in Fig. 1.

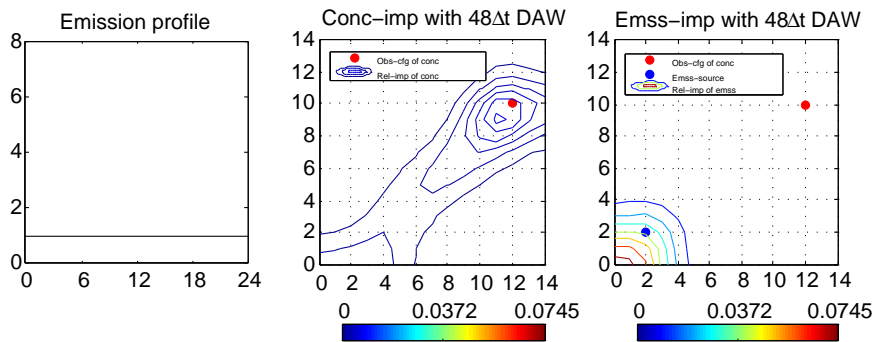


Figure 8. Emission signal test (weak) with $35\Delta t$ DAW and southwesterly wind ($v_x = 1$ and $v_y = 1$). Plotting conventions are as in Fig. 1.

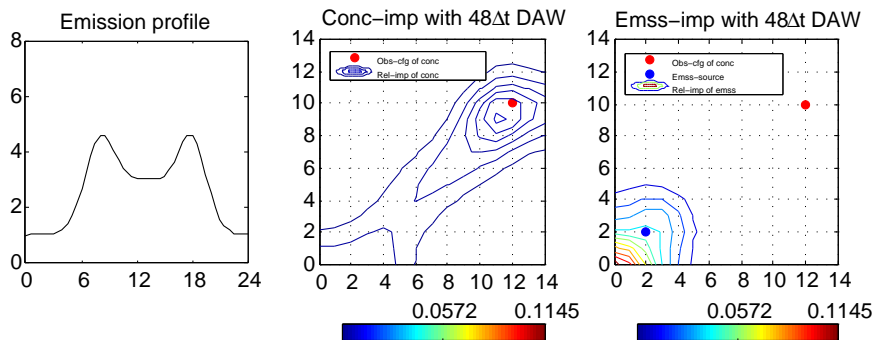


Figure 9. Emission signal test (strong) with $48\Delta t$ DAW and southwest wind ($v_x = 1$ and $v_y = 1$). Plotting conventions are as in Fig. 1.

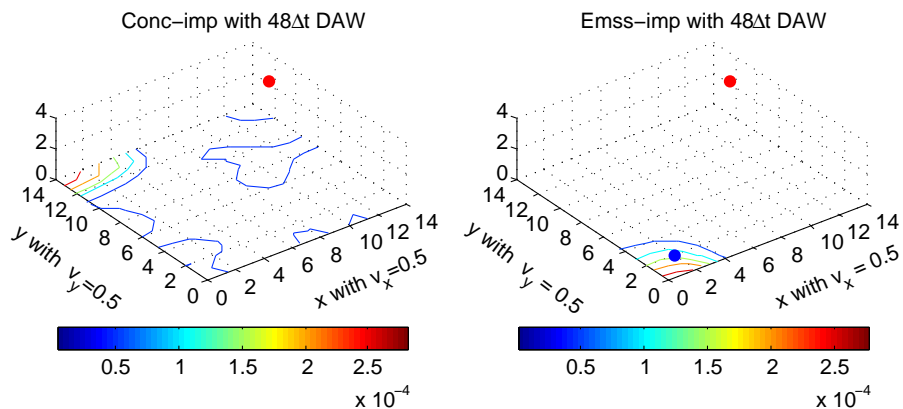


Figure 10. Diffusion test (weak) with $48\Delta t$ DAW and southwesterly wind. Plotting conventions are as in Fig. 1.

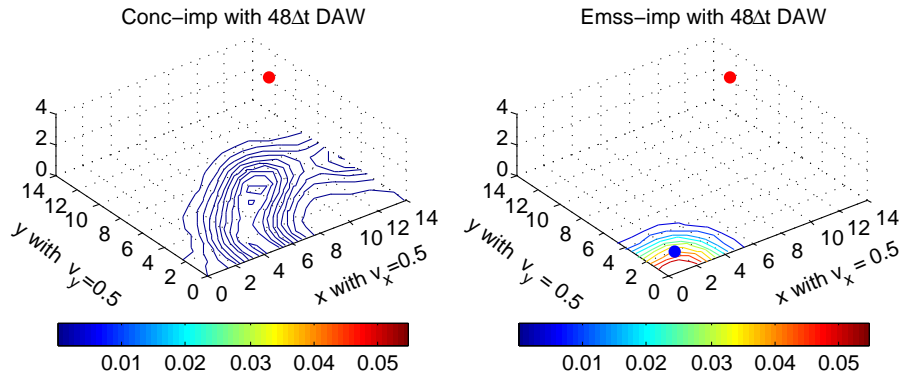


Figure 11. Diffusion test (strong) with $48\Delta t$ DAW and southwesterly wind. Plotting conventions are as in Fig. 1.

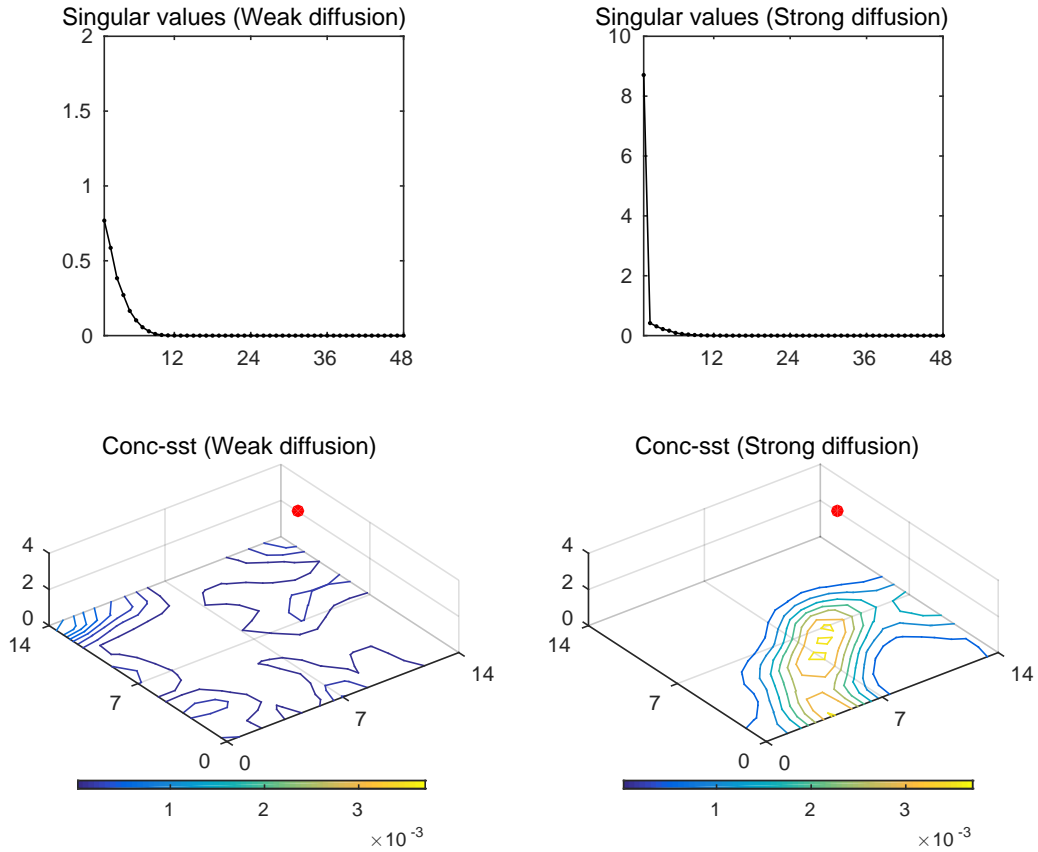


Figure 12. Singular values of Fig. 10 and Fig. 11 and sensitivities of initial states approximated by 5 leading singular values.

Table 1. Ensemble relative ratios of the initial value and emission rate at the lowest layer.

	Fig. 1	Fig. 2	Fig. 3	Fig. 5	Fig. 6	Fig. 7	Fig. 8	Fig. 9	Fig. 10	Fig. 11
\bar{p}^c	0.9920	0.4548	0.2848	0.9937	0.9939	0.9939	0.2248	0.1808	0.9928	0.1905
\bar{p}^e	0.0080	0.5452	0.7152	0.0063	0.0061	0.0061	0.7752	0.8192	0.0072	0.8095

Table 2. The ensemble degrees of freedom for signal of the initial value and emission rate at the lowest layer.

	\bar{P}_{low}^c	\bar{P}_{low}^e
Fig. 3	0.2767	0.8851
Fig. 10	0.0102	0.0030
Fig. 11	0.0500	0.7892

7 Conclusions and Outlooks

This study demonstrates the quantification of the sensitivity of a given measurement network to impact initial trace gas state and emission rates for transport-diffusion models forced by emission. The indicators adopt the degrees of freedom for signal concept. Resting on a Kalman smoother, the contribution to the degree of freedom for signal is derived as the criterion to evaluate the potential improvement of the extended state vector as calculated by singular value decomposition. With a statistical interpretation, we can apply it to determine in advance, which parameters can be optimized by the data assimilation procedure. The degree of freedom for signal and a number of metrics provide us with the quantitative solutions to measure to what extend the parameters can be optimized. Due to its normalization, it is uniformly available for any prior initial values of invertible background covariances. Further, the proposal of the ensemble based relative improvement covariance, based on EnKS, gives us a computationally feasible access to assess the degree of freedom for signal.

The sensitivity of observational networks was formulated by seeking the fastest directions of the perturbation ratio between initial states and observation configurations during the entire time interval. An elementary advection-diffusion example illustrated the significance of relative improvements covariances and their various metrics in different situations and compared them with the results of the sensitivity analysis.

In the future, it is planned to apply the efficiency analysis into the real atmospheric transport model to solve practical problems, also for nonlinear reactive chemistry-transport-diffusion models, as far as the validity of the tangent linear assumption holds, exactly as in atmospheric chemistry data assimilation problems. It is expected that we will get deeper insight to the sensitivity analysis for wider applications. For example, in order to evaluate the impact of observations in some certain locations, the local projection operator introduced by Buizza and Montani (1999) can be applied into approaches presented in Section 4 and Section 5.

8 Code availability

The codes are available via "<https://www.geosci-model-dev-discuss.net/gmd-2017-220/gmd-2017-220-supplement.zip>", or from the corresponding author on request.

Appendix A

- 5 In this appendix, we show a more general case to extend the control vector by emissions. As we know, the initial state and emission rates do not have the same dimension in some practical cases, as there can be more than one different kinds of emissions for one species. Compared with (2), the general situation leads us to consider the following model

$$\frac{d\delta c}{dt} = \mathbf{A}\delta c + B(t)\delta e(t), \quad (61)$$

where $B(t)$ is an operator transforming the emission state vector into the concentration-state space. Combining with (8), we

- 10 obtain the extended model

$$\begin{pmatrix} \delta c(t) \\ \delta e(t) \end{pmatrix} = \begin{pmatrix} M(t, t_0) & \int_{t_0}^t M(t, s)B(s)M_e(s, t_0)ds \\ 0 & M_e(t, t_0) \end{pmatrix} \begin{pmatrix} \delta c(t_0) \\ \delta e(t_0) \end{pmatrix}. \quad (62)$$

Appendix B

In this appendix, we will introduce the singular vector approach to identify the sensitive directions of initial values and emission rates to the observation networks and show the association between the efficiency and sensitivity of observation networks.

- 15 We define the magnitude of the perturbation of the initial state by the norm in the state space with respect to a positive definite matrix W_0 , typically the forecast error covariance matrix.

$$\|\delta x(t_0)\|_{W_0}^2 = \langle \delta x(t_0), W_0 \delta x(t_0) \rangle. \quad (63)$$

Likewise, we define the magnitude of the related observations perturbation in the time interval $[t_0, \dots, t_N]$ by the norm with respect to a sequence of positive definite matrices $\{W(t_k)\}_{k=1}^N$

$$20 \quad \|\delta y_c\|_{\{W(t_k)\}}^2 = \sum_{k=0}^N \langle \delta y_c(t_k), W(t_k) \delta y_c(t_k) \rangle, \quad (64)$$

where $\delta y_c = (\delta y_c^\top(t_0), \dots, \delta y_c^\top(t_N))^\top$.

In order to find the direction of observation configuration which minimizes the perturbation of the initial states, the ratio

$$\frac{\|\delta x(t_0)\|_{W_0}^2}{\|\delta y_c\|_{\{W(t_k)\}}^2}, \quad \delta y \neq 0_{m \times 1}. \quad (65)$$

must be minimized. It is equivalent to maximize the ratio of the magnitude of observation perturbation and the initial perturba-

- 25 tion

$$\frac{\|\delta y_c\|_{\{W(t_k)\}}^2}{\|\delta x(t_0)\|_{W_0}^2}, \quad \delta x(t_0) \neq 0_{n \times 1}. \quad (66)$$

Thus, we define the measure the perturbation growth as

$$\begin{aligned}
g^2 &= \frac{\|\delta y_c\|_{\{W(t_k)\}}^2}{\|\delta x(t_0)\|_{W_0}^2} \\
&= \sum_{k=0}^N \frac{\langle \delta y_c(t_k), W(t_k) \delta y_c(t_k) \rangle}{\langle \delta x(t_0), W_0 \delta x(t_0) \rangle} \\
&= \sum_{k=0}^N \frac{\langle H(t_k) \delta x(t_k), W(t_k) H(t_k) \delta x(t_k) \rangle}{\langle \delta x(t_0), W_0 \delta x(t_0) \rangle} \\
5 \quad &= \sum_{k=0}^N \frac{\langle \delta x(t_k), H(t_k)^\top W(t_k) H(t_k) \delta x(t_k) \rangle}{\langle \delta x(t_0), W_0 \delta x(t_0) \rangle} \\
&= \sum_{k=0}^N \frac{\langle \delta x(t_0), M(t_k, t_0)^\top H(t_k) W(t_k) H(t_k) M(t_k, t_0) \delta x(t_0) \rangle}{\langle \delta x(t_0), W_0 \delta x(t_0) \rangle} \\
&= \frac{\langle \delta x(t_0), \sum_{k=0}^N M(t_k, t_0)^\top H(t_k)^\top W(t_k) H(t_k) M(t_k, t_0) \delta x(t_0) \rangle}{\langle \delta x(t_0), W_0 \delta x(t_0) \rangle} \\
&= \frac{\langle \delta x(t_0), \mathcal{G}^\top \mathcal{W} \mathcal{G} \delta x(t_0) \rangle}{\langle \delta x(t_0), W_0 \delta x(t_0) \rangle}, \quad \delta x(t_0) \neq 0,
\end{aligned} \tag{67}$$

where \mathcal{G} has the same definition as given in Section 3, and

$$10 \quad \mathcal{W} = \begin{pmatrix} W(t_0) & & \\ & \ddots & \\ & & W(t_N) \end{pmatrix}. \tag{69}$$

We arrive at the eigenvalue problems (Liao et al., 2006)

$$W_0^{-\frac{1}{2}} \mathcal{G}^\top \mathcal{W} \mathcal{G} W_0^{-\frac{1}{2}} v_k = s_k^2 v_k, \quad W^{\frac{1}{2}} \mathcal{G} W_0^{-1} \mathcal{G}^\top W^{\frac{1}{2}} u_k = s_k^2 u_k, \tag{70}$$

where $s_1 \geq s_2 \geq \dots \geq s_n \geq 0$ are singular values, $\{v_k\}_{k=1}^n$ and $\{u_k\}_{k=1}^n$ are the corresponding orthogonal singular vectors.

Then, $\max_{\delta x(t_0) \neq 0} g^2 = s_1^2$.

15 Especially, if the perturbation norms are provided by the choice $W_0 = P^{-1}(t_0|t_{-1})$ and $\mathcal{W} = \mathcal{R}^{-1}$,

$$g^2 = \frac{\langle \delta x(t_0), \mathcal{G}^\top \mathcal{R}^{-1} \mathcal{G} \delta x(t_0) \rangle}{\langle \delta x(t_0), P^{-1}(t_0|t_{-1}) \delta x(t_0) \rangle}, \quad \delta x(t_0) \neq 0. \tag{71}$$

Correspondingly, we consider the singular value decomposition of the following matrices

$$P^{\frac{1}{2}}(t_0|t_{-1}) \mathcal{G}^\top \mathcal{R}^{-1} \mathcal{G} P^{\frac{1}{2}}(t_0|t_{-1}) v_k = s_k^2 v_k; \tag{72}$$

$$\mathcal{R}^{-\frac{1}{2}} \mathcal{G} P(t_0|t_{-1}) \mathcal{G}^\top \mathcal{R}^{-\frac{1}{2}} u_k = s_k^2 u_k, \quad k = 1, \dots, n, \tag{73}$$

20 where s_k , v_k and u_k , $k = 1, \dots, n$ denotes the singular values and corresponding singular vectors to simplify the notation. At the same time, it can be seen that the above singular value decomposition problem coincides with (20) in Section 3.

Acknowledgements. This study was supported by HITEC (Helmholtz Interdisciplinary Doctoral Training in Energy and Climate Research). HITEC is the Graduate School of Forschungszentrum Jülich and the five partner universities Aachen, Bochum, Cologne, Düsseldorf and Wuppertal focusing on energy and climate research. Technical and computational support was provided by the Jülich Supercomputer Centre (JSC) of the Research Centre Jülich. A large and central part of the case studies has been computed on JSC supercomputer JUECA. Access
5 given to these computational resources is highly appreciated.

References

- Abida, R. and Bocquet, M.: Targeting of observations for accidental atmospheric release monitoring, *Atmos. Env.*, 43, 6312-6327, 2009.
- Baker, N. L., Daley, R.: Observation and background adjoint sensitivity in the adaptive observation-targeting problem, *Q. J. R. Meteorol. Soc.*, 126, 1431-1454, 2000.
- Barkmeijer, J., Gijzen M. V. and Bouttier F.: Singular vectors and estimates of the analysis-error covariance metric, *Q. J. Roy. Meteorol. Soc.*, 124, 1695-1713, 1998.
- Belsky, T., Kostelich, E. J. and Mahalov, A.: Kalman filter data assimilation: targeting observations and parameter estimation, *Chaos*, 24(2):024406, doi: 10.1063/1.4871916, 2014.
- Berliner, L. M., Lu, Z., and Snyder, C.: Statistical design for adaptive weather observations, *J. Atmos. Sci.*, 56, 2536-2552, 1998.
- Bishop, C. H. and Toth, Z.: Ensemble transformation and adaptive observations, *J. Atmos. Sci.* 56, 1748-1765, 1999.
- Bocquet, M.: Parameter-field estimation for atmospheric dispersion: application to the Chernobyl accident using 4D-Var, *Q. J. R. Meteorol. Soc.*, 138, 664-681, 2012.
- Bocquet, M., Sakov, P.: Joint state and parameter estimation with an iterative ensemble Kalman smoother, *Nonlinear proc geoph.*, 20, 803-818, 2013.
- Brockett R. W.: Finite dimensional linear systems, John Wiley and Sons, Inc, 1970.
- Buizza, R.: Localization of optimal perturbations using a projection operator, *Quart. J. Roy. Meteor. Soc.*, 120, 1647-1681, 1994.
- Buizza, R. and Montani, A.: Targeting observations using singular vectors, *J. Atmos. Sci.* 56, 2965-2985, 1999.
- Buizza, R., Barkmeijer, J., Palmer, T. N. and Richardson, D. S.: Current status and future developments of the ECMWF Ensemble Prediction System, *Meteorol. Appl.*, 7, 163-175, 2000.
- Buizza, R. and Palmer, T. N.: The singular-vector structure of the atmospheric global circulation, *J. Atmos. Sci.*, 52, 1434-1456, 1995.
- Cardinali, C., Pezzulli, S. and Andersson, E.: Influence-matrix diagnostic of a data assimilation system, *Q. J. R. Meteorol. Soc.*, 130, 2767-2786, 2004.
- Cardinali, C.: Monitoring the observation impact on the short-range forecast, 135, 239-250, 2009.
- Catlin, D. E.: Estimation, control, and the discrete Kalman filter, Springer-Verlag, 1989.
- Cioaca A. and Sandu A.: Low-rank approximations for computing observation impact in 4D-Var data assimilation, *Compu. & Math. with Appl.*, 67(12), 2112-2126, 2014.
- Cioaca, A. and Sandu, A.: An optimization framework to improve 4D-Var data assimilation system performance, *J. Compu. Phy.*, 275, 377-389, 2014.
- Daescu, D.: On the sensitivity equations of four-dimensional variational (4D-Var) data assimilation, *Mon. Wea. Rev.*, 136, 3050-3065, 2008.
- Daescu, D. and Navon, I. M.: Adaptive observations in the context for 4D-Var data assimilation, *Meteorol. Atmos. Phys.*, 55, 205-236, 2004.
- Daley, R.: Atmospheric data analysis, Cambridge University Press, 1991.
- Dubovik, O., Lapyonok, T., Kaufman, Y. J., Chin, M., Ginoux, P., Kahn, R. A., and Sinyuk, A.: Retrieving global aerosol sources from satellites using inverse modeling, *Atmos. Chem. Phys.*, 8, 209-250, 2008.
- Elbern, H. and Schmidt, H.: A four-dimensional variational chemistry data assimilation scheme for Eulerian chemistry transport modeling, *J. Geophys. Res.*, 104, 583-598, 1999.

- Elbern, H., Schmidt, H., Talagrand, O. and Ebel, A.: 4D-variational data assimilation with an adjoint air quality model for emission analysis, *Environ modell softw*, 15, 539-548, 2000.
- Elbern, H., Strunk, A., Schmidt, H. and Talagrand, O.: Emission rate and chemical state estimation by 4-dimension variational inversion, *Atmos. Chem. Phys.*, 7, 3749-3769, 2007.
- 5 Evensen, G.: *Data assimilation: The ensemble Kalman filter*, 2th Edition, Springer, Norway, 2009.
- Eyre, J. R.: The Information Content of Data From Satellite Sounding Systems: A Simulation Study, *Quart. J. Roy. Meteor. Soc.*, 116, 401-434, 1990.
- Fisher, M.: Estimation of Entropy reduction and Degrees of freedom for signal for large variational analysis systems, ECMWF Technical Memorandum,397, 2003.
- 10 Fourrié, N. and Thépaut, J. N.: Validation of the NESDIS Near Real Time AIRS Channel Selection, *Quart. J. Roy. Meteor. Soc*, 2003.
- A. Gelb, A. (ed.): *Applied Optimal Estimation*, Cambridge, Mass.: M.I.T. Press, 1974.
- Gillijns, S. and De Moor, B.: Model error estimation in ensemble data assimilation, *Nonlin. Processes Geophys.*,14, 59-71, 2007.
- Goris, N. and Elbern, H.: Singular vector decomposition for sensitivity analyses of tropospheric chemical scenarios, *Atmos. Chem. Phys.*, 13, 5063-5087, 2013.
- 15 Goris, N. and Elbern, H.: Singular vector based targeted observations of chemical constituents: description and first application of EURAD-IM-SVA, *Geosci. Model Dev*, 8, 3929-3945, 2015.
- Ham, F. M. and Brown, R. G.: Observability, eigenvalues and Kalman filtering, *IEEE Tran. Aero. and Electronic sys.* AES-19(2), 269-273, 1983.
- Higham, N. J.: *Accuracy and stability of numerical algorithms*, Second edition, SIAM, 2002.
- 20 Johnson, C. D.: Optimization of a certain quality of complete controllability and observability for linear dynamic systems, *Trans. ASME*, 91, series D, 228-238,1969.
- Johnson, C.: *Information content of observations in variational data assimilation*, Ph. D thesis, 2003.
- Kalman, G. R.: A new approach to linear filtering and prediction problems, *J. Basic Engineering*, 82(D), 35-45, 1960.
- Kalman, R. E. and Bucy, R. S.: New results in linear filtering and prediction theory, *J. Basic Engineering*, 83(D), 95-108, 1961.
- 25 Kang, W. and Xu, L.: Optimal placement of mobile sensors for data assimilations, *Tellus A*, 64, 2012.
- Khattatov, B. B., Gille, J., Lyjak, L., Brasseur, G., Dvortsov, V., Roche, A. and Waters, J.: Assimilation of photochemically active species and a case analysis of UARS data, *J. Geophys. Res.*, 22, 18715-18738, 1999.
- Langland, R. H., Gelaro, R., Rohaly, G. D., and Shapiro, M. A.: Targeted observations in FASTEX: Adjoint based targeting procedures and data impact experiments in IOP17 and IOP18, *Q. J. R. Meteorolog. Soc.* 125, 3241-3270, 1999.
- 30 Leutbecher, M.: A reduced rank estimate of forecast error variance changes due to intermittent modifications of the observing network, *J. Atmos. Sci.*, 60, 729-742, 2003.
- Li, H., Kalnay, E., Miyoshi, T., Danforth, C. M.: Accounting for model errors in ensemble data assimilation, *Mon. Weather Rev.*, 137, 3407-3419, 2009.
- Li, Z., Navon, I. M.: Optimality of variational data assimilation and its relationship with the Kalman filter and smoother, *Q. J. R. Meteorol. Soc.*, 127, 661-683, 2001.
- 35 Liao, W., Sandu, A., Carmichael, G. R. and Chai, T.: Singular vector analysis for atmospheric chemical transport models, *Month. Weather Rev.*, 134, 2443-2465, 2006.

- Liu, J. and Kalnay, E.: Estimating observation impact without adjoint model in an ensemble Kalman filter, *Q. J. R. Meteorol. Soc.*, 134, 1327–1335, 2008.
- Liu, J., Kalnay, E., Miyoshi, T., Cardinali, C.: Analysis sensitivity calculation in ensemble Kalman filter, *Q. J. R. Meteorol. Soc.*, 135, 1842-1851, 2009.
- 5 Lorenz, E.N.: A study of the predictability of a 28 variable atmospheric model, *Tellus*, 17, 321-333, 1965.
- Martynenko, D., Holzer-Popp, T., Elbern, H. and Schroedter-Homscheidt, M.: Understanding the aerosol information content in multi-spectral reflectance measurements using a synergetic retrieval algorithm, *Atmos. Meas. Tech.*, 3, 2579-2602, 2010..
- Maybeck, P. S.: Stochastic models, estimation, and control, Volume 2, Academic press, New York, 1979.
- Miyazaki, K., Eskes, H. J., Sudo, K., Takigawa, M., vanWeele, M., and Boersma, K. F.: Simultaneous assimilation of satellite NO₂, O₃,
 10 CO, and HNO₃ data for the analysis of tropospheric chemical composition and emissions, *Atmos. Chem. Phys.*, 12, 9545-9579, 2012.
- Morrison, D. F.: *Multivariate Statistical Methods*, McGraw-Hill, New York, 1967.
- Navon, I. M.: Practical and theoretical aspects of adjoint parameter estimation and identifiability in meteorology and oceanography, *Dyn. Atmos. Oceans*, 27, 55-79, 1997.
- Peters, W., Miller, J. B., Whitaker, J., Denning, A. S., Hirsch, A., Krol, M. C., Zupanski, D., Bruhwiler, L. and Tans, P. P.: An ensemble data
 15 assimilation system to estimate CO₂ surface fluxes from atmospheric trace gas observations, *J. Geophys. Res.*, 110(D24), 2005.
- Rabier, F., Fourrié, N., Chafaï, D., and Prunet, P.: Channel selection methods for infrared atmospheric sounding interferometer radiances. *Quart. J. Roy. Meteor. Soc.*, 128, 1011-1027, 2002.
- Rodgers, C. D.: *Inverse methods for atmospheric sounding: theory and practice* (Series on atmospheric, oceanic and planetary physics—Vol.2, World scientific, Singapore, 2000.
- 20 Sandu, A., Cioaca, A. and Rao, V.: Dynamic sensor network configuration in InfoSymbiotic systems using model singular vectors, *Procedia Computer Science*, 18, 1909-1918, 2013.
- Singh, K., Sandu, A., Jardak, M., Bowman, K. W.: Information Theoretic Metrics to Characterize Observations in Variational Data Assimilation, *Procedia Computer Science*, 9, 1047-1055, 2012.
- Singh, K., Sandu, A., Jardak, M., Bowman, K.W. and Lee, M.: A practical method to estimate information content in the context of 4D-Var
 25 data assimilation, *SIAM/ASA J. Uncertainty quantification*, 1(1), 106-138, 2013.
- Smith, P. J., Thornhill, G. D., Dance, S. L., Lawless, A. S., Mason, D. C. and Nichols, N. K.: Data assimilation for state and parameter estimation: application to morphodynamic modelling, *Q. J. R. Meteorol. Soc.*, 139, 314-327, 2013.
- Szunyogh, I., Toth, Z., Emanuel, K. A., Bishop, C.H., Woolen, J., Marchok, T., Morss, R., and Snyder, C.: Ensemble based targeting experiments during FASTEX: The impact of dropsonde data from the Lear jet, *Q. J. R. Meteorolog. Soc.* 125, 3189-3218, 1999.
- 30 Szunyogh, I., Toth, Z., Zimin, A. V., Majumdar, S. J., and Persson, A.: Propagation of the effect of targeted observations: The 2000 Winter Storm Reconnaissance Program, *Mon. Weather Rev.* 130, 1144-1165, 2002.
- Tang, X., Zhu, J., Wangand, Z. F. and Gbaguidi, A.: Improvement of ozone forecast over Beijing based on ensemble Kalman filter with simultaneous adjustment of initial conditions and emissions, *Atmos. Chem. Phys.*, 11, 12901-12916, 2011.
- Tang, X., Zhu, J., Wang, Z. F., Wang, M., Gbaguidi, A., Li, J., Shao, M., Tang, G. Q., and Ji, D. S.: Inversion of CO emissions over Beijing
 35 and its surrounding areas with ensemble Kalman filter, *Atmos. Environment*, 81, 676-686, 2013.
- Tremolet, Y.: Model-error estimation in 4D-Var, *Q. J. R. Meteorol. Soc.*, 133, 1267-1280, 2007.

- Winiarek, V., Bocquet, M., Duhanyan, N., Roustan, Y., Saunier, O., and Mathieu, A.: Estimation of the caesium-137 source term from the Fukushima Daiichi nuclear power plant using a consistent joint assimilation of air concentration and deposition observations, *Atmos. Environment*, 82, 268-279, 2014.
- 5 Wu, X., Jacob, B. and Elbern, H.: Optimal control and observation locations for time-varying systems on a finite-time horizon, *SIAM J. Control Optim.*, 2016.
- Woodbury, M. A.: Inverting modified matrices, Memorandum Rept. 42, Statistical Research Group, Princeton University, Princeton, NJ, 1950.
- Yanenko, N. N.: The method of fractional steps: solution of problems of mathematical physics in several variables, Springer, Berlin, Heidelberg and New York, 1971.
- 10 Zupanski, D., Hou, Y. A., Zhang, Q. S., Zupanski, M., Kummerow, D. C., Cheung, H. S.: Applications of information theory in ensemble data assimilation, *Q. J. R. Meteorol. Soc.*, 133, 1533-1545, 2007.