Metrics for evaluating the "quality" in linear atmospheric inverse problems: a case study of a trace gas inversion

Vineet Yadav[1], Subhomoy Ghosh[2,3], and Charles E. Miller[1]

[1]Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive, Pasadena, CA, USA
[2]University of Notre Dame, Notre Dame, IN, USA
[3]National Institute of Standards and Technology, Gaithersburg, MD, USA

Correspondence: Subhomoy Ghosh (sghosh4@nd.edu)

Abstract. Several metrics have been proposed and utilized to diagnose the performance of linear Bayesian and geostatistical atmospheric inverse problems. These metrics ~~are mostly related to assessing reduction~~ primarily assess reductions in prior uncertainties, ~~comparing~~ compare modeled observations to true observations, and ~~checking~~ check distributional assumptions. ~~These metrics, though important,~~ Although important, these metrics should be augmented with sensitivity analysis to obtain a comprehensive understanding of ~~the performance of atmospheric inversions and critically~~ atmospheric inversion performance and improve the quality ~~of an atmospheric inverse model~~ and confidence in the ~~estimated fluxes~~inverse estimates. In this study, we derive ~~analytical forms of the local sensitivities of the estimated fluxes with respect to the number of inputssuch as~~ closed-form expressions of local sensitivities for various inputs, including measurements, covariance parameters, covariates, and a forward operator. ~~These local sensitivities have different units and vastly different magnitudes. To this end, we also propose a technique to rank local sensitivities. In addition to local sensitivity , we provide~~ To further enhance our understanding, we complement local sensitivity analysis with a framework for global sensitivity analysis ~~for linear atmospheric inversion that shows the apportionment of the uncertainty in different~~ that can apportion the uncertainty in inputs to the uncertainty ~~of estimated fluxes. Prior to performing an inversion, we also~~ associated with inverse estimates. Additionally, we propose a mathematical framework to construct nonstationary correlation matrices from a pre-computed forward operator~~that encompasses non-stationary structures. This~~ , which is closely tied to the overall quality of ~~estimated fluxes. We show~~ inverse estimates. We demonstrate the application of our methodology in the context of an atmospheric inverse problem for estimating methane fluxes in Los Angeles, California. ~~The proposed framework is applicable to any other domain that employs linear Bayesian and geostatistical inverse methods.~~

1 Introduction

~~Inverse models within the context of atmospheric applicationsare often used for constraining~~ In atmospheric applications, inverse models are frequently used to estimate global to regional scale fluxes of trace gases ~~(for discussion see, Enting, 2002). At~~ from atmospheric measurements (Enting, 2002). At a global scale, data assimilation ~~(for further details on data assimilation, see Wikle and Berliner, 2007) that sequentially assimilates observations~~ remains the primary inverse modeling framework, which assimilates observations sequentially and updates the prior estimates of fluxes by utilizing an atmospheric model cou-

pled with chemistry ~~remains the primary inverse modeling framework. This framework at regional scaleis complimented by inversions that assimilates~~ (for further details on data assimilation, see Wikle and Berliner, 2007). At a regional scale, inversions that assimilate all observations simultaneously by utilizing a ~~precomputed forward operator (Lin et al., 2003)~~ pre-computed forward operator (Lin et al., 2003) that describes the relationship between observations and fluxes are commonly used (for details, see Enting, 2002). This work focuses on ~~these latter class of inverse methods. It specifically~~ the use of pre-computed forward operators for atmospheric inverse modeling and addresses sensitivity analysis and correlation in the forward operator in the context of Bayesian (~~for~~ e.g., ~~see~~ Lauvaux et al., 2016) and geostatistical inverse methods (~~see~~ e.g., Kitanidis, 1996).

The ~~senstivity analysis in context of this study~~ sensitivity analysis in this work is covered under local and global themes. Primarily, we focus on local sensitivity analysis (LSA)~~that~~ , which measures the effect of a given input on a given output ~~.~~ ~~This~~ and is obtained by computing partial derivatives of an output ~~of interest with respect to~~ quantity of interest for an input factor (see See Rabitz, 1989, and Turányi, 1990). Within ~~global theme~~the global theme (designated as Global Sensitivity Analysis), we focus on how uncertainty in the model output can be apportioned to different ~~sources of uncertainty with respect to corresponding model input~~ model inputs (Saltelli et al., 2008).

~~Previously, many methods have been proposed and utilized to perform sensitivity analysis. These can be categorized as global and local sensitivity analyses. Global sensitivity analysis (GSA) includes Morris's (e.g. Morris, 1991) one step at a time method (OAT), Polynomial Chaos Expansion (PCE) (e.g. Sudret, 2008), Fourier amplitude sensitivity test (FAST) (e.g. Xu and Gertner, 2011), Sobol's method (e.g. Sobol, 2001) and Derivative based global sensitivity measures (DGSM) (e.g. Sobol and Kucherenko, 2010) among others. These existing GSA methods either: (1) assume independence of parameters (e.g., FAST and OAT), or are (2) computationally expensive (e.g., Sobol's method), or (3) require knowledge of the joint probability distribution of the parameter space (e.g., DGSM, PCE). Therefore, these traditional methods cannot be directly applied in linear atmospheric inverse problems, which consists of tens of thousands of non-normal, spatio-temporally correlated parameters (includes observations). Recently proposed active subspace based GSA (Constantine and Diaz, 2017) uses low dimensional approximation of the parameter space. In its current form, it is still computationally expensive for problems that consists of thousands of parameters (see case study in Constantine and Diaz, 2017).~~

~~In comparison to GSA, local sensitivity method like Bayesian Hyper Differential Sensitivity Analysis (HDSA) computes partial derivatives with respect to maximum a posteriori probability estimates (MAP) of a quantity of interest. Our method for LSA is similar to Bayesian HDSA, except for the fact that it directly finds analytical derivatives of the MAP solution with respect to the input parameters in linear atmospheric inverse problems. This is possible when we know analytical closed form solutions of the estimated fluxes. In this study, we leverage a framework that is not only one of the most commonly adopted forms in atmospheric inversions but also admit closed form solutions. Thus, unlike the previous work on Bayesian HDSA, we do not generate samples from the prior to compute multiple MAP points. As we have limited knowledge of the prior distribution of the spatio-temporally correlated parameters. We derive exact functional form of the local sensitivity equations based on the closed form analytical MAP solution. Our method is simple and amenable to tens of thousands of parameters.~~

~~Note as in all linear atmospheric inverse problems one of the key goals of this work is to study the importance of thousands of spatio-temporally varying parameters by ranking them and computation of the local sensitivities is a means to achieve that goal.~~

Overall, in atmospheric trace gas inversions, mostly LSA is performed. Within this context, LSA ~~asseses~~ assesses how sensitive the posterior estimates of fluxes are ~~with reference to~~ regarding the underlying choices or assumptions, like (1) observations included, (2) model-data error covariance, (3) the input prior information and its error, and (4) the forward operator (for discussion~~see,~~ , see Michalak et al., 2017). This task is sometimes performed to arrive at a robust estimate of fluxes and their uncertainties~~. It is achieved~~ , by running an inverse model multiple times ~~by~~ while varying the inputs and assessing their impact on the estimated fluxes and uncertainties. Another ~~complimentary~~ complementary way to do LSA is by computing local partial derivatives ~~with respect to these quantities down to an individual entry that go in~~ of inputs that go into an inversion.

LSA can be grouped with standard information content approaches such as ~~averaging kernel or model resolution matrix~~ an averaging kernel and degrees of freedom for signal (DOFS; for details, see Sec. 2.2.1 of this manuscript, Rodgers, 2000, and Brasseur and Jacob, 2017). ~~Averaging kernel matrix shows how the estimated fluxes are related or sensitive to the true fluxes. Thus, it belongs to the LSA category.~~ However, LSA is more informative than ~~DOFS and averaging kernel aloneas it goes after~~ these approaches alone, as it examines individual components (see Sec. 2.2) that determine DOFS ~~. Furthermore, DOFS is a measure that provides an estimate of the information resolved by an inversion. In comparison, LSA focuses on quantifying the impact and the~~ and quantifies the impact and relative importance of various components of an inversion~~in governing the estimates of fluxes~~.

In this study, we focus on the quality of the inverse estimates of the fluxes, which means providing diagnostic metrics to ~~better characterize~~ improve our understanding of the impact of input choices on the inverse estimates of fluxes and thus improve the quality of the inverse model. Specifically, in this technical note~~we provide:~~ , we provide (1) ~~analytical~~ closed-form expressions to conduct ~~post hoc (that is after an inversion has been performed)~~ LSA by computing partial derivatives, (2) a scientifically interpretable framework for ranking thousands of ~~spatio-temporally~~ spatiotemporally correlated input parameters with the same or different units of measurement, (3) a mathematical schema for conducting ~~GSA. However, GSA is considerably difficult to perform in the absence of the knowledge about the uncertainties associated with all the inputs that go in an inversion,~~ global sensitivity analysis (GSA), and (4) a technique to assess ~~spatio-temporal~~ the spatiotemporal correlation between forward operators of two or multiple observations~~. This~~ , which is tied to the overall diagnostics of the estimated fluxes ~~as fluxes are strongly sensitive to the forward operator and improvement in understanding the representation of the atmospheric transport model error through spatio-temporal association in the forward operators~~ and can lead to ~~significant improvement in designing the components of an atmospheric inversion framework.~~improved representation of errors in the forward operator.

2 ~~Organization of the study~~Methods and derivation

In a generic form, a linear inverse problem can be written as:

$$\mathbf{z} = \mathbf{H}\mathbf{s} + \boldsymbol{\epsilon}, \tag{1}$$

where $\mathbf{H}$ is a forward operator that maps model parameters ~~s~~ (fluxes in the context of this work) to measurements $\mathbf{z}$ and encapsulates our understanding of the physics of the measurements. The error $\boldsymbol{\epsilon}$ in Eq. (1) describes the mismatch between measurements and the modeled measurements (see Sec. ~~2~~3).

In a typical linear atmospheric inverse problem (see Fig. 1), the estimates of ~~the~~ fluxes (box 8 of Fig. 1) are obtained in a classical ~~one stage~~ one-stage batch Bayesian setup (for details, see Enting, 2002; Tarantola, 2005)~~, where~~ . In this setup, the a priori term (box 3 in Fig. 1) is based on a fixed flux pattern~~at a prescribed spatio-temporal resolution~~, and errors (box 6 in Fig. 1) are either assumed to be independent or are governed by a ~~prescribed~~ pre-defined covariance structure (for details, see Gurney et al., 2003; Rödenbeck et al., 2003, 2006).

Within the previously mentioned setup, the choice of the input parameters, including the forms of error structures~~have profound impact on~~ , profoundly impacts the quality of the inverse estimates of fluxes. Understanding the impact of these inputs is critical for evaluating the quality of the estimated fluxes. Thus, ~~in the first part of this work~~ first (Sec. 2.1), we utilize the understanding of the physics of the ~~measurement that is encapsulated in $\mathbf{H}$ to generate correlation matrices that are scientifically interpretable in the context of estimated fluxes and to build an interpretable non-stationary model of the residual covariance structure~~ measurements, encapsulated in $\mathbf{H}$, to generate scientifically interpretable correlation matrices (box 6 in Fig. 1). ~~This is described in Sec. 2.1. In the second part of this work~~ Second, we assess and rank the importance of the inputs ~~mentioned~~ (Sec. 2.2) shown in the middle column (the green background box) of Fig. 1 ~~in governing the estimates of fluxes~~ (box 8 of Fig. 1)~~. This is covered in Sec. 2.2. These two parts are followedby a methane (CH_4~~, which is finally followed, by methane (CH_4) case study that demonstrates the applicability of our methods (see Sec. ~~3~~2).

~~To maintain maximum transparency, facilitate assessment, and show applicability of our methods in Sec. 2 we also provide two well documented interactive MATLAB Livescripts (for details on Livescript see MatlabLivescript), one for each methodological part. These Livescripts contain equations, code, and visualizations as it relates to the real-data case study described in Sec. 3, and are included as supplementary material. Separate pdfs of these Livescripts are also included for the readers who do not have access to MATLAB.~~

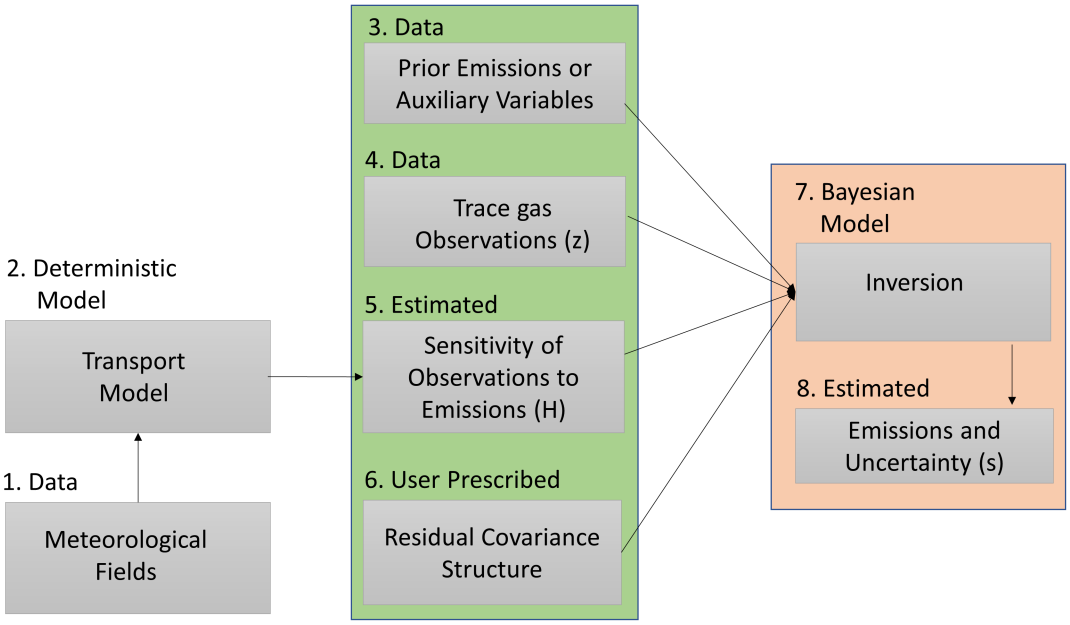

Figure 1. The schema for performing a linear atmospheric inversion to obtain estimates of the fluxes of greenhouse gases. The middle column (the green background box) lists all the inputs that are required for performing an inversion whereas the right column (the orange background box) lists the modeling process (box 7) and the output obtained after performing an inversion (box 8). Note this work focuses on understanding and ranking the impact of the inputs (box 3, 4, and 6 in the middle column) on the estimates of fluxes (box 8) and developing correlation structures from the forward operator (box 5).

3 ~~Methods and derivation~~

2.1 Analysis of the forward operator

In inversions that ~~assimilates~~ assimilate all observations simultaneously, ~~first~~ a forward operator for each observation ~~that would be~~ included in an inversion is obtained from a transport model. These observations ~~of trace gases~~ can be obtained from multiple platforms~~that include~~ , including an in-situ network of fixed locations on the surface, intermittent aircraft flights, and satellites. In most situations, the ~~spatio-temporal~~ spatiotemporal coverage of these forward operators ~~are~~ is visually assessed by plotting an aggregated sum or mean of their values over a ~~map of the spatial domainof the study~~spatial domain. However, standard quantitative metrics to ~~assess~~ evaluate their coverage and intensity in space and time ~~remains completely~~ remain absent. In this study, we present two metrics for this assessment~~and these~~ , which are defined below. These metrics conform to triangular inequality and ~~therefore can be defined as distance function~~ are distances in their respective metric spaces.

Note that sometimes in the published literature on trace gas inversions, the forward operator obtained from a transport model is referred to as a sensitivity matrix, Jacobian, or footprint. Henceforth, ~~to avoid misinterpretation,~~ we always refer to the Jacobian/sensitivity matrix ~~/footprint as forward operator~~or footprint as a forward operator to avoid misinterpretation. We show our application through forward operators constructed by running a Lagrangian transport model. However, ~~our~~ the proposed methods can also be applied in ~~analytical~~ the Eulerian framework (see Brasseur and Jacob, 2017 for details).

2.1.1 Integrated area overlap measurement index (IAOMI)

The Integrated Area Overlap Measurement Index (IAOMI) summarizes the shared information content between two forward operators and hence indirectly between two observations. ~~It istherefore~~ It is, therefore, a measure of the uniqueness of the flux signal associated with an observation ~~in comparison~~ compared to other observations.

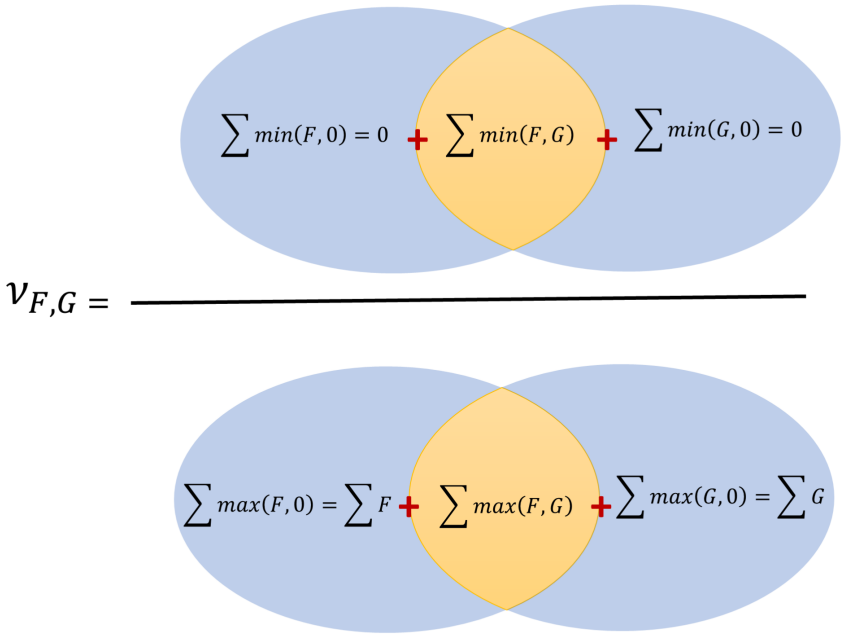

Figure 2. Venn diagram that defines ~~IOAMI~~ IAOMI in terms of two hypothetical forward operators $\mathbf{F}$ and $\mathbf{G}$

Intuitively, IAOMI can be better understood spatially. For a given time point, consider two forward operators $\mathbf{F}$ and $\mathbf{G}$ as two vector-valued functions over an area. ~~IOAMI~~ Index IAOMI is the proportion of the common contribution of the two forward operators from the intersected area with respect to the overall contribution of the two forward operators. This is demonstrated through a Venn diagram in Fig. 2. Thus, IAOMI can be defined as:

$$\nu_{\mathbf{F},\mathbf{G}} = \frac{\Sigma_{A_\mathbf{F} \cap A_\mathbf{G}} \mathbf{f}_1(\mathbf{F},\mathbf{G})}{\Sigma_{A_\mathbf{F} \cup A_\mathbf{F}} \mathbf{f}_2(\mathbf{F},\mathbf{G})} \frac{\Sigma_{A_\mathbf{F} \cap A_\mathbf{G}} \mathbf{f}_1(\mathbf{F},\mathbf{G})}{\Sigma_{A_\mathbf{F} \cup A_\mathbf{G}} \mathbf{f}_2(\mathbf{F},\mathbf{G})}, \tag{2}$$

~~Where~~ where for any forward operator $\mathbf{S}$, the corresponding set $A_{\mathbf{S}}$ on which forward operator is always positive, is defined as $A_{\mathbf{S}} = \{\boldsymbol{x} : \mathbf{S}(\boldsymbol{x}) > 0\}$ and the two vector-valued functionals $\mathbf{f}_1$ and $\mathbf{f}_2$ can be given as:

$$\mathbf{f}_1(\mathbf{F},\mathbf{G}) = \begin{cases} min(\mathbf{F},\mathbf{G}) & \text{on } A_{\mathbf{F}} \cap A_{\mathbf{G}} \\ 0 & \text{otherwise} \end{cases} \quad \text{and} \quad \mathbf{f}_2(\mathbf{F},\mathbf{G}) = \begin{cases} max(\mathbf{F},\mathbf{G}) & \text{on } A_{\mathbf{F}} \cap A_{\mathbf{G}} \\ \mathbf{F} & \text{on } A_{\mathbf{F}} \cap A_{\mathbf{G}}^c \\ \mathbf{G} & \text{on } A_{\mathbf{F}}^c \cap A_{\mathbf{G}} \end{cases} \tag{3}$$

Note that the IAOMI defined in Eq. (2) can also be written as a ratio of the sum of minimums over sum of the maximums as:

$$\nu_{\mathbf{F},\mathbf{G}} = \frac{\Sigma_{A_{\mathbf{F}} \cup A_{\mathbf{G}}} min(\mathbf{F},\mathbf{G})}{\Sigma_{A_{\mathbf{F}} \cup A_{\mathbf{G}}} max(\mathbf{F},\mathbf{G})} \tag{4}$$

IAOMI ν can also be thought as a measure of similarity between two forward operators. It is evident from Eq. (4) that this is a weighted Jaccard similarity index or Ruzicka index (Cha, 2007) which describes similarity between two forward operators $\mathbf{F}$ and $\mathbf{G}$. It follows that ν is closed and bounded in $[0, 1]$ and accounts for both the ~~spatio-temporal~~ spatiotemporal spread and the intensity of the forward operator. A stronger ν implies larger overlap of intensity in space and time~~and~~ , is analogous to finding the common area within two curves~~. The corresponding~~ , and is indicative of the magnitude of overlapping information, a knowledge beneficial in the context of satellite observations with a higher potential for sharing information content.

A measure of dissimilarity can be obtained from ν and can be defined by $1 - \nu$. The smaller the overlap or the larger the value of $1 - \nu$, the ~~larger is the dissimilarity~~more significant the disparity. Note the ν metric is only indicative of the overlap in the ~~spatio-temporal~~ spatiotemporal intensity between two forward operators. To measure how much of the shared intensity has come from either forward operator, we use a metric $\upsilon_{\mathbf{F}|(\mathbf{F},\mathbf{G})}$ defined as:

$$\upsilon_{\mathbf{F}|(\mathbf{F},\mathbf{G})} = \frac{\Sigma_{A_{\mathbf{F}} \cap A_{\mathbf{G}}} \mathbf{f}_1(\mathbf{F},\mathbf{G})}{\Sigma_{A_{\mathbf{F}}} \mathbf{f}_3(\mathbf{F})}, \tag{5}$$

~~Where~~ where $\mathbf{f}_3(\mathbf{F}) = F$ on $A_{\mathbf{F}}$ and 0 everywhere else. Likewise, we can define $\upsilon_{\mathbf{G}|(\mathbf{F},\mathbf{G})}$ which shows proportional contribution of the forward operator G on the shared intensity. Both ν and υ can be computed from observations taken from same or different platforms, at same or different time or for two different in-situ measurement sites over a specified time-interval.

2.1.2 Spatio-temporal Area of Dominance (STAD)

~~The notion of the spatio-temporal~~

The spatiotemporal area of dominance (STAD) stems naturally from IAOMI. For any two forward operators $\mathbf{F}$, and $\mathbf{G}$, we can find out the left-over dominant contribution of $\mathbf{F}$ and $\mathbf{G}$ by computing quantities $\mathbf{F} - \mathbf{G}$ and $\mathbf{G} - \mathbf{F}$ that ~~leads to~~ lead to the determination of the ~~area~~ areas where $\mathbf{F}$ or $\mathbf{G}$ is dominant.

For two forward operators $\mathbf{F}$ and $\mathbf{G}$, STAD of $\mathbf{F}$ with respect to $\mathbf{G}$ is defined as:

$$\text{STAD}_{\mathbf{F}}(\mathbf{F},\mathbf{G}) = \begin{cases} \mathbf{F} - min(\mathbf{F},\mathbf{G}) & \text{on } A_{\mathbf{F}} \cap A_{\mathbf{G}} \\ \mathbf{F} & \text{otherwise} \end{cases}$$

IAOMI and STAD of any forward operator $\mathbf{F}$ with respect to the forward operators $\mathbf{F}$ and $\mathbf{G}$ are linked by the following equation:

$$\nu_{\mathbf{F},\mathbf{G}} \Sigma_{A_{\mathbf{F}} \cup A_{\mathbf{G}}} H_2(\mathbf{F},\mathbf{G}) + \Sigma_{A_{\mathbf{F}} \cup A_{\mathbf{G}}} STAD_{\mathbf{F}}(\mathbf{F},\mathbf{G}) = \Sigma_{A_{\mathbf{F}}} \mathbf{F} \tag{6}$$

Given a number of forward operators $\{\mathbf{F}, \mathbf{G}_1, \mathbf{G}_2, \cdots\}$, STAD for any particular forward operator $\mathbf{F}$ with respect to all other forward operators can be generalized from Eq. (6) as $\mathbf{F}_{\text{STAD}}(\mathbf{F}, \mathbf{G}_{\text{max}})$ where $\mathbf{G}_{\text{max}} = \max_i \mathbf{G}_i$ on $A_{\mathbf{G}}$; $A_{\mathbf{G}} = \cup_k A_{\mathbf{G}_k}$ and $A_{\mathbf{G}_k}$ is the set on which forward operator $\mathbf{G}_k$ is always positive (see Sec. 2.1.1 for its definition). STAD can be aggregated over any time-periods. Intuitively, STAD determines areas in space-time where one forward operator dominates over other forward operators~~. This~~ , which is especially useful in locating the primary ~~sources of fluxes that influences~~ flux sources that influence an observation.

2.1.3 ~~Jensen-Shannon distance (JSD) for forward operators~~

~~Dissimilarity between forward operators can also be measured via entropy (for definition, see MacKay et al., 2003) based distances. Entropy distances are sensitive in capturing differences between two distributions that are similar in the first order (e.g. mean, or median) and second order moments (e.g. variance, or quartile deviation) but differ in higher order moments (e.g. Kurtosis) or modes (e.g. unimodal vs. multimodal). Entropy based distance metrics that adhere to triangular inequality can also be combined with spatio-temporal coverage to measure the probabilistic divergence between two forward operators. One such metric is Jensen-Shanon~~ One can use 1-IAOMI or distance metric like Jensen-Shannon distance (JSD~~) (Nielsen, 2019) which can be used to compute distance between two distributions generated by the forward operators. Normalized forward operators can be seen as samples from an underlying high-dimensional probability distribution such that total sum is one. For any vector-valued forward operator $\mathbf{F}$, normalization by the total sum can be given as:~~

$$P_{F_k} = \frac{F_k}{\sum_k F_k}$$

~~where F_k denotes k^{th} entry of $\mathbf{F}$ and index k spans over the entire domain. The symbol P denotes normalized forward operator. We can then use JSD to compute distance between two normalized forward operators. Thus, JSD can be computed as:~~

$JSD(P_{\mathbf{F}}||P_{\mathbf{G}}) =$

$]\frac{1}{2}D(P_{\mathbf{F}}||M) + \frac{1}{2}D(P_{\mathbf{G}}||M)$.

where D stands for Kulback-Leibler (KL) divergence (see MacKay et al., 2003 for details) . KL divergence D of any probability distribution p with respect to another probability distribution q is defined as: $D(p||q) = \sum p\log(p/q)$ and M is defined as: $M = \frac{1}{2}(P_{\mathbf{F}} + P_{\mathbf{G}})$. The symbol $||$ is used to indicate that $D(P_{\mathbf{F}}||M)$ and $D(P_{\mathbf{G}}||M)$ are not conditional entropies (see MacKay et al., 2003). JSD is closed and bounded in $[0, 1]$ when KL divergence is computed with base 2 logarithm. Intuitively, JSD and $1 - \nu$ (i.e. 1-IAOMI) are comparable since both of them are measures of dissimilarity.

Note that, one can use JSD or 1-IAOMI ; see Appendix B) matrix of all pairwise forward operators as a representative distance matrix for describing correlations in model-data errors (i.e., $\mathbf{R}$ in Eq. (7)). These correlation matrices need to be at least positive semi-definite. Since As JSD or 1-IAOMI matrices are real, symmetric, and admit orthogonal decomposition, the entry-wise exponential of such symmetric diagonalizable matrices is positive-semidefinite . Thus, they and can be incorporated in $\mathbf{R}$ via the commonly adopted exponential kernel of the distance matrix model data mismatch matrix $\mathbf{R}$ (see Ghosh et al., 2021). Furthermore, the IAOMI matrix itself is a positive semidefinite (Bouchard et al., 2013) matrix and can also be directly incorporated in $\mathbf{R}$ as a measure of correlation. This is an example of how IAOMI or 1- IAOMI could be particularly useful for satellite data based inversions with higher degree of spatial overlap of the forward operators. However, we do not explore this area of research in this manuscript.

2.2 Local sensitivity analysis (LSA) in inversions

For linear Bayesian and geostatistical inverse problem, the solutions (see, Tarantola, 2005 for the batch Bayesian and Kitanidis, 1996 for the geostatistical case) can be obtained by minimizing their respective objective functions. These objective functions can be given as:

$$L(\mathbf{s}|\mathbf{y}, \mathbf{s}_{\text{prior}}, \mathbf{H}, \mathbf{Q}, \mathbf{R}) = \frac{1}{2}(\mathbf{z} - \mathbf{H}\mathbf{s})^t\mathbf{R}^{-1}(\mathbf{z} - \mathbf{H}\mathbf{s}) + \frac{1}{2}(\mathbf{s} - \mathbf{s}_{\text{prior}})^t\mathbf{Q}^{-1}(\mathbf{s} - \mathbf{s}_{\text{prior}}) \tag{7}$$

$$L(\mathbf{s}|\mathbf{y}, \mathbf{H}, \mathbf{Q}, \mathbf{R}, \boldsymbol{\beta}) = \frac{1}{2}(\mathbf{z} - \mathbf{H}\mathbf{s})^t\mathbf{R}^{-1}(\mathbf{z} - \mathbf{H}\mathbf{s}) + \frac{1}{2}(\mathbf{s} - \mathbf{X}\boldsymbol{\beta})^t\mathbf{Q}^{-1}(\mathbf{s} - \mathbf{X}\boldsymbol{\beta}), \tag{8}$$

where lower case where lowercase symbols represent vectors and the uppercase symbols represent matrices, and this same approach of exact representation is adopted throughout the manuscript. In Eq. (7) and (8), $\mathbf{z}$ is an $(n \times 1)$ vector of available measurements with unit of each entry being ppm. The forward operator $\mathbf{H}$ is an $(n \times m)$ matrix with unit of each entry being ppm μmoles^{-1}m^{2}sec. The matrix $\mathbf{H}$ is obtained from a transport model that describes the relationship between measurements and unknown fluxes. Unknown flux $\mathbf{s}$ is an $(m \times 1)$ vector with unit of entries being μmoles m^{-2}sec^{-1}. The covariance matrix $\mathbf{R}$ of the model-data errors is an $(n \times n)$ matrix with unit of the entries being ppm^2. The covariate matrix $\mathbf{X}$ is an $(m \times p)$ matrix of known covariates related to $\mathbf{s}$. The unit of each of the entries in every column of the covariate matrix $\mathbf{X}$ is the unit

of its measurement or if it is standardized (e.g. subtract ~~a covariate by its mean~~ the mean from the covariate and divide by its standard deviation) then it is unitless. For further discussion on standardization and normalization see Gelman and Hill, 2006. The units of $(p \times 1)$ vector $\boldsymbol{\beta}$ are such that $\mathbf{X}\boldsymbol{\beta}$ and $\mathbf{s}$ have the same units. The prior error covariance matrix $\mathbf{Q}$ is an $(m \times m)$ matrix that represents the errors between $\mathbf{s}$ and $\mathbf{X}\boldsymbol{\beta}$ with unit of the entries being $\left(\mu\text{moles m}^{-2}\text{sec}^{-1}\right)^2$.

The analytical solutions for the unknown fluxes $\mathbf{s}$ in the Bayesian case (denoted by the subscript B) and the geostatistical case (denoted by the subscript G) can be obtained from Eq. (9) and (10) as given below.

$$\hat{\mathbf{s}}_B = \mathbf{s}_{\text{prior}} + \mathbf{Q}\mathbf{H}^t\left(\mathbf{H}\mathbf{Q}\mathbf{H}^t + \mathbf{R}\right)^{-1}\left(\mathbf{z} - \mathbf{H}\mathbf{s}_{\text{prior}}\right) \tag{9}$$

$$\hat{\mathbf{s}}_G = \mathbf{X}\boldsymbol{\beta} + \mathbf{Q}\mathbf{H}^t\left(\mathbf{H}\mathbf{Q}\mathbf{H}^t + \mathbf{R}\right)^{-1}\left(\mathbf{z} - \mathbf{H}\mathbf{X}\boldsymbol{\beta}\right) \tag{10}$$

~~Eq. (10) is often expressed as $\mathbf{s}_G = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ where $\mathbf{X}\boldsymbol{\beta}$ is the mean and $\boldsymbol{\epsilon} = \mathbf{Q}\mathbf{H}^t\left(\mathbf{H}\mathbf{Q}\mathbf{H}^t + \mathbf{R}\right)^{-1}\left(\mathbf{z} - \mathbf{H}\mathbf{X}\boldsymbol{\beta}\right)$ is the stochastic part of the estimated fluxes . As the~~ In linear Bayesian and geostatistical inverse problems described by equations 7 and 8, the estimated fluxes can be expressed as the sum of the prior information and the update obtained from the observations. In equations 9 and 10, the second term represents the observational constraint, while the first term describes the prior information (in Eq. 9) and the information about fluxes (through $\mathbf{X}$ in Eq. 10). When there is no additional information, the solution corresponds to the prior knowledge. Since the estimate of $\mathbf{s}_G$ in Eq. (10) depends on the unknown $\boldsymbol{\beta}$, it ~~needs to be estimated prior to~~ requires prior estimation of $\boldsymbol{\beta}$ before obtaining $\hat{\mathbf{s}}_G$. The solution for the $\hat{\boldsymbol{\beta}}$ can be obtained from pre-determined quantities as described earlier in the context of Eq. (8) and can be given as:

$$\hat{\boldsymbol{\beta}} = \boldsymbol{\Omega}^{-1}\mathbf{A}^t\boldsymbol{\Psi}^{-1}\mathbf{z}, \tag{11}$$

~~Plugging~~ plugging in $\hat{\boldsymbol{\beta}}$ in Eq. (10) leads to Eq. (12) where all symbols are defined previously or in Eq. (13).

$$\hat{\mathbf{s}}_G = \mathbf{X}\boldsymbol{\Omega}^{-1}\mathbf{A}^t\boldsymbol{\Psi}^{-1}\mathbf{z} + \mathbf{Q}\mathbf{H}^t\boldsymbol{\Psi}^{-1}\left(\mathbf{z} - \mathbf{A}\boldsymbol{\Omega}^{-1}\mathbf{A}^t\boldsymbol{\Psi}^{-1}\mathbf{z}\right), \quad \text{where} \tag{12}$$

$$\mathbf{A} = \mathbf{H}\mathbf{X}, \boldsymbol{\Psi} = \left(\mathbf{H}\mathbf{Q}\mathbf{H}^t + \mathbf{R}\right), \boldsymbol{\Omega} = \left(\mathbf{H}\mathbf{X}\right)^t\left(\mathbf{H}\mathbf{Q}\mathbf{H}^t + \mathbf{R}\right)^{-1}\mathbf{H}\mathbf{X} \tag{13}$$

Note that, $\hat{\mathbf{s}}_B$ and $\hat{\mathbf{s}}_G$ in Eq. (9) and (10) are essentially functions ~~which~~ that are represented by equations. ~~This~~ It is a commonly adopted nomenclature that is used by researchers working in the field of atmospheric inversions. We differentiate Eq. (9) with respect to $\mathbf{s}_{\text{prior}}$, $\mathbf{R}$, $\mathbf{Q}$, $\mathbf{z}$ and Eq. (12) with respect to $\mathbf{X}$, $\mathbf{R}$, $\mathbf{Q}$, $\mathbf{z}$ to obtain the local sensitivities. There are two ways to differentiate $\hat{\mathbf{s}}$ with respect to $\mathbf{z}$, $\mathbf{X}$, $\mathbf{H}$, $\mathbf{Q}$, and $\mathbf{R}$. In the first case, every entry in $\mathbf{z}$, $\mathbf{X}$, $\mathbf{H}$, $\mathbf{Q}$, and $\mathbf{R}$ can be considered as a parameter that results in differentiation of $\hat{\mathbf{s}}$ with respect to these quantities. An "entry" refers to each element of the matrix denoted by ij, where i represents the row number and j represents the column number. On the other hand, if the structures of the covariance matrices $\mathbf{Q}$ and $\mathbf{R}$ are determined by parameters then $\hat{\mathbf{s}}$ can be differentiated just with respect

to these parameters. In the former case, Eq. (9) and (12) are used to differentiate $\hat{\mathbf{s}}$ with respect to an entry at a time in $\mathbf{z}$, $\mathbf{X}$, $\mathbf{H}$, $\mathbf{Q}$, and $\mathbf{R}$. Such an approach of entry-by-entry differentiation is useful if the computational cost in terms of memory constraint is important or if we would like to know the influence of a single entry on $\hat{\mathbf{s}}$. We provide both sets of equations in this ~~work~~manuscript.

2.2.1 LSA with respect to observations, priors, scaling factors, and forward operators

Local sensitivity of $\hat{\mathbf{s}}$ with respect to observations ($\mathbf{z}$) can be given as:

$$\frac{\partial \hat{\mathbf{s}}_B}{\partial \mathbf{z}} = \mathbf{Q}\mathbf{H}^t\mathbf{\Psi}^{-1} \tag{14}$$

$$\frac{\partial \hat{\mathbf{s}}_G}{\partial \mathbf{z}} = \mathbf{X}\mathbf{\Omega}^{-1}\mathbf{A}^t\mathbf{\Psi}^{-1} + \mathbf{Q}\mathbf{H}^t\mathbf{\Psi}^{-1} - \mathbf{Q}\mathbf{H}^t\mathbf{\Psi}^{-1}\mathbf{A}\mathbf{\Omega}^{-1}\mathbf{A}^t\mathbf{\Psi}^{-1}, \tag{15}$$

where all quantities are as defined earlier. The units of the entries in $\frac{\partial \hat{\mathbf{s}}}{\partial \mathbf{z}}$ are μmoles^{-1}m^{2}sec^{-1}ppm^{-1} and the matrices are of dimension $(m \times n)$. These units are inverse of the units of $\mathbf{H}$. Local sensitivities with respect to an observation z_i for both the Bayesian and the geostatistical case can be written as a vector of sensitivities times an indicator for the i^{th} entry i.e. $\frac{\partial \hat{\mathbf{s}}}{\partial \mathbf{z}}\mathbf{e}_i$ where $\mathbf{e}_i = \frac{\partial \mathbf{z}}{\partial z_i}$ is a vector of zeros with the i^{th} entry ~~equals~~ equal to 1.

Note by utilizing $\frac{\partial \hat{\mathbf{s}}}{\partial \mathbf{z}}$, we can also obtain an averaging kernel (or model resolution matrix) and DOFS (see Rodgers, 2000). The averaging kernel matrix for any linear inverse model can be written as:

$$\mathbf{V} = \frac{\partial \hat{\mathbf{s}}}{\partial \mathbf{z}} \times \mathbf{H}, \tag{16}$$

where $\mathbf{V}$ of dimension $(m \times m)$ ~~,~~is the local sensitivity of $\hat{\mathbf{s}}$ with respect to the true unknown fluxes. Then the DOFS can be computed by taking the trace of the averaging kernel matrix $\mathbf{V}$. DOFS represents the amount of information resolved by an inverse model when a set of observations have been assimilated (for a detailed discussion, see Rodgers, 2000 and Brasseur and Jacob, 2017). Theoretically, the value of DOFS cannot exceed the number of observations (n) in ~~case of~~ an underdetermined system and the number of fluxes (m) in ~~case of~~ an overdetermined system.

We can directly compute local sensitivity of $\hat{\mathbf{s}}$ with respect to the prior mean flux $\mathbf{s}_{\text{prior}}$ in the Bayesian case. In the geostatistical case, the prior mean is modeled by two quantities $\mathbf{X}$ and $\boldsymbol{\beta}$. In this scenario, we need to find sensitivities with respect to $\mathbf{X}$ as well as $\boldsymbol{\beta}$. These local sensitivities can be given as:

$$\frac{\partial \hat{\mathbf{s}}_B}{\partial \mathbf{s}_{\text{prior}}} = \mathbf{I} - \mathbf{CH} \tag{17}$$

$$\frac{\partial \hat{\mathbf{s}}_G}{\partial \mathbf{X}} = \mathbf{K}_z \otimes \left(\mathbf{I} + \left(\mathbf{MA}^t - \mathbf{X\Omega}^{-1}\mathbf{A}^t - \mathbf{QH}^t\right)\mathbf{\Psi}^{-1}\mathbf{H}\right) + \left(\mathbf{X\Omega}^{-1} - \mathbf{M}\right) \otimes \left(\mathbf{F}_z - \mathbf{K}_z\mathbf{A}^t\Psi^{-1}\mathbf{H}\right) \tag{18}$$

$$\frac{\partial \hat{\mathbf{s}}_G}{\partial \hat{\boldsymbol{\beta}}} = \mathbf{X} - \mathbf{CA}, \tag{19}$$

where $\mathbf{A} = \mathbf{HX}, \mathbf{B} = \mathbf{QH}^t, \mathbf{C} = \mathbf{B\Psi}^{-1}, \mathbf{\Omega} = \mathbf{A}^t\mathbf{\Psi}^{-1}\mathbf{A}, \mathbf{K}_z = \mathbf{z}^t\mathbf{\Psi}^{-1}\mathbf{A\Omega}^{-1}, \mathbf{M} = \mathbf{CA\Omega}^{-1}$, and $\mathbf{F}_z = \mathbf{z}^t\mathbf{\Psi}^{-1}\mathbf{H}$. The symbol $\otimes$ represents the Kronecker product. The quantity $\frac{\partial \hat{\mathbf{s}}_B}{\partial \mathbf{s}_{\text{prior}}}$ is of dimension $(m \times m)$ and its entries are unitless. The quantity $\frac{\partial \hat{\mathbf{s}}_G}{\partial \hat{\boldsymbol{\beta}}}$ is of dimension $(m \times p)$ and units of the entries in each column of $\frac{\partial \hat{\mathbf{s}}_G}{\partial \hat{\boldsymbol{\beta}}}$ are of the form $(\mu\text{moles}^{-1}\text{m}^2\text{sec}^{-1})(\text{unit of } \beta_i)^{-1}$. The sensitivity matrix $\frac{\partial \hat{\mathbf{s}}_G}{\partial \mathbf{X}}$ is of dimension $(m \times mp)$ where every i^{th} block of m columns $((i-1)m + A : im)$ of $\frac{\partial \hat{\mathbf{s}}_G}{\partial \mathbf{X}}$ has units of the form $(\mu\text{moles}^{-1}\text{m}^2\text{sec}^{-1})(\text{unit of } \mathbf{X}_i)^{-1}$ where $\mathbf{X}_i$ is the i^{th} column of $\mathbf{X}$. Note that ~~,~~the sensitivity matrix $\frac{\partial \hat{\mathbf{s}}_B}{\partial \mathbf{s}_{\text{prior}}}$ in Eq. (17) can also be ~~thought as~~ considered as a proportion of posterior uncertainty to that of the prior uncertainty. In context of the Bayesian case, proportional uncertainty reduction becomes averaging kernel.

Sometimes, it is ~~important~~ essential to know the influence of the prior of any particular grid point or an area consisting of few ~~points on~~ grid-cells within $\hat{\mathbf{s}}$. Local sensitivity of $\hat{\mathbf{s}}$ with respect to the i^{th} entry in $\mathbf{s}_{\text{prior}}$ and $\hat{\beta}_i$ is a matrix of dimension $(m \times 1)$ and can be written as $\frac{\partial \hat{\mathbf{s}}_B}{\partial \mathbf{s}_{\text{prior}}}\mathbf{e}_i$ and $\frac{\partial \hat{\mathbf{s}}_G}{\partial \hat{\boldsymbol{\beta}}}\mathbf{e}_i$ respectively. However, the entry-wise $\frac{\partial \hat{\mathbf{s}}_G}{\partial \mathbf{X}_{ij}}$ is more complex and can be given by:

$$\frac{\partial \hat{\mathbf{s}}_G}{\partial X_{ij}} = \left(\mathbf{I} - \mathbf{CH}\right)\left(\left(\mathbf{I} - \mathbf{X\Omega}^{-1}\mathbf{X}^t\mathbf{H}^t\Psi^{-1}\mathbf{H}\right)\frac{\partial \mathbf{X}}{\partial \mathbf{X}_{ij}}\mathbf{\Omega}^{-1}\mathbf{X}^t + \mathbf{X\Omega}^{-1}\frac{\partial \mathbf{X}^t}{\partial \mathbf{X}_{ij}}\left(\mathbf{I} - \mathbf{H}^t\mathbf{\Psi}^{-1}\mathbf{HX\Omega}^{-1}\mathbf{X}^t\right)\right)\mathbf{F}_z^t, \tag{20}$$

where $\frac{\partial \mathbf{X}^t}{\partial X_{ij}} = \mathbf{E}_{ij}$ is a single-entry matrix with a one for a X_{ij} for which differentiation is being performed and zero everywhere else. For $\mathbf{z}$, entry-by-entry differentiation can be easily performed ~~,~~since both Eq. (9) and (12) result from linear models and are functions of the form $\mathbf{\Phi z} + \mathbf{n}$ where $\mathbf{\Phi}$ and $\mathbf{n}$ are independent of $\mathbf{z}$. For example, $\mathbf{\Phi}$ and $\mathbf{n}$ for Eq. (9) are $\mathbf{QH}^t\left(\mathbf{HQH}^t + \mathbf{R}\right)^{-1}$ and $\mathbf{s}_{\text{prior}} - \mathbf{QH}^t\left(\mathbf{HQH}^t + \mathbf{R}\right)^{-1}\mathbf{Hs}_{\text{prior}}$ respectively and are independent of $\mathbf{z}$. In this case, $\frac{\partial \hat{\mathbf{s}}_B}{\partial z_i}$ can be written as $\mathbf{\Phi e_i}$ where $\boldsymbol{e_i}$ is a single-entry vector with a one for a $\boldsymbol{z}_i$ for which differentiation is being performed and zero everywhere else. Local sensitivity $\frac{\partial \hat{\mathbf{s}}_G}{\partial z_i}$ can similarly be defined for the respective $\mathbf{\Phi}$. Here both the quantities $\frac{\partial \hat{\mathbf{s}}_G}{\partial X_{ij}}$ and $\frac{\partial \hat{\mathbf{s}}_B}{\partial z_i}$ are matrices of dimension $(m \times 1)$.

Local sensitivity of $\hat{\mathbf{s}}$ with respect to an entry in the forward operator has units of the form $\left(\mu\text{moles}^{-1}\text{m}^2\text{sec}^{-1}\right)^2\text{ppm}^{-1}$. In the Bayesian case, this sensitivity can be written as:

$$\frac{\partial \hat{\mathbf{s}}_B}{\partial \mathbf{H}} = \mathbf{Q} \otimes \mathbf{P}_z - \mathbf{BP}_z \otimes \mathbf{C}^t - \mathbf{BC}^t \otimes \mathbf{P}_z - \mathbf{Q} \otimes \mathbf{D} + \mathbf{BD} \otimes \mathbf{C}^t + \mathbf{BC}^t \otimes \mathbf{D} - \mathbf{s}_{\text{prior}} \otimes \mathbf{C}^t, \tag{21}$$

where $\frac{\partial \hat{\mathbf{s}}_B}{\partial \mathbf{H}}$ is a sensitivity matrix of dimension $(m \times mn)$. In the geostatistical case, this sensitivity can be partitioned into two components i.e., $\frac{\partial \hat{\boldsymbol{\beta}}}{\partial \mathbf{H}}$ and $\frac{\partial \hat{\boldsymbol{\epsilon}}}{\partial \mathbf{H}}$ as shown in Eq. (22) where $\frac{\partial \hat{\boldsymbol{\beta}}}{\partial \mathbf{H}}$ and $\frac{\partial \hat{\boldsymbol{\epsilon}}}{\partial \mathbf{H}}$ are obtained in an orderly sequence from Eq. (23) and (24).

$$\frac{\partial \hat{\mathbf{s}}_G}{\partial \mathbf{H}} = \mathbf{X}\frac{\partial \hat{\boldsymbol{\beta}}}{\partial \mathbf{H}} + \frac{\partial \hat{\boldsymbol{\epsilon}}}{\partial \mathbf{H}} \quad \text{where} \tag{22}$$

$$\frac{\partial \hat{\boldsymbol{\beta}}}{\partial \mathbf{H}} = -\mathbf{L} \otimes \mathbf{G}_z - \mathbf{P}_z^t \mathbf{A} \boldsymbol{\Omega}^{-1} \mathbf{X}^t \otimes \mathbf{K}^T + \mathbf{G}_z \mathbf{H}\mathbf{Q} \otimes \mathbf{K}^t + \mathbf{N} \otimes \mathbf{G}_z + \mathbf{L} \otimes \mathbf{P}_z^T - \mathbf{P}_z^T \mathbf{H}\mathbf{Q} \otimes \mathbf{K}^t - \mathbf{N} \otimes \mathbf{P}_z^t \tag{23}$$

$$\frac{\partial \hat{\boldsymbol{\epsilon}}}{\partial \mathbf{H}} = \mathbf{Q} \otimes \mathbf{P}_z - \mathbf{C}\mathbf{z} \otimes \mathbf{C}^t - \mathbf{C}\mathbf{H}\mathbf{Q} \otimes \mathbf{P}_z - \mathbf{X}\mathbf{K}^t\mathbf{z} \otimes \mathbf{C}^T - \mathbf{C}\mathbf{A}\frac{\partial \hat{\boldsymbol{\beta}}}{\partial \mathbf{H}} \tag{24}$$

The expanded form of some of the symbols in Eq. (21) through (24), which have not been expanded yet can be written as $\mathbf{D} = \boldsymbol{\Psi} - \mathbf{H}\mathbf{s}_{\text{prior}}$, $\mathbf{G}_z = \mathbf{z}^t \boldsymbol{\Psi}^{-1} \mathbf{A} \Omega^{-1} \mathbf{A}^t \boldsymbol{\Psi}^{-1}$, $\mathbf{L} = \boldsymbol{\Omega}^{-1} \mathbf{X}^t$, $\mathbf{N} = \boldsymbol{\Omega}^{-1} \mathbf{A}^t \boldsymbol{\Psi}^{-1} \mathbf{H}\mathbf{Q}$, $\mathbf{P}_z = \boldsymbol{\Psi}^{-1}\mathbf{z}$, and $\mathbf{K} = \boldsymbol{\Psi}^{-1}\mathbf{A}\Omega^{-1}$. The quantities $\frac{\partial \hat{\mathbf{s}}_G}{\partial \mathbf{H}}$, $\frac{\partial \hat{\boldsymbol{\beta}}}{\partial \mathbf{H}}$, and $\frac{\partial \hat{\boldsymbol{\epsilon}}}{\partial \mathbf{H}}$ are sensitivity matrices of dimensions $(m \times mn)$, $(p \times mn)$, and $(m \times mn)$ respectively. The units of the entries of $\frac{\partial \hat{\mathbf{s}}}{\partial \mathbf{H}}$ are of the form $(\mu\text{moles}^{-1}\text{m}^2\text{sec}^{-1})^2\text{ppm}^{-1}$.

There might be times when we would like to know the sensitivity of the transport ($\mathbf{H}$) with respect to certain source locations only. In this case, we can use ij form of Eq. (21) through (24) to obtain $\frac{\partial \hat{\mathbf{s}}_B}{\partial H_{ij}}$ in parts. In this formulation, $\frac{\partial \hat{\mathbf{s}}_B}{\partial H_{ij}}$ can be given as:

$$\frac{\partial \hat{\mathbf{s}}_B}{\partial H_{ij}} = \mathbf{C}\frac{\partial \mathbf{H}}{\partial H_{ij}}\left(\mathbf{C}(\mathbf{H}\mathbf{s}_{\text{prior}} - \mathbf{z}) - \mathbf{s}_{\text{prior}}\right) + (\mathbf{Q} - \mathbf{C}\mathbf{H}\mathbf{Q})\left(\frac{\partial \mathbf{H}}{\partial H_{ij}}\right)^t \boldsymbol{\Psi}^{-1}(\mathbf{z} - \mathbf{H}\mathbf{s}_{\text{prior}}) \tag{25}$$

$$\frac{\partial \hat{\mathbf{s}}_G}{\partial H_{ij}} = \mathbf{X}\frac{\partial \hat{\boldsymbol{\beta}}}{\partial H_{ij}} + \frac{\partial \hat{\boldsymbol{\epsilon}}}{\partial H_{ij}}, \quad \text{where} \tag{26}$$

$$\frac{\partial \hat{\boldsymbol{\beta}}}{\partial H_{ij}} = \left(-\mathbf{K}^t \frac{\partial \mathbf{H}}{\partial H_{ij}}\left(\mathbf{X}\mathbf{N} - \mathbf{C}\mathbf{A}\mathbf{S} + \mathbf{Q}\mathbf{H}^t\right) + \mathbf{K}^t\mathbf{H}\mathbf{Q}\frac{\partial \mathbf{H}^t}{\partial H_{ij}}\left(\boldsymbol{\Psi}^{-1}\mathbf{A}\mathbf{S}^t - \mathbf{I}\right) + \boldsymbol{\Omega}^{-1}\mathbf{X}^t\frac{\partial \mathbf{H}^t}{\partial H_{ij}}\left(\mathbf{I} - \boldsymbol{\Psi}^{-1}\mathbf{A}\mathbf{S}\right)\right)\boldsymbol{\Psi}^{-1}\mathbf{z} \tag{27}$$

$$\frac{\partial \hat{\boldsymbol{\epsilon}}}{\partial H_{ij}} = \left(\mathbf{Q}\frac{\partial \mathbf{H}^t}{\partial H_{ij}} - \mathbf{C}\frac{\partial \mathbf{H}}{\partial H_{ij}}\mathbf{Q}\mathbf{H}^t - \mathbf{C}\mathbf{H}\mathbf{Q}\frac{\partial \mathbf{H}^t}{\partial H_{ij}}\right)\boldsymbol{\Psi}^{-1}\left(\mathbf{z} - \mathbf{A}\hat{\boldsymbol{\beta}}\right) - \mathbf{C}\left(\frac{\partial \mathbf{H}}{\partial H_{ij}}\mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{A}\frac{\partial \hat{\boldsymbol{\beta}}}{\partial H_{ij}}\right), \tag{28}$$

where $\mathbf{S} = \mathbf{A}\boldsymbol{\Omega}^{-1}$ and the matrix $\frac{\partial \mathbf{H}}{\partial H_{ij}}$ is a single-entry matrix with a one for a H_{ij} entry for which the differentiation is being performed and zero everywhere else. The quantities $\frac{\partial \hat{\mathbf{s}}_B}{\partial H_{ij}}$, $\frac{\partial \hat{\mathbf{s}}_G}{\partial H_{ij}}$, $\frac{\partial \hat{\boldsymbol{\beta}}}{\partial H_{ij}}$, and $\frac{\partial \hat{\boldsymbol{\epsilon}}}{\partial H_{ij}}$ are sensitivity matrices of dimensions $(m \times 1)$, $(m \times 1)$, $(p \times 1)$, and $(m \times 1)$ respectively. Units of $\frac{\partial \hat{\mathbf{s}}_B}{\partial H_{ij}}$ and $\frac{\partial \hat{\mathbf{s}}_G}{\partial H_{ij}}$ are the same as their kronecker product counterparts.

2.2.2 LSA with respect to error covariance matrices~~and prior information~~

In order to compute the local sensitivities of $\hat{\mathbf{s}}$ with respect to $\mathbf{Q}$ and $\mathbf{R}$, consider that they are parametrized as $\mathbf{Q}(\boldsymbol{\theta}_\mathbf{Q})$ and $\mathbf{R}(\boldsymbol{\theta}_\mathbf{R})$ where $\boldsymbol{\theta}_\mathbf{Q}$ and $\boldsymbol{\theta}_\mathbf{R}$ are the parameter vectors. The differentiation with respect to error covariance parameters in $\mathbf{Q}$ and $\mathbf{R}$ can be accomplished from Eq. (29) through (32) where the subscript i indicates the i^{th} covariance parameter for which differentiation is being performed.

$$\frac{\partial \hat{\mathbf{s}}_B}{\partial \theta_{Q_i}} = (\mathbf{I} - \mathbf{CH}) \frac{\partial \mathbf{Q}}{\partial \theta_{Q_i}} \mathbf{H}^t \mathbf{\Psi}^{-1} (\mathbf{z} - \mathbf{H}\mathbf{s}_{\text{prior}}) \tag{29}$$

$$\frac{\partial \hat{\mathbf{s}}_G}{\partial \theta_{Q_i}} = \left(-\mathbf{X}\mathbf{\Omega}^{-1}\mathbf{A}^T\mathbf{\Psi}^{-1}\mathbf{H} + \mathbf{I} - \mathbf{Q}\mathbf{H}^T\mathbf{\Psi}^{-1}\mathbf{H} + \mathbf{Q}\mathbf{H}^T\mathbf{\Psi}^{-1}\mathbf{A}\mathbf{\Omega}^{-1}\mathbf{A}^T\mathbf{\Psi}^{-1}\mathbf{H} \right) \frac{\partial \mathbf{Q}}{\partial \theta_{Q_i}} \mathbf{H}^T\mathbf{\Psi}^{-1}(\mathbf{z} - \mathbf{A}\mathbf{\Omega}^{-1}\mathbf{A}^T\mathbf{\Psi}^{-1}\mathbf{z}) \tag{30}$$

$$\frac{\partial \hat{\mathbf{s}}_B}{\partial \theta_{R_i}} = -\mathbf{C} \frac{\partial \mathbf{R}}{\partial \theta_{R_i}} \mathbf{\Psi}^{-1} (\mathbf{z} - \mathbf{H}\mathbf{s}_{\text{prior}}) \tag{31}$$

$$\frac{\partial \hat{\mathbf{s}}_G}{\partial \theta_{R_i}} = (-\mathbf{X}\mathbf{\Omega}^{-1}\mathbf{A}^T - \mathbf{B} + \mathbf{C}\mathbf{A}\mathbf{\Omega}^{-1}\mathbf{A}^T)\mathbf{\Psi}^{-1} \frac{\partial \mathbf{R}}{\partial \theta_{R_i}} \mathbf{\Psi}^{-1}(\mathbf{z} - \mathbf{A}\mathbf{\Omega}^{-1}\mathbf{A}^T\mathbf{\Psi}^{-1}\mathbf{z}) \tag{32}$$

All the quantities $\frac{\partial \hat{\mathbf{s}}_B}{\partial \theta_{Q_i}}$, $\frac{\partial \hat{\mathbf{s}}_G}{\partial \theta_{Q_i}}$, $\frac{\partial \hat{\mathbf{s}}_B}{\partial \theta_{R_i}}$, and $\frac{\partial \hat{\mathbf{s}}_G}{\partial \theta_{R_i}}$ are sensitivity matrices of dimension $(m \times 1)$ and the units of the entries of $\frac{\partial \hat{\mathbf{s}}}{\partial \theta_{Q_i}}$ and $\frac{\partial \hat{\mathbf{s}}}{\partial \theta_{R_i}}$ are of the form $(\mu\text{moles}^{-1}\text{m}^2\text{sec}^{-1})(\text{unit of } \theta_{Q_i} \text{ or } \theta_{R_i})^{-1}$. It is also possible to find $\frac{\partial \hat{\mathbf{s}}}{\partial \mathbf{Q}}$ and $\frac{\partial \hat{\mathbf{s}}}{\partial \mathbf{R}}$ directly as shown in Eq. (33) through (36).

$$\frac{\partial \hat{\mathbf{s}}_B}{\partial \mathbf{Q}} = \mathbf{H}^t\mathbf{\Psi}^{-1}(\mathbf{z} - \mathbf{H}\mathbf{s}_{\text{prior}}) \otimes \left(\mathbf{I} - \mathbf{H}^t\mathbf{\Psi}^{-1}\mathbf{B}^t\right) \tag{33}$$

$$\frac{\partial \hat{\mathbf{s}}_G}{\partial \mathbf{Q}} = \left(\mathbf{G}_z - \mathbf{z}^t\right)\mathbf{\Psi}^{-1}\mathbf{H} \otimes \left(\left(\mathbf{B} - \mathbf{M}\mathbf{A}^t + \mathbf{L}^t\mathbf{A}^t\right)\mathbf{\Psi}^{-1}\mathbf{H} - \mathbf{I}\right) \tag{34}$$

$$\frac{\partial \hat{\mathbf{s}}_B}{\partial \mathbf{R}} = \mathbf{\Psi}^{-1}(\mathbf{z} - \mathbf{H}\mathbf{s}_{\text{prior}}) \otimes \mathbf{\Psi}^{-1}\mathbf{H}\mathbf{Q} \tag{35}$$

$$\frac{\partial \hat{\mathbf{s}}_G}{\partial \mathbf{R}} = \left(\mathbf{G}_z - \mathbf{z}^t\right)\mathbf{\Psi}^{-1} \otimes \left(\mathbf{B} - \mathbf{M}\mathbf{A}^t + \mathbf{L}^t\mathbf{A}^t\right)\mathbf{\Psi}^{-1} \tag{36}$$

First two quantities $\frac{\partial \hat{\mathbf{s}}_B}{\partial \mathbf{Q}}$ and $\frac{\partial \hat{\mathbf{s}}_G}{\partial \mathbf{Q}}$ are sensitivity matrices of dimension $(m \times m^2)$. The second set of quantities $\frac{\partial \hat{\mathbf{s}}_B}{\partial \mathbf{R}}$ and $\frac{\partial \hat{\mathbf{s}}_G}{\partial \mathbf{R}}$ are sensitivity matrices of dimension $(m \times n^2)$. Equations (33) through (36) are useful when $\mathbf{Q}$ and $\mathbf{R}$ are fully or partially non-parametric. However, dimensions of these matrices can be quite large and users needs to be careful in realizing the full matrix.

2.3 Global sensitivity analysis (GSA): a variance-based approach

GSA is a process of apportioning the uncertainty in ~~an output estimate~~ output to the uncertainty in ~~each input parameter~~the input parameters. The term "global" stems from ~~the idea of~~ accounting for the effect of all input parameters simultaneously. This is different from ~~"local" sensitivity analysis where the effect of a small~~ LSA, where the impact of a slight change in each parameter on the functional output is considered separately while keeping all other parameters constant. Although quite ~~important, a~~ significant, detailed GSA is challenging as it requires knowledge of the probabilistic variations of all possible combinations (also known as covariance) of the input parameters~~. In atmospheric inverse problems,it is hard to know the joint variation of all input parameters~~,which in most situations is unavailable. However, sometimes it might be possible to know the approximate joint variation of a small subset of the input parameters (e.g. the covariance between $\mathbf{Q}$ and $\mathbf{R}$ parameters). ~~In such case, we can use a variance based approximate method to find the relative contribution of their uncertainties with respect to~~

~~the total flux uncertainty. Note it is also possible to use DGSM (see Sobol and Kucherenko, 2010)~~ Besides the variance-based method, derivative-based global sensitivity measures or the active-subspace technique (see ~~Constantine and Diaz, 2017) in such a scenario. Since the variance based method proposed here doesn't require any~~ Appendix A for discussion) can also be used to conduct GSA. However, this work uses the variance-based method as it does not require sampling and can ~~levearage previously computed derivatives, we adher to this method in this study as an easy extension after LSA.~~

~~The GSA method presented here leverages local sensitivities but actually belongs to the class of variance based methods. This is an approach that addresses the contribution to the total variance of the estimated fluxes. This is an approximate method unlike the exact decomposition technique of Sobol using conditional variances. It applies a simple~~ leverage previously computed partial derivatives. It uses a first-order Taylor~~'s approximation around~~ 's approximation of parameter estimates to ~~obtain an approximate representation. This approach~~ compute global sensitivities. This technique has been used in many research works, including environmental modeling (e.g., Hamby, 1994) and life cycle assessment (Groen et al., 2017; Heijungs, 1996), among others.

Broadly, we can consider $\hat{\mathbf{s}}$ as a function of the covariates $\mathbf{Q}, \mathbf{R}, \mathbf{H}, \mathbf{X}(\text{or } \mathbf{s}_{\text{prior}})$, and $\mathbf{z}$ i.e. $\hat{\mathbf{s}} = \mathbf{f}(\mathbf{Q}, \mathbf{R}, \mathbf{H}, \mathbf{X}(\text{or } \mathbf{s}_{\text{prior}}), \mathbf{z})$. We can then compute how uncertainties of the individual components of $\mathbf{f}$ are accounted for in the overall uncertainty of $\hat{\mathbf{s}}$ by applying multivariate Taylor series expansion of $\hat{\mathbf{s}}$ about its mean. Approximation up to first-order polynomial of the Taylor series expansion leads to the equation:

$$\text{Var}(\hat{\boldsymbol{s}}) = \left(\frac{\partial \hat{\mathbf{s}}}{\partial \boldsymbol{\theta}}^{t} \mathbf{W}_{\boldsymbol{\theta}} \frac{\partial \hat{\boldsymbol{s}}}{\partial \boldsymbol{\theta}} \right)_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} + \text{Error, where}$$

where $\boldsymbol{\theta} = (\boldsymbol{\theta}_{\boldsymbol{Q}}, \boldsymbol{\theta}_{\boldsymbol{R}}, \boldsymbol{\theta}_{\boldsymbol{H}}, \boldsymbol{\theta}_{\mathbf{X}}(\text{or } \mathbf{s}_{\text{prior}}), \boldsymbol{\theta}_{\boldsymbol{z}})$ is the vector of parameters and $\mathbf{W} = \text{Var}(\boldsymbol{\theta})$ is the covariance matrix of the parameters.~~It is however,~~

It is challenging to estimate ~~some of the individual~~ covariance quantities such as the cross-covariance between $\boldsymbol{\theta}_{\boldsymbol{R}}$ and $\boldsymbol{\theta}_{\boldsymbol{H}}$ or between $\boldsymbol{\theta}_{\boldsymbol{H}}$, and $\boldsymbol{\theta}_{\boldsymbol{Q}}$ to get the best possible ~~decomposition~~ estimate of the total uncertainty of $\hat{\mathbf{s}}$. Assuming no cross-covariance between $\mathbf{Q}$ and $\mathbf{R}$ and ignoring other parameters not related to the variance parameters, the diagonal of the variance of the posterior fluxes can be approximated as:

$$\text{Var}(\hat{s}_i) = \sum_{j=1}^{L} \left(\frac{\partial \hat{\boldsymbol{s}}}{\partial \theta_{Q_j}} \right)_i^2 \text{Var}\left(\theta_{Q_j}\right) + \sum_{k=1}^{M} \left(\frac{\partial \hat{\boldsymbol{s}}}{\partial \theta_{R_k}} \right)_i^2 \text{Var}\left(\theta_{R_k}\right) \Bigg|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}, \tag{37}$$

~~Where~~ where the subscript i on the right-hand side of Eq. (37) refers to the i^{th} entry of the derivative vector, which is a scalar and parameters θ_{Q_j} and θ_{R_k} refer to the j^{th} and k^{th} parameters of the sets $\boldsymbol{\theta}_{\boldsymbol{Q}}$ and $\boldsymbol{\theta}_{\boldsymbol{R}}$ respectively. From Eq. (37), we can see how uncertainty in the flux estimate is apportioned ~~into variance components~~ between variance of $\boldsymbol{\theta}_{\boldsymbol{Q}}$ and $\boldsymbol{\theta}_{\boldsymbol{R}}$~~of an inversion framework~~. No normalization is necessary in such a framework ~~of GSA since~~ as, variance components on the right hand side

of Eq. (37) ~~, the variance components~~ are naturally weighted~~in such a way that both sides have same units~~ , resulting in the same units of measurement. Once the two ~~components~~ parts of $V_{\hat{s}_i}$ (i.e. Eq. (37)) are computed, they can also be summed over the solution space (e.g. number of gridcells $\times$ number of ~~time-periods~~periods) of $\hat{\mathbf{s}}$ and ranked to find the relative importance of the parameters.

Even after simplification, implementation of Eq. (37) is ~~difficult~~ complex as it requires knowledge of the uncertainties associated with the parameters of $\mathbf{Q}$ and $\mathbf{R}$ that are generally not known. ~~Note that, it is also possible to have a complete apportionment of the variance of $\hat{\mathbf{s}}$ for all the parameters of $\mathbf{f}$ at least up to the first-order polynomial in the Taylor's series. However, its implementation is difficult since it requires knowledge of the covariances of all the parameters.~~ We do not further discuss GSA in the context of the case study presented in this work, but we have shown its application with respect to $\mathbf{Q}$ and $\mathbf{R}$ in the MATLAB Livescript.

~~Other than the variance based Taylor series methoddescribed above~~ Besides the variance-based method, there are many ~~other approaches to perform GSA~~different approaches for performing GSA, as described in ~~the introductory section but either they are~~ Appendix. A. However, they are either computationally expensive or assume independence of the input parameters, which is not the case in atmospheric inverse problems. We do not pursue other approaches for quantifying GSA associated with $\mathbf{Q}$ and $\mathbf{R}$ as they would lead to similar results and would not add anything substantial to the contributions of this study.

2.4 Ranking importance of covariates, covariance parameters, and observations from LSA

In atmospheric inverse modeling, we encounter two situations while ranking the importance of parameters. These are ranking of parameters when they have the same or different units. The situation of ranking ~~of parameters with same units arise~~ parameters with the same units arises when we want to study the influence of a group of parameters~~like observations that have~~ , like observations with the same units. Comparatively, the ~~situation of~~ ranking of parameters with different units ~~arise~~ occurs when we want to ~~study the influence~~ explore the impact of groups of parameters ~~that have different units~~ with dissimilar units of measurements, like observations in $\mathbf{z}$ in comparison to the variance of observations in $\mathbf{R}$. Both these situations can be accounted ~~through GSA that is~~ for in GSA described in Sec. 2.3. However, GSA in most scenarios in atmospheric inverse modeling cannot be ~~fully~~ performed due to the reasons mentioned earlier. Therefore, in this work~~we adopted~~ , we adopt a regression-based approach to rank the importance of parameters. The proposed approach utilizes output from LSA, accounts for multicollinearity, and results in importance scores that are bounded between 0 to 1. We define the regression model for ranking as:

$$\hat{\mathbf{s}} = \mathbf{E}\boldsymbol{\gamma} + \boldsymbol{\xi}, \tag{38}$$

where $\hat{\mathbf{s}}$ are fluxes obtained from an inversion, and $\mathbf{E}$ is an $(m \times \text{number of derivatives})$ matrix of the previously estimated sensitivities. The vector of unknown coefficients $\boldsymbol{\gamma}$ is of dimension $(\text{number of derivatives} \times 1)$, and $\boldsymbol{\xi}$ is an $(m \times 1)$ vector of unobserved errors associated with the regression model. To exemplify, $\mathbf{E}$ in Eq. (38) can be arranged as:

$$\mathbf{E} = \left[\frac{\partial \hat{\mathbf{s}}}{\partial \mathbf{z}} \; \frac{\partial \hat{\mathbf{s}}}{\partial \mathbf{Q}} \; \frac{\partial \hat{\mathbf{s}}}{\partial \mathbf{R}} \; \cdot \; \cdot \right] \tag{39}$$

In a regression-based approach, as described in Eq. (38), multicollinearity between independent variables in $\mathbf{E}$ can pose a problem for determining the importance of independent variables in influencing $\boldsymbol{\Gamma}$. To avoid this problem, we ~~computed~~ compute relative importance weights by using the method outlined in Johnson, 2000. These weights are ~~computed~~ obtained by first deriving uncorrelated orthogonal counterparts of the covariates in $\mathbf{E}$ and then regressing $\hat{\mathbf{s}}$, on $\mathbf{E}$ to get importance weights for each covariate. The ~~weights are standardized by the~~ coefficient of determination then standardizes the weights, i.e., R^2 such that they range between 0 to 1 with the ~~sum of all the weights being~~ aggregated sum of 1. Implementation of this method is included in the Livescript submitted with this manuscript.

Note Least Absolute Shrinkage and Selection Operator (LASSO) or Principal Component Analysis (PCA) can also ~~be employed to compute ranking~~ rank parameters under multicollinearity. However, both these methods result in ~~weights that are unbounded~~ unbounded weights. Furthermore, "inference after selection" is ambiguous ~~in linear regression which is the case~~ for LASSO coefficients (see Berk et al., 2013 or chapter 6 of Hastie et al., 2015 for details). Consequently, interpreting the LASSO coefficients as ranks may not be the best approach.

The regression-based approach described above can ~~be employed when we want to~~ rank parameters with ~~both~~ the same and different units of measurement. However, an additional normalization step is required ~~if we are interested in getting~~ to get the overall rank of the parameters ~~that have different units~~ with varying units of measure, like in $\mathbf{z}$, $\mathbf{Q}$, and $\mathbf{R}$. To perform this normalization, first, each column in every sensitivity matrix (e.g. $\frac{\partial \hat{\mathbf{s}}}{\partial \mathbf{z}}$, $\frac{\partial \hat{\mathbf{s}}}{\partial \mathbf{Q}}$, and so forth) that is to be ranked is normalized (min-max normalization; see Vafaei et al., 2020) between ~~0 to 1. Following which~~0 to 1. After which, all columns for a sensitivity matrix are summed and renormalized to vary between ~~0 to 1. This results~~ 0 to 1, resulting in one column ~~that is representative of~~ representing a sensitivity matrix for a particular group. We denote this by the subscript "grouped" (e.g. $\frac{\partial \hat{\mathbf{s}}}{\partial \mathbf{z}_{\text{grouped}}}$) in latter sections.

Once the normalized sensitivity vectors are obtained for each group, the regression methodology as described above can be used to rank the importance of each group. The ranking methodology proposed above does not account for the non-linear relationship between estimates of the fluxes and the derivatives. If this is a concern, then the strength of the ~~nonlinear~~ non-linear relationship among the derivative vectors can be first obtained by computing distance correlation between fluxes and the local derivatives of the parameters. ~~After which we can employ variable transformation (e.g.,~~ If necessary, variable transformation techniques such as Box-Cox transformation ~~;~~ (see Sakia, 1992) ~~before applying~~ can be employed before adopting the regres-

sion methodology described above.

Note that ~~most analytical inversions use DOFS to diagnose information content of an inversion. DOFS = 0~~ in most batch inversion methods, DOFS is used to assess the information content provided by observations. DOFS $= 0$ in these inversions implies that no informational gain happened~~in an inversion~~. In this case, the estimated flux reverts ~~back~~ to prior. In Eq. (38), this means that the $\boldsymbol{\gamma}$ coefficient that corresponds to $\mathbf{Q}$ would have the ~~largest~~ most significant impact. Likewise if DOFS is large, then the $\boldsymbol{\gamma}$ coefficients for $\mathbf{z}$ and $\mathbf{R}$ would be larger (and likely correlated). We show this correspondence in Sec. 3.

Finally, all ~~different kinds of diagnostic methods that are~~ diagnostic methods applied in the context of any regression-based model can be used ~~for understanding~~ to understand the relationship between dependent and independent variables~~. However~~; however, what covariates to include in $\mathbf{E}$ depends on the specific case study under consideration.

3 Results~~: Los Angeles methane inversion case study~~

To demonstrate the applicability of our methods, we utilize data from our published work on CH_4 fluxes in the Los Angeles megacity (see Yadav et al., 2019). In this previous work, fluxes were estimated for South Coast Air Basin (SoCAB) region (see Fig. 3) at 0.03° spatial (1826 grid-cells) and 4-day temporal resolution from the Jan 27, 2015 through Dec 24, 2016. However, in the current work, we utilize input data from Oct 23, 2015, through Oct 31, 2015~~that~~ , which is a single inversion period, to contextualize the applicability of our methods. This period overlaps with the beginning of the well-studied Aliso Canyon gas leak (Conley et al., 2016). ~~We do not extend our analysis for the full duration of the previous study as this is not the objective of this workand all the details associated with computing the inverse flux estimates can be found in that work. Furthermore, in the Livescript we present our sensitivity based equations with respect to the geostatistical approach to inverse modeling as this was the approach adopted in the previous study~~As in previous work, $\mathbf{R}$ and $\mathbf{Q}$ are assumed to be diagonal with separate parameter for each site in $\mathbf{R}$ and a single parameter that governs the scaling of errors in $\mathbf{Q}$. Similarly, $\mathbf{X}$ is a column vector consisting of the prior estimates of CH_4 fluxes.

For each observation included in the case study, a forward operator was obtained by using Weather Research Forecasting-Stochastic Time Inverted Lagrangian Model (see Yadav et al., 2019). These forward operators are used to demonstrate the application of the methodology for building ~~IOAMI and JSD based~~ IAOMI and JSD-based correlation matrices in the MATLAB Livescript. They are also used ~~in conjunction with measurements ,~~ with measurements and prior information to estimate the fluxes and perform LSA.

3.1 STAD from the forward operators

In this work, we identify STAD for the 4-day period for which the inversion was performed. The spatial domain of the study over this ~~time~~ period is uniquely disaggregated by STAD, as shown in Fig. 3. The STAD for different sites ~~are~~ is mostly

spatially contiguous~~but for some sites~~. Still, for some monitoring sites, we found isolated grid cells ~~which~~ that were not within the ~~contiguous~~ adjacent zones. We ~~have~~ manually combined these with STAD for the nearest site to create a spatially continuous map, as shown in Fig. 3. The discontinuous version of the STAD shown in Fig. 3 is included in the Livescript. The discontinuities in the STAD result ~~mostly from~~ mainly from an unequal number of observations across sites and ~~indicates~~ indicate that aggregation over ~~longer time-period~~ a more extended period is required to ~~completely identify a noise free~~ identify a noise-free STAD. We do not investigate the ~~time-period~~ period of this aggregation as this is beyond the scope of this work.

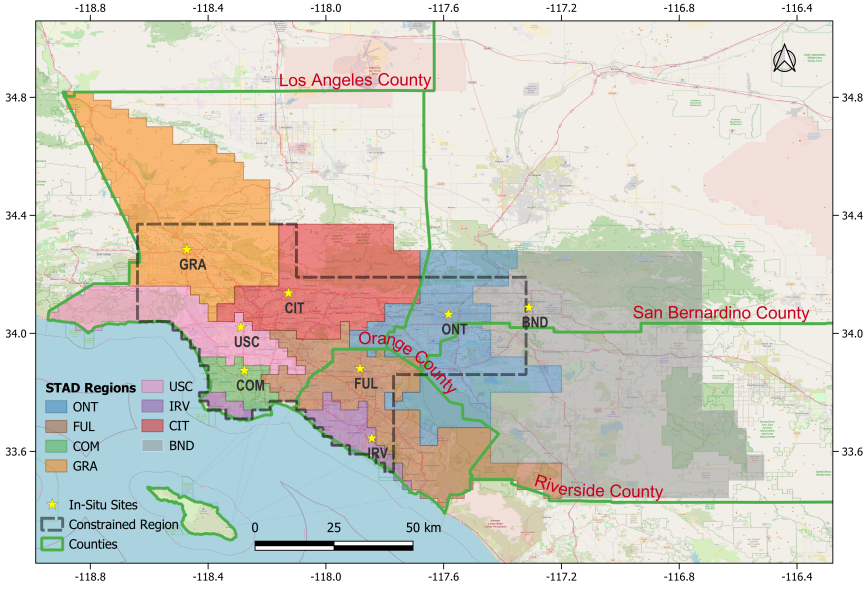

Figure 3. Study area with county boundaries, measurement locations and the Spatio-temporal Area of Dominance of measurement locations. The black dotted line shows the area constrained by observations, as shown in Yadav et al., 2019.

Overall, ~~the~~ STAD for each site indicates spatial regions of fluxes ~~that contributes~~ over a period that contribute most to the observational ~~(e.g. CH4 enhancement) signal . This in turn allows us to sub-divide the spatio-temporal variations in fluxes or enhancements by the STAD regions.~~ signal observed at a site allowing us to associate the change in fluxes to the specific area in the basin where reductions or increases in emissions are likely to have occurred. Some information in the observational signal is shared between observations from different sites. This shared information (though not shown) can be computed as part of STAD and forms part of overall basin-scale estimates of fluxes that combines measurements from all sites. Note that STAD does not represent the network's coverage, i.e., regions of emissions constrained by observations. These regions are shorter than STAD (see the grey outline in Fig. 3). They are obtained before performing an inversion by identifying areas of continuous spatiotemporal coverage as provided by atmospheric transport (Fig. 4) or by assessing the model resolution after performing an inversion (for an explanation, see Yadav et al., 2019).

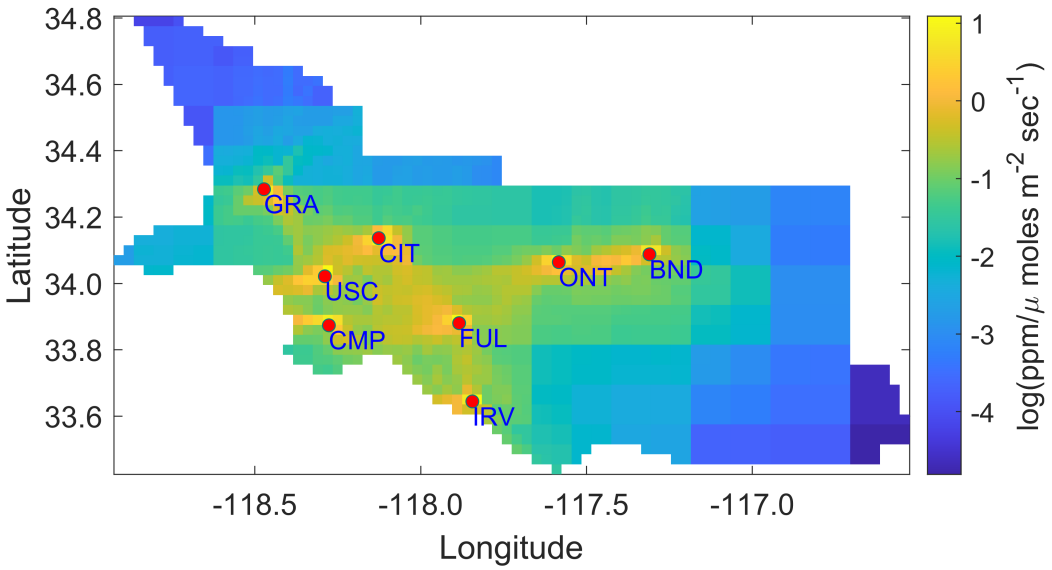

Figure 4. ~~Study area with county boundaries, measurement locations and the Spatio-Temporal Area of Dominance of measurement locations~~Heatmap of the aggregated forward operators for the case study period.

3.2 Sensitivity analysis

One of the main goals of the sensitivity analysis after performing an inversion is to identify the observations that had the most influence on the flux estimates. Other than observations, it is also ~~important~~ essential to explore the importance of ~~other~~ different inputs to an inversion, like variance parameters in $\mathbf{R}$. We describe the process of performing this analysis within the context of the case study mentioned in Sec. 3~~. This section~~ , which discusses the relative importance of the input quantities in influencing $\hat{\mathbf{s}}$~~by utilizing the~~ , by utilizing local sensitivities.

3.2.1 Comparison and ranking of the observations

Importance of ~~the~~ individual measurements in influencing $\hat{\mathbf{s}}$, can be easily computed through the relative importance methodology described in section 2.4. Although ~~,~~ all entries of $\frac{\partial \hat{\mathbf{s}}}{\partial \mathbf{z}}$ are in same units of measurement, direct ranking of observations or sites without employing the relative importance technique can lead to misleading results~~. This~~ , which happens due to the presence of large negative and positive values in $\frac{\partial \hat{\mathbf{s}}}{\partial \mathbf{z}}$ that are governed by the overall ~~spatio-temporal spread,~~ spatiotemporal spread, the intensity of forward operators, and ~~observations with large~~ high enhancements.

For the case study in this work, we find that observations collected at the GRA site that is located nearest to the source of the Aliso Canyon gas leak are most influential in governing $\hat{\mathbf{s}}$, as shown by site-based rankings in Table 1. These rankings primarily show the importance of observations from a site in influencing the estimated fluxes for the ~~time~~ period in consideration

Site	Importance Score	Rank
GRA	0.26	1
ONT	0.24	2
COM	0.13	3
IRV	0.11	4
BND	0.10	5
CIT	0.07	6
FUL	0.07	7
USC	0.06	8

Table 1. The importance scores and ranking of 8 sites based on the sensitivity of the estimated fluxes ($\hat{\mathbf{s}}$) to observations ($\mathbf{z}$).

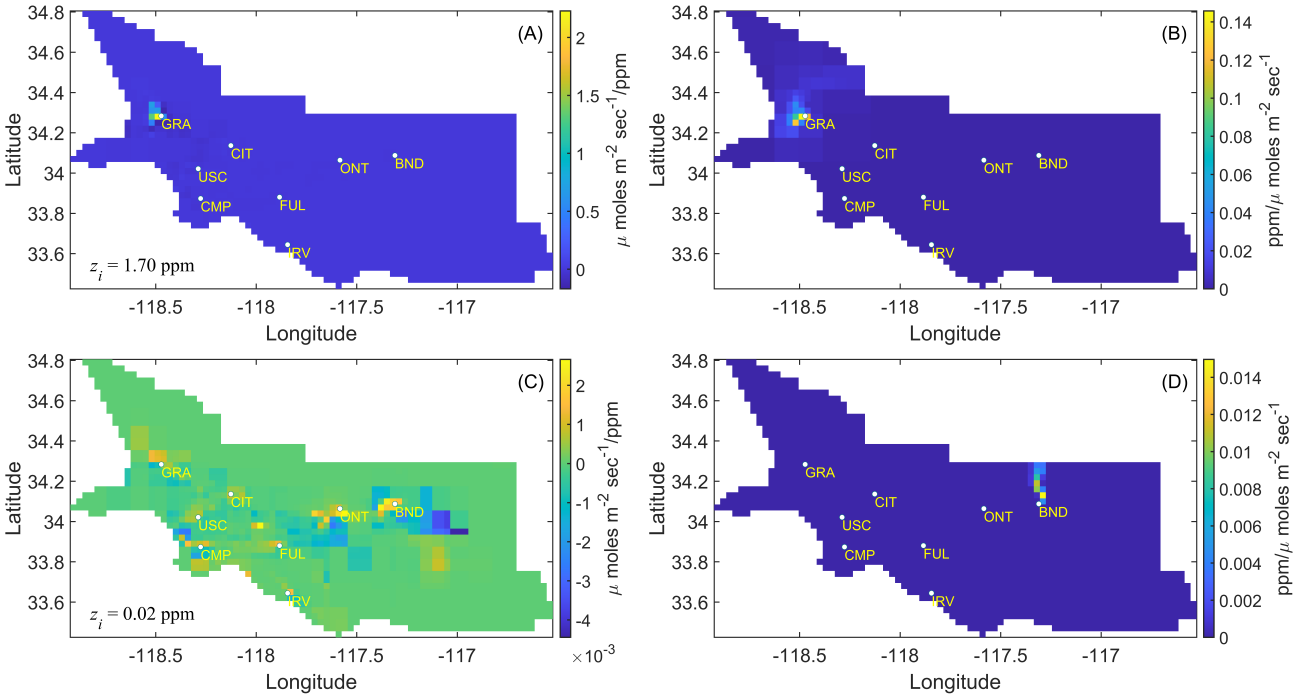

Figure 5. The sensitivities ($\frac{\partial \hat{\mathbf{s}}}{\partial \mathbf{z}_i}$) and forward ~~operator~~ operators of the most and least important ~~observation in inversions. Subplot~~ observations are shown here. Subplots A and C ~~show~~ depict the sensitivity of $\hat{\mathbf{s}}$ with respect to the most (A) and least ~~important~~ (C) ~~observation~~important observation, respectively, during the case study period. The CH_4 enhancement ~~associated with~~ corresponding to these observations is shown in the bottom left corner of the subplots and ~~identified~~ denoted by the symbol z_i. The right subplots, B and D~~show~~, display the forward operators associated with the sensitivities shown in subplots A and C, respectively.

. Observation based assessment of and are obtained by summing the weights for each observation by employing the relative importance methodology.

Outliers have a significant impact on these rankings. The high weight associated with even one observation from a site can make that site more important compared to other sites. For example, if we remove the observation with the highest weight from each site, ONT is the most important site, followed by GRA, CMP, IRV, CIT, FUL, BND, and USC. As part of sensitivity analysis, examining the influence of the observations associated with high weights is crucial because they are likely to have an enormous impact on the flux estimates. Site level importance should be judged not only by examining the aggregated ranking as presented in Table 1 but also by looking at the distribution of weights shown through the boxplot in the Livescript associated with section 3.2. A site with evenly distributed weights is more important than one whose importance is just due to the presence of a few observations with high weights.

The ranking of each observation in influencing the estimates of fluxes can be obtained by examining the weights of the column vectors of $\frac{\partial \hat{\mathbf{s}}}{\partial \mathbf{z}}$ resulted in ranking an observation with , and is provided in the Livescript. To exemplify, this ranking of weights showed that observation from the GRA site with the enhancement of 1.7 ppm was most important, whereas an observation from the largest enhancement of 1.7 ppm to be most important. Contrarily, an observation for the BND site that had with an enhancement of 0.02 ppm is 0.02 ppm was found to be least important in influencing $\hat{\mathbf{s}}$s. Note this is not an observation with the lowest enhancement but with the lowest influence. The most and least important observation along with their corresponding forward operators are shown in least influence (Fig. 5).

3.2.2 Relative importance of $\mathbf{Q}, \mathbf{R}, \mathbf{X}, \boldsymbol{\beta}$, and $\mathbf{z}$

After the two-step normalization of $\frac{\partial \hat{\mathbf{s}}}{\partial \mathbf{z}}$, $\frac{\partial \hat{\mathbf{s}}}{\partial \mathbf{X}}$, $\frac{\partial \hat{\mathbf{s}}}{\partial \mathbf{H}}$, $\frac{\partial \hat{\mathbf{s}}}{\partial \boldsymbol{\beta}}$, $\frac{\partial \hat{\mathbf{s}}}{\partial \mathbf{Q}}$, and $\frac{\partial \hat{\mathbf{s}}}{\partial \mathbf{R}}$ as described in section 2.4 , the spatial plots of all these grouped quantities that we call as $\frac{\partial \hat{\mathbf{s}}}{\partial \mathbf{z}}_{\text{grouped}}$, $\frac{\partial \hat{\mathbf{s}}}{\partial \mathbf{X}}_{\text{grouped}}$, $\frac{\partial \hat{\mathbf{s}}}{\partial \mathbf{H}}_{\text{grouped}}$, $\frac{\partial \hat{\mathbf{s}}}{\partial \boldsymbol{\beta}}_{\text{grouped}}$, $\frac{\partial \hat{\mathbf{s}}}{\partial \mathbf{Q}}_{\text{grouped}}$, and $\frac{\partial \hat{\mathbf{s}}}{\partial \mathbf{R}}_{\text{grouped}}$ can be created to explore the regions of the low and high weights (see Fig.5 6) at the grid scale.

Figure 6 shows that the weights of $\frac{\partial \hat{\mathbf{s}}}{\partial \mathbf{X}}_{\text{grouped}}$ is lower in the regions well constrained by the observations. However, opposite is true in the case of $\frac{\partial \hat{\mathbf{s}}}{\partial \mathbf{Q}}_{\text{grouped}}$ and Some of these quantities are correlated and should be seen in conjunction. For example, $\mathbf{R}$ describes errors in $\mathbf{z}$, among other errors, and implies that $\frac{\partial \hat{\mathbf{s}}}{\partial \mathbf{R}}_{\text{grouped}}$. This implies , that data constrained regions have lower posterior uncertainty thereby increasing the influence of prescribed or estimated uncertainty parameters. There is smoothness in the weights of $\frac{\partial \hat{\mathbf{s}}}{\partial \mathbf{Q}}_{\text{grouped}}$ in the domain except around some sites (ONT, FUL, and IRV), which is an indication that the estimates of $\hat{\mathbf{s}}$ remain insensitive to and $\frac{\partial \hat{\mathbf{s}}}{\partial \mathbf{z}}_{\text{grouped}}$ should be evaluated together to understand their importance in influencing flux estimates. Similarly $\mathbf{Q}$ describes errors in $\mathbf{s} - \mathbf{X}\boldsymbol{\beta}$ implying that $\frac{\partial \hat{\mathbf{s}}}{\partial \mathbf{Q}}_{\text{grouped}}$ and $\frac{\partial \hat{\mathbf{s}}}{\partial \mathbf{X}}_{\text{grouped}}$, should be assessed together to understand their importance in influencing flux estimates. A larger value of $\frac{\partial \hat{\mathbf{s}}}{\partial \mathbf{z}}_{\text{grouped}} + \frac{\partial \hat{\mathbf{s}}}{\partial \mathbf{R}}_{\text{grouped}}$ is likely to be found around in-situ sites due to increased model resolution. However, if around these locations $\frac{\partial \hat{\mathbf{s}}}{\partial \mathbf{R}}_{\text{grouped}}$ is larger in comparison of $\frac{\partial \hat{\mathbf{s}}}{\partial \mathbf{z}}_{\text{grouped}}$ then it suggests that errors in $\mathbf{R}$ should be adjusted and therefore observations should be more important in governing the

flux estimates around in-situ sites. In this case study, this is due to the ~~$\mathbf{Q}$ parameter in these regions. These relationships can be quantified by assessing correlation between local sensitivities and $\hat{\mathbf{s}}$ as shown in Fig.~~ large variability in the enhancement caused by the Aliso Canyon leak and the presence of large point sources near in-situ sites. Overall, for the exact location, a larger $\frac{\partial \hat{\mathbf{s}}}{\partial \mathbf{z}_{\text{grouped}_i}}$ should be accompanied by a lower $\frac{\partial \hat{\mathbf{s}}}{\partial \mathbf{R}_{\text{grouped}_i}}$, as confirmed by the correlation subplots A and B of Fig. 7.

~~There is strong evidence of multicollinearity among covariates in explaining (e.g. see first column of the Fig. 7). The direction of the best fit line appears to be in sync with the expectation regarding CH_4 fluxes in the region during that time period. Thus, $\frac{\partial \hat{\mathbf{s}}}{\partial \mathbf{z}_{\text{grouped}}}$ is positively correlated with $\hat{\mathbf{s}}$, which implies that higher enhancement in $\mathbf{z}$ leads to an increase in the estimated fluxes. Similarly $\frac{\partial \hat{\mathbf{s}}}{\partial \boldsymbol{\beta}_{\text{grouped}}}$ is also positively correlated with $\hat{\mathbf{s}}$ implying that any increase in the scaling factor increases the estimated fluxes. The negative relationship of $\frac{\partial \hat{\mathbf{s}}}{\partial \mathbf{X}_{\text{grouped}}}$ and $\hat{\mathbf{s}}$ just indicates that an increase in $\frac{\partial \hat{\mathbf{s}}}{\partial \mathbf{X}_{\text{grouped}}}$ inversely influences the magnitude of the estimated fluxes. This occurs as $\hat{\mathbf{s}}$ reverts to $\mathbf{X}$ in regions unconstrained by observations whereas opposite happens in areas constrained by observations that in the context of the case study includes sources of largest fluxes.~~

The increased model resolution also results in lower importance of $\frac{\partial \hat{\mathbf{s}}}{\partial \mathbf{X}_{\text{grouped}}}$ and $\frac{\partial \hat{\mathbf{s}}}{\partial \mathbf{Q}_{\text{grouped}}}$, around sites. However, areas unconstrained by observations are likely to have larger $\frac{\partial \hat{\mathbf{s}}}{\partial \mathbf{X}_{\text{grouped}}} + \frac{\partial \hat{\mathbf{s}}}{\partial \mathbf{Q}_{\text{grouped}}}$ as seen in Fig. 6 for $\frac{\partial \hat{\mathbf{s}}}{\partial \mathbf{X}_{\text{grouped}}}$ and $\frac{\partial \hat{\mathbf{s}}}{\partial \mathbf{Q}_{\text{grouped}}}$, quantities. If in locations constrained by observations, $\frac{\partial \hat{\mathbf{s}}}{\partial \mathbf{Q}_{\text{grouped}_i}}$ is larger in comparison to $\frac{\partial \hat{\mathbf{s}}}{\partial \mathbf{X}_{\text{grouped}_i}}$ then $\mathbf{X}$in these locations is incorrect and needs adjustment. Likewise, in the case of $\frac{\partial \hat{\mathbf{s}}}{\partial \mathbf{R}_{\text{grouped}}}$ a larger $\frac{\partial \hat{\mathbf{s}}}{\partial \mathbf{X}_{\text{grouped}_i}}$ is generally accompanied by lower $\frac{\partial \hat{\mathbf{s}}}{\partial \mathbf{z}_{\text{grouped}_i}}$ and vice versa, which is also visible in the correlation subplots C and D in Fig. 7. Quantity $\frac{\partial \hat{\mathbf{s}}}{\partial \boldsymbol{\beta}_{\text{grouped}}}$ provides information about the grid-cells that are determining the value of $\hat{\beta}$ and in this case study as expected this is around Aliso Canyon leak whose X_i is being adjusted due to the larger flux from that region. This can also be seen in subplot E in Fig. 7 where it is positively correlated with $\hat{\mathbf{s}}$.

4 Discussion

This study lays out techniques to assess the quality of the inferred estimates of fluxes. Sensitivity analysis is an important diagnostic tool to understand the impact of the choices made with respect to inputs on the estimated fluxes. However, it is not a recipe for selecting the proper forms of $\mathbf{X}$ or the structure of $\mathbf{Q}$ or $\mathbf{R}$ before performing an inversion. Other tools or methods such as Bayesian Information Criterion, Variance Inflation Factor should be used to perform this task.

The case study in this work is designed only to demonstrate the methodologies described in Sec. 2. We do not impose non-negativity constraints to obtain positive CH_4 fluxes as was done in the original 2019 study (Yadav et al., 2019). This is done because posterior likelihood changes its functional form under non-negativity constraints ~~and~~ that invalidate the analytical forms of ~~senstivity~~ sensitivity equations presented in this work~~become invalid~~. Thus, some CH_4 fluxes obtained in this study have negative values as can be seen in the map of $\hat{\mathbf{s}}$ in the MATLAB Livescript. ~~However, even~~ Even in these situations assessing sensitivity through an inversion without the imposition of non-negativity is ~~useful~~ helpful as it provides insights into the role of $\mathbf{z}$, $\mathbf{R}$, $\mathbf{Q}$, and $\mathbf{X}$ in governing estimates of non-negative $\hat{\mathbf{s}}$.

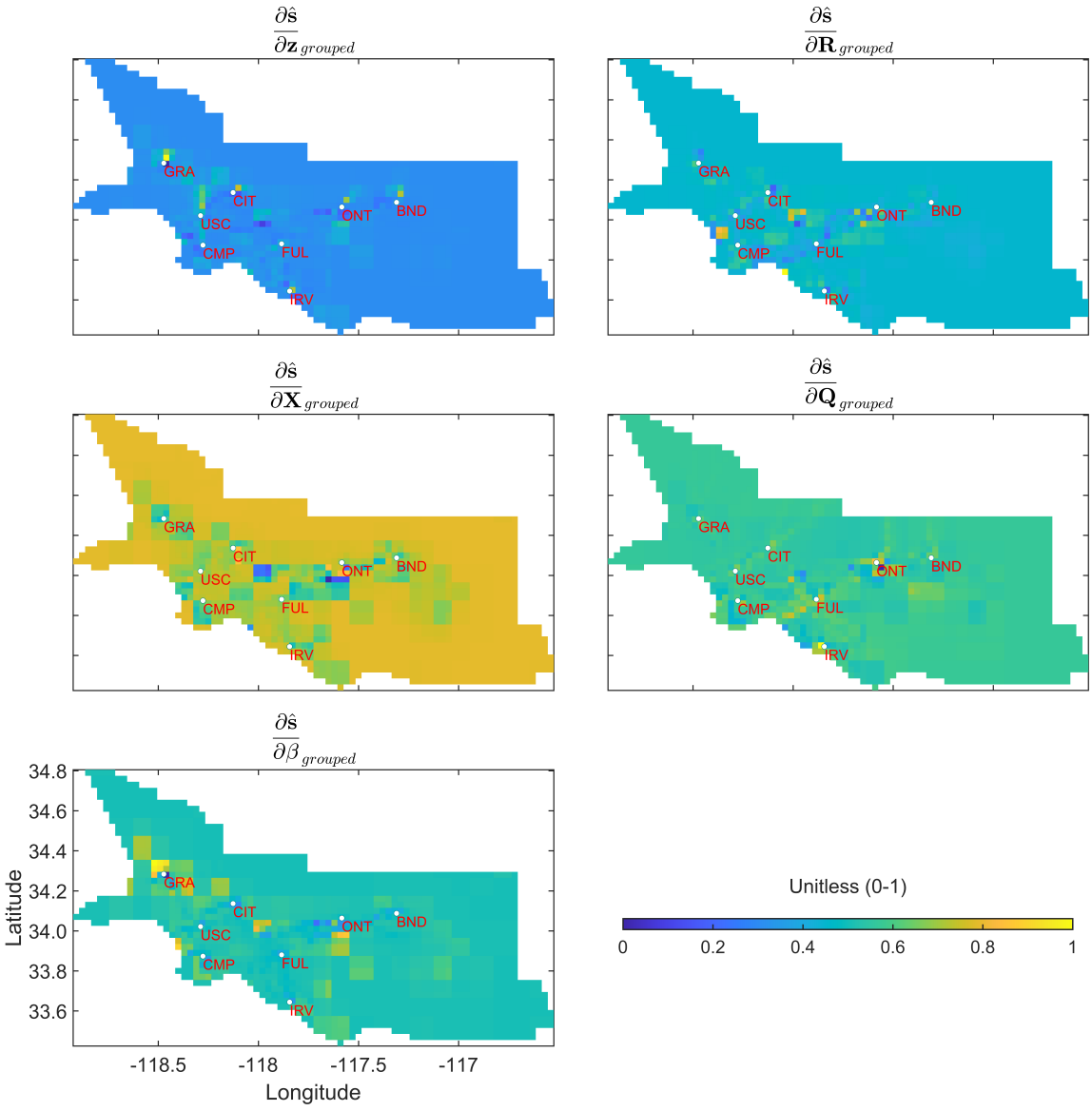

Figure 6. Grouped local sensitivities of the estimated fluxes ($\hat{\mathbf{s}}$) with respect to $\mathbf{z}$, $\mathbf{R}$, $\mathbf{X}$, $\mathbf{Q}$, and $\boldsymbol{\beta}$ from top-left to bottom-right respectively. Note, in the case of $\frac{\partial \hat{\mathbf{s}}}{\partial \mathbf{z}}_{\text{grouped}}$, $\frac{\partial \hat{\mathbf{s}}}{\partial \mathbf{R}}_{\text{grouped}}$, and $\frac{\partial \hat{\mathbf{s}}}{\partial \mathbf{X}}_{\text{grouped}}$ two-step normalization is performed to generate subplots associated with these quantities. Derivatives with respect to: (1) observations in $\mathbf{z}$, (2) parameters in $\mathbf{R}$, and (3) entries in $\mathbf{X}$ are normalized between 0 and 1 and then after aggregating these for every grid-cell another Min-Max normalization is performed to limit their ranges between 0 and 1. Only single normalization is performed in case of $\frac{\partial \hat{\mathbf{s}}}{\partial \mathbf{Q}}_{\text{grouped}}$ and $\frac{\partial \hat{\mathbf{s}}}{\partial \boldsymbol{\beta}}_{\text{grouped}}$ as they consist of only one parameter.

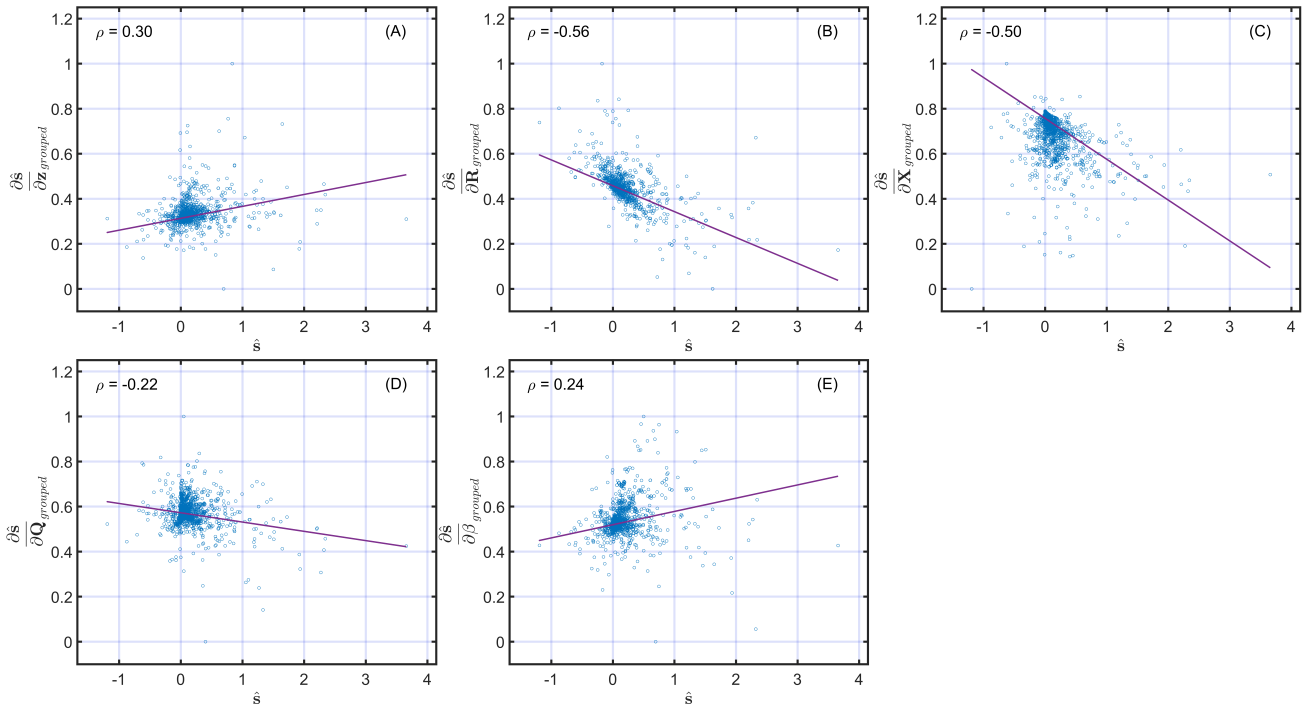

Figure 7. Scatterplots of relationships between $\hat{\mathbf{s}}$ and $\frac{\partial \hat{\mathbf{s}}}{\partial \mathbf{z}}_{\text{grouped}}$, $\frac{\partial \hat{\mathbf{s}}}{\partial \mathbf{R}}_{\text{grouped}}$, $\frac{\partial \hat{\mathbf{s}}}{\partial \mathbf{X}}_{\text{grouped}}$, $\frac{\partial \hat{\mathbf{s}}}{\partial \mathbf{Q}}_{\text{grouped}}$, $\frac{\partial \hat{\mathbf{s}}}{\partial \boldsymbol{\beta}}_{\text{grouped}}$. Note as before in Fig. 6 all the derivatives are normalized to limit their range between 0 and 1. The correlation coefficient of the relationships shown in each scatterplot is reported on the top right corner of the subplots. The least square line of best fit is shown in red color in every subplot.

Like $\mathbf{z}$, the importance of $\mathbf{Q}$ and $\mathbf{R}$ parameters can be directly obtained when all parameters have the same units ~~. This happens~~ of measurement as in the case study presented in this ~~work~~study. However, this is not guaranteed as $\mathbf{R}$ can be a function of variance parameters and ~~spatio-temporal correlation length~~ spatiotemporal correlation lengths expressed in the distance units in space and time. Furthermore, a nonstationary error covariance $\mathbf{R}$ can have parameters that have even more complicated units. This situation is not limited to $\mathbf{R}$ ~~and~~ but also applies to the prior error covariance $\mathbf{Q}$ and $\mathbf{X}$. Under these conditions, ~~a comparison between~~ comparing the sensitivity matrices is only possible after normalization. Therefore, ~~for comparative assessment we recommend use of~~ we recommend using a multiple linear ~~regression based~~ regression-based relative importance method to rank these quantities for comparative assessment.

The overall importance of $\frac{\partial \hat{\mathbf{s}}}{\partial \mathbf{z}}$ is best explored by performing ~~column based~~ column-based normalization and then employing the relative importance method. Additionally, column based normalization can be augmented by row-based normalization to assess and rank the influence of observations in governing ~~gridscale~~ grid-scale estimates of $\hat{\mathbf{s}}$. Qualitatively, column and row-based assessment increase our understanding about the ~~spatio-temporal~~ spatiotemporal estimates of $\hat{\mathbf{s}}$~~. This~~ ^, which is especially important when point sources are the dominant sources of emissions. Moreover, it ~~also provides an insight into~~ provides

insight into the temporal aggregation error (e.g. Thompson et al., 2011) as the information encoded in an instantaneous measurement can get lost over the coarser ~~time-period of inversion~~ inversion period. This aggregation error also manifests spatially and is determined by the resolution at which fluxes are obtained. ~~Note in many situations~~In many situations, these aggregation errors are unavoidable as the choice of the ~~spatio-temporal~~ spatiotemporal resolution of inversions is governed by the density of observations in space and time.

Other than aggregation error, the aggregation of the estimated fluxes also has profound implications as it affects the robustness of the estimated fluxes. It can be proved (see Appendix C1) that aggregation of $\hat{\mathbf{s}}$ in space and time from an inversion conducted at finer resolution leads to reduction in uncertainty. However, even though ratio of observations to the estimated fluxes increases, the number of fluxes uniquely resolved declines at coarser resolution (see Appendix C1).

The computational cost to calculate analytical partial derivatives is minimal as it is a onetime operation and is bounded by the computational cost to perform matrix multiplications, which at max is $O(n^3)$. For the case study presented~~in this work~~, we can compute analytical derivatives and rank ~~approximately 4000~~ for approximately 4000 parameters in few minutes on a laptop. Computing derivatives by using the Kronecker form of equations (Eq. (18), (21) through (24), and (33) though (36)) is faster for ~~small~~ smaller problems. However~~for large inverse problems~~, for large problems, the storage costs associated with these equations can become prohibitive. In these situations, we propose the use of ij form of the equations (Eq. (20), (25) through (28), and (29) though (32)) for assessment. Furthermore, computational problems can also arise in ranking the inputs if we have ~~large number~~ numerous derivatives (e.g. greater than 10,000), as the ranking method used in this work relies on ~~eigen value~~ eigenvalue decomposition that has $O(n^3)$ computational complexity. To overcome this problem, we advise grouping of derivatives to reduce the dimension of the problem.

Finally, the estimation of STAD and the importance of sites can be influenced by data gaps~~therefore~~ ; therefore, it is not advised in presence of vast differences in the number of observations between sites.

5 Conclusions

Our work makes ~~novel and major~~ a novel and significant contributions that can ~~significantly improve~~ improve the understanding of linear atmospheric inverse problems. It provides ~~:~~ (1) a framework for post hoc analysis of the impact of inputs on the estimated fluxes and (2) a way to understand the correlations in the forward operators or atmospheric transport model. The authors are ~~not aware~~ unaware of any work where local sensitivities with different units of measurement are compared to rank the importance of inputs in a linear atmospheric inverse model.

~~With respect to~~ Concerning forward operators, we provide mathematical foundations for ~~IOAMI, and Jensen-Shannon based~~ IAOMI and JSD-based metrics. These two metrics can be used to construct ~~and accommodate a non-stationary~~ a nonstationary

error covariance for the atmospheric transport component of the model-data mismatch matrix $\mathbf{R}$. Furthermore, ~~IOAMI based~~ IAOMI-based assessments can be extended to identify STAD from forward operators that can help in disaggregating regions of influence of the observations over a chosen temporal duration. This assists in understanding the connection between the sources of fluxes and observations from a particular measurement location.

The ~~IOAMI and JSD based~~ IAOMI and JSD-based metrics provide an ~~important~~ essential insight into the two critical and only required components for an inversion~~that is~~ : observations and forward operators (e.g., ~~influence of an~~ the influence of observation to the sources of fluxes through STAD)~~. This task~~ , which can be accomplished ~~prior to~~ before conducting an inversion and should be ~~complimented~~ complemented by post hoc LSA, which is ~~a necessity~~ necessary for understanding the behavior of an inverse model. Overall, LSA can answer questions like for which locations and in what order of precedence was an observation important in influencing the estimated fluxes. This kind of analysis is entirely different from estimating uncertainty~~that tells us reduction in~~ , which tells us the prior uncertainty reduction due to observations.

LSA is not a replacement for statistical tests that check ~~the~~ inverse models' underlying assumptions and model specifications~~in inverse models~~. Neither is it a recipe for selecting inputs to an inverse model. However, as explained above, it has an ~~important role as explained above~~ essential role that can lead to an improved understanding of an atmospheric inverse model.

Code and data availability. All the code and data utilized in this study are submitted as supplementary material.

Appendix

Appendix A: Review of previously employed methods to conduct sensitivity analyses

Earlier, many methods have been proposed and utilized to perform sensitivity analysis. These can be categorized as global and local sensitivity analyses. Global sensitivity analysis (GSA) includes Morris's (e.g. Morris, 1991) one step at a time method (OAT), Polynomial Chaos Expansion (PCE) (e.g. Sudret, 2008), Fourier amplitude sensitivity test (FAST) (e.g. Xu and Gertner, 2011), Sobol's method (e.g. Sobol, 2001) and Derivative based global sensitivity measures (DGSM) (e.g. Sobol and Kucherenko, 2010) among others. These existing GSA methods (1) assume independence of parameters (e.g., FAST and OAT), or (2) computationally expensive (e.g., Sobol's method), or (3) require knowledge of the joint probability distribution of the parameter space (e.g., DGSM, PCE). Therefore, these traditional methods cannot be directly applied in linear atmospheric inverse problems, which consists of tens of thousands of non-normal, spatiotemporally correlated parameters (including observations). Constantine and Diaz, 2017 proposed an active subspace-based GSA that uses a low-dimensional approximation of the parameter space. But it is still

computationally expensive for problems with thousands of parameters (see case study in Constantine and Diaz, 2017).

Compared to GSA, a local sensitivity method like Bayesian Hyper Differential Sensitivity Analysis (HDSA) (Sunseri et al., 2020) computes partial derivatives concerning maximum a posteriori probability (MAP) estimates of a quantity of interest. However, unlike Bayesian HDSA, we do not generate samples from the prior estimate to compute multiple MAP points since we have limited knowledge of the prior distribution of the spatiotemporally correlated parameters. We derive the functional form of the local sensitivity equations based on the closed-form MAP solution. Our method is simple and amenable to tens of thousands of parameters. Note that, like all linear atmospheric inverse problems, one of the critical goals of this work is to study the importance of thousands of spatiotemporally varying parameters by ranking them, and computation of the local sensitivities is a means to achieve that goal.

Appendix B: Jensen-Shannon distance (JSD) for forward operators

The dissimilarity between forward operators can also be measured via entropy (for definition, see MacKay et al., 2003) based distances, which can capture differences between two probability distributions. One such metric is Jensen-Shanon distance (JSD) (Nielsen, 2019), which can be used to compute the distance between two forward operators after normalizing them by their total sum. For a forward operator F this can be given as:

$$P_{F_k} = \frac{F_k}{\sum_k F_k} \tag{B1}$$

where F_k denotes k^{th} entry of $\mathbf{F}$ resulting in normalized forward operator P. We can then use JSD to compute distance between two normalized forward operators from equation B:

$$JSD\left(P_{\mathbf{F}}||P_{\mathbf{G}}\right) =$$

$$]\tfrac{1}{2}D\left(P_{\mathbf{F}}||M\right) + \tfrac{1}{2}D\left(P_{\mathbf{G}}||M\right)$$

where D stands for Kulback-Leibler (KL) divergence (see MacKay et al., 2003 for details). KL divergence D of any probability distribution p with respect to another probability distribution q is defined as: $D(p||q) = \sum p\log(p/q)$ and M stands for $\frac{1}{2}\left(P_{\mathbf{F}} + P_{\mathbf{G}}\right)$. The symbol $||$ is used to indicate that $D\left(P_{\mathbf{F}}||M\right)$ and $D\left(P_{\mathbf{G}}||M\right)$ are not conditional entropies (see MacKay et al., 2003). JSD is closed and bounded in $[0, 1]$ when KL divergence is computed with base 2 logarithm. Intuitively, JSD and $1 - \nu$ (i.e. 1-IAOMI) are comparable since both of them are measures of dissimilarity.

Appendix C: Uncertainty and model resolution under aggregation

Here we show the proofs of two mathematical statements on the robustness and quality of the estimated fluxes as mentioned in Sec. 4. First, we show why marginal variance of the estimated fluxes (which is the diagonal of covariance matrix of $\hat{\mathbf{s}}$) decrease when estimated fluxes are post aggregated to a coarser scale or upscaled (A). Second, we show why in such case the model resolution (also termed as, total information resolved by the observations) also decreases (B). Note that, the nomenclature used in the appendix should not be confused with the nomenclature introduced in Sec. 2. The abbreviations and symbols used here are independent of what are used in the Sec. 2.

Appendix D: ~~Proof of the reduction of marginal variance of $\hat{\mathbf{s}}$ when upscaling is performed~~

C1 Proof of the reduction of marginal variance of $\hat{\mathbf{s}}$ when aggregation is performed

Post inversion aggregation or upscaling of any flux field $\mathbf{s}$ is equivalent to pre-multiplication by a weight matrix (in fact, a row stochastic matrix). This can be written as:

$$\tilde{\mathbf{s}} = \mathbf{J}\hat{\mathbf{s}}, \tag{C1}$$

~~Where~~ where $\mathbf{J}$ is a row stochastic (i.e. row-sums are all unity) $k \times m$ weight matrix ($k < m$). Variance of $\tilde{\mathbf{s}}$ can be written as $\mathbf{J\Sigma J}^t$ where var($\tilde{\mathbf{s}}$) = $\mathbf{J}$var($\hat{\mathbf{s}}$)J^t=$\mathbf{J\Sigma J}^t$. The general structure of $\mathbf{J}$ is as follows:

$$J = \begin{bmatrix} 0 & j_{12} & j_{13} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ j_{21} & \mathbf{0} & j_{2r+1} & j_{2r+2} & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & j_{km} & \mathbf{0} & \mathbf{0} \end{bmatrix} = \begin{bmatrix} \mathbf{j}_1^t \\ \mathbf{j}_2^t \\ \vdots \\ \mathbf{j}_k^t \end{bmatrix} \tag{C2}$$

However, $\mathbf{J}$ is mostly sparse~~and values in~~, with non-zero values in only a few places. ~~Rest~~ The rest of the entries are zeros. Essentially, $\mathbf{J}$ can have any number of non-zero entries in a row that may or may not be consecutive. This is because~~although on a map,~~, although adjacent grids are averaged on a map, they may not be adjacent upon vectorization. Moreover, the geometry of the map may not be exactly square or rectangular. ~~This means~~Therefore, depending on the aggregation or upscaling factor and geometry, ~~for any particular grid,~~ there may or may not be any neighboring grid for averaging around a particular grid. However, the rows are linearly independent, as nearby grids are considered only once for averaging. The properties of $\mathbf{J}$ are as follows:

1. $\mathbf{J1} = 1$ or $\mathbf{j}_i^t\mathbf{1} = 1 \quad \forall i = 1, 2, \cdot, \cdot, k$
2. $\mathbf{j}_i^t\mathbf{j}_r = 0$ for $i \neq r$

We can rearrange the columns of $\mathbf{J}$ and the rows of $\mathbf{\Sigma}$ accordingly without loss of any structure such that non-zero entries are consecutive for each row of $\mathbf{J}$. Matrix $\mathbf{J\Sigma J'}$ under column permutation can be written as:

$$\mathbf{J\Sigma J^t} = \mathbf{J}_\pi \mathbf{\Sigma}_\pi \mathbf{J}_\pi^t = \begin{bmatrix} \mathbf{l}_1^t & 0 & \dots & 0 \\ 0 & \mathbf{l}_2^t & \dots & 0 \\ \vdots & \vdots & \ddots & . \\ 0 & 0 & \dots & \mathbf{l}_k^t \end{bmatrix}^{k\times m} \begin{bmatrix} \mathbf{\Xi}_{11} & \mathbf{\Xi}_{12} & \dots & \mathbf{\Xi}_{1k} \\ \mathbf{\Xi}_{21} & \mathbf{\Xi}_{22} & \dots & . \\ \vdots & \vdots & \ddots & . \\ \mathbf{\Xi}_{k1} & . & \dots & \mathbf{\Xi}_{kk} \end{bmatrix}^{m\times m} \begin{bmatrix} \mathbf{l}_1 & 0 & \dots & 0 \\ 0 & \mathbf{l}_2 & \dots & 0 \\ \vdots & \vdots & \ddots & . \\ 0 & 0 & \dots & \mathbf{l}_k \end{bmatrix}^{p\times k} \tag{C3}$$

$$= \begin{bmatrix} \mathbf{l}_1^t \mathbf{\Xi}_{11} \mathbf{l}_1 & . & \dots & \mathbf{l}_1^t \mathbf{\Xi}_{1k} \mathbf{l}_k \\ . & \mathbf{l}_2^t \mathbf{\Xi}_{22} \mathbf{l}_2 & \dots & . \\ \vdots & \vdots & \ddots & . \\ \mathbf{l}_k^t \mathbf{\Xi}_{k1} \mathbf{l}_1 & . & \dots & \mathbf{l}_k^t \mathbf{\Xi}_{kk} \mathbf{l}_k \end{bmatrix}^{k\times k} \tag{C4}$$

where $\mathbf{J}_\pi$ and $\mathbf{\Sigma}_\pi$ are the permuted $\mathbf{J}$ and Σ respectively. However, for notational clarity, we use $\mathbf{l}$ and $\mathbf{\Xi}$ as the sub-vector and sub-block-matrix of the $\mathbf{J}_\pi$ and $\mathbf{\Sigma}_\pi$ respectively. Note that, any $\mathbf{l}_i^t$ is a row-vector of dimension $(1, d_i)$, and $\mathbf{\Xi}_{ii}$ is a square matrix of dimension (d_i, d_i) where $\sum_{i=1}^k d_i = m$. Thus, diagonal entry $\mathbf{l}_i^t \mathbf{\Xi}_{ii} \mathbf{l}_i$ is a scalar quantity. For any i^{th} diagonal entry, the corresponding scalar quantity can be written as $\sum_{jrl} l_{ij} l_{ir} \Xi_{jr}$. By symmetry of $\mathbf{\Xi}$, this reduces to

$$\mathbf{l}_i^t \mathbf{\Xi}_{ii} \mathbf{l}_i = \sum_r l_{ir}^2 \Xi_{lr}^2 + 2 \sum_{j>r} l_{ij} l_{ir} \Xi_{jr} \tag{C5}$$

By Cauchy Squartz inequality on Ξ_{jr}, this can be written as

$$\sum_r l_{ir}^2 \sigma_{lr}^2 - 2 \sum_{j>r} l_{ij} l_{ir} \sigma_{jj} \sigma_{rr} \leq \sum_r l_{ir}^2 \sigma_{rr}^2 + 2 \sum_{j>r} l_{ij} l_{ij} \sigma_{jr} \leq \sum_r l_{ir}^2 \sigma_{rr}^2 + 2 \sum_{j>r} l_{ij} l_{ij} \sigma_{jj} \sigma_{rr} \tag{C6}$$

$$\left(l_{ir} \sqrt{\sigma_{ir}} - \sum_{r\geq 2} l_{ir} \sqrt{\sigma_{ir}} \right)^2 \leq \sum_r l_{ir}^2 \sigma_{rr}^2 + 2 \sum_{j>r} l_{ij} l_{ij} \sigma_{jj} \sigma_{rr} \leq \left(\sum_{ir} l_{ir} \sqrt{\sigma_{rr}} \right)^2 \tag{C7}$$

$$\min_r \sigma_{rr} \left(l_{ir} - \sum_{r\geq 2} l_{ir} \right)^2 \leq \sum_r l_{ir}^2 \sigma_{rr}^2 + 2 \sum_{j>r} l_{ij} l_{ij} \sigma_{jj} \sigma_{rr} \leq \max_r \sigma_{rr} \left(\sum_{ir} l_{ir} \right)^2 \tag{C8}$$

This implies (by property 1 of the weight matrix $\mathbf{J}$) that the i^{th} diagonal entry is bounded by:

$$\min_r \sigma_{rr} \left(l_{ir} - \sum_{r\geq 2} l_{ir} \right)^2 \leq \mathbf{J_i' \Sigma_{ii} J_i} \leq \max_r \sigma_{rr} \leq \sum_{r=1}^{d_i} \sigma_{rr} \tag{C9}$$

where $\sum_{r=1}^{d_i} \sigma_{rr}$ is the sum of the marginal variance of the ~~ith block of un-averaged~~ i^{th} block of unaveraged $\hat{\mathbf{s}}$. Thus, sum of the marginal variance of $\tilde{\mathbf{s}}$ which is the sum of the i^{th} diagonal $\mathbf{J}_i^t \mathbf{\Sigma}_{ii} \mathbf{J}_i$ is also smaller or ~~equals~~ equal to the sum total of marginal variance of $\hat{\mathbf{s}}$. ~~Clearly, we see that under upscaling or averaging, diagonal of the variance matrix shrinks in magnitude from the un-averaged one. As a consequence, it implies that~~ This implies that the marginal variance of the posterior mean decreases as a result of the diagonal of the variance matrix shrinking in magnitude upon averaging.

Appendix D: ~~Proof of the reduction in model resolution when upscaling is performed~~

C1 Proof of the reduction in model resolution when aggregation is performed

~~Upscaled~~ Aggregated forward operator $\tilde{\mathbf{H}}$ can be written as:

$$\tilde{\mathbf{H}} = \mathbf{HB} \text{where is the upscaling matrix,} \tag{C1}$$

where $\mathbf{B}$ is the upscaling matrix. Dimension of $\mathbf{B}$ has the dimension of transpose of $\mathbf{J}$. ~~Structual~~ Structural form of $\mathbf{B}$ is similar to the form of $\mathbf{J}$ explained in C2. Non-zero entries of $\mathbf{B}$ are in the same place as $\mathbf{J}'$ with magnitude replaced by unity. This is evident from the fact that forward operator is summed instead of being averaged for ~~upscaling~~aggregation. Properties of $\mathbf{B}$ are as follows:

1. $\mathbf{B1} = \mathbf{1}$
2. $\mathbf{JB} = diag(\mathbf{N})^{k\times k}$ where $\mathbf{N}$ is the vector of number of neighboring ~~gridcells~~ grid-cells for any particular ~~gridcell~~ grid-cell i.e. $\mathbf{N} = (N_1, \ldots, N_k)$
3. $\mathbf{BJ} = \begin{bmatrix} \mathbf{C_1} & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{0} & \mathbf{C_2} & \ldots & \mathbf{0} \\ \vdots & \vdots & \ddots & . \\ \mathbf{0} & . & \ldots & \mathbf{C_k} \end{bmatrix}^{m\times m}$ is a block diagonal matrix. Any block $\mathbf{C}_i$ of $\mathbf{JA}$ can be expressed as a varying dimension (depending on the number of neighboring grids of any particular ~~gridcell~~grid-cell) matrix of form:

$$\mathbf{C}_i = \begin{bmatrix} \frac{1}{N_i} & \cdots & \frac{1}{N_i} \\ \vdots & \ddots & \vdots \\ \frac{1}{N_i} & \cdots & \frac{1}{N_i} \end{bmatrix}^{N_i\times N_i} = \frac{1}{N_i}\mathbf{1}\mathbf{1}^t \tag{C2}$$

4. $\mathbf{BJ}$ is symmetric and positive semi-definite

First three properties are simple observations from the construction. So, here we provide proof of the fourth property.

Proof. By construction, $Det(\mathbf{BJ} - \lambda\mathbf{I}) = Det(\mathbf{C_1} - \lambda\mathbf{I}) \ldots Det(\mathbf{C_k} - \lambda\mathbf{I})$. So, ~~eigen values~~ eigenvalues of $\mathbf{BJ}$ are the list of ~~eigen values~~ eigenvalues of the block matrices. It can be proved that 1 and 0 are the only two distinct ~~eigen values~~ eigenvalues of $\mathbf{C}_i$ for any i. Below here is a brief argument on that:

$\left(\frac{1}{N_i}\mathbf{1}\mathbf{1}^t\right)\mathbf{1} = \frac{1}{N_i}\mathbf{1}N_i = 1\cdot\mathbf{1}$ implies one ~~eigen value~~ eigenvalue of $\mathbf{C_i}$ is 1. Observe that, $rank\left(\frac{1}{N_i}\mathbf{1}\mathbf{1}^t\right) = rank(\mathbf{1}) = 1$. Hence, dimension of null space $dim\left(\mathcal{N}\left(\frac{1}{N_i}\mathbf{1}\mathbf{1}^t\right)\right) = k - rank\left(\frac{1}{N_i}\mathbf{1}\mathbf{1}^t\right) = k - 1$. This implies that the other ~~eigen value~~ eigenvalue of $\mathbf{C_i}$ is 0 with multiplicity $k-1$.

So, not only $\mathbf{C}_i$ is symmetric but also the ~~eigen values~~ eigenvalues $\mathbf{C}_i$ are always non negative. Consequently, all ~~eigen values~~ eigenvalues of $\mathbf{BJ}$ are of similar form i.e. $\mathbf{BJ}$ is symmetric positive semidefinite. □

Finally, model resolution matrix for inversion can be written as $\frac{\partial \hat{\mathbf{s}}}{\partial \mathbf{z}}\mathbf{H}$ where $\mathbf{H}$ is the forward operator operator. Post inversion aggregated model-resolution can be written as:

$$\frac{\partial \tilde{\mathbf{s}}}{\partial \mathbf{z}}\tilde{\mathbf{H}} = \mathbf{A}\frac{\partial \hat{\mathbf{s}}}{\partial \mathbf{z}}\mathbf{HB} \quad \text{By Eq. (C1) and C1} \tag{C3}$$

The question is what happens to the trace of the model-resolution under the ~~upscaled case~~aggregated scenario? We provide a proof for the simple batch Bayesian case in lemma C1. Proof for the geostatistical case is similar and left for the enthusiastic readers.

Lemma 1.

$$\mathbf{Mres} = \mathbf{QH'}\boldsymbol{\psi}^{-\mathbf{1}}\mathbf{H}$$

$$\mathbf{Mres}_{aggregated} = \mathbf{JQH'}\boldsymbol{\psi}^{-\mathbf{1}}\mathbf{HB} \quad \textit{then}$$

$$trace\,(\mathbf{Mres}_{aggregated}) \leq trace(\mathbf{Mres}) \tag{C4}$$

Proof. odel resolution for the aggregated ~~case~~ scenario can be written as:

$$\text{trace}\,(\mathbf{Mres}_{\text{aggregated}}) = \text{trace}(\mathbf{JQH'}\boldsymbol{\psi}^{-\mathbf{1}}\mathbf{HB}) = \text{trace}(\mathbf{BJQH'}\boldsymbol{\psi}^{-\mathbf{1}}\mathbf{H}) = \text{trace}(\mathbf{WS}) \text{ where } \mathbf{W} = \mathbf{BJ}, \mathbf{S} = \mathbf{QH'}\boldsymbol{\psi}^{-\mathbf{1}}\mathbf{H}, \tag{C5}$$

~~Where~~ where $\mathbf{S}$ and $\mathbf{W}$ are both of dimension $(m \times m)$. $\mathbf{S}$ is a positive semidefinite matrix since both $\mathbf{Q}$ and $\mathbf{H'}\boldsymbol{\psi}^{-\mathbf{1}}\mathbf{H}$ are positive semidefinite. For $\mathbf{W}^{m\times m}$ and $\mathbf{S}^{m\times m}$ positive semidefinite, trace of their product can be bounded by the following quantities (see Kleinman and Athans, 1968 and discussion in Fang et al., 1994):

$$\lambda_{min}(\mathbf{W})trace(\mathbf{S}) \leq trace(\mathbf{WS}) \leq \lambda_{min}(\mathbf{W})trace(\mathbf{S}) \tag{C6}$$

By Property 4 of the weight matrix $\mathbf{B}$, we know that $\lambda_{min}(\mathbf{W}) = 0$ and $\lambda_{max}(\mathbf{W}) = 1$, hence the above reduces to $0 \leq trace(\mathbf{WS}) \leq 1 \cdot trace(\mathbf{S})$. Hence is the proof by C5.

□

Author contributions. V.Y., and S.G. contributed equally in preparing the manuscript.

Competing interests. The authors declare no competing interest.

Acknowledgements. The authors thank Anna Karion, Kimberly Mueller, James Whetstone (National Institute of Standards and technology, NIST), and Daniel Cusworth (University of Arizona, UA) for their review and advice on the manuscript. This work was partially funded by NIST's Greenhouse Gas Measurements Program. Support to University of Notre Dame provided by NIST grant 70NANB19H132. Support for JPL was provided via an interagency agreement between NIST and NASA. A portion of this research was carried out at JPL, California Institute of Technology, under a contract with NASA (80NM0018D0004).

References

Berk, R., Brown, L., Buja, A., Zhang, K., and Zhao, L.: Valid post-selection inference, The Annals of Statistics, pp. 802–837, 2013.

Bouchard, M., Jousselme, A.-L., and Doré, P.-E.: A proof for the positive definiteness of the Jaccard index matrix, International Journal of Approximate Reasoning, 54, 615–626, 2013.

Brasseur, G. P. and Jacob, D. J.: Modeling of atmospheric chemistry, Cambridge University Press, 2017.

Cha, S.-H.: Comprehensive survey on distance/similarity measures between probability density functions, City, 1, 1, 2007.

Conley, S., Franco, G., Faloona, I., Blake, D. R., Peischl, J., and Ryerson, T.: Methane emissions from the 2015 Aliso Canyon blowout in Los Angeles, CA, Science, 351, 1317–1320, 2016.

Constantine, P. G. and Diaz, P.: Global sensitivity metrics from active subspaces, Reliability Engineering & System Safety, 162, 1–13, 2017.

Enting, I. G.: Inverse problems in atmospheric constituent transport, Cambridge University Press, 2002.

Fang, Y., Loparo, K. A., and Feng, X.: Inequalities for the trace of matrix product, IEEE Transactions on Automatic Control, 39, 2489–2490, 1994.

Gelman, A. and Hill, J.: Data analysis using regression and multilevel/hierarchical models, Cambridge university press, 2006.

Ghosh, S., Mueller, K., Prasad, K., and Whetstone, J.: Accounting for transport error in inversions: An urban synthetic data experiment, Earth and Space Science, 8, e2020EA001 272, 2021.

Groen, E. A., Bokkers, E. A., Heijungs, R., and de Boer, I. J.: Methods for global sensitivity analysis in life cycle assessment, The International Journal of Life Cycle Assessment, 22, 1125–1137, 2017.

Gurney, K. R., Law, R. M., Denning, A. S., Rayner, P. J., Baker, D., Bousquet, P., Bruhwiler, L., Chen, Y.-H., Ciais, P., Fan, S., et al.: TransCom 3 CO2 inversion intercomparison: 1. Annual mean control results and sensitivity to transport and prior flux information, Tellus B: Chemical and Physical Meteorology, 55, 555–579, 2003.

Hamby, D. M.: A review of techniques for parameter sensitivity analysis of environmental models, Environmental monitoring and assessment, 32, 135–154, 1994.

Hastie, T., Tibshirani, R., and Wainwright, M.: Statistical learning with sparsity, Monographs on statistics and applied probability, 143, 143, 2015.

Heijungs, R.: Identification of key issues for further investigation in improving the reliability of life-cycle assessments, Journal of Cleaner Production, 4, 159–166, 1996.

Johnson, J. W.: A heuristic method for estimating the relative weight of predictor variables in multiple regression, Multivariate behavioral research, 35, 1–19, 2000.

Kitanidis, P. K.: On the geostatistical approach to the inverse problem, Advances in Water Resources, 19, 333–342, 1996.

Kleinman, D. and Athans, M.: The design of suboptimal linear time-varying systems, IEEE Transactions on Automatic Control, 13, 150–159, 1968.

Lauvaux, T., Miles, N. L., Deng, A., Richardson, S. J., Cambaliza, M. O., Davis, K. J., Gaudet, B., Gurney, K. R., Huang, J., O'Keefe, D., et al.: High-resolution atmospheric inversion of urban CO2 emissions during the dormant season of the Indianapolis Flux Experiment (INFLUX), Journal of Geophysical Research: Atmospheres, 121, 5213–5236, 2016.

Lin, J., Gerbig, C., Wofsy, S., Andrews, A., Daube, B., Davis, K., and Grainger, C.: A near-field tool for simulating the upstream influence of atmospheric observations: The Stochastic Time-Inverted Lagrangian Transport (STILT) model, Journal of Geophysical Research: Atmospheres, 108, 2003.

MacKay, D. J., Mac Kay, D. J., et al.: Information theory, inference and learning algorithms, Cambridge university press, 2003.

MatlabLivescript: version 9.9.0 (R2020b), The MathWorks Inc., Natick, Massachusetts, https://www.mathworks.com/help/matlab/matlab_prog/what-is-a-live-script-or-function.html.

Michalak, A. M., Randazzo, N. A., and Chevallier, F.: Diagnostic methods for atmospheric inversions of long-lived greenhouse gases, Atmospheric Chemistry and Physics, 17, 7405–7421, 2017.

Morris, M. D.: Factorial sampling plans for preliminary computational experiments, Technometrics, 33, 161–174, 1991.

Nielsen, F.: On the Jensen–Shannon symmetrization of distances relying on abstract means, Entropy, 21, 485, 2019.

Rabitz, H.: Systems analysis at the molecular scale, Science, 246, 221–226, 1989.

Rödenbeck, C., Houweling, S., Gloor, M., and Heimann, M.: Time-dependent atmospheric CO2 inversions based on interannually varying tracer transport, Tellus B: Chemical and Physical Meteorology, 55, 488–497, 2003.

Rödenbeck, C., Conway, T., and Langenfelds, R.: The effect of systematic measurement errors on atmospheric CO 2 inversions: a quantitative assessment, Atmospheric Chemistry and Physics, 6, 149–161, 2006.

Rodgers, C. D.: Inverse methods for atmospheric sounding: theory and practice, vol. 2, World scientific, 2000.

Sakia, R. M.: The Box-Cox transformation technique: a review, Journal of the Royal Statistical Society: Series D (The Statistician), 41, 169–178, 1992.

Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., and Tarantola, S.: Global sensitivity analysis: the primer, John Wiley & Sons, 2008.

Sobol, I. and Kucherenko, S.: Derivative based global sensitivity measures, Procedia-Social and Behavioral Sciences, 2, 7745–7746, 2010.

Sobol, I. M.: Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates, Mathematics and computers in simulation, 55, 271–280, 2001.

Sudret, B.: Global sensitivity analysis using polynomial chaos expansions, Reliability engineering & system safety, 93, 964–979, 2008.

Sunseri, I., Hart, J., van Bloemen Waanders, B., and Alexanderian, A.: Hyper-differential sensitivity analysis for inverse problems constrained by partial differential equations, Inverse Problems, 36, 125 001, 2020.

Tarantola, A.: Inverse problem theory and methods for model parameter estimation, SIAM, 2005.

Thompson, R., Gerbig, C., and Rödenbeck, C.: A Bayesian inversion estimate of N 2 O emissions for western and central Europe and the assessment of aggregation errors, Atmospheric Chemistry and Physics, 11, 3443–3458, 2011.

Turányi, T.: Sensitivity analysis of complex kinetic systems. Tools and applications, Journal of mathematical chemistry, 5, 203–248, 1990.

Vafaei, N., Ribeiro, R. A., and Camarinha-Matos, L. M.: Selecting normalization techniques for the analytical hierarchy process, in: Doctoral Conference on Computing, Electrical and Industrial Systems, pp. 43–52, Springer, 2020.

Wikle, C. K. and Berliner, L. M.: A Bayesian tutorial for data assimilation, Physica D: Nonlinear Phenomena, 230, 1–16, 2007.

Xu, C. and Gertner, G.: Understanding and comparisons of different sampling approaches for the Fourier Amplitudes Sensitivity Test (FAST), Computational statistics & data analysis, 55, 184–198, 2011.

Yadav, V., Duren, R., Mueller, K., Verhulst, K. R., Nehrkorn, T., Kim, J., Weiss, R. F., Keeling, R., Sander, S., Fischer, M. L., et al.: Spatio-temporally resolved methane fluxes from the Los Angeles Megacity, Journal of Geophysical Research: Atmospheres, 124, 5131–5148, 2019.