atmosphere

Supplementary Materials:

Use of Assimilation Analysis in 4D-Var Source Inversion: Observing System Simulation Experiments (OSSEs) with GOSAT Methane and Hemispheric CMAQ

Sina Voshtani [1,2,3,*], Richard Ménard [2], Thomas W. Walker [1] Amir Hakami [1]

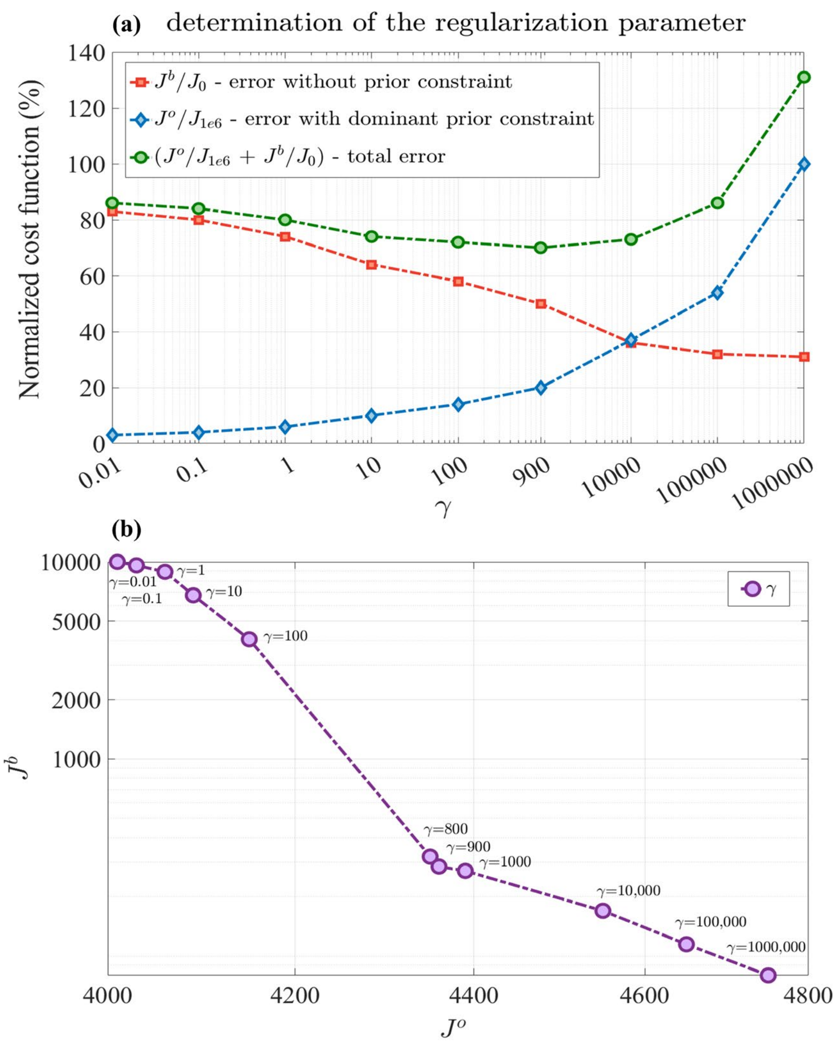

Figure S1. Given the cost function of Equation 15 in the main manuscript, the prior and observation term of the cost function has the form $J^b = \frac{1}{2}(x-x_b)^T \mathbf{B}^{-1}(x-x_b)$ and

$$J^o = \sum_{t=0}^{n} \tfrac{1}{2}\left(y_t^o - H_t(c_1^a, x)\right)^T \left(H^o \mathbf{P}_t^f(\mathbf{A}_1, \mathbf{Q}) H^{o\,T} + \mathbf{R}_t\right)^{-1} \left(y_t^o - H_t(c_1^a, x)\right),$$ respectively. (a) Show a traditional method of estimating γ that minimizes the sum of a normalized cost function [67]. J_0 is the magnitude of the total cost function ($J = \gamma J^b + J^o$) once γ = 0, indicating an optimization without prior constraint; and J_{1e6} is the magnitude of the total cost function at γ = 10^6, showing an optimization with a dominant prior constraint. In this method, we aim at a γ among a few selected values that minimize the total normalized error. It shows that γ = 900 is the appropriate choice, although, for a wider range of this parameter (e.g., 500-2000), the choice of γ has little impact on the overall optimization (inversion) solution. (b) The L-curve method for the determination of the regularization parameter shows a comparison between the prior term of the cost function (J^b) in the y-axis and the observation term of the cost function (J^o) in the x-axis for different choices of γ. According to the method of Hansen (1999) [70], γ = 900 is an optimal (balanced) choice for the regularization parameter. In principle, the optimal γ is obtained when the solution tends to change in nature from beings dominated by the prior cost (or perturbation error, where a small variation of γ causes rapid changes in J^b) to being dominated by the observation cost (or regularization/smoothing error where a large variation of γ makes a slow improvement in J^b).

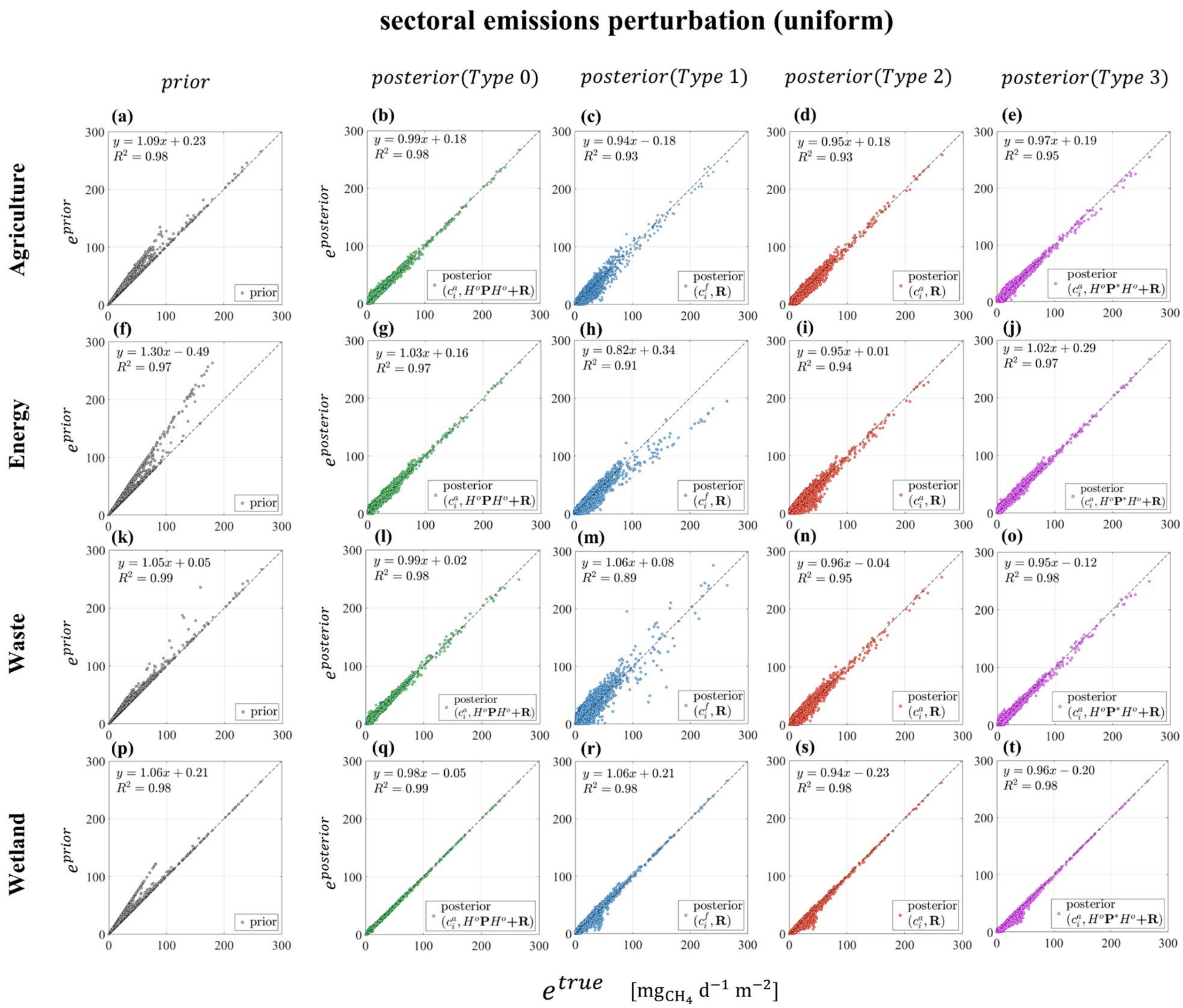

Figure S2. (a-e) statistical comparison of the prior and the posterior emissions against the true emissions in scatter plots for the only perturbed agriculture sector, (f-j) only perturbed energy sector, (k-o) only perturbed waste sector, and (p-t) only perturbed wetland sector. The prior emissions are generated using 50% uniform perturbation of each sector individually.

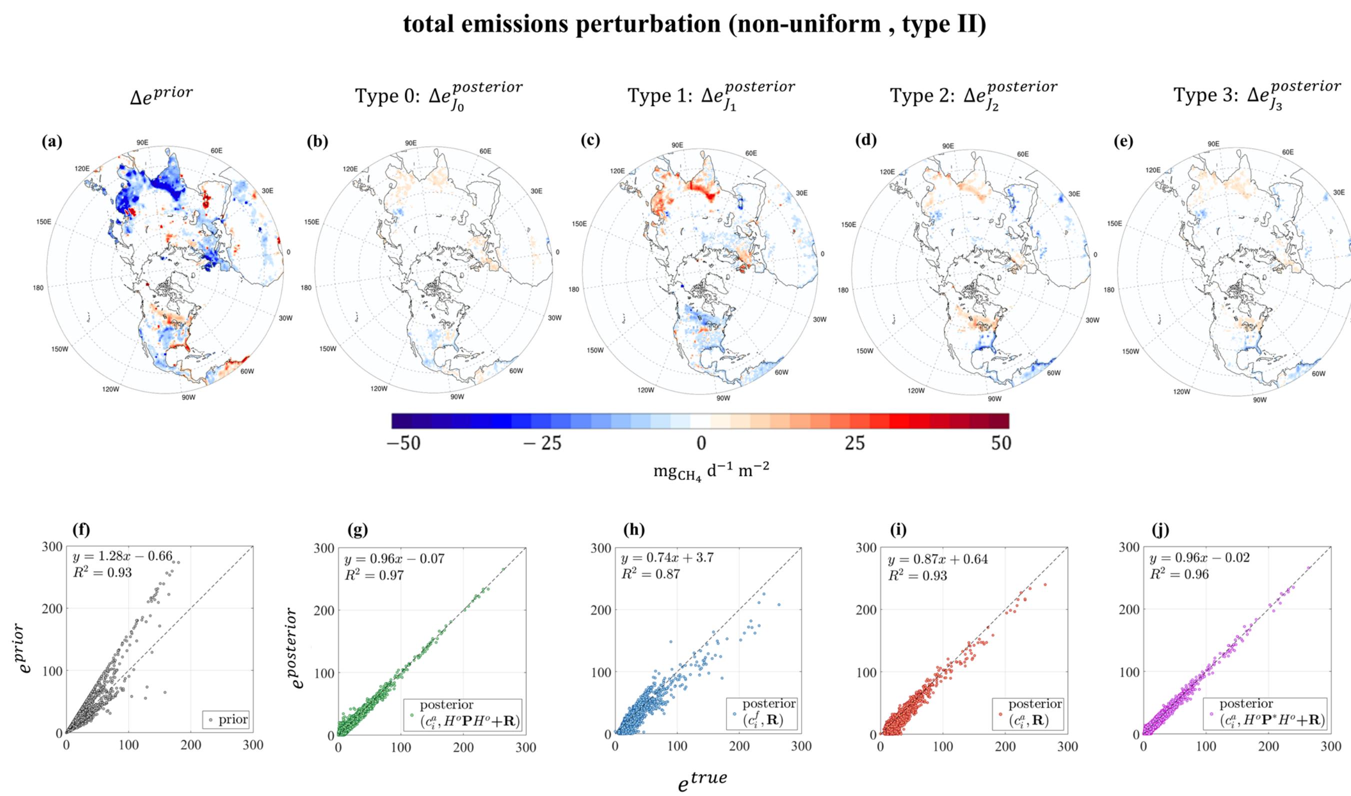

Figure S3. (**a**) The prior – true emissions (± 25-50% variable perturbation); (**b**) the posterior – true emissions in Type 0 inversion using analysis initial (c_1^a) and both observation **R** and model propagated analysis error covariance $H^o\mathbf{P}_t^f(\mathbf{A}_1,\mathbf{Q})H^{oT}$; (**c**) the posterior – true emissions in Type 1 inversion using forecast initial (c_1^f) and observation error covariance **R**, (**d**) the posterior – true emissions in Type 2 inversion using analysis initial (c_1^a) and observation error covariance **R**; (**e**) the posterior – true emissions in Type 3 inversion using analysis initial (c_1^a) and both observation **R** and model propagated analysis error covariance $H^o\mathbf{P}_t^f(\mathbf{A}_1)H^{oT}$, but with no model error. Statistical comparison of the (**f**) the prior emissions and (**g-j**) the posterior emissions of Type 0-3 inversion, respectively. x-axis and y-axis represent the true and the prior/posterior emissions, respectively. In (**f-j**), $\mathbf{P}^f(\mathbf{A}_1,\mathbf{Q})$ is shown as $\mathbf{P}$, and $\mathbf{P}^f(\mathbf{A}_1)$ is shown as $\mathbf{P}^*$. Synthetic observations are generated using the nature run initialized by the analysis, and a 2-week spin-up is used for the initialization.

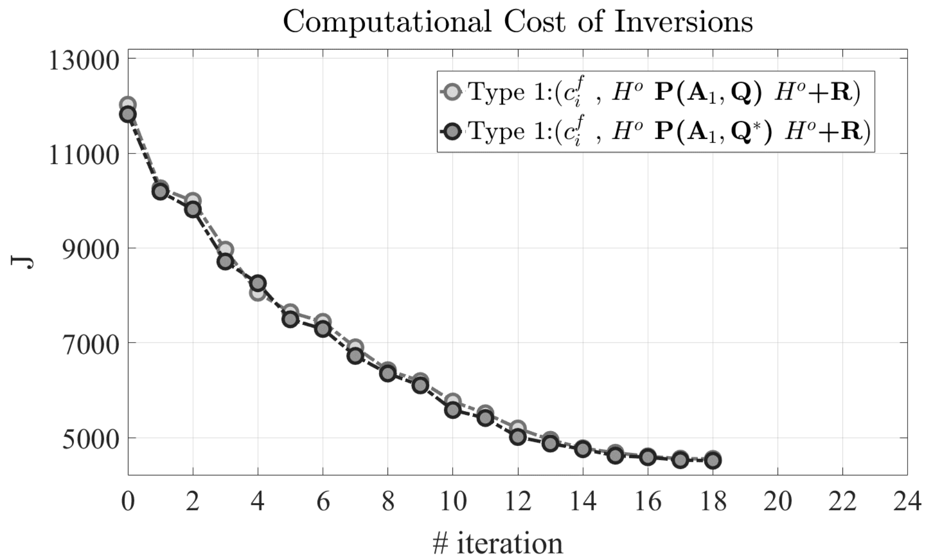

Figure S4. Comparison between the computational cost of two inversions in which only model transport error is different in the cost function. $\mathbf{Q}^*$ is the updated form of model transport error ($\mathbf{Q}^* = \mathbf{Q} + \mathbf{Q}_{diffusion}$). $\mathbf{Q}$ is the estimated model error during the PvKF assimilation (or $\mathbf{Q}_{PvKF}$) and $\mathbf{Q}_{diffusion}$ is approximated model error due to neglecting propagation of error correlations by diffusion.

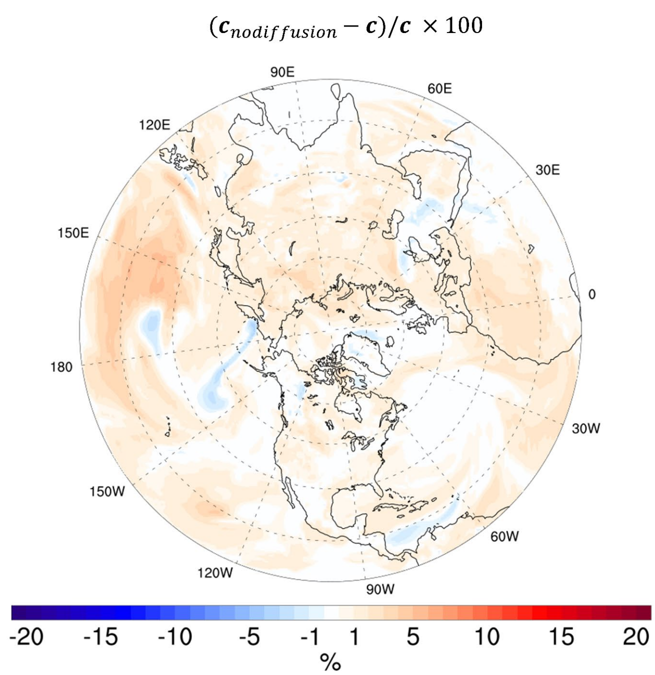

Figure S5. Normalized difference of concentrations between two cases where in the first one, the model diffusion scheme is deactivated, and in the second one, it is activated. It shows the distribution at the model's first layer after one month of simulation. Except for this difference, all other inputs and configurations between the two cases are the same.