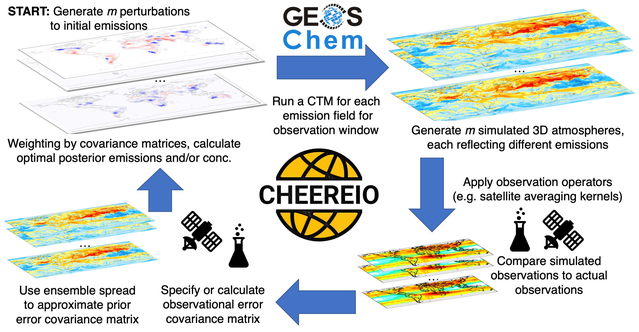
Continuous sub-weekly monitoring of global methane emissions at 2°×2.5° degrees from an ensemble Kalman filter using TROPOMI observations

Drew C. Pendergrass[1], Daniel J. Jacob[1], Nicholas Balasus[1], Lucas Estrada[1], Daniel J. Varon[1], Todd A. Mooring[1], James D. East[1], Elise Penn[2], and Hannah O. Nesser[3]

[1]School of Engineering and Applied Sciences, Harvard University, Cambridge, MA, USA. [2]Department of Earth and Planetary Sciences, Harvard University, Cambridge, MA, USA
[3]NASA Jet Propulsion Laboratory, Pasadena, CA, USA.

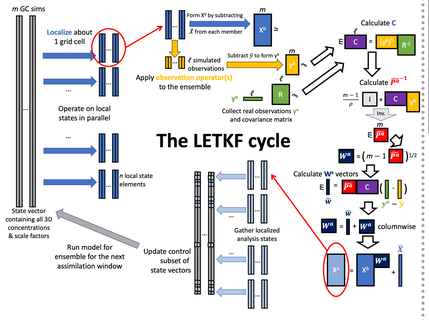
Abstract. Methane is a strong greenhouse gas, contributing 0.6 °C of warming from the pre-industrial baseline. Because of the potency and short lifetime of methane, decreasing methane emissions is an effective way to mitigate climate change in the near-term while also achieving air quality co-benefits from reduced tropospheric ozone. Bottom-up methane emissions inventories link emissions to processes, providing information necessary for mitigation strategies, but inventory construction takes several years and is subject to error. Methane can be detected by satellites through backscatter of solar radiation, including by the TROPOMI instrument, which can then be used to correct inventories via a Bayesian synthesis approach. Global top-down methane emissions estimates with high temporal resolution are desirable because of the variability of methane emissions in time, such as from intermittent point sources or from seasonally-varying sources including wetlands and rice; both may be incorrectly specified in bottom-up inventories but cannot be corrected with annual-scale inversions. In our preliminary results, we use the localized ensemble transform Kalman filter (LETKF) algorithm to demonstrate that TROPOMI observations can be used to estimate global methane emissions at 2°×2.5° degrees spatial resolution and 5-day temporal resolution. To examine the importance of methane seasonality on our inversion, we use two different inventories for wetland emissions with divergent spatial and temporal patterns. We evaluate the quality of the LETKF emissions adjustments and discuss the emissions trends inferred by the inversion, as well as the role of wetland seasonality in the interpretation of top-down emissions estimates at high temporal resolution.

The LETKF algorithm



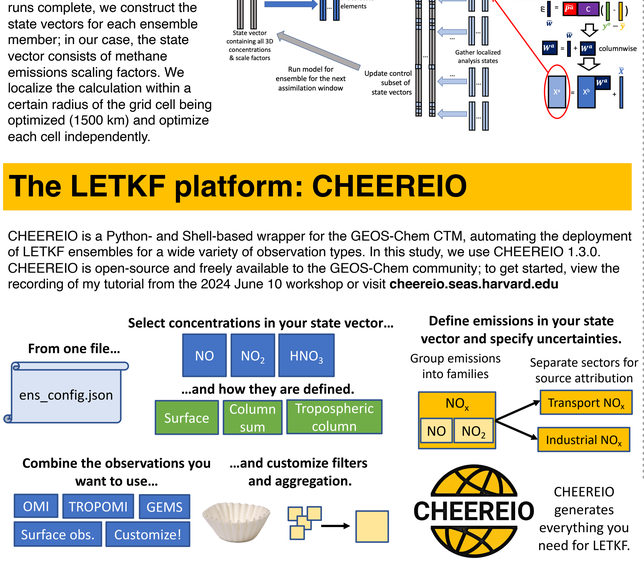
The Localized Ensemble Transform Kalman Filter (LETKF) is a Bayesian algorithm that can optimize emissions or concentrations of chemical species, weighting observations and prior knowledge (e.g. inventories) by relevant uncertainties. LETKF avoids the need for the adjoint of the chemical transport model (CTM) because it is powered by an ensemble of CTM simulations which capture the nonlinearity of the system. Each ensemble member is initialized with random perturbations applied to emissions of interest, capturing prior uncertainty. The CTM enables comparison of the ensemble with observations, allowing the prior to be updated and the simulation/LETKF cycle to repeat, **as shown in the above figure**.

The LETKF algorithm optimizes a state vector x of emissions and/or concentrations following the linear algebra **shown at right**. In the LETKF, m ensemble members are initialized at time t_0 and the forward CTM (GEOS-Chem) is run in parallel for a user-specified time (termed the assimilation window) for each of these ensemble members. After the runs complete, we construct the state vectors for each ensemble member; in our case, the state vector consists of methane emissions scaling factors. We localize the calculation within a certain radius of the grid cell being optimized (1500 km) and optimize each cell independently.



The LETKF platform: CHEEREIO

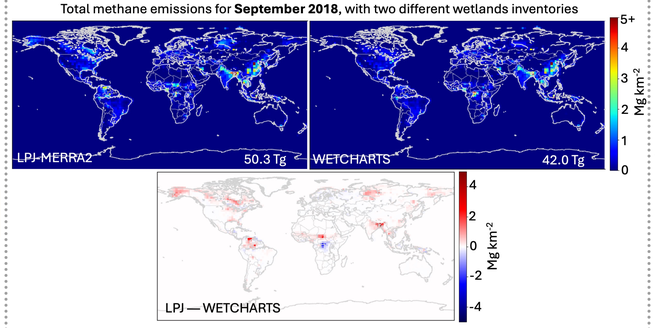
CHEEREIO is a Python- and Shell-based wrapper for the GEOS-Chem CTM, automating the deployment of LETKF ensembles for a wide variety of observation types. In this study, we use CHEEREIO 1.3.0. CHEEREIO is open-source and freely available to the GEOS-Chem community; to get started, view the recording of my tutorial from the 2024 June 10 workshop or visit **cheereio.seas.harvard.edu**



Prior emissions: the wetlands challenge

Prior emissions from oil, gas, and coal are from the 2010-19 Global Fuel Exploitation Inventory (GFEI) version 2.0 and other anthropogenic emissions are from the 2012 EDGARv6 inventory. Both inventories are overwritten for the continental US, Mexico, and Canada by national inventories. Anthropogenic emissions are assumed to be aseasonal, except for manure management and rice for which we apply seasonal scaling factors. For natural emissions, fires are from the Global Fire Emissions Database (GFED4), termites from Fung et al. (1991), and geological seeps from Etiope et al. (2019). Methane is primarily lost due to oxidation by OH, represented by fields archived from a GEOS-Chem full-chemistry simulation scaled so that methane's tropospheric lifetime due to loss to OH matches the best estimate derived from methyl chloroform observations.`

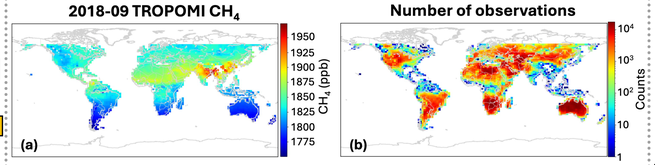
Total methane emissions for **September 2018,** with two different wetlands inventories



For wetland emissions, we run two different inversions for two separate inventories. We first use the high-performance subset of WetCHARTs v1.3.1, but this inventory underestimates the global seasonal cycle of methane as observed by the GOSAT-corrected TROPOMI product (East et al., 2024). For comparison, we use the LPJ-wsl dynamic global vegetation model driven by MERRA-2 which better matches observed global methane seasonality (East et al., 2024). **The above figure** shows total methane emissions (anthropogenic and natural) for September 2018 in the top row, with LPJ-MERRA2 wetlands at left and WETCHARTS at right. Total emissions for September are inset. The difference in the wetland inventories (LPJ minus WETCHARTS) for this month is shown in the bottom panel.

Observations: GOSAT-corrected TROPOMI

TROPOMI retrieves dry-column methane mixing ratios (X_{CH_4}) at 5.5x7 km^2 nadir pixel resolution at 13:30 local solar time via a full-physics algorithm. Balasus et al. (2023) trained a machine learning algorithm to bias-correct TROPOMI X_{CH_4} according to observations from the GOSAT instrument, which was launched in 2009. We filter out observations with a quality assurance value of 0.5 or lower and also remove retrievals over coastlines and oceans, but otherwise keep all GOSAT-corrected TROPOMI X_{CH_4}retrievals, improving coverage of key source areas (Amazon, Congo, southeast Asia). **The below figure** shows the mean GOSAT-corrected TROPOMI X_{CH_4}retrievals for September 2018 as well as the number of observations in each grid cell.

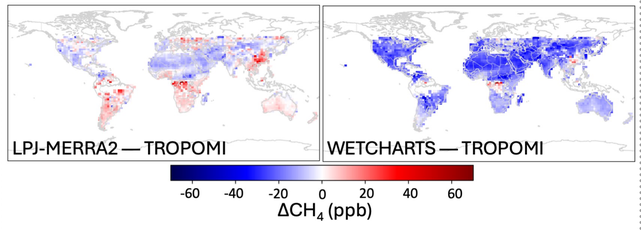


When ingesting the GOSAT-adjusted TROPOMI observations into the LETKF, we aggregate the data into "super-observations" by averaging onto the 2.0°×2.5° GEOS-Chem grid. To model the reduction in observational error variance due to averaging and obtain the super-observation error standard deviation σ_{super}, we follow a two-component error variance equation which separates contributions due to forward model transport error variance ($\sigma^2_{\text{transport}}$) from satellite error variance for a single retrieval (σ_i^2):

$$\sigma_{\text{super}} = \sqrt{\left[\left(\frac{1}{n}\sum_{i=1}^{n}\sigma_i\right)\cdot\left(\frac{1-c}{n}+c\right)\right]^2 + \sigma^2_{\text{transport}}}$$

Here n is the number of observations aggregated into a super-observation and c is the error correlation between the individual retrievals within a super-observation. Error correlation is expected due to shared retrieval parameters such as surface reflectance. Chen et al. (2023) used the residual error method to empirically determine the parameters for TROPOMI methane and obtained σ_i = 17 ppb, transport error $\sigma_{\text{transport}}$ = 6.1 ppb, and error correlation c = 0.28; we use these values in this study

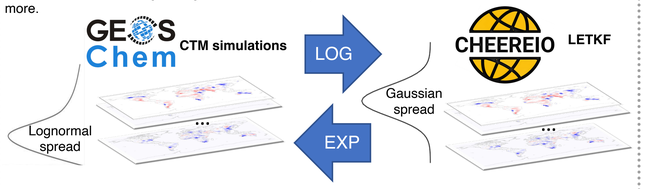
Bias between prior model and TROPOMI after six months (2019-01)



The above figure shows the importance of prior wetland inventories in obtaining a seasonally-robust simulation that can reproduce observational trends. Here we initialize GEOS-Chem to match TROPOMI global means for May 2018, then run for six months with either LPJ-MERRA2 wetlands (left) or WETCHARTS (right). Because LPJ-MERRA2 has more wetland emissions in northern hemisphere summer and autumn, it produces a globally unbiased estimate of global methane (regional biases remain pronounced), while the WETCHARTS simulation is biased low nearly everywhere. These different methane budgets can affect the ability for LETKF to offer continuous methane monitoring, as discussed in the rightmost column.

Lognormal errors in LETKF

Lognormal errors better capture the upper tail of the methane emissions distribution than does a Gaussian; by forbidding negative values, it also prevents unphysical negative emissions in analytical inversions. However, imposing a lognormal distribution across ensemble members violates the assumptions of the LETKF equations. We solve this problem by sampling methane emissions scaling factors for each ensemble member according to a lognormal distribution centered on 1 (prior emissions inventory) and run GEOS-Chem for each ensemble member with these scaling factors applied. When it is time for the LETKF calculation, we apply a logarithmic transform to the methane scaling factor distributions and thus obtain a normal distribution (centered on 0) for the construction of the background perturbation matrix $\boldsymbol{X^b}$. We perform the LETKF and once it is complete we apply an exponential to transform back to the original lognormal distribution, which is then used to evolve GEOS-Chem once more.



Inflation: relaxation to prior spread

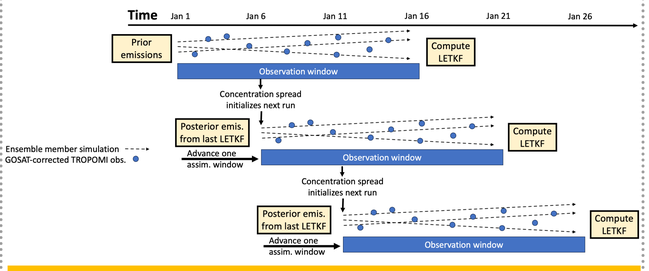
Because ensemble-based methods undersample the possibility space of the data assimilation problem, they suffer from shrinking dispersion between ensemble members which can lead to vanishingly small prior error estimation and thus for later observations to be discarded; an error inflation method is necessary to prevent ensemble collapse. Following Bisht et al. (2023), we use the Relaxation to Prior Spread (RTPS) inflation method. RTPS inflates the standard deviation $\boldsymbol{\sigma^a}$ of the analysis perturbation matrix $\boldsymbol{X^a}$ (see "The LETKF cycle" diagram in left column) such that it partially reflects the standard deviation $\boldsymbol{\sigma^b}$ of the background perturbation matrix $\boldsymbol{X^b}$:

$$\boldsymbol{X^a_{\text{infl}}} = \left(\frac{\alpha_{\text{RTPS}}\boldsymbol{\sigma^b} + (1-\alpha_{\text{RTPS}})\boldsymbol{\sigma^a}}{\boldsymbol{\sigma^a}}\right)\boldsymbol{X^a}$$

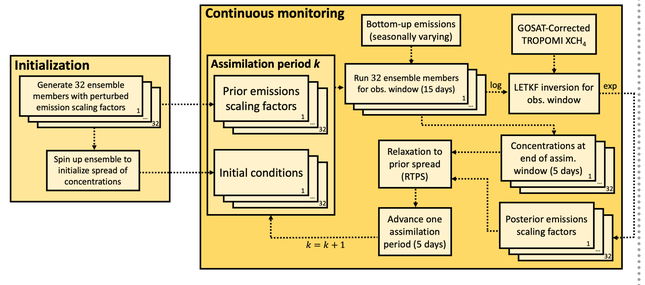
Here $\boldsymbol{\alpha_{\text{RTPS}}}$ is a parameter between 0 and 1 which represents the weighted contribution of the background standard deviation $\boldsymbol{\sigma^b}$ in inflating the analysis ensemble to obtain the final analysis perturbation matrix $\boldsymbol{X^a_{\text{infl}}}$. We take $\boldsymbol{\alpha_{\text{RTPS}}}$ to be 0.7. We additionally apply RTPS to 3D methane concentrations in the ensemble members even though we do not formally include concentrations in the state vector; this is because we find inflating emissions alone is not necessarily enough to prevent the convergence of concentrations across ensemble members in the short term.

Run-in-place: handling long-lived species

We modify the treatment of the LETKF assimilation window using the run-in-place (RIP) method (Kalnay and Yang, 2010). For the Gaussian and linearity assumptions of an ensemble Kalman filter algorithm to be satisfied, a short assimilation window must be used, but the estimation of parameters such as methane emissions benefit from a longer observational record. With RIP, we calculate the LETKF assimilation update using a long period of observations (15 days, called the observation window), but then advance the assimilation window forward for a shorter period (5 days, the assimilation window). RIP thus maintains linear growth in posterior perturbations and allows the system more time to correct assimilation errors. Importantly, after advancing the assimilation window forward, we do not reinitialize the ensemble for new runs. Instead, the assimilated state of the previous observation window becomes the initial background state of the next assimilation window, as shown in **the below figure**.



Workflow for continuous monitoring

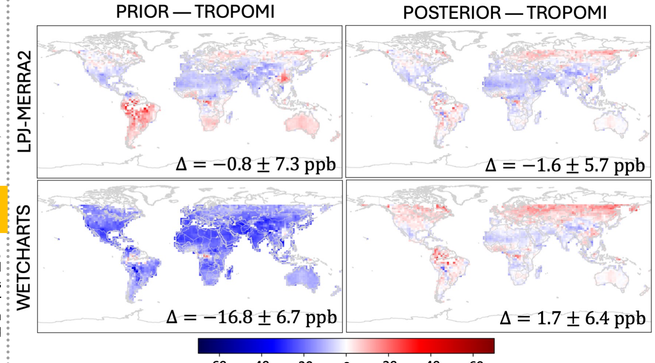


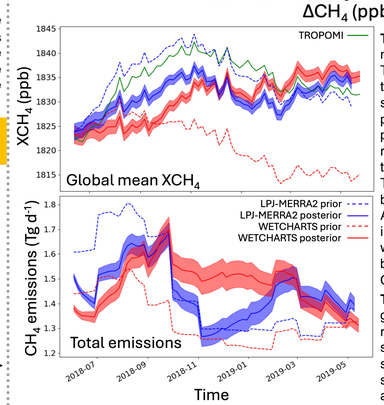
Above, flowchart of our LETKF inversion procedure for continuous global methane emissions monitoring. We initialize with randomized multiplicative perturbations to bottom-up emissions inventories, applied to each of the 32 ensemble members. For assimilation period k, we run GEOS-Chem for the observation window (15 days) for each ensemble member, then conduct the LETKF inversion by comparing the ensemble sampled with the TROPOMI observation operator to the GOSAT-corrected TROPOMI CH_4 over the observation window. Posterior emission scaling factors and concentrations are then inflated to reflect the prior spread using the RTPS procedure (above). The posterior emission estimates and inflated concentrations then become the prior for the k+1 assimilation period, beginning 5 days after the prior.

Preliminary results

We create two thirty-two member ensembles of GEOS-Chem within the CHEEREIO environment, one for each wetland prior, following the architecture in the "Workflow for Continuous Monitoring" section. For localizations, we weight observations according to their distance from the grid cell in question according to the Gaspari-Cohn function, a piecewise polynomial resembling a bell curve with a value of 1 at the grid cell and 0 at 1500 km away, which leads to smoother assimilation results. In line with analytical inversions of methane emissions, we additionally apply a regularization factor of $\gamma = 0.1$ to effectively increase observational errors and avoid overfit to TROPOMI. Planned simulations for this work (2018 through 2024) are currently running. Only the first year (June 2018 through May 2019) was available at the time of this presentation. **Results shown below are based on this first year of results**.

Impact of assimilation on bias for 2019-01 through 2019-05 period

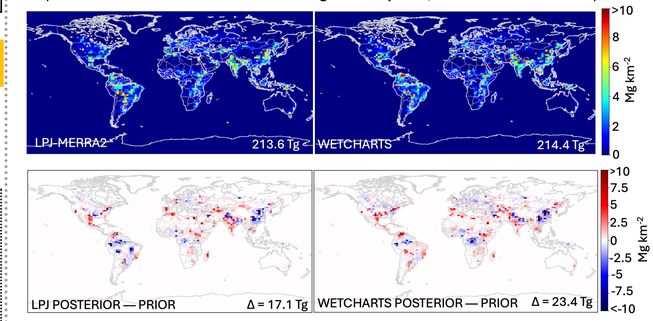




The above figure compares simulated methane columns from GEOS-Chem with TROPOMI observations for January 2019 through May 2019. Top row shows the simulation with the LPJ-MERRA2 wetland prior and the bottom with the Wetcharts prior; posterior simulations are shown in the right column as compared with the priors in the left, with global biases relative to TROPOMI inset. Local biases are reduced by assimilation in both cases (South America in LPJ) but global bias is only improved in the WETCHARTS simulation, which has a globally imbalanced methane budget with the methyl chloroform-adjusted OH sink.

The figure at left shows, in the top panel, global mean dry-column methane mixing ratios for TROPOMI (green), the LPJ simulation (blue), and the Wetcharts simulation (red). Ensemble mean and standard deviation are shown in solid lines and envelopes, with priors shown by dotted lines. After a year of runtime, both simulations approach TROPOMI global mean. Despite converging concentrations, emissions (bottom panel) show different temporal patterns as the Wetcharts posterior shows high methane emissions through winter, apparently to "catch up" with bias introduced from the budget imbalance. For our full results, we will likely discard the first year of simulation as a "burn-in" period, a practice common in CO_2 inversions. **The below figure** shows posterior methane emissions for the January 2019 through May 2019 period. Both LPJ-MERRA2 and Wetcharts show remarkable agreement on total methane emissions for this period (annualized: 521 Tg, slightly lower than other budgets but methane emissions peak in the second half of the year due to wetland activity; East et al., 2024).

Total posterior methane emissions for **2019-01 through 2019-05 period**, with two different wetlands priors



Acknowledgements, contact, and links

This work was funded by the NASA Carbon Monitoring System. DCP was funded in part by an NSF Graduate Research Fellowship Program (GRFP) grant.

Contact Drew Pendergrass at **pendergrass@g.harvard.edu**
CHEEREIO website: **cheereio.seas.harvard.edu**
Download and use the CHEEREIO code: **bit.ly/cheereio**
CHEEREIO documentation: **bit.ly/CheereioDocs**

Scan for CHEEREIO website