# Background Error Statistics for Assimilation of Atmospheric $CO_2$


Abhishek Chatterjee[1,2], Richard, J. Engelen[3], Stephan R. Kawa[4], Colm Sweeney[5,6], Anna M. Michalak[2]

1 The University of Michigan, Ann Arbor, MI, USA
2 Carnegie Institution for Science, Stanford, CA, USA
3 European Centre for Medium-Range Weather Forecasts, Reading, UK
4 NASA Goddard Space Flight Center, Greenbelt, MD, USA
5 NOAA Earth System Research Laboratory, Boulder, CO, USA
6 CIRES, University of Colorado, Boulder, CO, USA


## Introduction

- Background error covariance plays an important role in the effectiveness of a data assimilation system, by prescribing the degree to which the model state can be updated by assimilated observations
- For $CO_2$ assimilation, available forecast or ensemble-based methods cannot: 1) optimally capture the spatial and temporal statistics of the errors associated with the carbon emissions and uptake, or 2) adequately characterize the errors associated with the $CO_2$ transport
- Here, we propose that the difference between predicted $CO_2$ fields from two models with different flux and transport specifications ($\Delta CO_2$) can be used as a surrogate field for quantifying the background error covariances

*The novel background error covariance is used in an operational 4DVAR system, constrained with satellite $CO_2$ observations, to generate global 4D fields of atmospheric $CO_2$ concentrations*

## Four-dimensional Variational Data Assimilation

The ECMWF atmospheric 4DVAR system is a practical formulation of Bayesian estimation theory for time-evolving three-dimensional atmospheric fields. The analysis correction ($\delta\mathbf{x}(t_0)$) to the atmospheric model initial state ($\mathbf{x}^b$) is sought as the minimizer to the following cost function:

$$J(\delta\mathbf{x}(t_0)) = \frac{1}{2}\left((\delta\mathbf{x}(t_0))^T \mathbf{B}^{-1}(\delta\mathbf{x}(t_0))\right) + \frac{1}{2}\sum_{i=0}^{n}\left(\left(H_i[\delta\mathbf{x}(t_i)] - \mathbf{d}_i\right)^T \mathbf{R}^{-1}\left(H_i[\delta\mathbf{x}(t_i)] - \mathbf{d}_i\right)\right)$$

where, $\mathbf{d}_i$ : observation departures between the observed radiances ($\mathbf{y}_i$) and the atmospheric model simulated equivalent radiances, or

$$\mathbf{d}_i = \mathbf{y}_i^o - H_i\left[\mathbf{x}^r(t_i)\right]$$

$H_i$ : non-linear observation operator in the form of a radiative transfer model
$\mathbf{R}$ : observation error covariance matrix
$\mathbf{B}$ : background error covariance matrix
$\mathbf{x}^r$ : reference model state at time $t_i$ evolved according to the full non-linear forecast model $M$

Once the cost function is minimized with respect to the increments of the initial state ($\delta\mathbf{x}(t_0)$) the increments are added to the background state ($\mathbf{x}^b$) to obtain the analysis $\mathbf{x}(t_0)$ :

$$\mathbf{x}(t_0) = \mathbf{x}^b + \delta\mathbf{x}(t_0)$$

Experimental setup-

- ground-based and satellite data → constrains relevant fields in forecast model including atmospheric temperature, vorticity, pressure etc.
- GOSAT-ACOS v2.9 → constrains $CO_2$ concentrations
- IFS-CASA → prescribes the $CO_2$ background to which the increments are added
- NMC (control run) or Δ-statistics → background error statistics

## Acknowledgments


TCCON data were obtained from the TCCON data archive, operated by the California Institute of Technology from the website at http://tccon.ipac.caltech.edu/. The GOSAT-ACOS data were produced by the ACOS/OCO-2 project at the Jet Propulsion Laboratory, California Institute of Technology, and obtained from the ACOS/OCO-2 data archive maintained at the NASA Goddard Earth Science Data and Information Services Center. Funding for Abhishek Chatterjee was provided by the NASA Earth and Space Science Fellowship. The work of R. Engelen was funded through the MACC project, which is funded by the European Commission under the Seventh Research Framework Programme. The work of S. Kawa was supported through the NASA Carbon Cycle Science and Atmospheric $CO_2$ Observations from Space opportunities. The authors also acknowledge the efforts of NOAA-ESRL group in collecting and sharing the aircraft data.


## Background Error Correlations: Δ–statistics

### Horizontal error correlations

CALCULATED

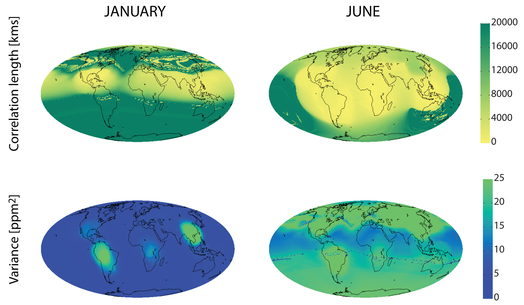
$$\gamma_{region}(h) = \frac{1}{2}\left[\left(\Delta CO_2(x_{region}) - \Delta CO_2(x_{region+h})\right)^2\right]$$

Regional variability ($\gamma_{region}$) of $\Delta CO_2$ can be captured by a two-point variance ($\sigma^2$), a correlation length ($3l$) and a positive-definite function describing the decay of auto-correlation as the separation distance ($h$) between two observations locations increase.

MODELED

$$\gamma_{theo}(h) = \sigma^2\left(1 - \exp\left(-\frac{h}{l}\right)\right)$$

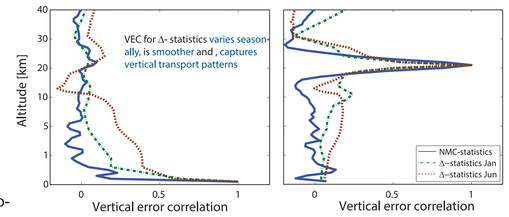
Shorter correlation length → greater spatial variability at small scales
Higher variance → greater overall variability




Sample correlation length and variance at ~975 hPa on the 15th of the month (6 pm UTC). Note the higher variances, and shorter correlation length, over the Northern Hemisphere in June, due to the highly active biosphere.

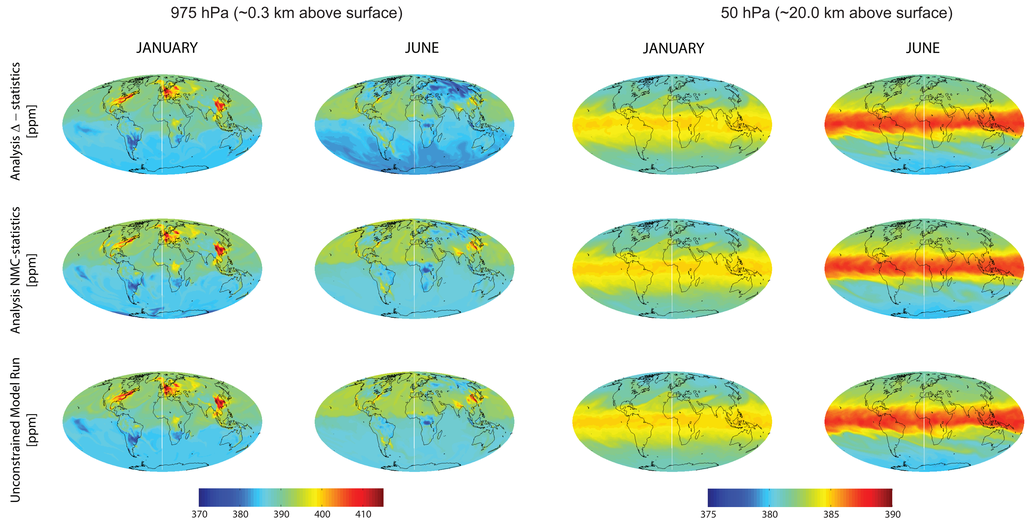
### Vertical error correlations

Calculated as the correlation coefficients of $\Delta CO_2$ between levels of the atmosphere. Shown below are the error correlations for the surface level (left) and at ~ 50 hPa (right).



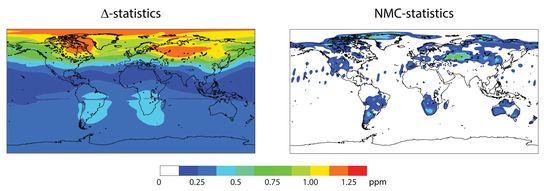

## Results

### Comparison of analyzed $CO_2$ fields at two different atmospheric levels

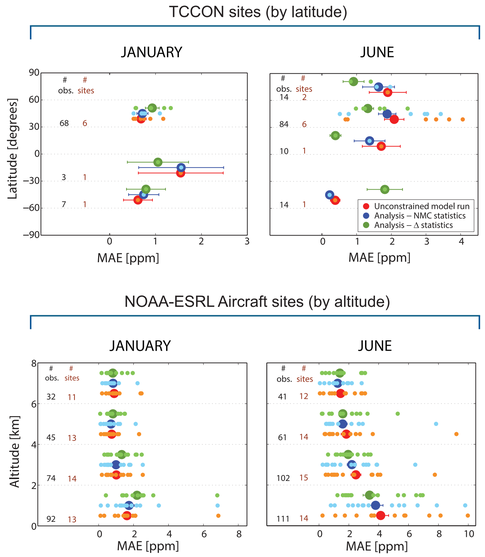



Top: Analyzed $CO_2$ concentrations on the 15th of the month (6 pm UTC) at ~975 hPa, and at ~50 hPa. Analysis with the Δ-statistics allows pronounced adjustments by the GOSAT observations to the background model, relative to the analysis with the NMC-statistics. Differences between the two analyses are greater near the surface (1) when high variability in the $CO_2$ processes (e.g. fossil fuel emissions, biospheric exchanges, influence of boundary layer variations etc.) are present, and (2) where GOSAT observations have larger influence than higher up in the atmosphere.

Bottom: Column-averaged $CO_2$ mixing ratio analysis increments averaged over the month of June. Longer error correlation lengths prescribed via the Δ-statistics result in broad-scale increments compared to the localized changes allowed by the NMC-statistics.




### Validation with independent $CO_2$ data




Mean Absolute Error (MAE) calculated between the three 4D-VAR analyses, and indepedent $CO_2$ observations. Performance over the Southern Hemisphere is difficult to judge, however, given the paucity of TCCON sites, and wintertime $CO_2$ data.

## Conclusions

- *Accounting for spatio-temporal variations in the $CO_2$ field, the Δ-statistics provides realistic background error statistics, which improves assimilated atmospheric $CO_2$ concentrations*
- *Significant improvement occurs during the summer over the Northern Hemisphere, when significant variability and uncertainty in underlying fluxes are present*
- *Longer error correlations in the $\Delta CO_2$ statistics allows observations to inform atmospheric $CO_2$ over larger areas, making the DA system more data-driven*