Geosci. Model Dev. Discuss.,
https://doi.org/10.5194/gmd-2020-246-RC2, 2021
© Author(s) 2021. This work is distributed under
the Creative Commons Attribution 4.0 License.

Geoscientific Model Development Discussions

Interactive comment on "Data reduction for inverse modeling: an adaptive approach v1.0" *by* Xiaoling Liu et al.

Anonymous Referee #2

Received and published: 12 January 2021

Interactive comments on "Data reduction for inverse modeling: an adaptive approach v1.0"

The authors attempt to reduce computation cost of inverse modeling by reducing satellite data size based on different regional variability in satellite observations. They also attempt to use variability in reduced data to determine to which level the satellite data should be reduced. This work is interesting. And indeed, a need for data reduction in inverse modeling emerges with increasing number of satellite observations. However, the following issues need to be addressed before publication.

Major comments:

1. Do you consider the DOF of the satellite observations when you interpolate satellite

data? I wonder how much of the satellite DOF is lost after you interpolating the satellite observations. Would it make sense more if you determine the data reduction level with additional consideration of satellite DOF loss? In inverse modeling, averaging kernels are usually applied to remove impacts from satellite a priori. How do you deal with averaging kernels of OCO-2 XCO2 when you use kriging to interpolate satellite CO2 observations?

2. In your case studies, did you apply observation operator? If not, then you assume all satellite observations have identical DOF, which is not true for real satellite observations. Applying observation operator or not can make large difference on your top-down results. Thus, this can impact the evaluation of your data reduction algorithm.

3. Section 4.4: Since your goal is to reduce computation cost for inverse modeling, I wonder how large the difference is between the computation costs of your case studies (full data vs reduced data)? like the difference between computation costs of Figure 3 (b) and (c)? Also, did you try running your inverse modeling with full data, but using parallelization (e. g. 24 cores)? I wonder how large the computation cost would be compared to that of inverse modeling with reduced data but without parallelization.

Specific comments: 1. Line 55-62: Several problems here:

a. "The cost of running the forward and adjoint models usually increases as the number of observations increases, particularly for the model adjoint." The cost of running the forward and adjoint models would not increase as the number of observations increases. It is the cost of running observation operators that increases with increasing number of satellite observations.

b. "..a single run of GEOS-Chem adjoint model . . .30 days of . . .from the "lite" file." This sentence can not prove your point. I guess you want to prove that the cost of running one iteration of adjoint inversion with OCO-2 observations increase as the number of satellite observation increases. Then you need to compare the computation costs from at least two runs: one is assimilating fewer satellite observations, and the other is assimilating more satellite observations.

c. "...many adjoint models for GHG applications are not designed to exploit parallel computing architecture." Can you pointing out what adjoint models do not use parallelization? As far as I know, most Eulerian chemical transport models and their adjoint models (if applicable) exploit parallelization, either OMP parallelization or MPI parallelization. For example, GEOS-Chem adjoint model exploits OMP parallelization.

d. "However, parallel versions are currently under development ... (Eastham et al., 2018)." What Eastham et al. (2018) developed is high-performance GEOS-Chem forward model (with MPI parallelization), not adjoint model.

2. Figure 4 caption: "...with modest data reduction (b, 1098 data points) ... with no reduction (a, 4183 data points). ... large data reduction (c, 251 data points) ...". I think it should be "...with modest data reduction (c, 1098 data points) ... with no reduction (b, 4183 data points). ...large data reduction (d, 251 data points) ..."

Interactive comment on Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2020-246, 2020.