Geosci. Model Dev. Discuss.,
https://doi.org/10.5194/gmd-2020-246-RC1, 2021

Geoscientific Model Development Discussions
Open Access
EGU

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

Anonymous Referee #1

Received and published: 12 January 2021

Before I begin: What does v1.0 mean in the title of the paper. Is there an approach v2.0 ?? If there is one than I would like to see it

This manuscript requires revisions before publication. The manuscript adds another method to a series of existing methods for data reduction and authors are aware of this as they mention multiple approaches adopted in different disciplines to do this. Like clustering or any other method, a subjective criterion (through exploratory analysis is still needed to select observations) like a choice of an inflection point is required.

This is not a bad thing and the authors can choose magnitude of the slope to suggest an inflection point (Note the variance curve is monotonically increasing). I would like to also see if it makes a difference to use a correlation length obtained from summer months for one year and applied to observations of another year and in the second

Printer-friendly version

Discussion paper

stage select observations based on the variograms for the same year for which inversions are being performed. Thus, if it does not lead to any major difference in RMSE than these variograms can be precomputed for a selected year and applied for inversions for subsequent or previous years.

The section on variance criterion needs to be further elaborated. It is not clear, how the variances (is it variance in strict sense) are being computed. Is it the square of a difference between the kriged point and the support of the kriged point? Please demonstrate this with a small example within the context of Figure 2. Authors also need to elaborate whether the uncertainty from Ordinary Kriging was used in inversions or not. Furthermore, rather than using an exponential covariance structure for Ordinary Kriging why not use exponential + nugget model for doing kriging as this way you can have OCO-2 errors along the diagonal of the nugget and exponential structure can account for correlations in observations in space and/or time. This would allow Ordinary Kriging uncertainty to be directly included in data assimilation and/or inversions. I also wonder what would happen if you try to simultaneously use observations from multiple instruments and which due to different biases or viewing geometry leads to different values of XCO2 for a same location and completely contaminates the correlation length obtained from the variogram. This per se is not a problem with the methodology proposed here but can create problems in real-data applications of the methodology.

I would also like to see what kind of reduction happens if a variogram that accounts for both space and time is used to reduce the data. This can be checked within the context of a synthetic study and would become necessary in case of geostationary missions like GEOCARB that may have multiple overpasses for a same location over a single day.

Specific comments Line 40: replace particle with Lagrangian. Line 49: Why 2 seconds? How would you deal with it when you have simultaneous observations from multiple instruments? Line 54: Which vector? Line 59: What is the context of 30 days of wall-clock time (what kind of computer?) Line 62: Parallel computing architecture. Please

briefly elaborate especially within the context of 4D-Var and ensemble approaches

---

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