EGU21-6036
https://doi.org/10.5194/egusphere-egu21-6036
EGU General Assembly 2021

Using machine learning techniques to generate analog ensembles for data assimilation

Lucia Yang[1] and Ian Grooms[2]
[1]University of Colorado Boulder, Applied Mathematics, Boulder, United States of America (lucia.yang@colorado.edu)
[2]University of Colorado Boulder, Applied Mathematics, Boulder, United States of America (groomsi@colorado.edu)

We propose to use analogs of the forecast mean to generate an ensemble of perturbations for use in ensemble optimal interpolation (EnOI) or ensemble variational (EnVar) methods. In addition to finding analogs from a library, we propose a new method of constructing analogs using autoencoders (a machine learning method). To extend the scalability of constructed analogs for use in data assimilation on geophysical models, we propose using patching schemes to divide the global spatial domain into digestable chunks. Using patches makes training the generative models possible and has the added benefit of being able to exploit parallel computing powers. The resulting analog methods using analogs from a catalog (AnEnOI), constructed analogs (cAnEnOI), and patched constructed analogs (p-cAnEnOI) are tested in the context of a multiscale Lorenz-`96 model, with standard EnOI and an ensemble square root filter for comparison. The use of analogs from a modestly-sized catalog is shown to improve the performance of EnOI, with limited marginal improvements resulting from increases in the catalog size. The method using constructed analogs is found to perform as well as a full ensemble square root filter, and to be robust over a wide range of tuning parameters. Lastly, we find that p-cAnENOI with larger patches produces the best data assimilation performance despite having larger reconstruction errors. All patch variants except for the variant that uses the smallest patch size outperform cAnEnOI as well as some traditional data assimilation methods such as the ensemble square root filter.