Geophysical Research Abstracts
Vol. 18, EGU2016-1733-2, 2016
EGU General Assembly 2016

Optimality in Data Assimilation

Grey Nearing and Soni Yatheendradas
NASA GSFC, Hydrologic Sciences Lab, Maryland, United States

It costs a lot more to develop and launch an earth-observing satellite than it does to build a data assimilation system. As such, we propose that it is important to understand the efficiency of our assimilation algorithms at extracting information from remote sensing retrievals. To address this, we propose that it is necessary to adopt completely general definition of "optimality" that explicitly acknowledges all differences between the parametric constraints of our assimilation algorithm (e.g., Gaussianity, partial linearity, Markovian updates) and the true nature of the environmetnal system and observing system.

In fact, it is not only possible, but incredibly straightforward, to measure the optimality (in this more general sense) of any data assimilation algorithm as applied to any intended model or natural system. We measure the information content of remote sensing data conditional on the fact that we are already running a model and then measure the actual information extracted by data assimilation. The ratio of the two is an efficiency metric, and optimality is defined as occurring when the data assimilation algorithm is perfectly efficient at extracting information from the retrievals. We measure the information content of the remote sensing data in a way that, unlike triple collocation, does not rely on any a priori presumed relationship (e.g., linear) between the retrieval and the ground truth, however, like triple-collocation, is insensitive to the spatial mismatch between point-based measurements and grid-scale retrievals. This theory and method is therefore suitable for use with both dense and sparse validation networks.

Additionally, the method we propose is *constructive* in the sense that it provides guidance on how to improve data assimilation systems. All data assimilation strategies can be reduced to approximations of Bayes' law, and we measure the fractions of total information loss that are due to individual assumptions or approximations in the prior (i.e. the model uncertainty distribution), and in the likelihood (i.e. the observation operator and observation uncertainty distribution). In this way, we can directly identify the parts of a data assimilation algorithm that contribute most to assimilation error in a way that (unlike traditional DA performance metrics) considers nonlinearity in the model and observation and non-optimality in the fit between filter assumptions and the real system.

To reiterate, the method we propose is theoretically rigorous but also dead-to-rights simple, and can be implemented in no more than a few hours by a competent programmer.

We use this to show that careful applications of the Ensemble Kalman Filter use substantially less than half of the information contained in remote sensing soil moisture retrievals (LPRM, AMSR-E, SMOS, and SMOPS). We propose that this finding may explain some of the results from several recent large-scale experiments that show lower-than-expected value to assimilating soil moisture retrievals into land surface models forced by high-quality precipitation data. Our results have important implications for the SMAP mission because over half of the SMAP-affiliated “early adopters” plan to use the EnKF as their primary method for extracting information from SMAP retrievals.