Lecture 19: Putting it all together (Everything, everywhere, all at once

Recap

We've not reached the end of the quarter. Last time we looked closely at objective mapping and other strategies for placing irregularly sampled data on a regular grid.

Course topics

This quarter we've covered a range of topics:

1. Basic statistics, probability density functions, autocovariance, and degrees of freedom
2. Least squares fitting, weighted least squared and constrained least squares (along with a review of linear algebra)
3. Eigensystems and empirical orthogonal functions
4. Objective mapping (and some variations and machine learning analogs)

In least-squares fitting, we minimize the misfit between a model fit to observations $\mathbf{y}$, finding parameters $\mathbf{m}$ for a model fit $\tilde{\mathbf{y}} = \mathbf{G}\mathbf{m}$ that lead to the smallest possible model–data misfit:

$$\mathcal{L} = \sum(\mathbf{y} - \mathbf{G}\mathbf{m})^{\in} = \sum(\mathbf{y} - \tilde{\mathbf{y}})^{\in} \tag{1}$$

In objective mapping, we aimed to minimize the differeence between a true value $\hat{y}$ and our estimated quantity $\tilde{y}$, finding the smallest misfit between a model estimate and the unknown truth, given data and assumptions about the covariances between the data and model. In this case for each mapped data value $\tilde{y}$, we minimize

$$\mathcal{L} = (\hat{\dagger} - \tilde{\dagger})^{\in} \tag{2}$$

Emipirical orthogonal functions provide us with a means to identify the dominant patterns of variability, with no prior knowledge about model functions or covariance.

And machine learning strategies can draw on all of these conceps to search for key behaviors in data.

A bit about state estimation (or data assimilation)

One challenge in data analysis is to keep track of the physics. In the ocean and atmosphere, the Navier–Stokes equations govern motions, so it's natural to ask if we can incorporate the constraints imposed by the Navier–Stokes equations into our analysis of observations. Can we allow estimated fields to move forward in time in a way that is consistent with the physics. This is the goal of data assimilation.

When we constrain a model with observations, we're effectively doing the same thing that we do in least-squares fitting, but we can imagine systems with much more complicated equations. Numerical weather prediction and ocean state estimation problems are effectively large least-squares fitting problems, aimed at minimizing model–data misfit, taking into account data uncertainties. We can define:

$$\mathcal{L} = \sum_{\rangle=\infty}^{\mathcal{N}} \frac{\sum(\tilde{\dagger}_{\rangle} - \dagger_{\rangle})^{\in}}{\sigma_{\rangle}^{\in}}, \tag{3}$$

where y represents data, $\tilde{y}$ represents a model fit, and σ is the data uncertainty. The model $\tilde{y}$ conforms to the Navier–Stokes equations.

How do you make model and data match. Wunsch et al. (2009) explain that two strategies have been conserved:

1. A time reversal process called a Kalman filter.
2. A method based on Lagrange multipliers (much like the examples we looked at in this class) which is the basis for 4-dimensional variational assimilation (4dVar).

In practice, 4dVar has proven practical for many weather forecasting and ocean data assimilation applications. In this process, the numerical model is run forward in time, misfits are identified, and the model boundary conditions and parameters are adjusted to reduce misfit. Since the model variables are not reset to new values, the physics of the system is preserved.

The process of minimization requires differeniate a cost function, which is carried out using automated differentiation to find the Lagrange multipliers, which can evolve in time. This is the adjoint of the model.

The slides contain a few examples from the ECCO project.

An exercise

The final segment of class is to put the resources that we've discussed to work in thinking about experiment design. What questions interest you? How can you use the tools that we've investigated to address these? What do you need to measure, and what will you calculate? In class, students worked in groups to consider examples aimed at probing air–sea exchange.

Reference

Wunsch, C., P. Heimbach, R. M. Ponte, I. Fukumori, and the ECCO-GODAE Consortium Members, 2009. The global general circulation of the ocean estimated by the ECCO Consortium, *Oceanography*, **22**, 88-103, doi:10.5670/oceanog.2009.41.