Introductory Data Assimilation

Jesse Dumas

September 14, 2018

This is a discussion on data assimilation and its use in climate modelling. The focus is in on building a general understanding of the concept. A data assimilation tutorial is presented, and overcoming model biases with data assimilation is considered.

Introduction and Overview

The ability of fully coupled climate simulations, like any numerical model, to make accurate predictions depends strongly on initial conditions and how accurately those initial conditions reflect the actual system of interest. Even if the model perfectly represented the Earth's climate, small errors in the initial conditions can drastically change the outputs of the simulation since the system is nonlinear and chaotic. Additionally, climate models can drift from reality due to their biases [1]. One way to counter biases in a numerical model is to make small corrections to the model output based on observations from a real physical system. Broadly, data assimilation (DA) is a group of methods for incorporating observational data into a model to improve the model prediction.

[1] S Zhang, Z Liu, A Rosati, and T Delworth. A study of enhancive parameter correction with coupled data assimilation for climate estimation and prediction using a simple coupled model. *Tellus A*, 64, 2012

Though DA applied to a single model can be effective, more often the method is applied to ensemble models. Running many instances of a model with varied initial conditions permits quantifying uncertainty in the outputs, and this promotes better probabilistic forecasts.

This project was a literature review of DA to form a basic understanding of the topic. To put my new knowledge to the test, I explored an example DA problem using the Data Assimilation Research Testbed (DART), developed and maintained by the Data Assimilation Research Section (DAReS) at the National Center for Atmospheric Research (NCAR). Lorenz's simplified convection model[2] was used to test an ensemble Kalman filter (EnKF). The example was a so-called perfect model since the observations correcting the ensemble were sampled from the same model and chosen as the "true state."

[2] Edward N Lorenz. Deterministic nonperiodic flow. *Journal of the atmospheric sciences*, 20(2):130–141, 1963

I will be discussing the basic procedure behind DA methods, but I will not go into detail on the workings of the EnKF, as that is beyond the scope of this project. A brief overview of the Lorenz model is also provided before discussing the results from the tutorial and recommendations for future work.

Theoretical Background

The Data Assimilation Framework

At it's heart, DA as applied to ensemble models is an iterative Bayesian process. As ensemble predictions of the system state evolve in time, observations with their associated probabilities of being true are incorporated with the model states using Bayes rule

$$P(T|T_O, C) = \frac{P(T_O|T,C)P(T|C)}{P(T_O|C)}, \tag{1}$$

where the color of the term indicates

- Posterior or update
- Likelihood
- Prior
- Normalization.

Figure 1 illustrates the DA process.

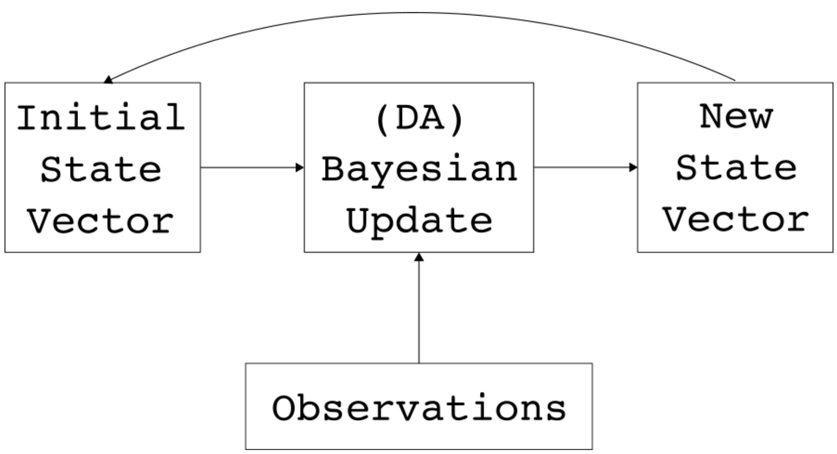

Figure 1: A schematic example of a Bayesian iterative process in DA.

The effect of the update from observations is to bring the value of an ensamble member state closer to the observation. This impact is shown in Figure 2.

DART

The DART tutorial I focused on utilizes the EnKF [3] and an ensemble of 20 members. The "true state" of the system is harvested from one instance of the model. Figure 3 shows an overview of their process.

[3] DAReS. Dart lab tutorial section 1: Ensemble data assimilation concepts in 1d. [Online; accessed 20-April-2017]

Figure 2: Ensemble members move closer to observed state estimates.

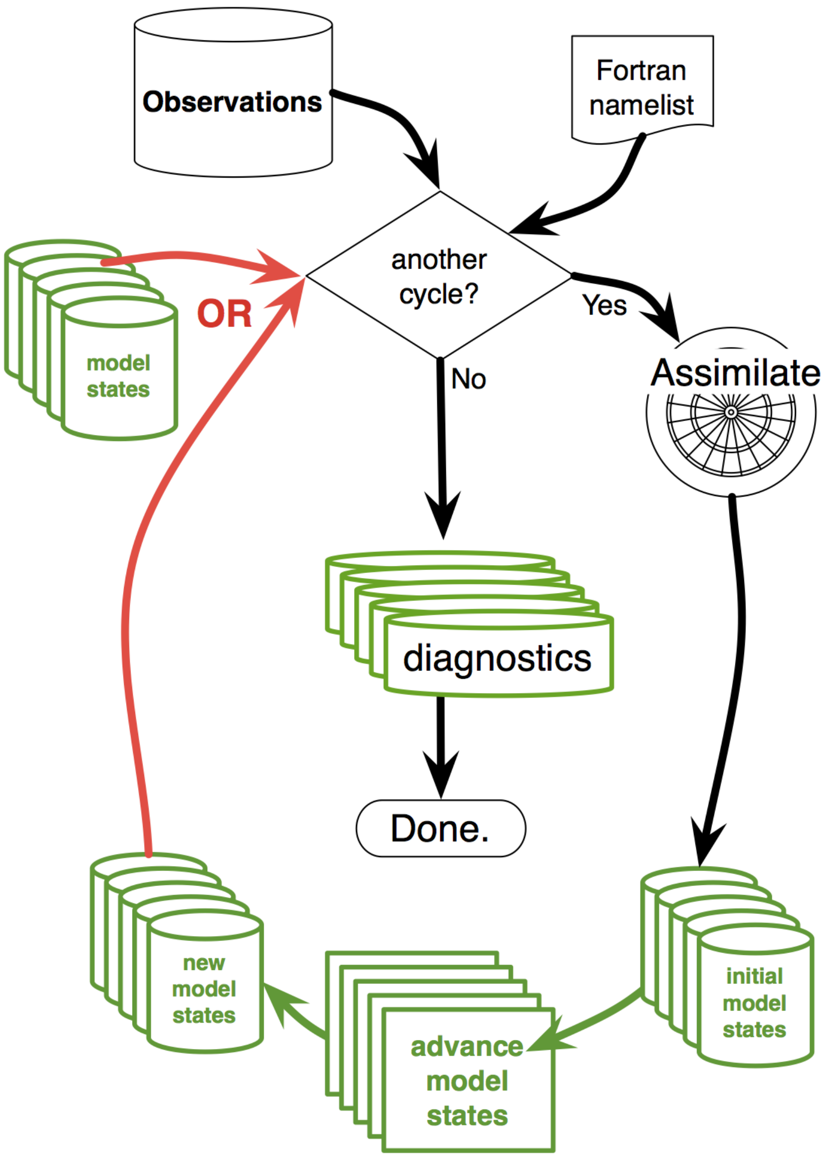

Figure 3: The DART process.

Lorenz's Model

Before moving forward with the details of data assimilation, it is helpful to understand the model Lorenz was interested in. The system is a simplification of Staltzman's[4] system, which is based on studies by Rayleigh[5]. Rayleigh's study focused on a flow of uniform depth H between two surfaces differing in temperature by a constant ΔT, analogous to the Earth's surface and the top of the atmosphere. In the steady state solution of the system, the temperature varies linearly with depth, and no motion occurs. In the interesting cases, the solution is unstable, and convection develops. For the case where all motion is parallel to the x-z plane and no motion takes place in the y-direction, the governing equations are

[4] Barry Saltzman. Finite amplitude free convection as an initial value problem. *Journal of the Atmospheric Sciences*, 19(4):329–341, 1962

[5] Lord Rayleigh. On convection currents in a horizontal layer of fluid, when the higher temperature is on the under side. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 32(192):529–546, 1916

$$\frac{\partial}{\partial t}\nabla^2\psi = -\frac{\partial(\psi, \nabla^2\psi)}{\partial(x,z)} + \nu\nabla^4\psi + g\alpha\frac{\partial\theta}{\partial x}\,, \tag{2}$$

$$\frac{\partial}{\partial t}\theta = -\frac{\partial(\psi,\theta)}{\partial(x,z)} + \frac{\Delta T}{H}\frac{\partial\psi}{\partial x} + \kappa\nabla^2\theta\,. \tag{3}$$

The constants ν, g, α, and κ are the kinematic viscosity, acceleration due to gravity, coefficient of thermal expansion, and thermal conductivity, respectively, and ψ is the stream function. Here, θ is the departure in temperature from the stable case. Saltzman, and subsequently Lorenz, expanded ψ and θ in double Fourier series in x and z. Lorenz truncated the series to three terms and let

$$a(1+a^2)^{-1}\kappa^{-1}\psi = X\sqrt{2}\sin(\pi a H^{-1}x)\sin(\pi H^{-1}z)\,, \tag{4}$$

$$\pi Ra_c^{-1} Ra \Delta T^{-1}\theta = Y\sqrt{2}\cos(\pi a H^{-1}x)\sin(\pi H^{-1}z) - Z\sin(2\pi H^{-1}z)\,, \tag{5}$$

where X, Y, and Z are strictly functions of time, and Ra and Ra_c are the Rayleigh number and critical Rayleigh number, respectively. Substituting equations 4 and 5 into 2 and 3 gives the system used by Lorenz (and DART)

$$\dot{X} = -\sigma X + \sigma Y\,, \tag{6}$$

$$\dot{Y} = -XZ + rX - Y\,, \tag{7}$$

$$\dot{Z} = XY - bZ\,, \tag{8}$$

where the derivative denoted by a dot is with respect to dimensionless time $\tau = \pi^2 H^{-2}(1+a^2)\kappa t$. In these equations $\sigma = \kappa^{-1}\nu$ is the the Prandl number (the ratio of momentum diffusivity to thermal diffusivity), $r = Ra_c^{-1}Ra$ indicates the departure from the critical Rayleigh number above which motion is triggered in the system, and $b = 4(1+a^2)^{-1}$). The variable X is proportional to the convective motion intensity, Y is proportional to the temperature difference between rising and falling currents, and Z is proportional to the deviation of the vertical temperature gradient from linearity.

Results

The DART Tutorial

The DART tutorial helped to build some intuition for how DA methods work, however, really digging into the model parameters required Fortran and NCAR command language familiarity somewhat beyond my abilities and time resources. I was nonetheless able to step through the tutorial with default values and put DA into practice.

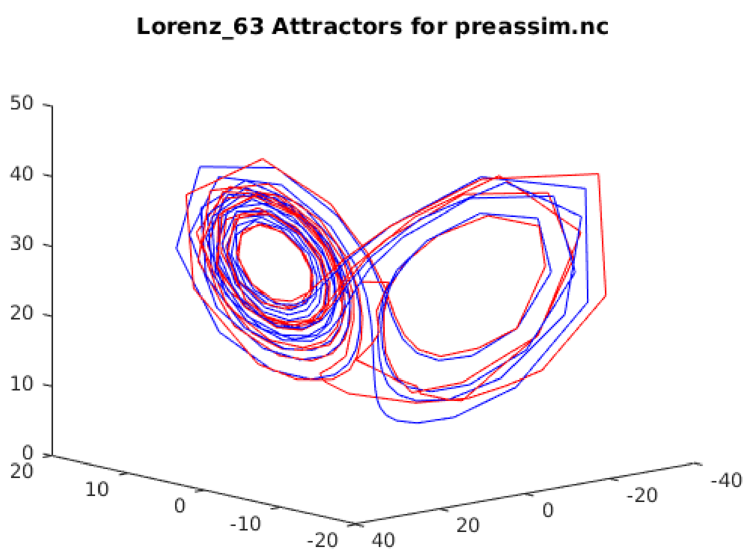

Figure 4: A plot of the X, Y, Z trajectory, showing the iconic Lorenz attractor. The "true state" is shown in blue, and the ensemble mean is in red.

The ability of the ensemble mean to track the true trajectory of the system is shown in Figure 4. Figure 5 illustrates the ensemble states of X, Y, and Z and the ensemble mean. At times when the ensembles start to diverge, assimilating observations clearly brings them closer to the true state.

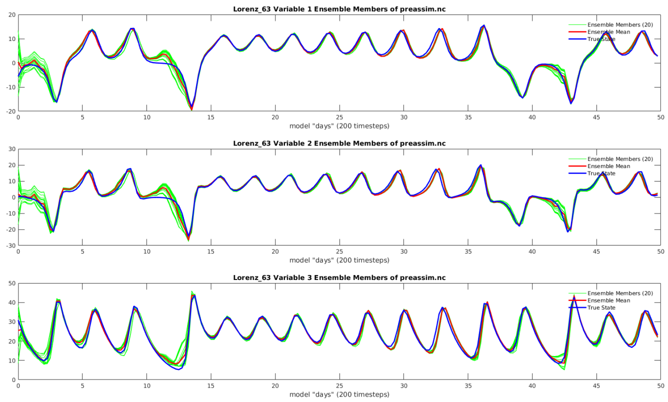

Figure 5: The ensemble runs of the model and the ensemble mean.

Summary and Conclusions

Data Assimilation methods are powerful tools for overcoming biases in climate models[6]. When coupled with ensemble methods, DA can narrow the bound on uncertainty for the modeled system. Current research in DA is working toward dealing with biases that can be introduced by DA methods[7]. Known weaknesses are being addressed, and bias-aware DA techniques are being developed in part by improving uncertainty quantification for observations.

I am now personally interested in applying DA in my own research. The computational simulations I run and the experimental work of my research group are both geared toward developing a 1-D model of a launch system. I believe incorporating observations using DA techniques could improve the accuracy of my simulations, which would feed back into better understanding experimental results.

[6] Dick P Dee. Bias and data assimilation. *Quarterly Journal of the Royal Meteorological Society*, 131(613):3323–3343, 2005

[7] Juanzhen Sun. Convective-scale assimilation of radar data: progress and challenges. *Quarterly Journal of the Royal Meteorological Society*, 131(613):3439–3463, 2005

References

[1] S Zhang, Z Liu, A Rosati, and T Delworth. A study of enhancive parameter correction with coupled data assimilation for climate estimation and prediction using a simple coupled model. *Tellus A*, 64, 2012.

[2] Edward N Lorenz. Deterministic nonperiodic flow. *Journal of the atmospheric sciences*, 20(2):130–141, 1963.

[3] DAReS. Dart lab tutorial section 1: Ensemble data assimilation concepts in 1d. [Online; accessed 20-April-2017].

[4] Barry Saltzman. Finite amplitude free convection as an initial value problem. *Journal of the Atmospheric Sciences*, 19(4):329–341, 1962.

[5] Lord Rayleigh. On convection currents in a horizontal layer of fluid, when the higher temperature is on the under side. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 32(192):529–546, 1916.

[6] Dick P Dee. Bias and data assimilation. *Quarterly Journal of the Royal Meteorological Society*, 131(613):3323–3343, 2005.

[7] Juanzhen Sun. Convective-scale assimilation of radar data: progress and challenges. *Quarterly Journal of the Royal Meteorological Society*, 131(613):3439–3463, 2005.