Outline of course

- Introduction to Data Assimilation
- Bayesian Data Assimilation
- Sequential Data Assimilation
  - Ensemble Kalman Filter
- Variational Data Assimilation
  - 3DVar and 4DVar

What is Data Assimilation (and why do we need it?)

- General circulation patterns are largely responsible for large-scale climate variations
  - This includes the positions of storm tracks, prevailing wind directions, monsoonal circulations, and other defining features of a region's climate
  - These patterns occur at scales large enough to be adequately reflected in the station data
  - These patterns are not explicitly accounted for by interpolation methods.
- Climate models predict how average conditions will change in a region over the coming decades.
  - Includes atmospheric, oceanic and land processes
  - Typically defined by partial differential equations discretized at thousands of data points
  - Simulate the transfer of energy and water that takes place in climate systems.
- Spatial and temporal climate data are often key drivers of computer models and statistical analyses
  - These analyses form the basis for scientific conclusions, management decisions, and other important outcomes.

- Complete data defining all of the states of a system at a specific time are, however, rarely available.
- Both the models and the available initial data contain inaccuracies and random noise that can lead to significant differences between the predicted states and the actual states of the system.

Data Assimilation: Observations of the system over time can be incorporated into the model equations to derive 'improved' estimates of the states and also to provide information about the 'uncertainty' in the estimates.

Note: The problem of state-estimation is an inverse problem, however, traditional control techniques are not practicable and 'data assimilation' schemes have been developed to generate accurate state-estimates.

Data Assimilation Schemes

- Bayesian - maximum likelihood estimation and Bayes rule
- Sequential data assimilation: there is a 'dynamic observer,'
  - Observations are 'fed-back' into the model at each time they are available, and a best estimate is produced that is used to predict future states.
- Variational data assimilation: there is a 'direct observer,'
  - A feasible state trajectory is found that best fits the observed data over a time window, and the estimated states at the end of the window are used to produce the next forecast.

Note: Under certain mathematical assumptions, these approaches solve the same 'optimal' state-estimation problem

Climate Models

1. energy balance models (EBMs);
2. one dimensional radiative-convective models (RCMs);
3. two-dimensional statistical-dynamical models (SDMs);
4. three-dimensional general circulation models (GCMs).

We will focus on the mathematical concepts of Data Assimilation using idealized models

Climate Models

Zero-dimensional models

$$S(1-\alpha) = 4\epsilon\sigma T^4$$

s- solar energy, α reflectivity of Earth System, ϵ - infrared transmissivity, σ-relates temperature to radiant emission, T- global mean surface temperature.

$$C_p \frac{\partial T}{\partial t} = S(1-\alpha) - 4\epsilon\sigma T^4$$

C_p - specific heat capacity of earth system.

Requires numerical methods, e.g.

$$\frac{\partial T}{\partial t} \approx \frac{T(t_{j+1}) - T(t_j)}{k}$$

One-dimensional energy balance model

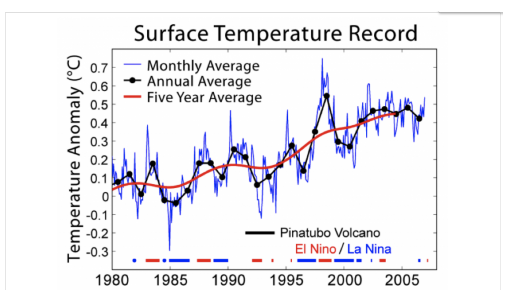
$$C_p \frac{\partial T_{lat}}{\partial t} = S(1-\alpha) - 4\epsilon\sigma T_{lat}^4 - k(T_{lat} - T_{global})$$

i.e. $T_{lat} = T(x) = T(x,t)$ so that we have

$$C_p \frac{\partial T(x,t)}{\partial t} = S(1-\alpha) - 4\epsilon\sigma T^4(x,t) - k\left(T(x,t) - T_{global}(t)\right)$$

This model requires an initial condition

You can see from the above example that if you have good weather stations spread uniformly across the planet (land and sea) and they have been recording continuously for a long time, then one can take the mean annual temperature of each station and calculate a simple global average for each year, and thus the history of temperature change for our planet. But, as you might imagine, the stations are not uniformly distributed — they are clustered in populated countries — and the number of stations declines as you go backward in time, so the actual process of assembling an instrumental record takes some care. The trick here is in how you combine the individual temperature records to come up with a global average. This is complicated by the fact that some weather stations may have problems related to things like the "urban heat island effect." Man-made materials retain heat better than open land and the lack of trees also amplifies warming in cities, which are currently warming at double the rate of the global average! Thus, if urban development encroaches on a weather station, the urban heat island effect will make the local temperature rise for reasons that are unrelated to any regional climate change. Researchers have found ways of ensuring that this effect does not skew the results, and many different groups come up with results that are nearly identical, giving us confidence that the data analysis is sound.



The last 25 years of globally averaged instrumental surface temperature measurements from the HadCRUT3 dataset.

Credit: Robert Rohde

It is interesting to see how similar the curves are given that they use different strategies for averaging the data, and some of the records are based on slightly different sets of weather stations. In particular, note that none of these estimates show a general cooling trend over this length of time — they all show warming. Back in the 1800s, there were fewer weather stations, and so it is more difficult to estimate global temperature back then (see figure below), but it gets steadily better as time goes on, and for the last few decades, we have excellent data due to the satellites that now circle the globe taking temperature measurements of every spot on Earth (more on this in a bit).