Comments about Reanalyses

Are Reanalyses Truth?

Reanalyses (CFSRR, NARR, R1, R2, ERA-40, ERA-15, ERA-interim, JRA-25, MERRA) are all similar

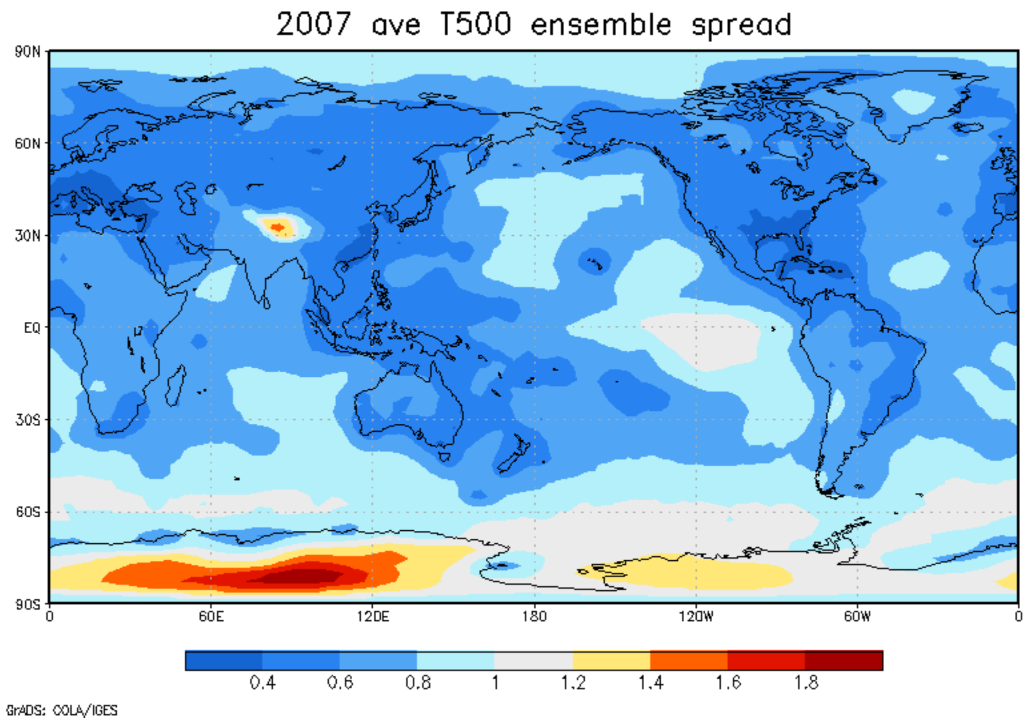
- products of numerical weather prediction systems
- use forecast model and data assimilation techniques
- start with an analysis (forecast-model basic state)
- 3D-Var: make a forecast with previous analysis, and find the atmospheric state that best fits the observations, the model forecast (first guess) and various constraints like geostrophic balance, thermal wind, etc.
- 4D-Var: find the time-evolving state that best fits the observations and the previous analysis
- Both 3D and 4D-Var depend on the forecast model and its adjoint model. Hence, areas where the model is weak will also be areas where the analyses would be of lower quality. For example, all models are reasonably good in the free troposphere but have difficulty in simulating the boundary layer and convection. Also some models have problems with the stratosphere.
- Observation density is an critical factor. Without enough observations, the analyses tends to drift towards the model climatology. The few observations that are present are often rejected by quality control as being too far from the first guess and unsupported by any neighboring observations (buddy check)
- U, V, T and Z are related quantities. Knowing T, for example, will tell you something about U, V and Z. These relationship mean that U,V,T and Z are higher quality.
- Humidity, on the other hand, has very few constraints. Given a point observation of humidity, there is no way to infer the gradient of the humidity. This is one of the reasons that the humidity analyses are of lower quality than the U,V,T and Z analyses.
- Some fields are model derived ("physics" terms)
- Difficult to assimilate surface variables (except pressure)

Are Observations Truth?
Representativeness error

People often consider in situ observations to be truth. They are the "truth" if you want a measurement at that point at that time; however, most people want a measurement that is representative of the environment. For example, the humidity from a sonde going through a cloud will different from a sonde launched 1 mile to the east and from the average humidity of a 1-degree grid box. The difference between the cell average and the point measurement is the “representativeness error”. Assimilation systems do not try to fit the observations exactly because of the instrument and representativeness errors. Consequently the difference between the analysis and assimilated observation is not a good measure of the quality of the analysis. Note, that if the analysis is tuned to fit the observations too closely, the forecast skill will be degraded which suggests a poorer analysis.

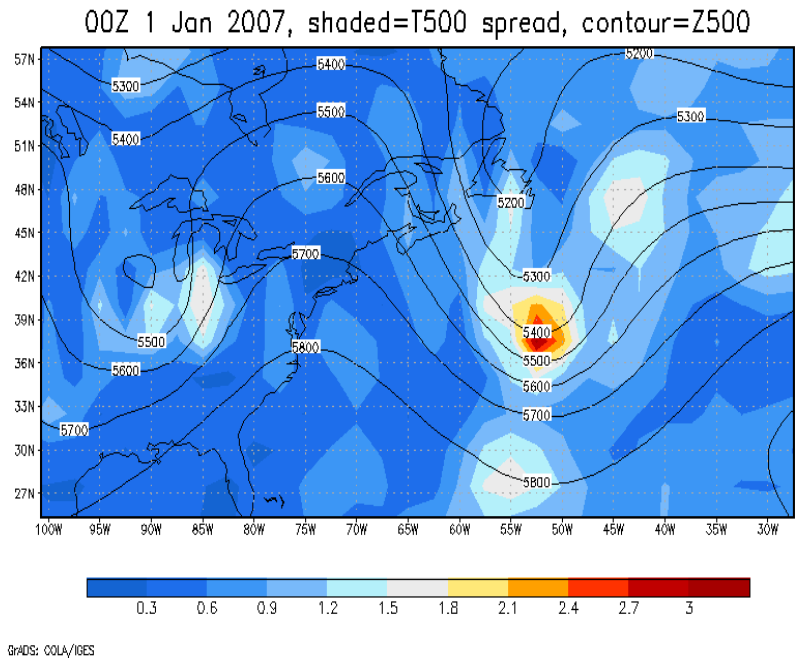
If there is no truth, we need error bars

Some approaches for calculating error bars include: comparison to independent (un-assimilated) observations, comparison to assimilated observations, and comparison to other (re)analyses. The latter can be done by obtaining a collection of operational analyses and various reanalyses (e.g., , NCEP-NCAR Analyses DVD, available from NCAR) and computing the standard deviation about the mean analysis. This, the ensemble spread, may underestimate the uncertainty because the analyses are not independent; however, it is easy and can be used in regions without corresponding observations.



In the following figure, the average 2007 spread for the 500 mb temperature shows lowest uncertainty over land as one would expect. The high values over the Himalaya and Antarctica are probably the result of the elevations approaching the 500 mb level and the few observations. (Most of the Antarctic observations are coastal.)

The uncertainty (ensemble spread) at any one day show much more small-scale features. These small-scale features are not stationary and are often associated with synoptic waves or tropical systems. Naturally, one would expect larger uncertainty in synoptic and convective systems. In the following plot the largest temperature uncertainties (shaded) are associated with the troughs in the 500 mb heights (contoured).




Monthly means have a smaller uncertainty than the average uncertainty for an individual analyses because the random errors are reduced when making the monthly mean as shown by the following plot (the uncertainty of the Jan 2007 monthly mean 500 mb temperature). The monthly and more so the seasonal means show the the uncertainty caused by biases in the model.

ensemble spread Jan 2007 monthly mean T500