

Analysis, reanalysis, forecast—what’s the difference?

This extends the discussion about the difference between an analysis and a forecast found in FNLvGFS.pdf.

<http://rda.ucar.edu/datasets/ds083.2/docs/FNLvGFS.pdf>

Read that first, and then come back.

“Reanalysis data provide a multivariate, spatially complete, and coherent record of the global atmospheric circulation. Unlike archived weather analyses from operational forecasting systems, a reanalysis is produced with a single version of a data assimilation system—including the forecast model used—and is therefore not affected by changes in method.” [Dee et al., 2011]

An analysis uses myriad observations on an irregular grid to produce a representation of the atmospheric state (values of the set of atmospheric parameters needed to specify the state) over a regular grid. Creators of analyses use a complex toolset that includes statistical measures of both the variability of the measurements and of the atmosphere itself (e.g. covariance matrices), physical models of how the atmosphere behaves (e.g. geostrophic and hydrostatic balance), and mathematical physics models (e.g. continuity equation). This requires teams of specialists working over long periods of time.

Subtle differences distinguish analyses from each other; they may ingest different input data and may be “tuned” for different objectives. Which one is “better” is often determined by the intended use. The RDA provides a variety of analyses and reanalyses. We encourage users to try more than one and then report back to us, which better met their objectives.

Physical forecast models propagate an atmospheric state forward in time. In doing so, they can spatially fill in data from data rich regions into data sparse regions. For instance, they can advect (blow) air masses downwind. In contrast, an analysis can statistically spread an observation with Kriging or a relaxed Gaussian fit—but the spread is usually isotropic in space. Ideally, your knowledge of an atmospheric state should flow downwind, not upwind.

A good physical forecast model also forms a “memory” of the measurement. For instance, if you know the temperature, heat capacity of the air, and energy flux at time t , then you can predict the temperature at time $t+dt$. Given the same initial analysis fields as input, the output of the more skilled forecast model will more closely resemble measurements of the atmospheric state at a future time.

This sounds very circular, because it is. That's why it's called an analysis-forecast cycle.

Physical models also spread information from observed variables to calculate estimates of unobserved variables. For instance, it's not possible to blanket a wide area with rain gauges. However, physical models can calculate orographic rain (when moist air masses are pushed uphill until the moisture precipitates out) using upwind temperature, pressure, wind and relative humidity measurements along with terrain elevation and land use from a database.

Analyses and forecasts share many of the same parameters in common. However, the overlap is not perfect. There is one very important difference. Analyses are a snapshot in time. Forecasts can contain accumulated parameters such as rainfall over a time period. The FNL analysis (FNL files ending with _00) starts out "dry", with zero accumulated rainfall. As the GFS forecast model runs, the amount of rain precipitated out of the atmosphere is collected and summed at each time step. The GFS forecast (FNL or GFS files ending with a number greater than _00) will then contain non-zero accumulated rainfall fields.

Any changes in software for an operational analysis can result in spurious signals or shifts. Operational systems frequently change their software as they uncover bugs or biases and fix them. The changes are usually not announced ahead of time. The change log may be difficult for the non-expert user to decipher. Thus, operational analyses are not appropriate for compiling a long time series to study changes over time.

Reanalyses are a special type of analysis done with a fixed software system. Both the data assimilation and the forecast model software are "frozen" for the time span of a reanalysis. This inoculates them from changes in software.

Both analyses and reanalyses experience shifts due to changes in observation systems (e.g. when a new satellite comes on-line or an old one is decommissioned). Researchers using reanalysis time series should carefully check times of observed shifts with times of observation system shifts.

Note that even in-situ station data may experience shifts. For instance, a station may move away from urban encroachment or receive a new instrument package.

After I've just told you a reanalysis is just a special type of analysis, you may ask, "How come so many reanalyses contain rainfall parameters when analyses don't?" Preparers of reanalyses understand research needs. They often include the accumulated rainfall from the forecast model (used to create the background field for the next data assimilation cycle) as a field in the analysis.

A very good reference to learn more:

Dee, D. P., S. M. Uppala, A. J. Simmons, P. Berrisford, P. Poli, S. Kobayashi, U. Andrae et al. (2011). The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. Quarterly Journal of the Royal Meteorological Society, 137(656), 553-597, doi: 10.1002/qj.828.

Read <http://rda.ucar.edu/datasets/ds083.2/docs/FNLvGFS.pdf>

And send any questions not answered in that document or this one to grace@ucar.edu.

Grace Peng, PhD

Atmospheric & Geoscience Research Data Archive

Computational & Information Systems Laboratory

National Center for Atmospheric Research

303-497-1218

4 December 2014