Climate reanalysis: progress and future prospects

Ron Gelaro
Global Modeling and Assimilation Office
NASA Goddard Space Flight Center

Third Symposium on Multi-scale Predictability: Data-model Integration and Uncertainty Quantification for Climate and Earth System Monitoring and Prediction

AMS 98th Annual Meeting, Austin, Texas, 7 – 11 January 2018

Acknowledgment

Special thanks to **Adrian Simmons** of ECMWF, who graciously provided the material that appears on several slides in this talk.

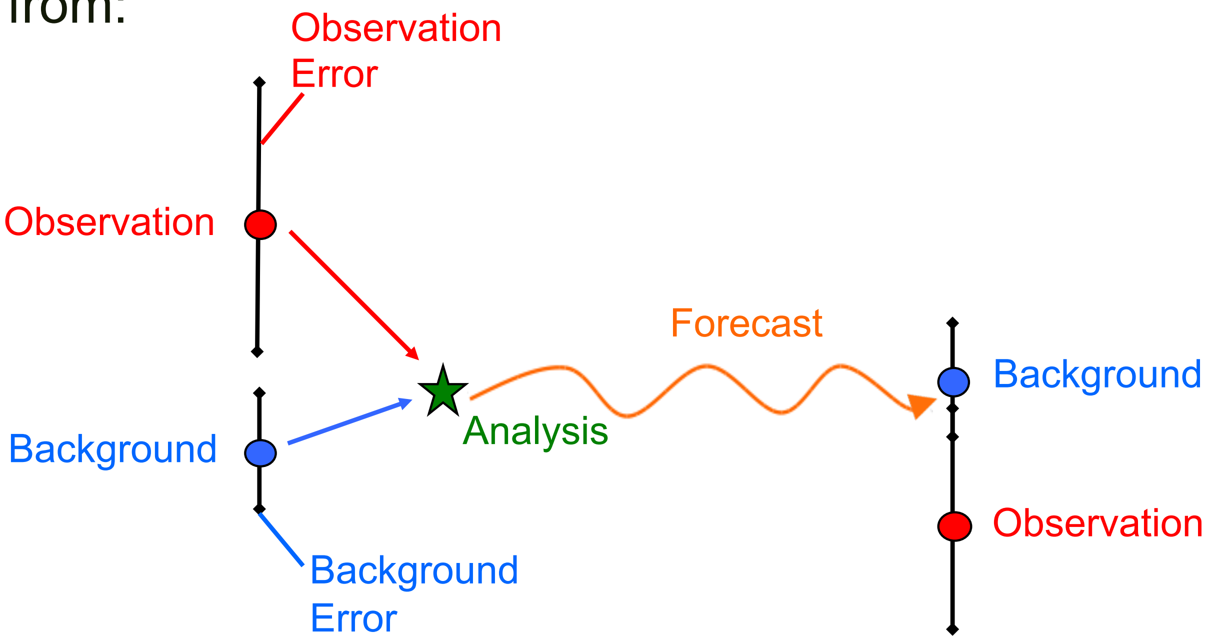
See his keynote address at the 5th International Conference on Reanalysis: https://climate.copernicus.eu/sites/default/files/repository/Events/ICR5/Talks/Simmons_keynote_ICR5_13pm.pdf

Reanalysis and data assimilation

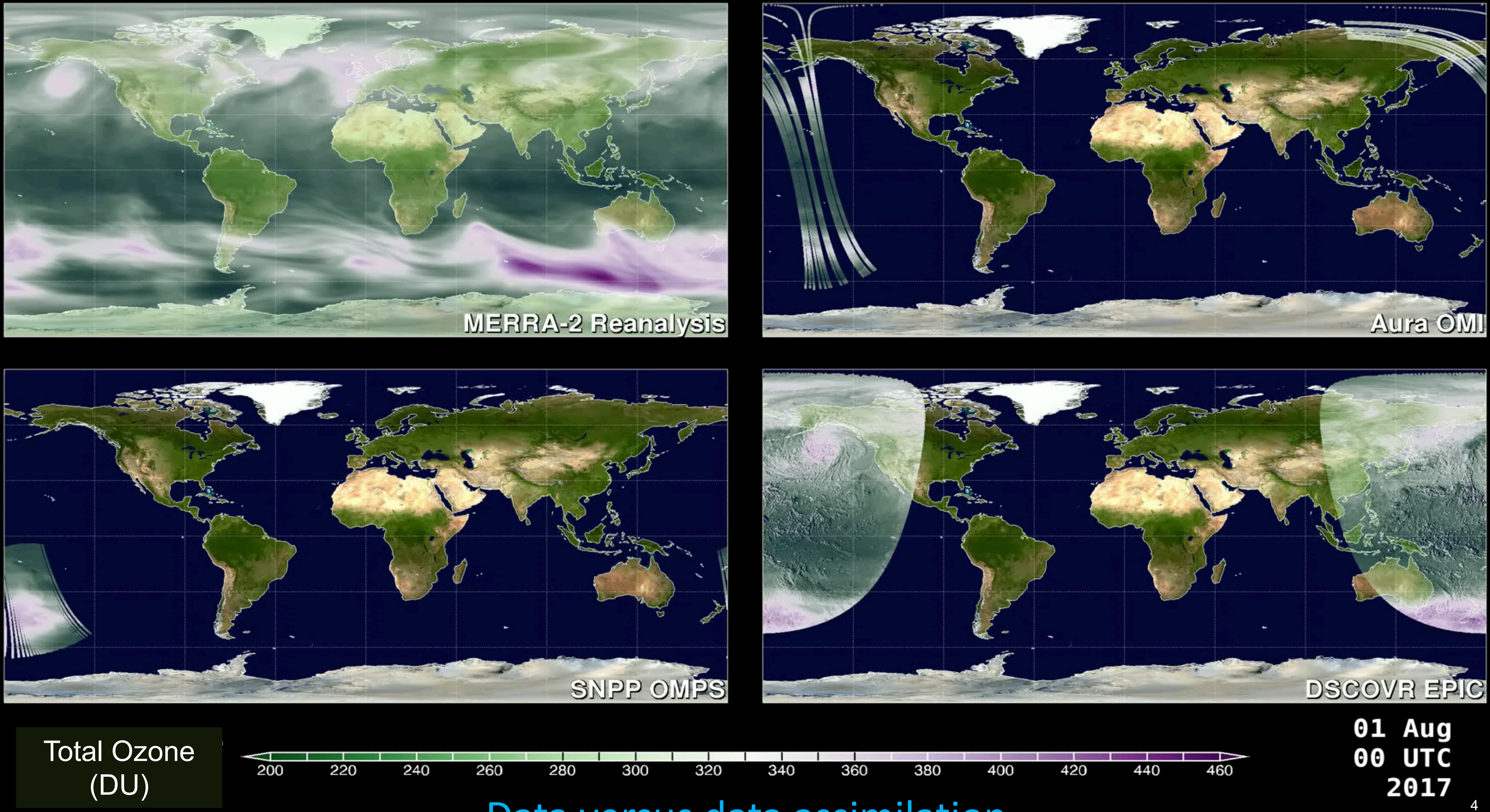
Reanalysis is the process whereby a *fixed* modern data assimilation system is used to provide a consistent reprocessing of observations, typically over an extended period

The **data assimilation** blends information from:

- many types of observations
- a short "background" model forecast
- estimates of observational and background errors, including biases
- dynamical relationships used in the representation of background errors



Data Assimilation



Data versus data assimilation

Development of atmospheric observations up to 1979

Early years	Growth of network of surface measurements Development of measurements from balloons
1940s	Establishment of network of radiosonde measurements from North Atlantic and North Pacific Weather Ships
1957	Radiosonde network enhanced in southern hemisphere for the International Geophysical Year
1972	Operational sounding of temperature and humidity from polar-orbiting satellite Some data from commercial aircraft
1979	Improved sounding from polar orbiters Winds from geostationary orbit Much more data from commercial aircraft Drifting buoys

Development of global atmospheric reanalyses

Global modeling became established for climate in the 1960s, and global systems for numerical weather prediction were introduced at NMC (NCEP) and ECMWF in the 1970s

ECMWF and GFDL produced analyses for 1979 from Global Weather Experiment data

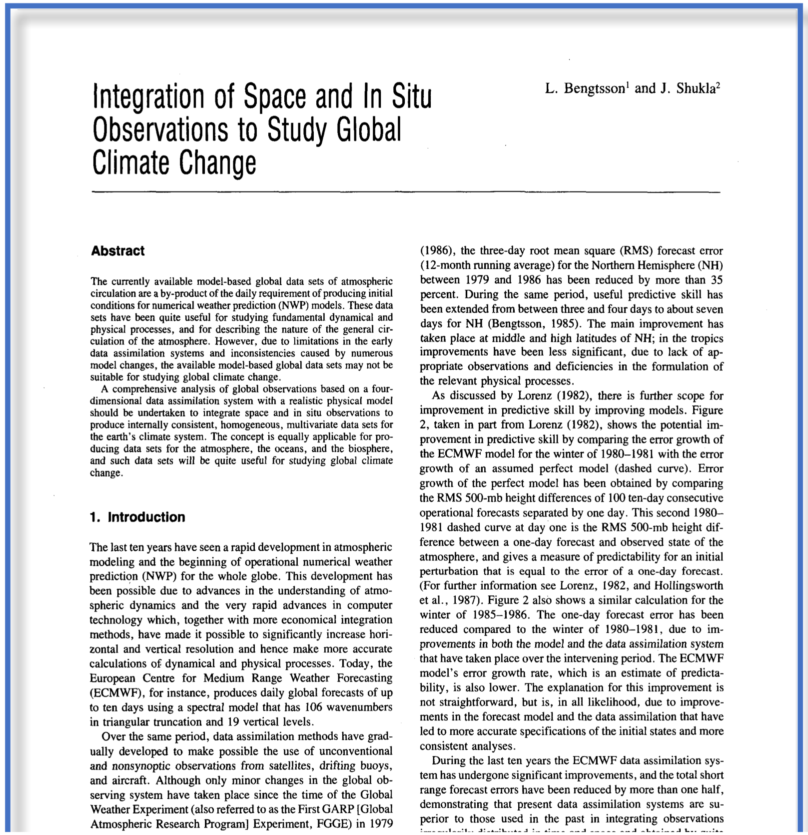
Analysis datasets were quite extensively used, but soon supplemented by global analyses from operational weather forecasting for multi-year studies

Frequent operational changes clouded the picture, leading to a call for reanalysis (Trenberth & Olson, 1988)

Bengtsson & Shukla (1988) made a more specific proposal for atmospheric reanalysis of the period 1979−1988 …

…and stated that the concept is equally applicable to the ocean and biosphere, and that reanalyses would be *"quite useful for studying global climate change"*

Bengtsson and Shukla, BAMS, 1988



Integration of Space and In Situ Observations to Study Global Climate Change

L. Bengtsson[1] and J. Shukla[2]

Abstract

The currently available model-based global data sets of atmospheric circulation are a by-product of the daily requirement of producing initial conditions for numerical weather prediction (NWP) models. These data sets have been quite useful for studying fundamental dynamical and physical processes, and for describing the nature of the general circulation of the atmosphere. However, due to limitations in the early data assimilation systems and inconsistencies caused by numerous model changes, the available model-based global data sets may not be suitable for studying global climate change.

A comprehensive analysis of global observations based on a four-dimensional data assimilation system with a realistic physical model should be undertaken to integrate space and in situ observations to produce internally consistent, homogeneous, multivariate data sets for the earth's climate system. The concept is equally applicable for producing data sets for the atmosphere, the oceans, and the biosphere, and such data sets will be quite useful for studying global climate change.

1. Introduction

The last ten years have seen a rapid development in atmospheric modeling and the beginning of operational numerical weather prediction (NWP) for the whole globe. This development has been possible due to advances in the understanding of atmospheric dynamics and the very rapid advances in computer technology which, together with more economical integration methods, have made it possible to significantly increase horizontal and vertical resolution and hence make more accurate calculations of dynamical and physical processes. Today, the European Centre for Medium Range Weather Forecasting (ECMWF), for instance, produces daily global forecasts of up to ten days using a spectral model that has 106 wavenumbers in triangular truncation and 19 vertical levels.

Over the same period, data assimilation methods have gradually developed to make possible the use of unconventional and nonsynoptic observations from satellites, drifting buoys, and aircraft. Although only minor changes in the global observing system have taken place since the time of the Global Weather Experiment (also referred to as the First GARP [Global Atmospheric Research Program] Experiment, FGGE) in 1979

(1986), the three-day root mean square (RMS) forecast error (12-month running average) for the Northern Hemisphere (NH) between 1979 and 1986 has been reduced by more than 35 percent. During the same period, useful predictive skill has been extended from between three and four days to about seven days for NH (Bengtsson, 1985). The main improvement has taken place at middle and high latitudes of NH; in the tropics improvements have been less significant, due to lack of appropriate observations and deficiencies in the formulation of the relevant physical processes.

As discussed by Lorenz (1982), there is further scope for improvement in predictive skill by improving models. Figure 2, taken in part from Lorenz (1982), shows the potential improvement in predictive skill by comparing the error growth of the ECMWF model for the winter of 1980–1981 with the error growth of an assumed perfect model (dashed curve). Error growth of the perfect model has been obtained by comparing the RMS 500-mb height differences of 100 ten-day consecutive operational forecasts separated by one day. This second 1980–1981 dashed curve at day one is the RMS 500-mb height difference between a one-day forecast and observed state of the atmosphere, and gives a measure of predictability for an initial perturbation that is equal to the error of a one-day forecast. (For further information see Lorenz, 1982, and Hollingsworth et al., 1987). Figure 2 also shows a similar calculation for the winter of 1985–1986. The one-day forecast error has been reduced compared to the winter of 1980–1981, due to improvements in both the model and the data assimilation system that have taken place over the intervening period. The ECMWF model's error growth rate, which is an estimate of predictability, is also lower. The explanation for this improvement is not straightforward, but is, in all likelihood, due to improvements in the forecast model and the data assimilation that have led to more accurate specifications of the initial states and more consistent analyses.

During the last ten years the ECMWF data assimilation system has undergone significant improvements, and the total short range forecast errors have been reduced by more than one half, demonstrating that present data assimilation systems are superior to those used in the past in integrating observations

Comprehensive atmospheric analyses

The first multi-year reanalyses were produced in the early to mid 1990s

ERA-15 (1979−93), NASA/DAO (1980−1993) and NCEP/NCAR (1948− …)

A second round of production followed

ERA-40 (1958−2001), JRA-25 (1979−2014) and NCEP/DOE (1979− …)

And a third

ERA-Interim (1979− …), JRA-55 (1958− …), NASA MERRA (1979−2016) and NOAA CFSR/CFSv2 (1979−2010/2011− …)

A fourth round is proceeding

MERRA-2 (1980− …) is now up-to-date and continues close to real time

ERA5 production is well advanced

JRA-3Q is planned to enter production in Japanese Fiscal Year 2018

CRA-40 will be produced by CMA using NOAA/NASA, NCAR systems; planned completion 2020

Atmospheric observations since 1979

Satellites become dominant…

Rain-sensitive **microwave imagery** data in substantial numbers since 1992

Surface wind information from **scatterometry** since 1992

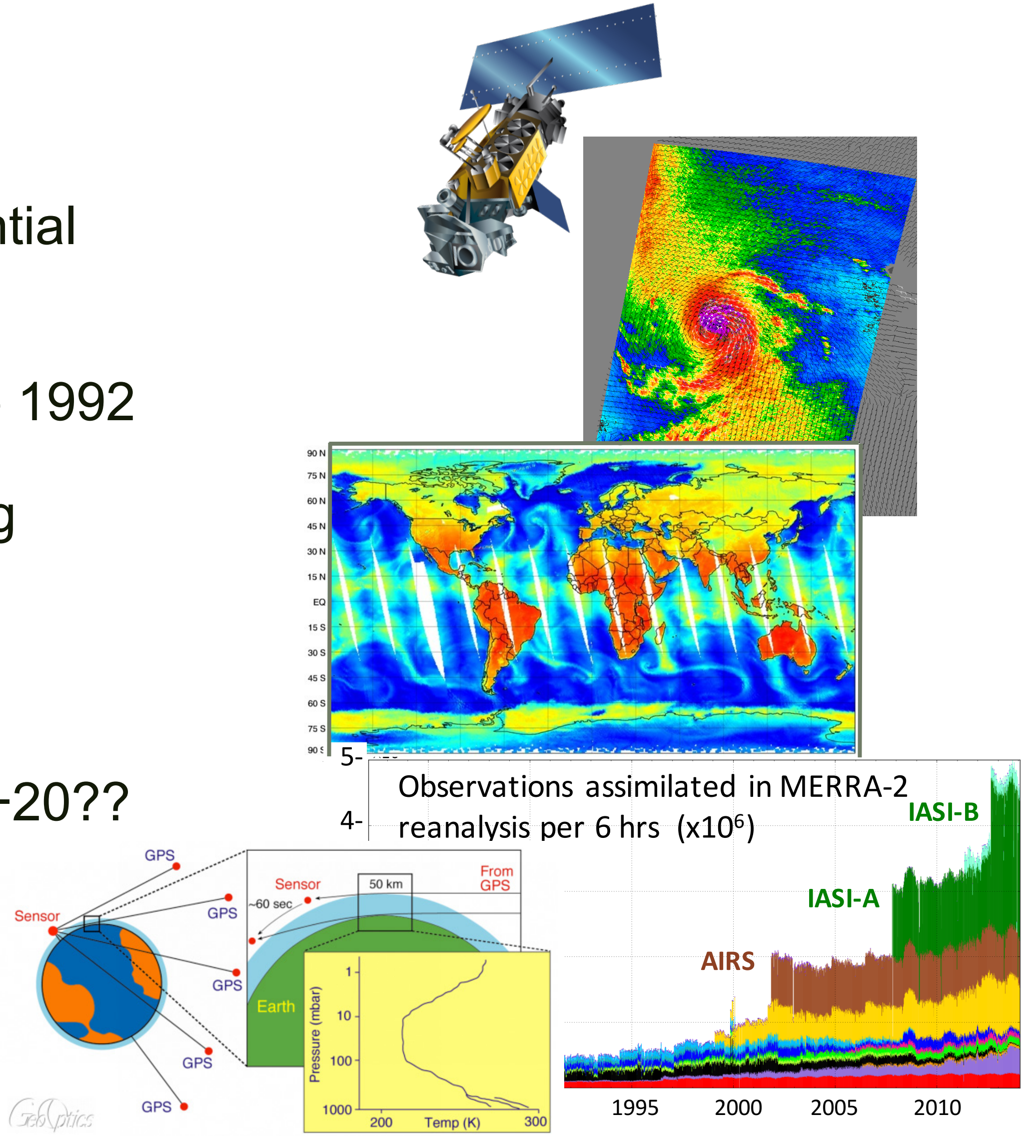
ATOVS (AMSU/MHS and improved HIRS) sounding starts in 1998; MSU & SSU end in 2006

Hyperspectral infrared sounding since 2002

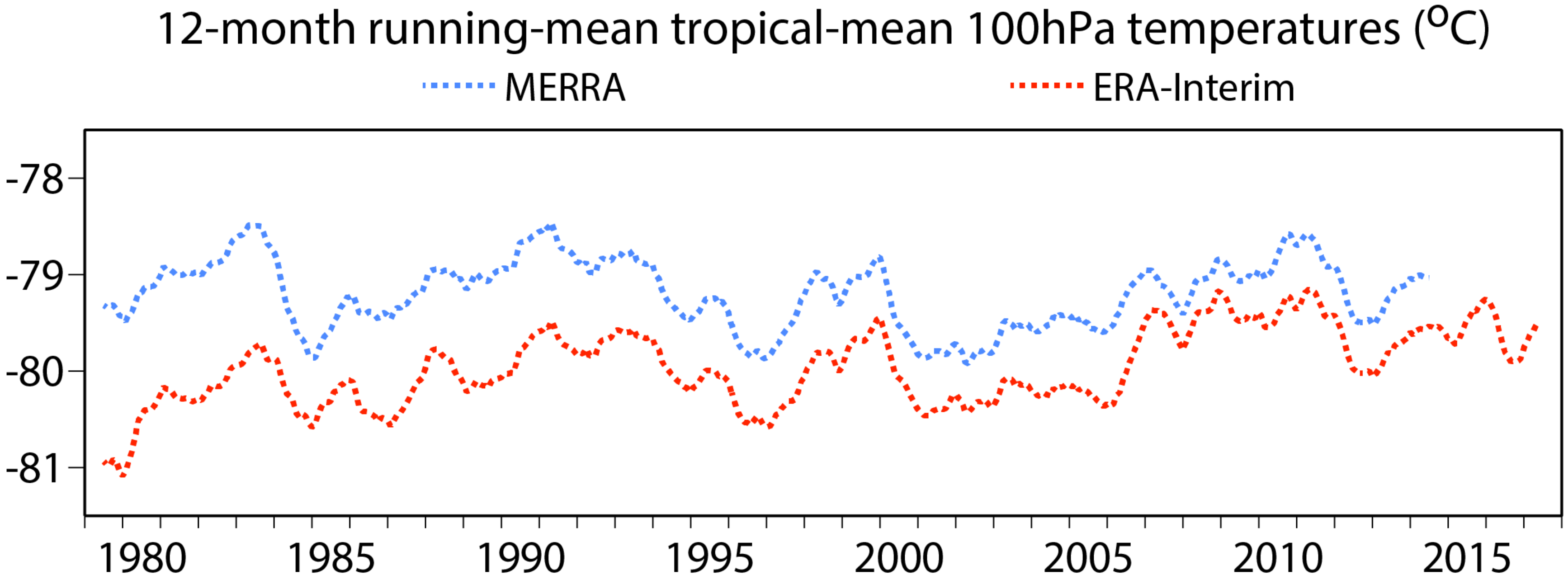
Microwave limb sounding (Aura MLS) from 2004−20??

GNSS (GPS) radio occultation data in substantial numbers since 2006

…



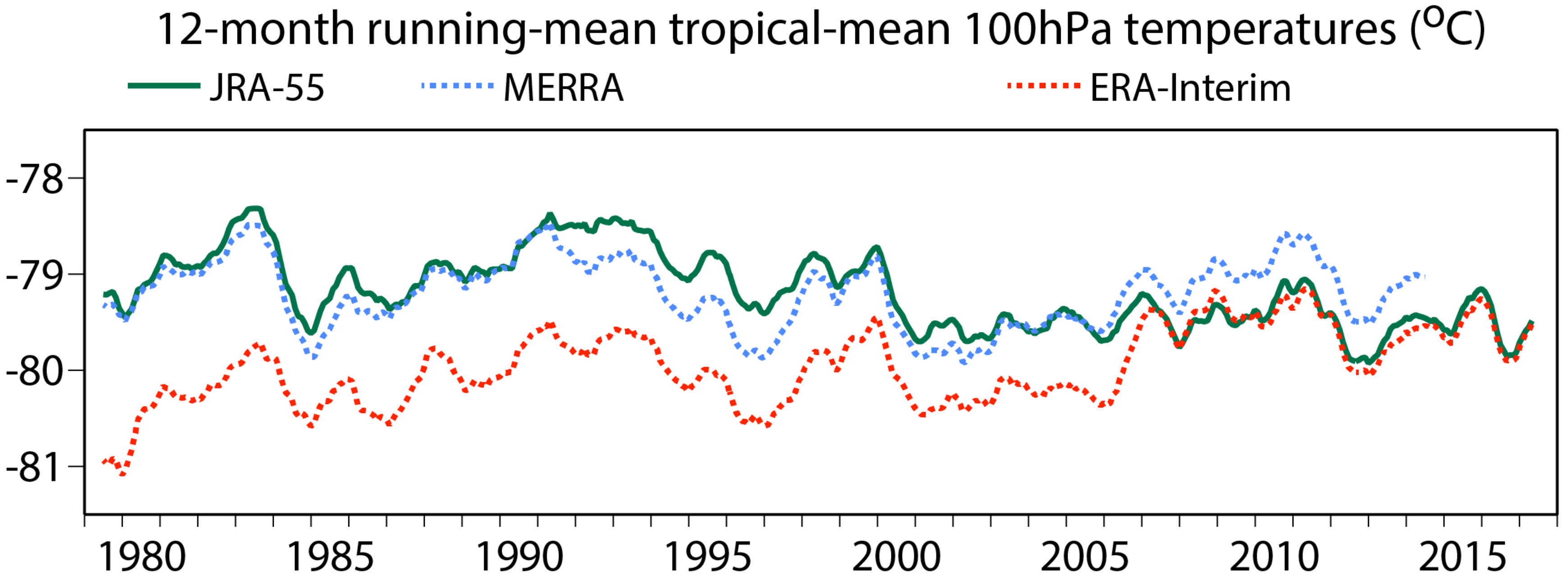
Tropical tropopause temperature



Significant amounts of GPSRO data assimilated in ERA-Interim but not in MERRA

A. Simmons, ECMWF

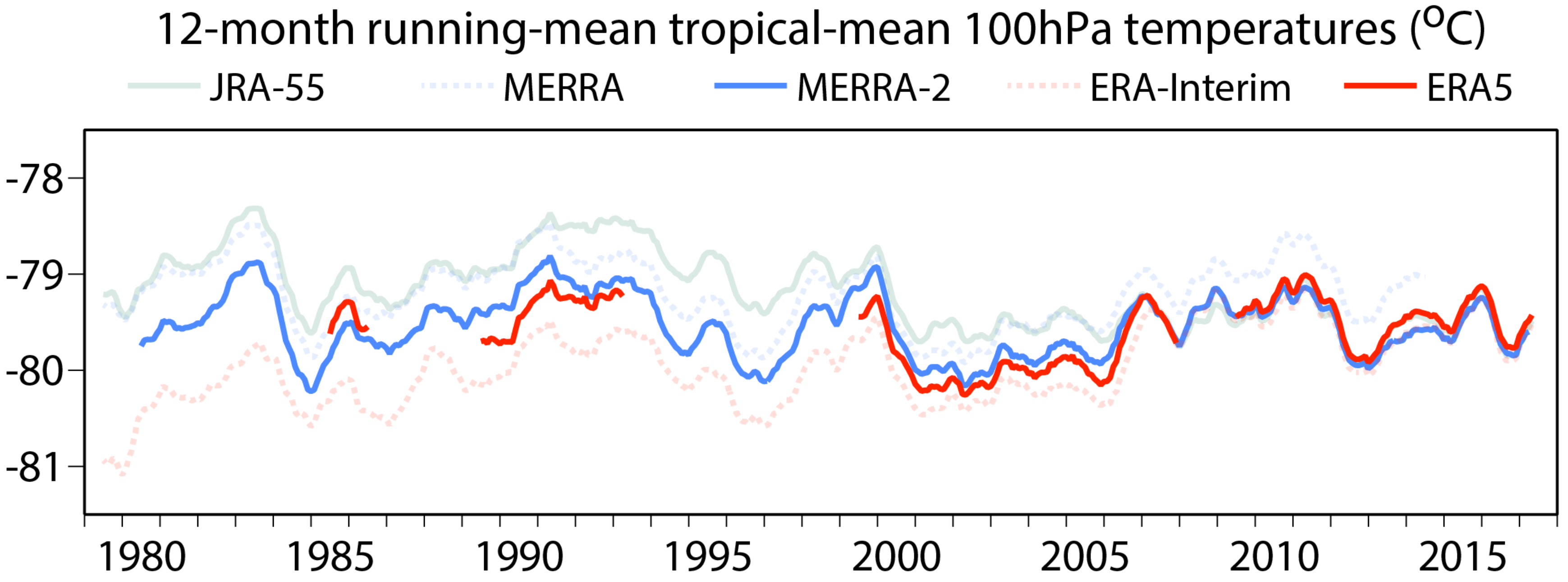
Tropical tropopause temperature



Significant amounts of GPSRO data assimilated in ERA-Interim and JRA-55

A. Simmons, ECMWF

Tropical tropopause temperature



Significant amounts of GPSRO data assimilated in all except MERRA

Assimilating GPSRO data brings reanalyses together from 2006, and gives confidence in the earlier values from ERA5 and MERRA-2

A. Simmons, ECMWF

Use of reanalyses in operational forecasting

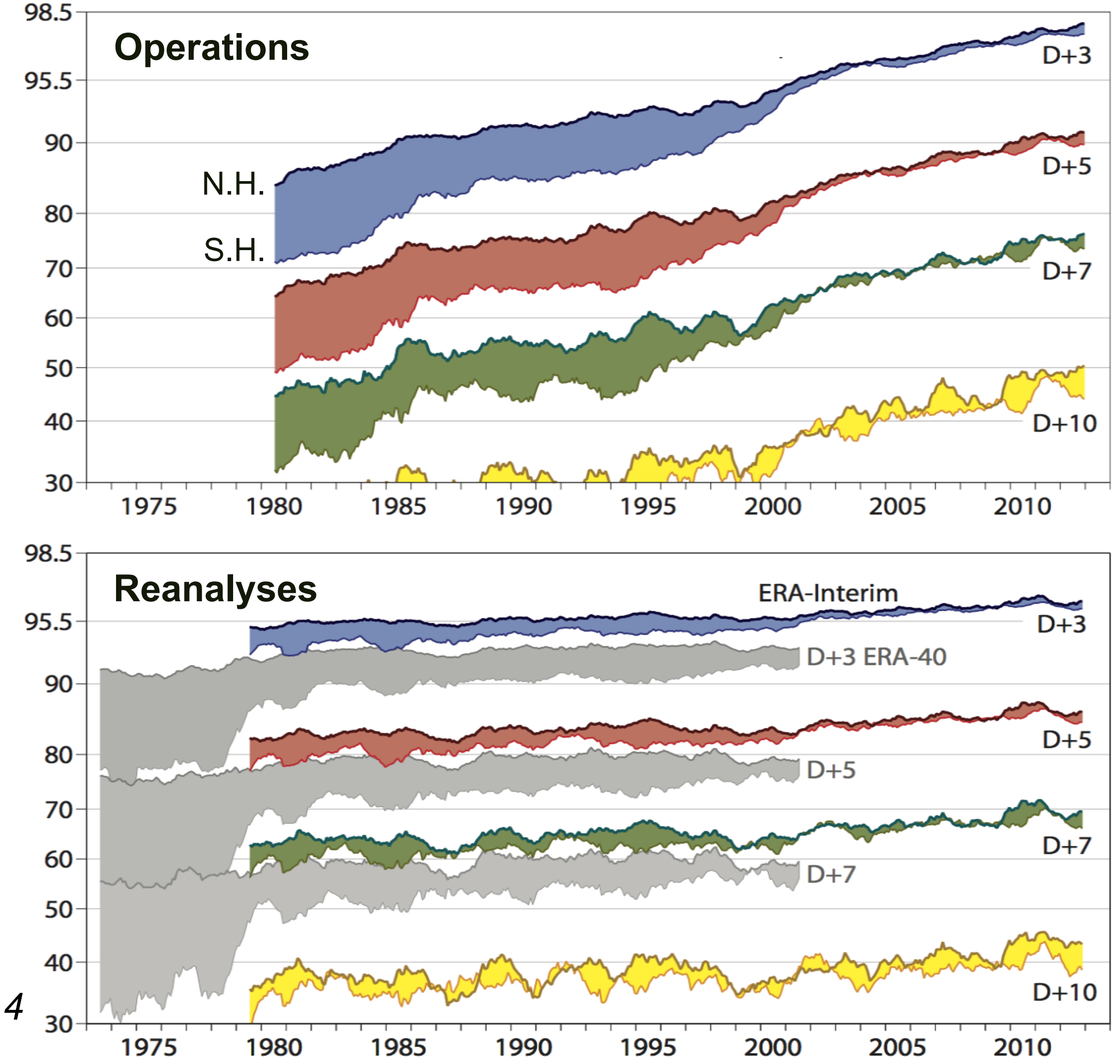
‘Reforecasts’ from reanalyses help assess and improve the performance of operational forecasts

- Quantify contributions to improved skill:
  - the overall observing system
  - satellite data coverage
  - satellite data assimilation

Also used to:

- Estimate the model climate distribution to *predict extremes as seen by the model*
- Calibrate seasonal forecasts by computing the model error relative to the reanalysis

Anomaly Correlation 500 hPa Height

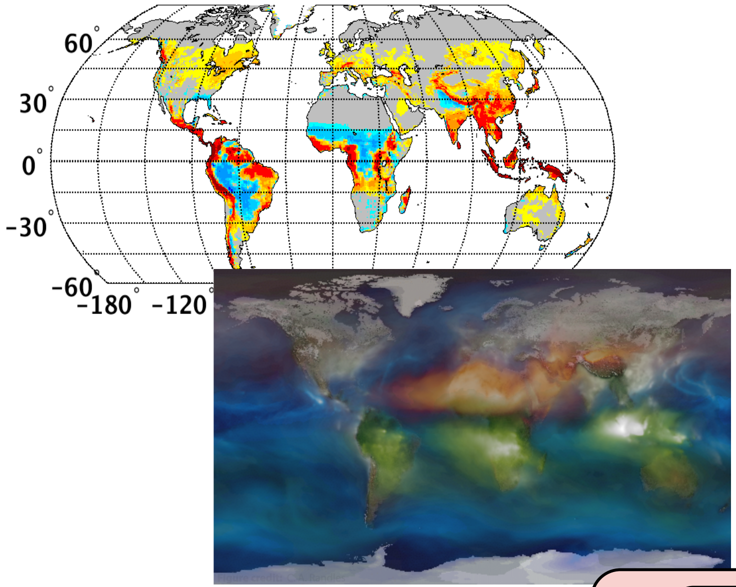


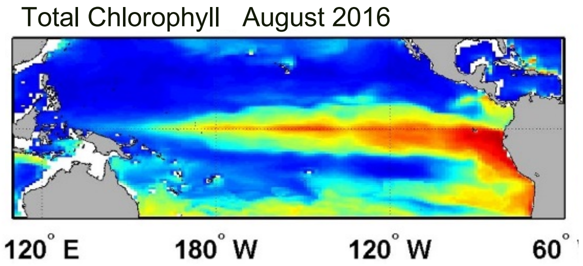
From Dee et al. 2014

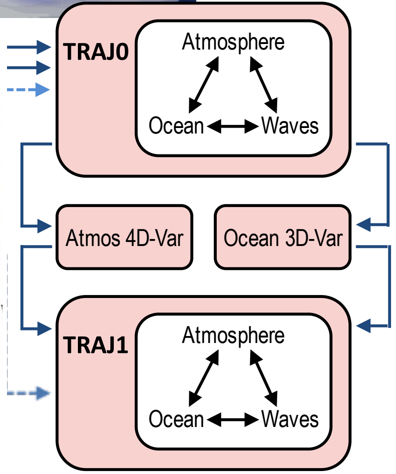
Reanalysis diversification and coupling

Reanalyses have become more diverse, with varied but generally increasing levels of coupling

- **Land**
  - coupled with the atmosphere, possibly using observed instead of model precipitation
  - stand alone and possibly downscaled
- **Atmospheric composition**
  - trace species in addition to ozone, driven by or coupled with the atmosphere
  - aerosols, reactive chemical species, GHGs
- **Ocean circulation**
  - possibly including sea-ice or biogeochemistry
- **Coupled atmosphere-ocean-land**
  - strongly or weakly coupled
  - moving closer to Earth-system reanalysis

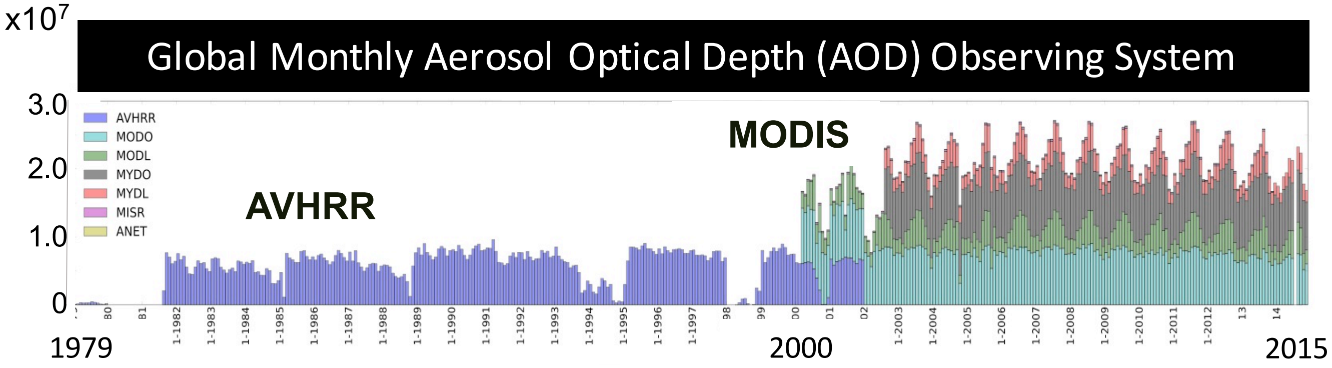



Total Chlorophyll August 2016

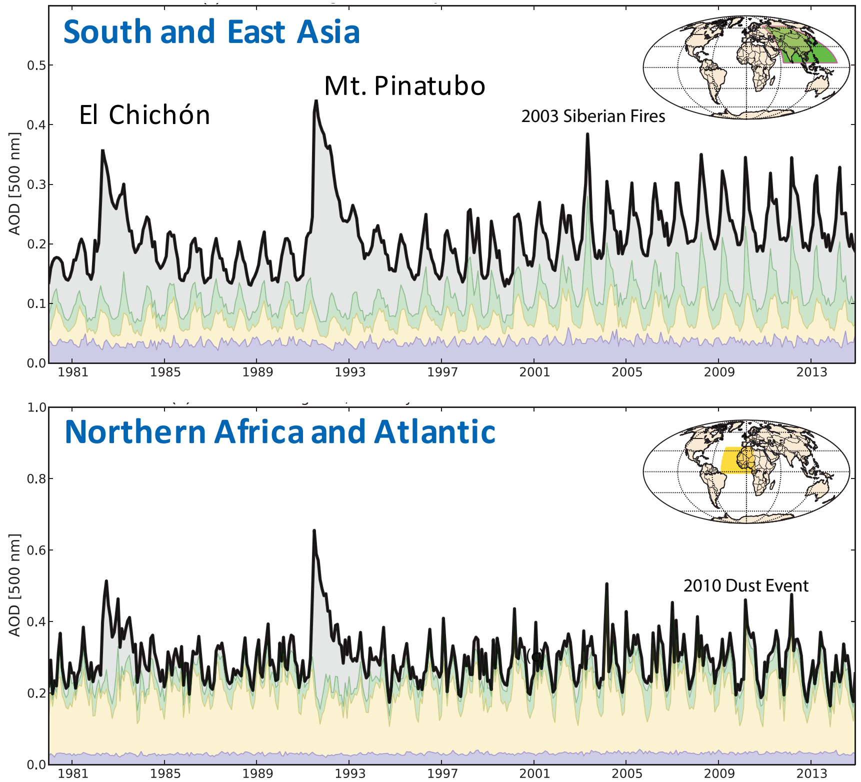


Aerosol data assimilation in MERRA-2

Observations of **total column AOD** number ~900K per day since 2001, (compared to ~100M per day currently available for NWP)

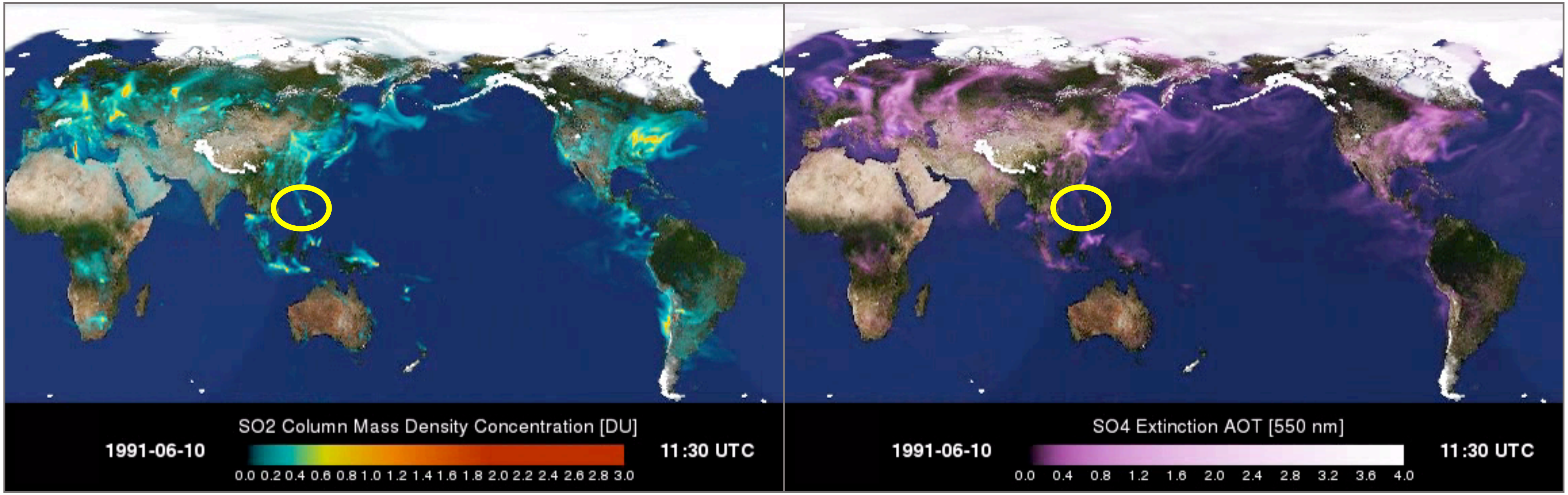


In MERRA-2 the meteorological and aerosol analyses are performed separately but aerosols feed back to meteorology via the AGCM radiation



MERRA-2 time series of aerosols over Asia and northern Africa, showing regionally dominant species and major volcanic eruptions

Mt. Pinatubo eruption in MERRA-2



Gaseous SO_2

Aerosol SO_4

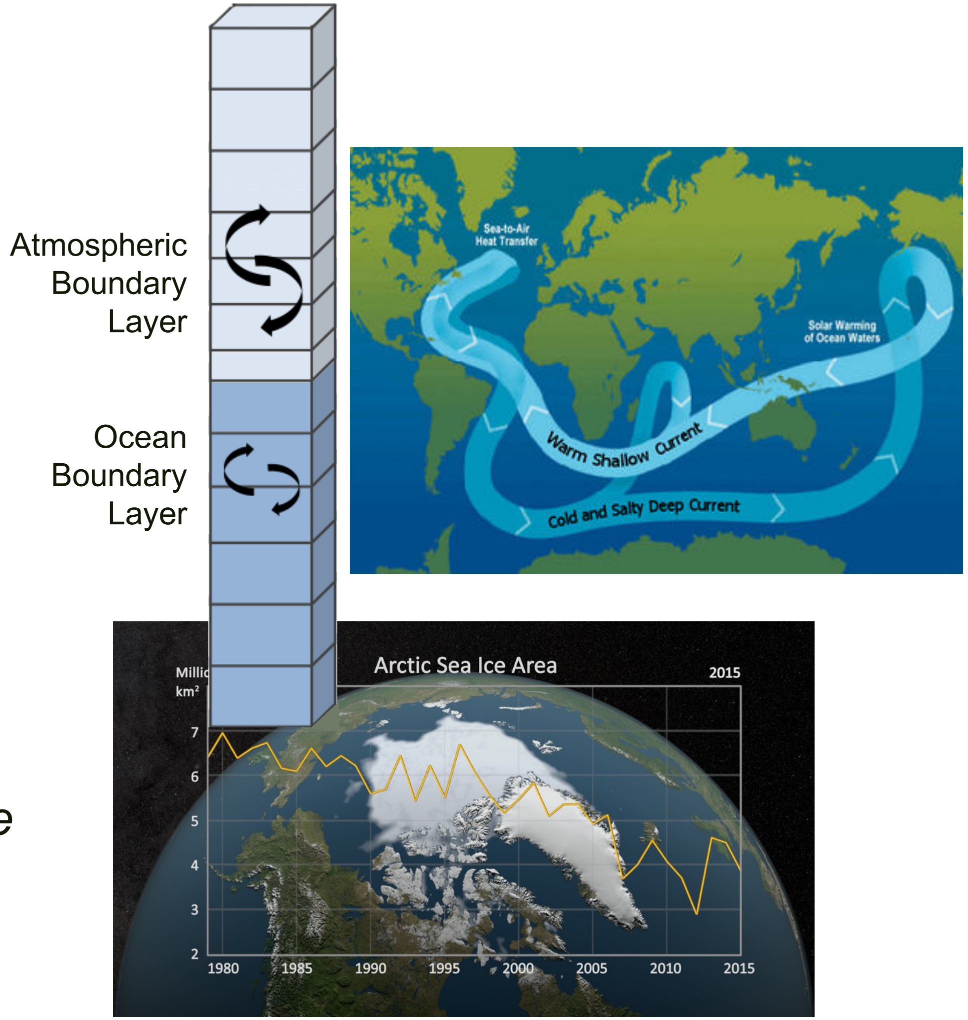
- Co-evolution of gaseous SO_2 emissions from Pinatubo (left) and formation of the sulfate aerosol plume (right) as SO_2 is converted into particles.
- SO_2 gas is from emissions inventories and unconstrained by assimilation. Sulfate aerosol AOD is impacted by the assimilation of total aerosol AOD.

Coupling the atmosphere and ocean

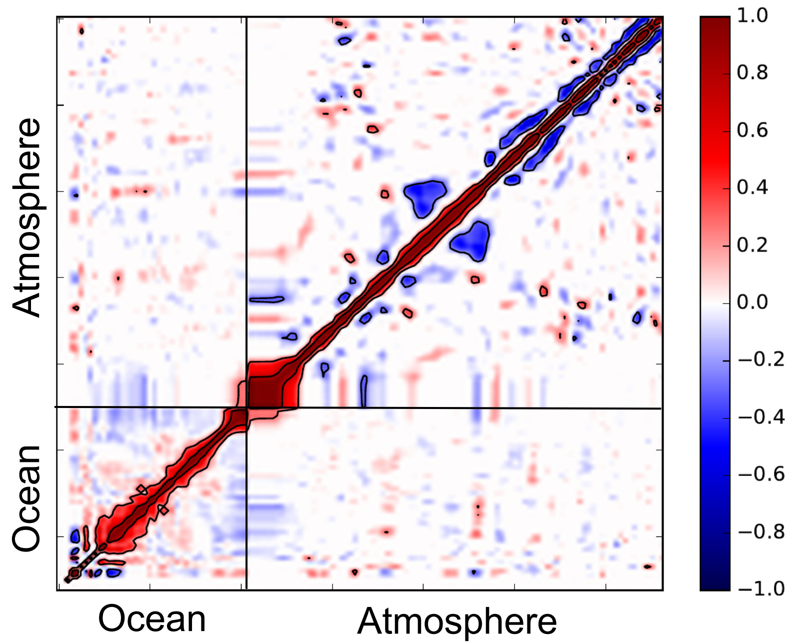
Essential for improving reanalysis and seasonal-to-decadal prediction, but also an increasing priority for NWP

- Physical consistency
- Improved use of observations, especially near the interface
- Reduced uncertainty

For reanalysis, consider the fact that the best available observational estimates of global sea surface temperature cannot be considered reliable on time scales less than a month or so in the pre-satellite era. (Dee et al. 2014)

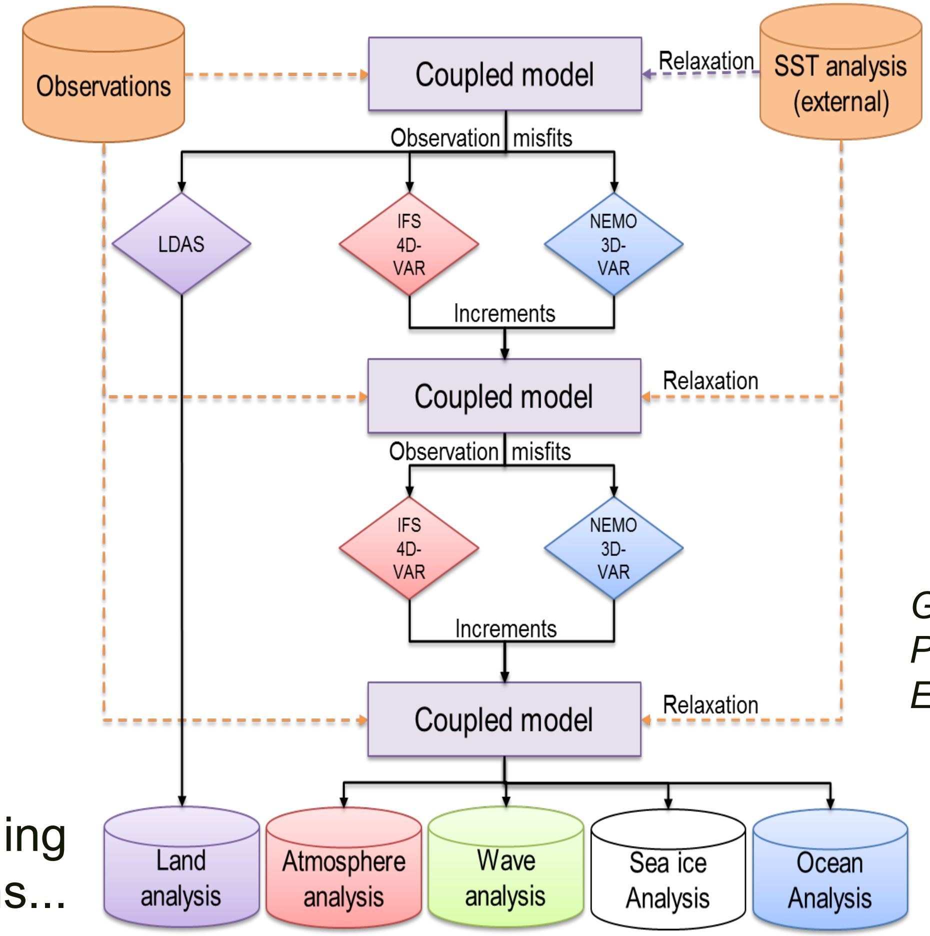


But how much coupling?



A **single analysis** with **explicit** coupled background error covariances...

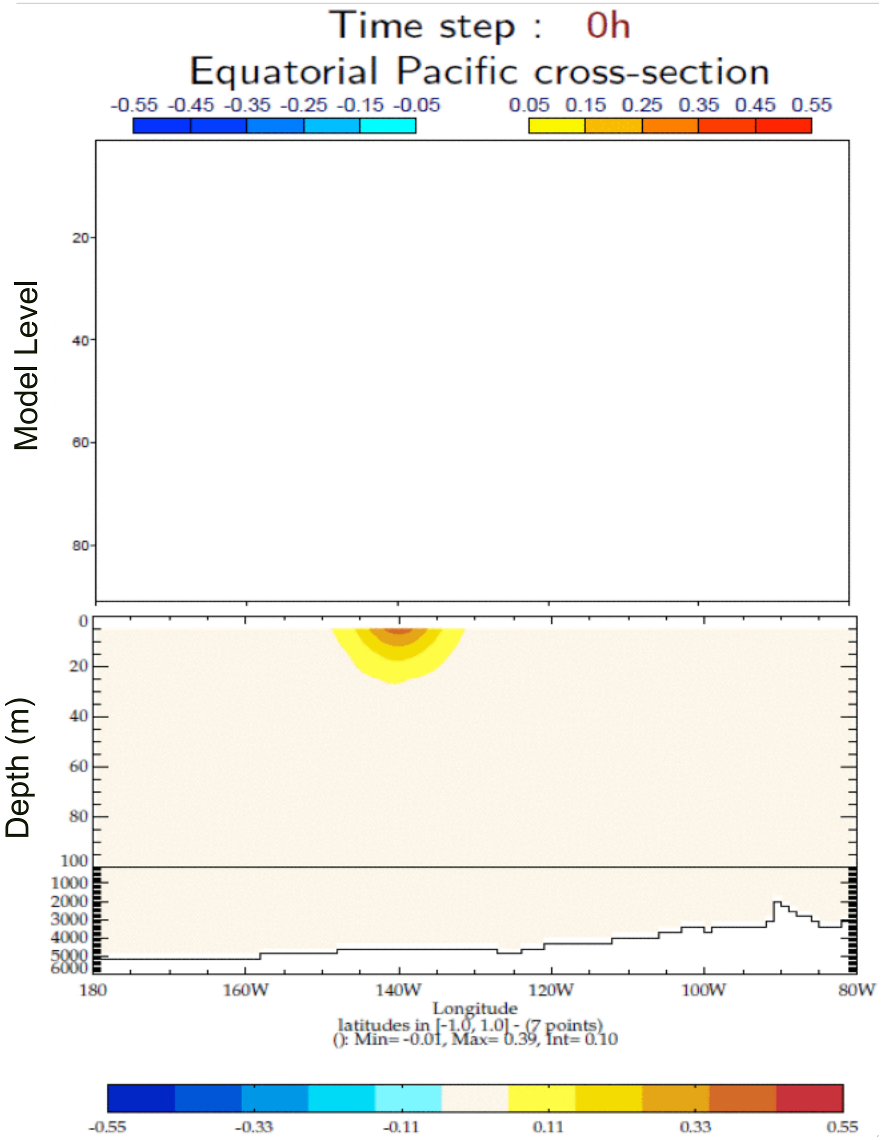
ECMWF Coupled Earth Reanalysis System (CERA)



Graphics P. Laloyaux, ECMWF

Separate analyses with outer loop (incremental) coupling to generate **implicit** cross-component correlations...

Information exchange in a coupled assimilation system



Temperature cross-section in the ECMWF CERA coupled ocean-atmosphere data assimilation system

Ocean increment (assimilation of one temperature observation at 5-meter depth) spreads in the atmosphere during the assimilation process

(*Laloyaux et al. 2016, QJRMS*)

More recent work shows the similarity of these cross-correlations to the explicit ones used in a simple coupled Kalman filter (previous slide)

Development of the ocean observing system

Sea surface temperature data evolve from buckets to engine intakes, drifting buoys and (from the early 1980s) satellites

Sea ice data from microwave imagery began in 1978

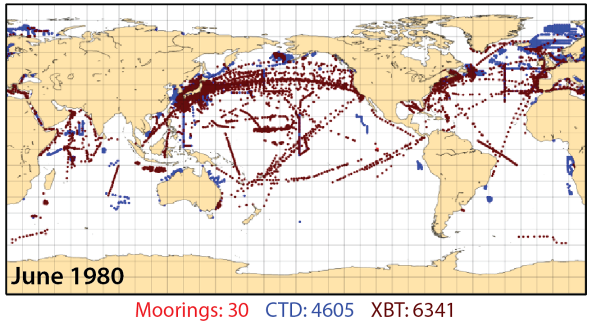
Temperature and salinity profiles from ship transects (XBTs and CDTs) increased from the 1960s to the 1990s

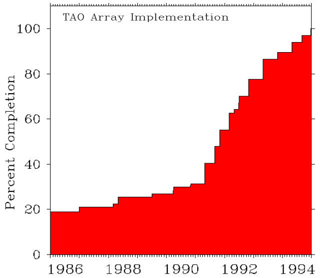
Tropical moored array was built up 1984−1994 under the WCRP TOGA programme

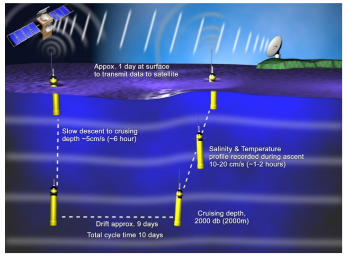
Sea level has been sensed from space since late 1992

Argo profiling-float network was established between 2003 and 2007, and expanded thereafter

… **and more** …







Prospects for improved forecast skill with coupling...

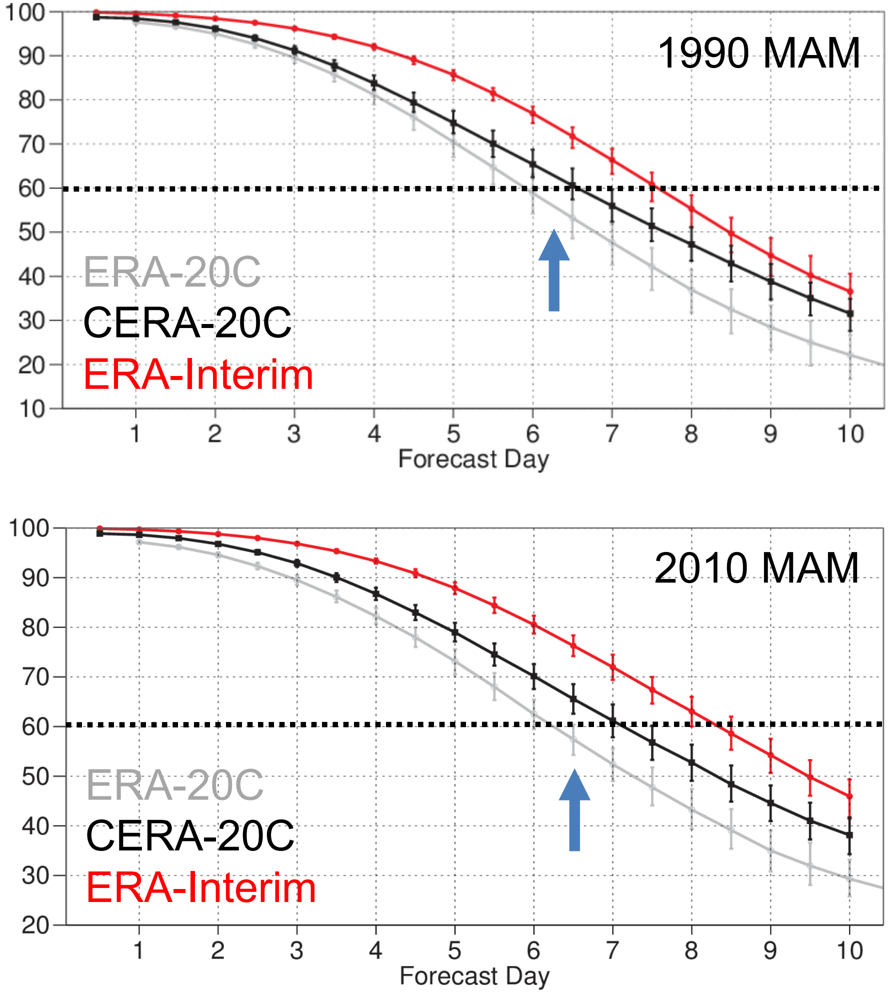
Comparison of forecast skill from reanalyses with and without coupled atmosphere-ocean data assimilation

- **ERA-20C**: ECMWF centennial atmospheric reanalysis of surface conventional observations only
- **CERA-20C**: Like ERA-20C atmosphere but with coupled assimilation of ocean salinity and temperature profiles

At the 60% threshold, forecast skill is improved by roughly 12 hours in 1990 and by almost 24 hours in 2010 between ERA-20C and CERA-20C

Courtesy P. Laloyaux, ECMWF

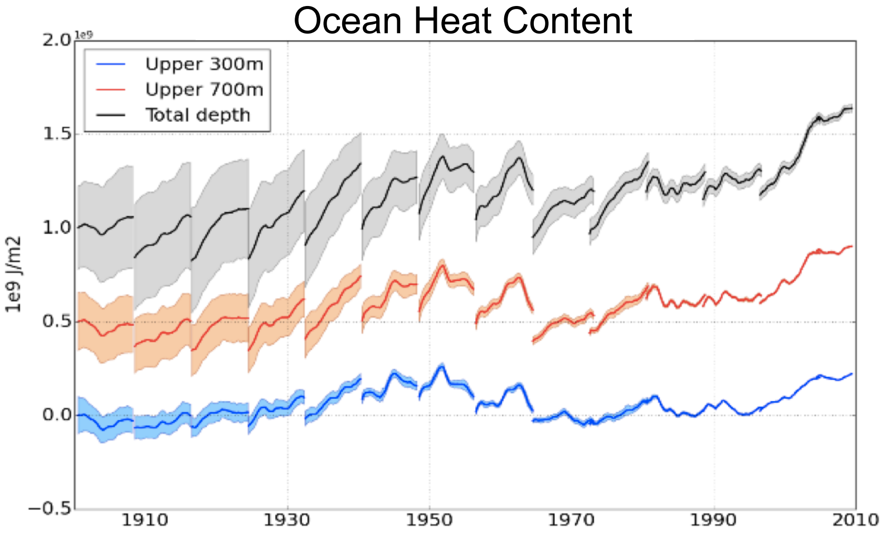
Anomaly Correlation for N.H. 500 hPa Height



...but plenty of challenges remain

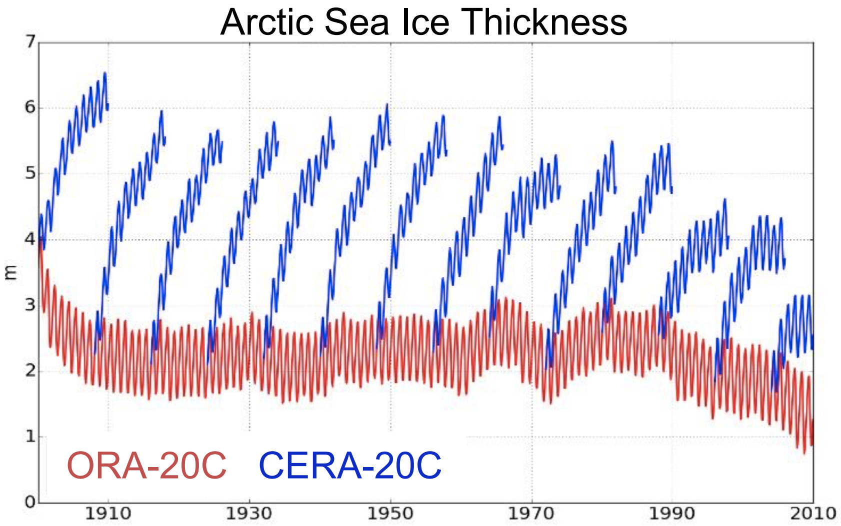
Lack of observational constraints on the sub-surface ocean

This leads to discontinuities in ocean heat content across the different reanalysis ”streams”, which worsen with increasing depth



Transfer of positive feedbacks, but also biases

In CERA-20C, sea-ice gets very thick in the Arctic, with an increase in the Antarctic as well - insufficient melting in summer (ORA-20C is an uncoupled ocean reanalysis)



Inclusion of reactive chemical species

Major role in radiative physics, with impact on temperature and dynamics, as well as air quality implications

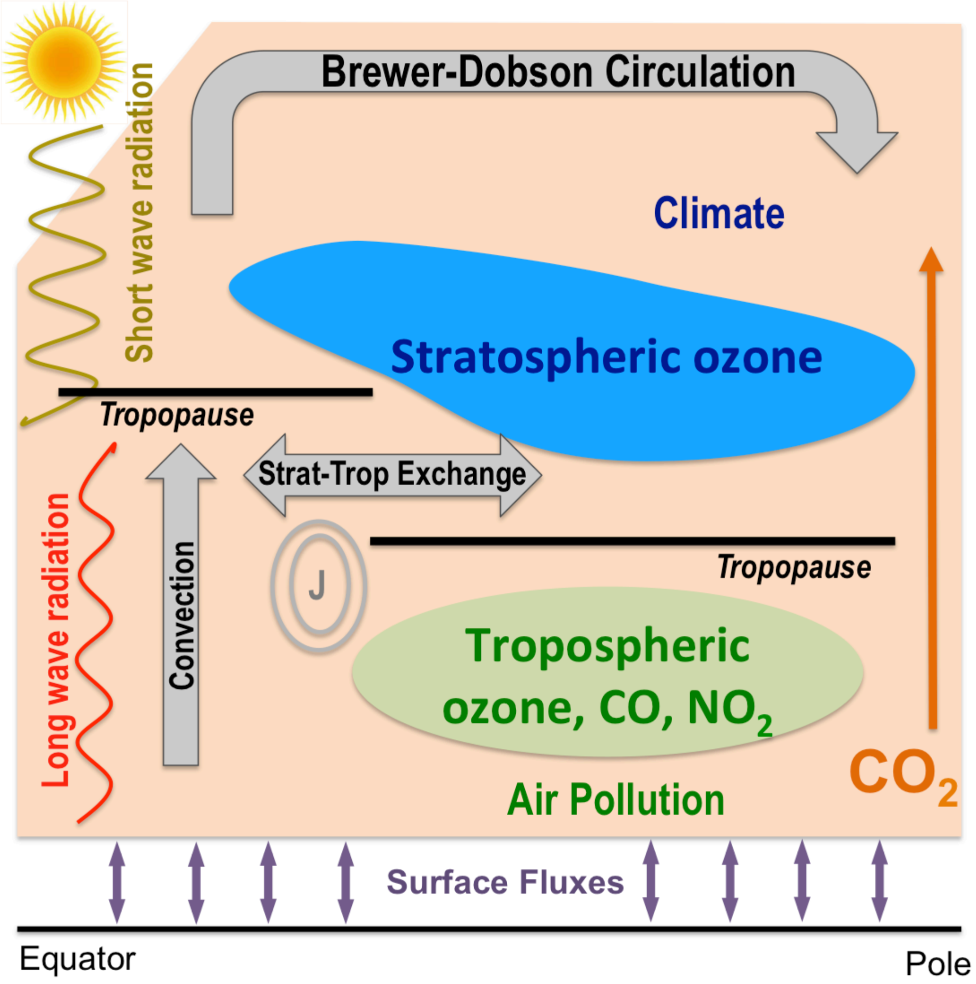
Climate-relevant species

CO_2, CH_4, N_2O, stratospheric O_3

Air quality-relevant species

CO, NOx, tropospheric O_3

Also **carbon cycle** applications, with the potential to inform meteorological analysis, but still large model and observational biases



The observing system for chemical data assimilation

Pre-EOS era

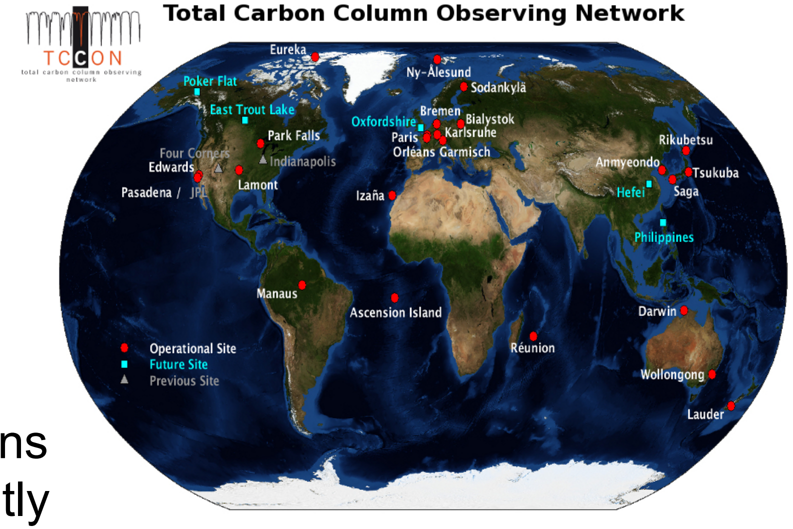
Few satellite observations, primarily stratospheric ozone

EOS era

The golden age of stratospheric chemistry begins, satellite observations of ozone, methane, CO, CO_2, ... greatly increase the prospects for chemical data assimilation

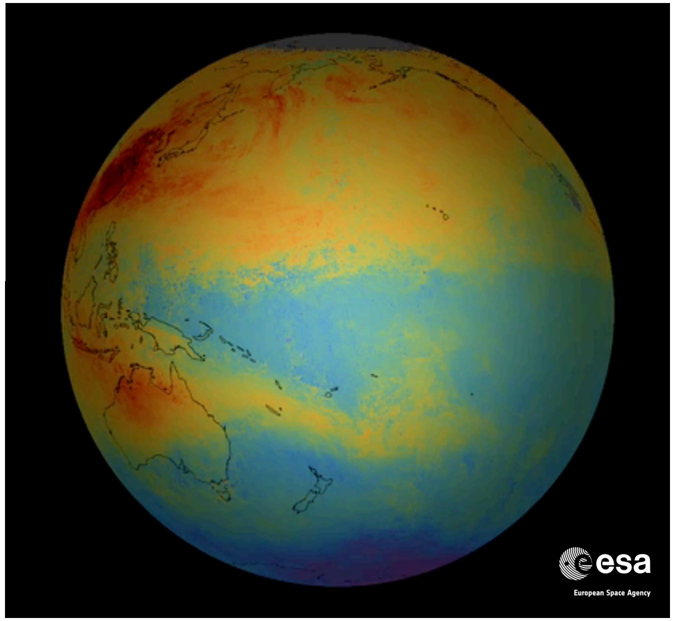
Current and near future

Observations for chemistry-climate remain plentiful, but increasing focus on observing systems for air-quality and carbon cycle applications



2009: Carbon observing network
Sparse, fixed observing network (TCCON and NOAA), long latency

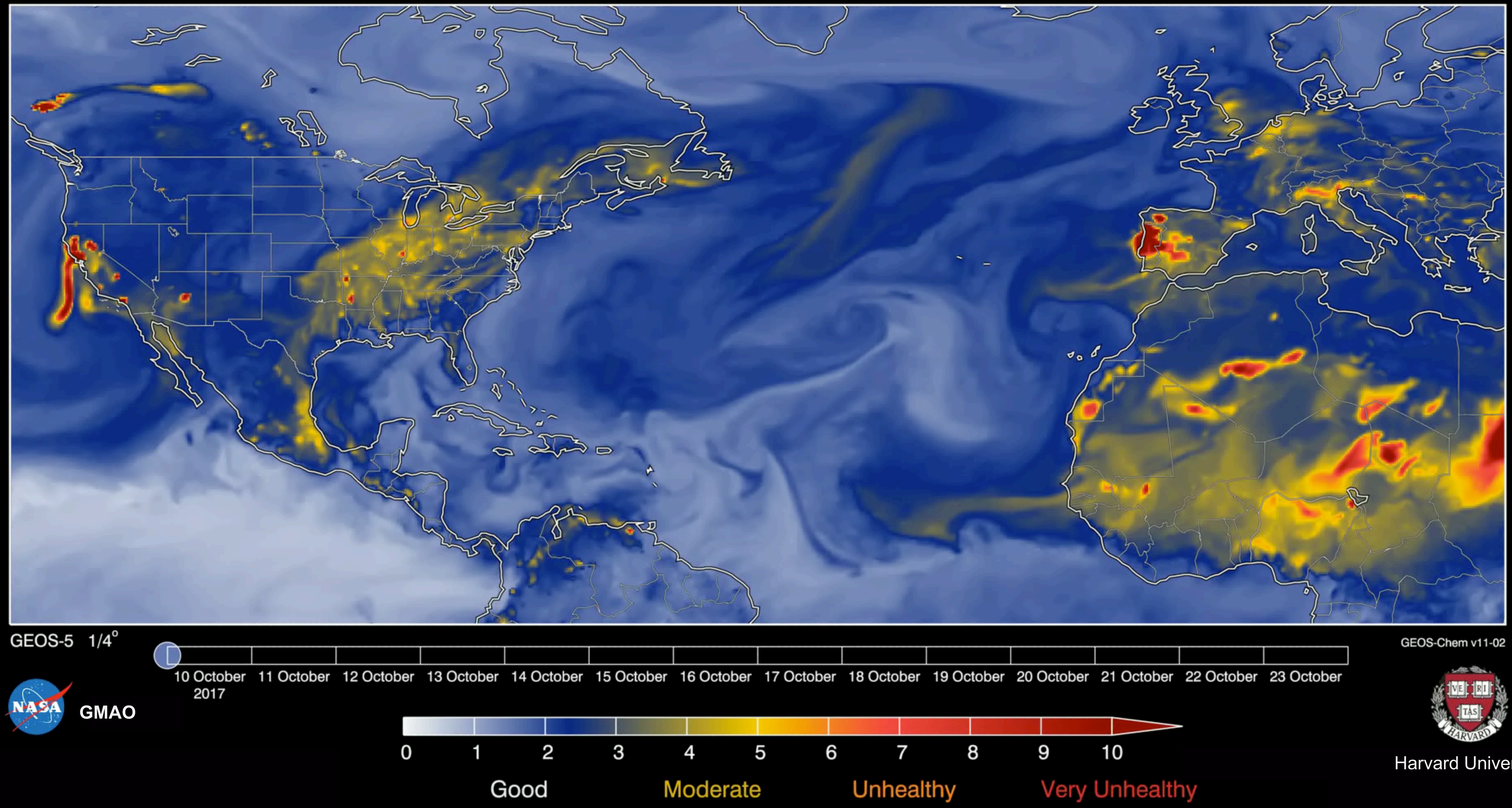
Observing carbon then
...and now



2017: Sentinel-5P TROPOMI
Launched 13 Oct 2017, global map of CO shows high levels over parts of Asia, Africa and South America

Analyzed Global Health Air Quality Index

Combines O_3, NO_2 and $PM_{2.5}$ based on GEOS and GEOS-Chem



Harvard University

Evaluation is the key to establishing confidence

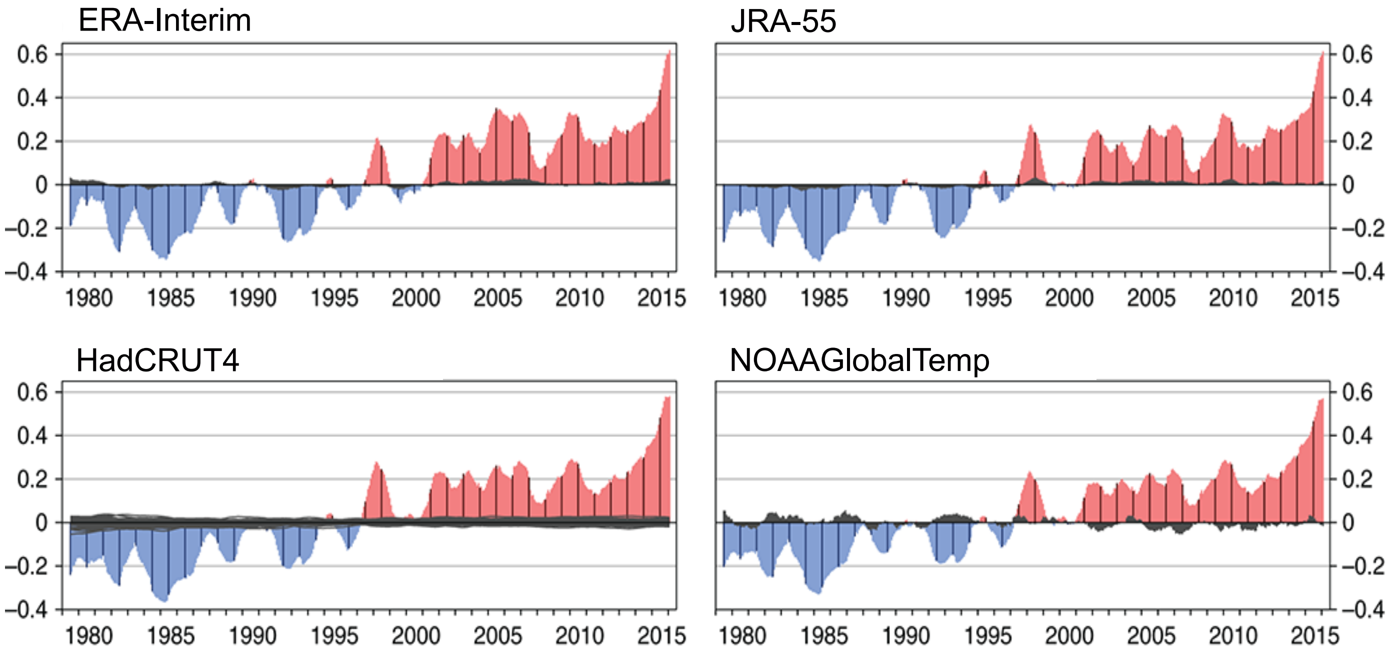
What data are assimilated? How well do the background and analysis fit these data? How does this change over time?

How large are the changes made each assimilation cycle? How do the changes vary in space and time?

How do reanalyses compare with one another and with any alternative observation-based datasets for a particular variable?

How well do reanalyses fit independent data? How well do they perform locally?

How consistent are their global budgets? How small are any adjustments made to ensure balance?



Twelve-month running averages of global mean surface temperature from 1979 onwards for two recent reanalyses and two conventional data sets show good agreement (Simmons et al. 2016)

Final remarks

Four generations of reanalyses with improving quality and diversity, now a staple of Earth science research, operational forecasting and, increasingly, business sectors such as energy, agriculture, water...

The recent extension of forecast systems that allow integrated modeling of meteorological, land, oceanic, and chemical variables provide the basic elements for fully coupled DA and offer the prospect of improved reanalyses (and forecasts) through better use of observations in all components, especially at their interfaces.

Increased system complexity will inevitably lead to additional assumptions and practical decisions to make implementation feasible but, ultimately, realistic results are possible only if the additional degrees of freedom can be adequately constrained by observations.

While not uncontroversial, reanalysis arguably offers the best potential for extracting maximum information about the recent climate from the total instrument record by using models to relate and combine information from otherwise disparate observations (Dee et al. 2014).

Quantifying uncertainty in reanalyses remains and important challenge for increasing their utility, especially as tool for climate change assessment. Evaluation (and more evaluation!) is the key to increasing confidence.