Reanalysis

Data assimilation training course 2019

Dinand Schepers,
Hans Hersbach, Joaquin Muñoz Sabater

Dinand.Schepers@ecmwf.int
“All steamers that came in yesterday were coated with ice from the tops of their masts down to the water line, and all had passed through mountainous waves.”
A broad view on reanalysis

Reanalysis process
Integration of an invariant, modern version of a data assimilation system and numerical weather prediction model, over a long time period, assimilating a selection of observations

Models

Observations

Understanding

- Polar-orbiting Satellite
- Argo Float
- Geostationary Satellite
- Bathythermograph
- Aircraft
- Buoy
- Balloon, Radiosonde
- Ship
- (Semi-) Automatic Station
- Observer, with instruments
Why not use simply observations?

“Observations-only” climatology

“Model only” integration

Reanalysis

Gross exaggeration towards discontinuity

"outliers"

Gross exaggeration towards continuity
Why not use simply operational NWP?
A consistent and complete picture of the past atmosphere

- **Complete**: combining vast amounts of observations into (global) fields

- **Consistent**: use the same physical model and DA system throughout

- **State-of-the-art**: use the best available observations and model at highest feasible resolution
A consistent and complete picture of the past atmosphere Earth system

...in Time
...in the Horizontal
...across Atmospheric Parameters
...in the Vertical
... across domains

ERA-Interim 2-metre temperature (°C)
15 August 2003 03 UTC

2-metre temperature anomaly (°C) over Africa

Southern Oscillation Index (hPa)

Standard deviation of differences between ERA-Interim and radiosondes temperature (°C) in the southern hemisphere
Two classes of Reanalysis

**Reanalyses of the modern observing period (~30-50 years):**

- Produce the best state estimate at any given time (as for NWP)
- Use as many observations as possible, including from satellites
- Closely tied to forecast system development and evaluation
- Can support product updates in near-real time

**Extended climate reanalyses (~100-200 years):**

- As far back as the instrumental record allows
- Pioneered by NOAA-CIRES 20th-Century Reanalysis Project
- Long perspective needed to assess current changes
- Main focus is on consistency, low-frequency variability
- Use only a restricted set of observations
Some important concepts

1) How reanalysis deals with “missing data”
   • Only assimilate observations when and where they exist
   • In between, the “best model available” (from NWP!) is used to “fill in the blanks”, from past and neighbouring information

2) Reanalyses produce fields are space- and physically-consistent
   • As specified by the underlying numerical model based on physical laws

3) Reanalyses use the widest variety of observations
   • Not just temperatures, or winds, or humidities in isolation of each other,
   • Also pressures, satellite observations, etc… = multi-variate approach
   • In fact, reanalyses are extremely data-rich products

4) Reanalysis uses and evaluates all observations in a consistent way
   • Accuracy (error bias) and precision (error std.dev.) explicitly taken into account
   • Quality control (QC) procedures apply across all observation types
   • The background prediction provides QC advantage w.r.t statistical reconstruction

5) Observation quality and quantity changes over time are not easily dealt with
   • LIKE ANY OTHER observations-based dataset.
   • Reanalyses can adjust the observation influence to take account of how much information is already known (background errors).

Observations-only datasets are the “observation limit” of reanalyses
They also point out deficiencies in reanalyses that further help improve understanding
Outline

**What is reanalysis?**
- General concepts
- Goals of reanalysis

**How are reanalyses made?**
- Observations
- Model
- Data assimilation

**Reanalysis products & applications**
- Users
- Applications

**Conclusions**
- Summary of concepts
- Challenges ahead
Constructing a history of the past with 4DVAR data assimilation

In a maximum-likelihood sense, which is equivalent to the minimum variance, provided that background and observation errors are Gaussian, unbiased, uncorrelated with each other; all error covariances are correctly specified; model errors are negligible within the analysis window.
Constructing a history of the past with 4DVAR data assimilation

Observational constraint
New or improved observation operators $h$ (along with tangent linear and adjoint) may be required.

Observation errors $R$ can be refined after assimilation to benefit next cycles.

Background constraint
May require work (new estimate or automatic updates) as function of available observation system.

Bias correction schemes
Bias model $b$ and coefficients may need expanding to tackle known biases in observations.

For each analysis, construct a cost function, and find its minimum:

$$J(x) = (x_b - x)^T B_b^{-1} (x_b - x) + [y - h(x)]^T R^{-1} [y - h(x)]$$
$$h(x) = h[M(x)]$$

$$J(z) = (z_b - z)^T B_z^{-1} (z_b - z) + [y - \tilde{h}(z)]^T R^{-1} [y - \tilde{h}(z)]$$
$$z^T = [x^T \beta^T]$$
$$\tilde{h}(z) = h(x) + b(x, \beta)$$
### Reanalysis components

#### Part 1: Observations

| Use as many observations as possible | - Goal being to produce the best estimate of the atmospheric state, at any given time and place  
- Question whether short datasets add long-lasting value |
|-------------------------------------|---------------------------------------------------------------------------------------------------|
| Use “good” observations | - Use corrected/reprocessed datasets when available  
- Focus efforts on long-term records  
- Consider the traceability of sources |
| Keep track of what goes in/comes out | - Monitoring the key steps:  
  - observation ingestion, blacklisting,  
  - thinning, assimilation |
| Keep that setup throughout | - A reanalysis production can take several years  
- Beware of large components of the observing system that suddenly disappear from the assimilation… bug? |
Evolution of the observing system

- **Manual stations, limited data exchange**
- **1890**: Surface observations
  - First radiosonde networks, systematic soundings
  - 1938
  - International Geophysical Year: radiosonde network enhanced, especially in the Southern Hemisphere
  - 1957
  - First operational satellite soundings (NOAA-2)
  - 1973
  - Improved sounding from polar orbiters; Winds from geostationary orbit; More data from commercial aircraft; First drifting buoys
  - 1979
  - More satellites, aircraft, buoys, ocean gliders and drifters. Fewer radiosondes, but probe higher. Better knowledge of instruments. More obs. per hour.
  - Today
  - Opportunistic sensors: cell phones, UAVs, vehicles, rooftops, ...
  - 2025
- **Satellites**
- **Upper-air soundings**

D. Dee
Data rescue

The FP7 ERA-CLIM project (2011-2013)

Preparing input observations, model data, and data assimilation systems for a global atmospheric reanalysis of the 20th century

The ERA5 observing system

Over 200 report types

Reprocessed data sets
Radiances: SSM/I brightness temp from CM-SAF
MSG from EUMETSAT
Atmospheric motion vector winds: METEOSAT, GMS/GOES-9/MTSAT, GOES-8 to 15, AVHRR METOP and NOAA
Scatterometers: ASCAT-A (EUMETSAT), ERS 1/2 soil moisture (ESA)
Radio Occultation: COSMIC, CHAMP, GRACE, SAC-C, TERRASAR-x (UCAR)
Ozone: NIMBUS-7, EP TOMS, ERS-2 GOME, ENVISAT SCIAMACHY, Aura MLS, OMI, MIPAS, SBUV
Wave Height: ERS-1, ERS-2, Envisat, Jason

Latest instruments
IASI, ASCAT, ATMS, CrIS, MWHS, Himawari, ...

Improved data usage
all-sky vs clear-sky assimilation,
latest radiative transfer function, corrections,
extended variational bias control

How to monitor all this?

Courtesy: Paul Poli
ERA5 data usage has increased from 0.75 million/day (1979) to 21 million/day (2018)

ERA-Interim is progressively getting outdated. It is not able to:
- use the latest instruments
- respond to changes in data format
  (like the ongoing transition to BUFR format for conventional data)
Reanalysis components
Part 2: forecast model

Use a fixed version
- Dynamics, physics etc…
- Resolution must be computationally affordable but satisfy user requirements

Use the “best” model around
- Use the near-latest, stable, model version operational at some point
- Not the time to start experimenting with new, untested configurations

Shop around for forcing data
- Ideally, one dataset per forcing, to cover the whole time period
- Consider standards such as CMIP5

Keep that setup throughout the production
- Be extra careful with forcing data – any problem will map into products!
- Be extra careful when changing machine, compiler….
Why not use simply operational NWP?

Models are essential tools to propagate the information and ensure consistency (over short) time scales between geophysical variables. The advances in models and in data assimilation help deliver improved products.
Forcings appropriate for climate

**CMIP5 recommended data sets**
Total solar irradiance, greenhouse gases, ozone, aerosols (including volcanic)

*(Prepared in the ERA-CLIM project, ERA-20CM, Hersbach et al., 2015)*

**SST and sea ice cover**
Carefully selected from OSTIA, OSI-SAF and HadISST2 (Hadley Centre, ERA-CLIM)

*(Hirahara et al., 2016)*
The FP7 ERA-CLIM2 project (2014-2017)

Production of a consistent 20th-century reanalysis of the coupled Earth-system: atmosphere, land surface, ocean, sea-ice, and the carbon cycle

CERA-20C a 20th century reanalysis using an ocean-atmosphere coupled model and DA.

CERA-SAT a modern day pilot reanalysis using an ocean-atmosphere coupled model and DA.
Coupled processes: Tropical instability waves

Tropical instability waves (TIW)
westward-propagating waves near the equator

CERA-20C (Coupled reanalysis)
- represents TIWs thanks to the ocean dynamics
- atmosphere responds accordingly (surface wind stress is sensitive to the ocean TIW)

ERA20C (Forced reanalysis)
- no TIWs or wind stress signals (forced by monthly SST)
Impact of scatterometer winds on ocean salinity

Improved fit to in-situ observations

NO SCATT - SCATT / Sea Surface Salinity / Jan-Jun 2014

Mean departure

rmse

SCATT
No SCATT

Courtesy of Giovanna De Chiara
Use a fixed data assimilation system (DAS)
- A blacklist to cover the entire reanalysis period
- Observation handling for all: operators, thinning, etc…
- Test the DAS with various amounts of observations

Errors in the background
- They change over time!
- Need to account for this in one way or another

Errors in the observations
- Homework to find out Gross errors, Biases, and Random errors (std. dev. = specified as ‘observation errors’)

Keep that setup and monitor it
- Be extra careful during run-time etc…
- Implement automated monitoring for all the key steps of the assimilation
Ensemble of 4DVAR data assimilations: Discretization of the PDF of uncertainties

Aims:
1. Estimate automatically our background errors, and update them
2. Provide users with some uncertainties estimates
Over the course of time, increased observations result in…

- **Smaller** background error variances, with **sharper** horizontal structures
- Analysis increments that are smaller, over smaller areas
Ensemble spread as a proxy for the background error

1971 CERA-20C: Surface pressure, marine wind, only

1971 ERA5: Upper-air data

1980 ERA5: Early-satellite era

2018 ERA5: Current observing system
Long-term evolution of the background covariance matrix

For early decades correlation lengths for the static part of $\mathbf{B}$ appear too short. **Solution:** 1) build dedicated $\mathbf{B}_{\text{cli}}$ for distinct periods in time

However, this risks introducing discontinuities in the reanalysis

**ERA5 scout, Modern $\mathbf{B}_{\text{cli}}$ vs ERA-Interim for 1979**

**ERA5 scout, 1979 $\mathbf{B}_{\text{cli}}$ vs ERA-Interim for 1979**
The transition from the 1979-B the ERA5 modern B in 2000 introduces a seam in the stratosphere. Extended usage of the 1979-B till the availability of COSMIC GNSS-RO could have alleviated this.
Another look at uncertainties: Analysis increments and trends

Mean temperature analysis increments in ERA-Interim (Jan 1979-Dec 2010)

Outline

What is reanalysis?
- General concepts
- Goals of reanalysis

How are reanalyses made?
- Observations
- Model
- Data assimilation

Reanalysis products & applications
- Users
- Applications

Conclusions
- Summary of concepts
- Challenges ahead
The ECMWF reanalysis landscape

<table>
<thead>
<tr>
<th>Atmosphere/land</th>
<th>including ocean waves</th>
<th>including sea ice</th>
<th>Ocean</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>5) 2016 - ... ERA5</td>
<td>2006 ORAS3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2010 - ... ORAS4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2016 - ... ORAS5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2013 - 2015 ERA-20CM/20C</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2016 CERA-20C</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2017 CERA-SAT</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2012 ERA-Int/Land</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2014 ERA-20C/Land</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2018 - ... ERA5L</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2008 - 2009 GEMS</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2010 - 2011 MACC</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2017 - ... CAMS</td>
</tr>
</tbody>
</table>

Centennial Coupled

Enhanced land

Atmospheric composition
ERA5 - Improvement of forecast skill and status

Up to one day gain with respect to ERA-Interim

**Phase 1:** production from 1979: to finished October 2018

**Phase 2:** production of 1950-1978 has recently started

Publicly available in the C3S climate data store:

Currently: **1979 to present,** 2-3 months behind real time

Soon: **ERA5T,** 2-5 days behind real time

**Note:** ERA-Interim production will not be supported after **mid 2019**
Usage of Reanalysis and its importance at ECMWF

Reanalysis provides consistent “maps without gaps”. Its role in climate monitoring is now widely recognized.

ERA-Interim has more than 30,000 users worldwide

Importance at ECMWF:
- Input to timely monthly C3S climate bulletins
- Required for re-forecasts
- Evaluation of progress in forecast skill
- Provision of climatologies: EFI, ACC, probabilistic events
- Benchmark for developments in R&D
- Many more
Linear trend in surface air temperature over land

*Kelvin/Decade for 1979-2017 (~0.18 globally)*

There is a good general consensus between various products (including ERA5), although there are differences in the details.
Climate monitoring and reporting

The long-term warming trend has continued in 2018, with the average global temperature set to be the fourth highest on record. The 20 warmest years on record have been in the past 22 years, with the top four in the past four years, according to the World Meteorological Organization (WMO).

Other tell-tale signs of climate change, including sea level rise, ocean heat and acidification and sea-ice and glacier melt continue, whilst extreme weather and a trail of devastation on all continents, according to the WMO provisional Statement on the State of the Climate in 2018. It includes details of impacts of climate change based on contributions from a wide range of United Nations partners.

The report shows that the global average temperature for the first ten months of the year was nearly 1°C above the pre-industrial baseline (1850-1900). This is based on five independently maintained global temperature data sets.

“We are not on track to meet climate change targets and rein in temperature increases,” said WMO Secretary-General Petteri Taalas. “Greenhouse gas concentrations are near-record at record levels, and if the current trend continues, the global average temperature increase will fall outside the 1.5°C and 2°C limits that the international community has set as stabilisation targets.”

ERA5-Land, a high-resolution downscaling of the land-surface component

Discharge time series correlation difference ERA5-Land vs. ERA5

Blue = improvement
Red = Deterioration

ERA-Interim
(~ 80 km)

ERA5
(~ 31 km)

ERAS5-Land
(~ 9 km)
ERA5-Land in a simple diagram

Climatology (static+dynamic fields) & Initial States

Next cycle States

Land Surface integration

Lapse-rate & forcing adjustment

ERA5L hourly meteorology

Temperature
Humidity
ERA5
Precipitation
Radiation

Applications

### Table 1: ERA5L monthly mean archiving

<table>
<thead>
<tr>
<th>$\bar{M}$</th>
<th>$M(x)$</th>
<th>$\sigma(x)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\bar{M}$</td>
<td>$M(x)$</td>
<td>$\sigma(x)$</td>
</tr>
<tr>
<td>$\bar{M}$</td>
<td>$M(x)$</td>
<td>$\sigma(x)$</td>
</tr>
<tr>
<td>$\bar{M}$</td>
<td>$M(x)$</td>
<td>$\sigma(x)$</td>
</tr>
<tr>
<td>$\bar{M}$</td>
<td>$M(x)$</td>
<td>$\sigma(x)$</td>
</tr>
</tbody>
</table>

- $\bar{M}$: monthly climatology
- $M(x)$: hourly meteorology
- $\sigma(x)$: standard deviation
## Outline

<table>
<thead>
<tr>
<th>Section</th>
<th>Topics</th>
</tr>
</thead>
<tbody>
<tr>
<td>What is reanalysis?</td>
<td>General concepts, Goals of reanalysis</td>
</tr>
<tr>
<td>How are reanalyses made?</td>
<td>Observations, Model, Data assimilation</td>
</tr>
<tr>
<td>Reanalysis products &amp; applications</td>
<td>Users, Applications</td>
</tr>
<tr>
<td>Conclusions</td>
<td>Summary of concepts, Challenges ahead</td>
</tr>
</tbody>
</table>
Summary of important concepts

Reanalysis neither produces “gridded observations” nor “model data”
But it enables to extract information from observations in one, unique, theoretically consistent framework, using the model to propagate the information in space and time, and across variables.

Reanalysis sits at the end of the (long) meteorological research and development chain
observation and measurement collection, processing, exchange modelling and data assimilation for numerical weather prediction.

Unlike NWP, a very important concern in reanalysis is the consistency in time, spanning several years, decades or even centuries.

Reanalysis is bridging slowly, but surely, the gap between the “weather datasets” and the “climate datasets”
Resolution gets finer, reanalyses cover longer time periods, extend to today
Reanalysis data assist the work of many users in many applications who do not necessarily place themselves in “weather or climate”
Current status of global reanalysis & Future outlook

**Reanalysis is worth repeating as all ingredients continue to evolve:**
Models, data assimilation, observation (re-)processing and data rescue
With each new reanalysis we improve our understanding of systematic errors in the various components of the observing system

**Uncertainties in products are hard to characterize**
We now have some initial framework to propagate error estimates with ensembles; including boundary conditions and observation errors
Yet the resulting uncertainties have signatures in all dimensions: low-frequency, spatial domain, vertical …

**More challenges for comprehensive reanalyses:**
Bringing in additional or reprocessed observations
Dealing with changing background quality over time
Dealing with model bias, tied to problems with trends interpretation
Inclusion of coupled processes in model *and data assimilation*
Bridging the gap with climate models
Selected further reading and data access


Desroziers et al. (2005), Diagnosis of observation, background and analysis-error statistics in observation space. Q.J.R. Meteorol. Soc. 131, 3385–3396. doi: 10.1256/qj.05.108

Global and regional reanalyses: http://www.reanalyses.org

Copernicus Climate Change Service (C3S) Climate Data Store: https://cds.climate.copernicus.eu/#!/home