

ECMWF ReAnalysis (ERA) Data assimilation aspects



ECMWF Data Assimilation Training Course Reanalysis

1



James Stagg, Chief Meteorologist



Hand-drawn analysis for 13 UTC 6 June 1944

8

7

Significant wave height [m]

1

0

ERA-20C 108-hr forecast for D-DAY



ERA-PreSAT 108-hr forecast for D-DAY



ERA-PreSAT Analysis for D-DAY



March 2014

Surface observations ECMWF Data Assimilation Training Course
 O Upper air (mostly pilot balloons) Reanalysis

Reanalysis course outline



The three pillars of geosciences



March 2014



Objective: Reconstruct the past



Reanalysis *products* and Reanalysis *process*

- Reanalysis *products*
 - Gridded fields of NWP model
 - Control variables: vorticity, divergence, temperature, humidity, ozone...
 - Derived variables: precipitation, radiation...
 - Fit to observations
 - Before, and after, assimilation
 - Before, and after, bias correction
- 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

Differences with observations-only gridded datasets

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 neighboring 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 the most data-rich products to date (30 billion obs. in ERA –Interim)

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). Example later with ERA-20C ensemble.

Observations-only datasets are the "observation limit" of reanalyses. They are extremely important for improving understanding.

Why re-analyze? Overall aim is a greater time-consistency of the products





Reconstructing the past more smoothly

RMS of differences between observations from radiosondes and short-term forecast (background)

Thin line for Northern Hemisphere extratropics **Thick line** for Southern Hemisphere, typically less well observed



ECMWF Data Assimilation Training Course Reanalysis



ECMWF Data Assimilation Training Course Reanalysis

Reanalysis course outline



The ubiquitous data assimilation slide:

Constructing a history of the past with (24-hour) 4DVAR data assimilation



March 2014

Reanalysis

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 your sources
Keep track of what goes in/comes out	 Monitoring the key steps: observation ingest, 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?

Increased satellite observation diversity

1962 1964 1966 1968 1970 1972 1974 1976 1978 1980 1982 1984 1986 1988 1990 1992 1994 1996 1998 2000 2002 2004 2006 2008 2010 2012



In blue: data that were assimilated in ERA-Interim

In grey: data that were not assimilated. ...For future reanalyses...

Note the timeline starts in 1969

1962 1964 1966 1976 1972 1974 1976 1978 1980 1982 1984 1986 1988 1990 1992 1994 1996 1988 2000 2002 2004 2006 2008 2010 2012

Observation timeline (atmosphere)

ECMWF Data Assimilation Training Course



Efforts to improve the historical upper-air data record: "data rescue" a.k.a "data recovery"



Stickler et al., 2014 : "ERA-CLIM: Historical Surface and Upper-Air Data for Future Reanalyses." *Bulletin of the American Meteorological Society*

ECMWF Data Assimilation Training Course

Reanalysis

Evolution of the observation coverage

1°1° average in 3 categories; > 1 per



Surface pressure network

60°W

EX A

60°E



1°1° average in 3 categories: > 1 ne

120°W

Reanalysis

Reanalysis components Part 2: forecast model

Use a fixed version	 Dynamics, physics etc Resolution must be computationally affordable Producing N decades in 1 year implies a factor N in run-time
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

Illustration of resolution improvements

25 km resolution





Model forcings for reanalysis

- ERA-20CM integrates the ECMWF model, but without data assimilation
- So far, the previous ECMWF reanalyses did not attempt to use so many "historical forcing" datasets.
- ERA-20CM uses the following forcings:
 - Sea-surface temperature and sea-ice cover (Hadley Centre)
 - Solar irradiance (CMIP5)
 - Greenhouse gases (CMIP5)
 - Ozone for radiation (CMIP5)
 - Tropospheric aerosols (CMIP5)
 - Volcanic aerosols (CMIP5)

Comparison between "model-only" ERA-20CM and ERA-Interim reanalysis



Reanalysis components Part 3: Data assimilation & errors

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

Reanalysis course outline



Ensemble of 4DVAR data assimilations: Discretization of the PDF of uncertainties



Reanalysis

1-year ensemble spread, throughout the century



From the ensemble spread, one can estimate background error variances



ECMWF Data Assimilation Training Course Reanalysis





Self-updating background error covariances, throughout the century _____



Over the course of the century, more observations result in...
 → Smaller background errors, with sharper horizontal structures
 → Analysis increments that are smaller, over smaller areas
 = ERA-20C ensemble system adapts itself to the information available





Impact of background error assumptions



Assimilation error assumptions: budget closure ... or ... "data assimilation reality check"



Showing only observations in the first 90 minutes of the 24-h window

ECMWF Data Assimilation Training Course Reanalysis



Estimates of observation errors





ERA-20C assumed time invariant observation errors. This does not seem to be the case...

ECMWF Data Assimilation Training Course Reanalysis



How useful are these revised (larger) observation error std. dev. estimates? Surface pressure at Montreal, Quebec Observations from ISPD 3.2.6, collection #3004 (Canadian Stations Environment Canada) ERA-20C ensemble [hPa] **Observations** 102000 with uncertainties 101500 (some could not be fitted they are VARQC rejected 101000-100500 Analysis, with uncertainties 100000 Background forecast, with uncertainties in the model and its forcings (HadISST2.1.0.0 ensemble) 99500 3 5 2 4 Jun [hPa] 1900 Increased observation error std. dev. assumptions 102300-Same observations but with LARGER uncertainties 101700-No obs. rejected 101100 100500 99900-5 2 Analysis, presents LARGER uncertainties Jun 1900 **Background forecast**, Initial state (and subsequent ones) presents LARGER uncertainties ECMWF Data Assimilation Training Course March 2014

Reanalysis

()bservation errors 2/3

Impact of a single bad time-series

30 March 1954, 00 UTC 31 March 1954, 00 UTC 31 March 1954, 03 UTC



30 March 1954, 06 UTC

31 March 1954, 09 UTC

31 March 1954, 12 UTC



S

0

ECMWF Data Assimilation Training Course Reanalysis





What about error growth within the 24-hour window?



ECMWF Data Assimilation Training Course Reanalysis

Reanalysis course outline



A (short) history of atmospheric reanalysis

- **1979**: **Observation datasets collected for the First GARP Global Atmospheric Research Program Experiment** (FGGE): used *a posteriori* for several years, to initialize models, track progress in NWP.
- **1983**: **Reanalysis concept proposed** by Daley for monitoring the impact of forecasting system changes on the accuracy of forecasts
- **1988**: **Concept proposed again, but for climate-change studies**, in two separate papers: by Bengtsson and Shukla, and by Trenberth and Olson
- **1990s**: First-generation comprehensive global reanalysis products (~OI-based)
 - NASA/DAO (1980 1993) from USA
 - NCEP/NCAR (1948 present) from USA
 - ERA-15 (1979 1993) from ECMWF with significant funding from USA
- Mid 2000s: Second-generation products (~3DVAR)
 - JRA-25 (1979 2004) from Japan
 - NCEP/DOE (1979 present) from USA
 - ERA-40 (1958 2001) from ECMWF with significant funding from EU FP5
- Today: Third generation of comprehensive global reanalyses (~better than 3DVAR)
 - NASA/GMAO-MERRA (1979 present) from USA (IAU)
 - NCEP-CFSRR (1979 2008) from USA (land/ocean/ice coupling)
 - JRA-55 (1958 2012) from Japan (4DVAR)
 - 20-CR from USA (Ensemble Kalman Filter, surface pressure observations only)
 - ERA-Interim (1979 present) from ECMWF (4DVAR)
 - ERA-20C (1900-2010) from ECMWF (4DVAR ensemble)



ECMWF Data Assimilation Training Course Reanalysis

How (outside) users exploit reanalysis data

- Monitor the observing system
 - Feedback on observational quality, bias corrections
 - Basis for homogenization studies of long data records
- Develop climate models
 - Use reanalysis products for verification, diagnosis, calibrating output,, ...
- Drive users' models/applications
 - Use reanalysis as large-scale initial or boundary conditions for smaller-scale models (global→regional; regional→local), in various fields: wind energy, ocean circulation, chemical transport and dispersion, crop yield, health indicators, ...
- Use climatologies derived from reanalysis for direct applications
 - Ocean waves, wind and solar power generation, insurance, ...
- Study short-term atmospheric processes and influences
 - Process of drying of air entering stratosphere, bird migration, ...
- Study of longer-term climate variability/trends
 - Requires caution due to changes in observations input
 - Lead to major findings in recent years in understanding variability

How ECMWF users exploit reanalysis data

- Baseline to track NWP score improvements
- Calibration for seasonal forecasting system
- Reference to diagnose changes brought by model improvements

ERA-Interim: more than 15,000 users

Growing recognition for climate application

BAMS State of the Climate in 2008



Plate 2.1. Global annual anomaly maps for those variables for which it was possible to create a meaningful anomaly estimate. Climatologies differ among variables, but spatial patterns should largely dominate over choices of climatology period. Dataset sources and climatologies are given in the form (dataset name/data source, start year-end year) for each variable. See relevant section text and figures for more details. Lower stratospheric temperature (RSS MSU 1981–90); lower tropospheric temperature (UAH MSU 1981–90); surface temperature (NCDC 1961–90); cloud cover (PATMOS-x 1982–2008); total column water vapor (SSM/I/GPS 1997–2008); precipitation (RSS/GHCN 1989–2008); mean sea level pressure (HadSLP2r 1961–90); wind speed (SSM/1988–2007); total column ozone (annual mean global total ozone anomaly for 2008 from SCIAMACHY. The annual mean anomalies were calculated from 1° x 1.25° gridded monthly data after removing the seasonal mean calculated form GOME (1996–2003) and SCIAMACHY (2003–07)]; vegetation condition [annual FAPAR anomalies relative to Jan 1998 to Dec 2008 from monthly FAPAR products at 0.5° x 0.5° (derived from SeaWIFS (NASA) and MERIS (ESA) data].

BAMS State of the Climate in 2009

BAMS State of the Climate in 2010

BAMS State of the Climate in 2011

Plate 2.1. Global annual anomaly maps for those variables for which it is possible to create a meaningful 2009 anomaly estimate. Climatologies differ among variables, but spatial patterns should largely dominate over choices of climatology period. Dataset courses (names are as follows: lower stratospheric temperature (RSS MSU); lower tropospheric temperature (ERA-interim) surface temperature (NOAA NCDC); cloudiness (PATMOS-x); total column water vapor (som/rover ocean, ground based GPS over land); precipitation (RSS over ocean, GHCN (gridded) over land); river discharge (authors); mean sea level pressure (HadSLP2r); wind speed (AMSR-E); ozone (GOME2); FAPAR (SeaWIFS); Biomass Burning (GEMS/MACC). See relevant section text and figures for more details.

Plate 2.1. Global annual anomaly maps for those variables for which it is possible to create a meaningful 2010 anomaly estimate. Reference base periods differ among variables, but spatial patterns should largely (encoded to the strategy of the strategy o

(SBUVs/OMI/TOMS/GOME1/SCIAMACHY/GOME2, base period data from the multi-sensor reanalysis, MSR); FAPAR [SeaWiFS (NASA) and MERIS (ESA) sensors]; biomass burning (GFAS). See relevant section text and figures for more details.

PLATE 2.1. (1) ERA-Interim 011 anomalies of MSU Channel 4 equivalent for the lower stratospheric temperature; b) ERA-Interim 2011 anomalies of MSU Channel 2LT equivalent for the lower tropospheric temperature; (c) NOAA-NCDC 2011 anomalies of surface temperature; (d) ARCLAKE 2011 summer season anomalies of lake surface temperature; (e) PATMOS-x 2011 anomalies of cloudiness; (f) SSMIS (Ocean) and radiosonde and ground-based GPS (circles) (Land) 2011 anomalies map of TCWV anomalies of total column water vapour; g) ERA-Interim 2011 anomalies of surface specific humidity; h) ERA-Interim 2011 anomalies of surface relative numinary; (1) RSS and GHCN precipitation; (j) Water Balance Model (WBM) analysis by authors showing 2011 anomalies of river discharge over continents and into oceans; (k) GRACE satellite observations of 2011 minus 2010 annual mean terrestrial water storage (the sum of groundwater, soil water, surface water, snow, and ice, as an equivalent height of water in cm); (I) WACMOS satellite observations of 2011 anomalies of soil moisture; (m) HadSLP2r 2011 anomalies of sea level pressure; (n) Satellite radiometer (ocean) and in situ (land; 1152 sites from ISD-Lite and Tim McVicar) 2011 anomalies of surface wind speed; () MACC reanalysis for 2011 anomalies of total aerosol optical depth; (p) GOME/SCIAMACHY/GOME2 2011 anomalies or stratospheric ozone; (g) MODIS White Sky broadband 2011 anomalies of land surface albedo from the visible spectrum; (r) MODIS White Sky broadband 2011 anomalies of land surface albedo from the near-infrared spectrum; (s) Combined SeaWiFS (NASA) and MERIS (ESA) 2011 anomalies of fraction of absorbed photosynthetically active radiation (FAPAR); (t) MACC GFAS processed MODIS observations for 2011

ECMWF Data Assimilation Training Contrained of biomass burning in terms of annual carbon emission per unit area.

Reanalysis

STATE OF THE CLIMATE

IN 2011

STATE OF THE CLIMATE

IN 2010

38

Reanalysis course outline



Summary of important concepts

- Reanalysis does <u>not</u> produce "gridded observations"
 - But it enables to extract information from observations in one, unique, theoretically consistent framework
- Reanalysis sits at the end of the (long) meteorological research and development chain that encompasses
 - observation and measurement collection,
 - observation processing and data exchange,
 - numerical weather prediction modelling and data assimilation
- Unlike NWP, a very important concern in reanalysis is the consistency in time, over several years
- Reanalysis is bridging slowly, but surely, the gap between the "weather datasets" and the "climate datasets"
 - Resolution gets finer
 - Reanalyses cover longer time periods, without gap
 - Helps different communities work together
 - Reanalysis has developed into a powerful tool for many users and applications

Current status of global reanalysis & Future outlook

- Reanalysis is worth repeating as all ingredients continue to evolve:
 - Models are improving

FRA-20C

March 2014

- Data assimilation methods are improving
- Observation (re-)processing is improving
- Old observations (paper & microfilm records) are being rescued
- The technical infrastructure for running & monitoring improves constantly
- With each new reanalysis we improve our understanding of systematic errors in the various components of the observing system

• Major challenges for a future comprehensive reanalysis project:

- Bringing in additional observations (not dealt with in ERA-Interim)
- Dealing with changing background quality over time
- Dealing with model bias, tied to problems with trends interpretation
- Coupling with ocean and land surface
- Making observations used in reanalysis more accessible to users

Providing first uncertainty estimates for the reanalysis products
 ERA-CLIM2

Further reading and on-line material

- Kalnay et al. (1996), "The NCEP/NCAR 40-Year Reanalysis Project", Bull. Am. Meteorol. Soc. 77 (3), 437-471
- Uppala et al. (2005), "The ERA-40 reanalysis", Q. J. R. Meteorol. Soc. 131 (612), 2961-3012, doi:10.1256/qj.04.176
- Bengtsson et al. (2007), "The need for a dynamical climate reanalysis", *Bull. Am. Meteor. Soc.* 88 (4), 495-501
- SciDAC Review (2008), "Bridging the gap between weather and climate", on the web at <u>http://www.scidacreview.org/0801/pdf/climate.pdf</u> with contributions from Compo and Whitaker
- Global and regional reanalyses twiki: <u>http://www.reanalyses.org</u>
- Dee et al. (2011), "The ERA-Interim reanalysis: configuration and performance of the data assimilation system", *Q. J. R. Meteorol. Soc.*, **137** (656), 553-597
- Hersbach et al. (2013), "ERA-20CM: a twentieth century atmospheric model ensemble", ERA Report Series 16, <u>http://www.ecmwf.int/publications/library/do/references/show?id=90989</u>
- Poli et al. (2013), "The data assimilation system and initial performance evaluation of the ECMWF pilot reanalysis of the 20th-century assimilating surface observations only (ERA-20C)", ERA Report Series 14, <u>http://www.ecmwf.int/publications/library/do/references/show?id=90833</u>
- Simmons et al. (2014), "Estimating low-frequency variability and trends in atmospheric temperature using ERA-Interim". *Q.J.R. Meteorol. Soc.* doi: 10.1002/qj.2317

ECMWF data server http://apps.ecmwf.int/datasets/

CECMWF	Home My room Contact Feedback Sitemap Search ECMWF *Log in
About us Products Services	Research Publications News & events
Navigation Datasets Batch access See also Data FAQ Data Servers Data Services GRIB decoder	 Downloadable datasets DEMETER Project ENSEMBLES project GEMS Reanalysis and Near Real-time MACC Reanalysis TIGGE TIGGE LAM YOTC
	 ERA-Interim (Jan 1979 - present) ERA-Interim/LAND (Jan 1979 - Dec 2010) ERA-40 (Sep 1957 - Aug 2002) ERA-15 (Jan 1979 - Dec 1993) ERA-20CM (Jan 1900 - Dec 2010): Climate Model Integration (experimental) Observation Feedback
	 ISPD v2.2 ICOADS v2.5.1 with interpolated 20CR feedback
▲ Top of page	copyright © ECMWI