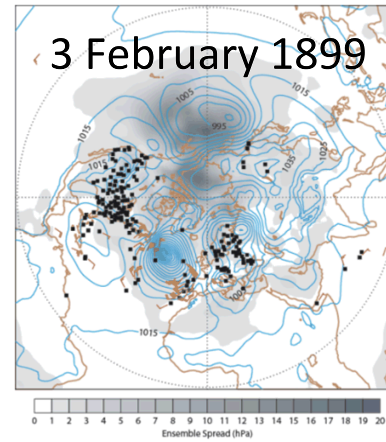
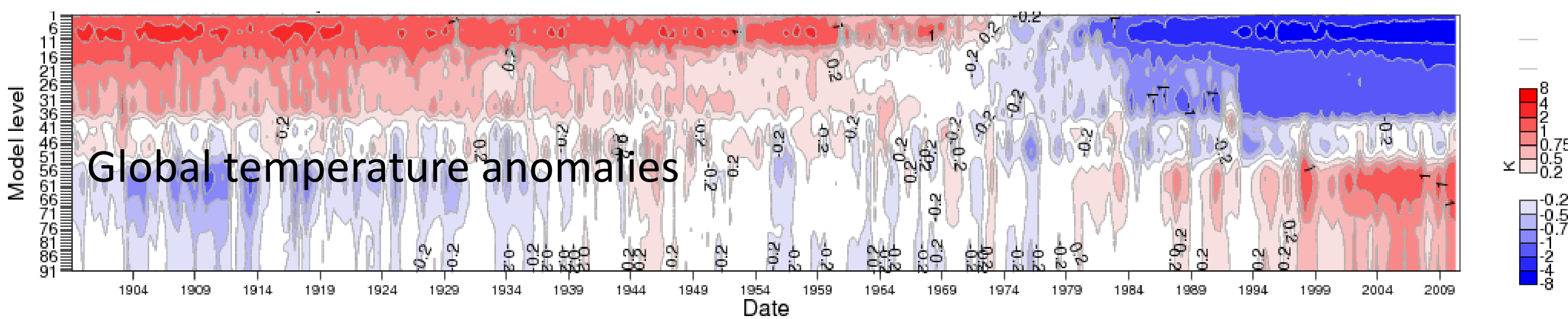
ECMWF ReAnalysis (ERA)
Data assimilation aspects

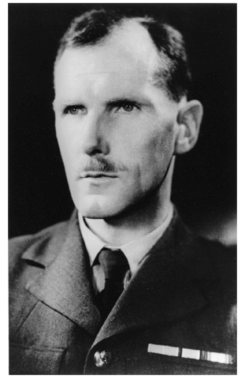


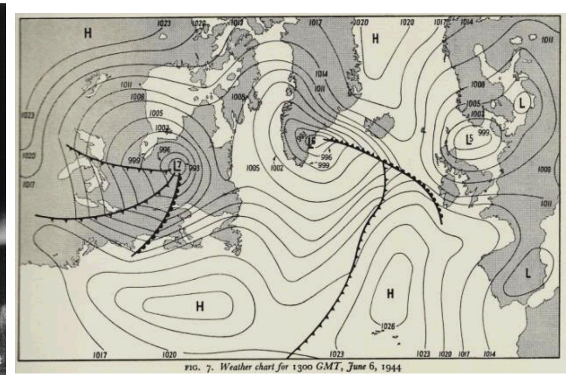


Paul Poli (paul.poli@ecmwf.int)

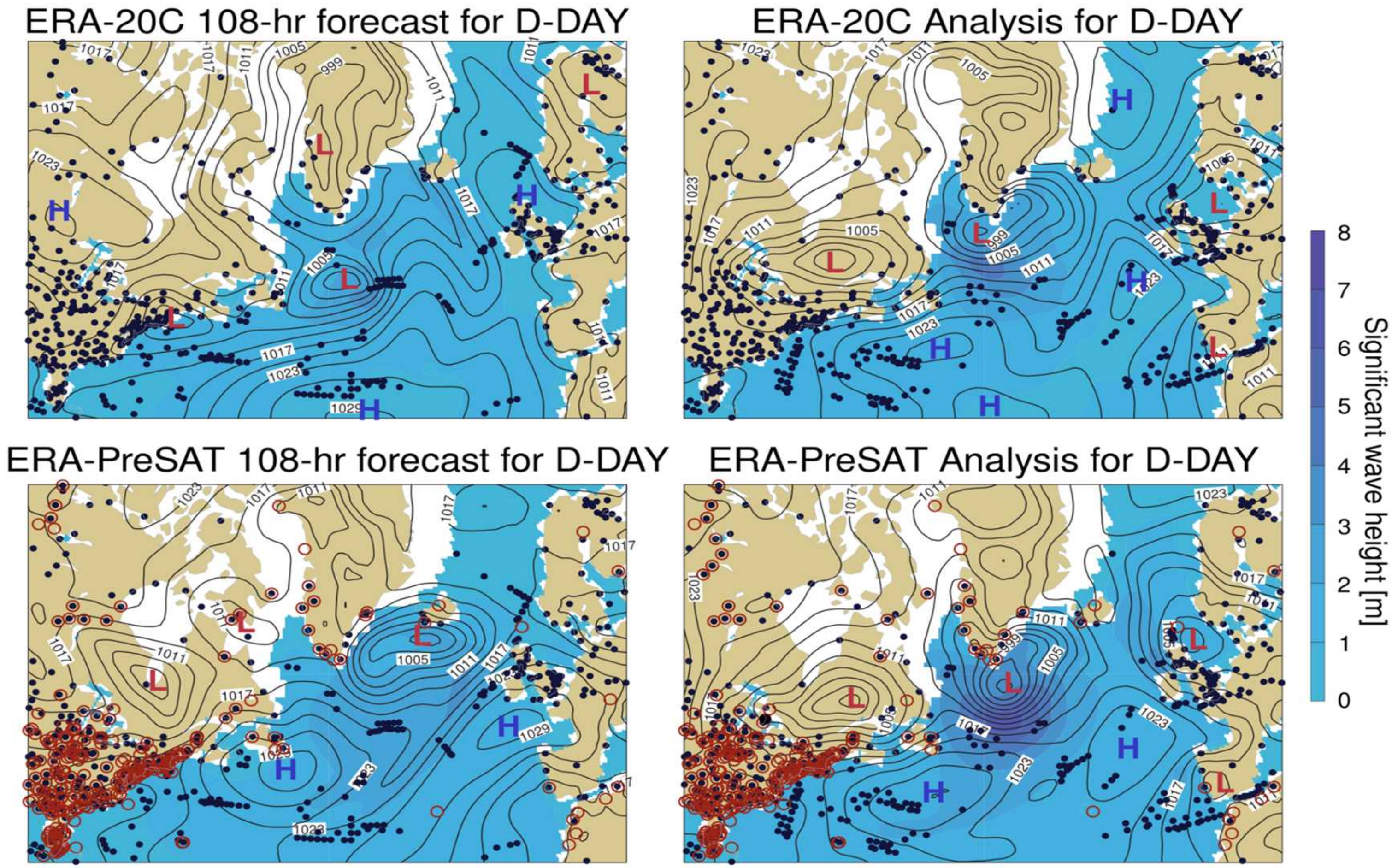
James Stagg,
Chief Meteorologist in 1944

http://www.ecmwf.int/en/about/media-centre/news/2014/ecmwf-revisits-meteorology-d-day-period





Hand-drawn analysis for 13 UTC 6 June 1944



- **Surface observations**

o **Upper air (mostly pilot balloons)**

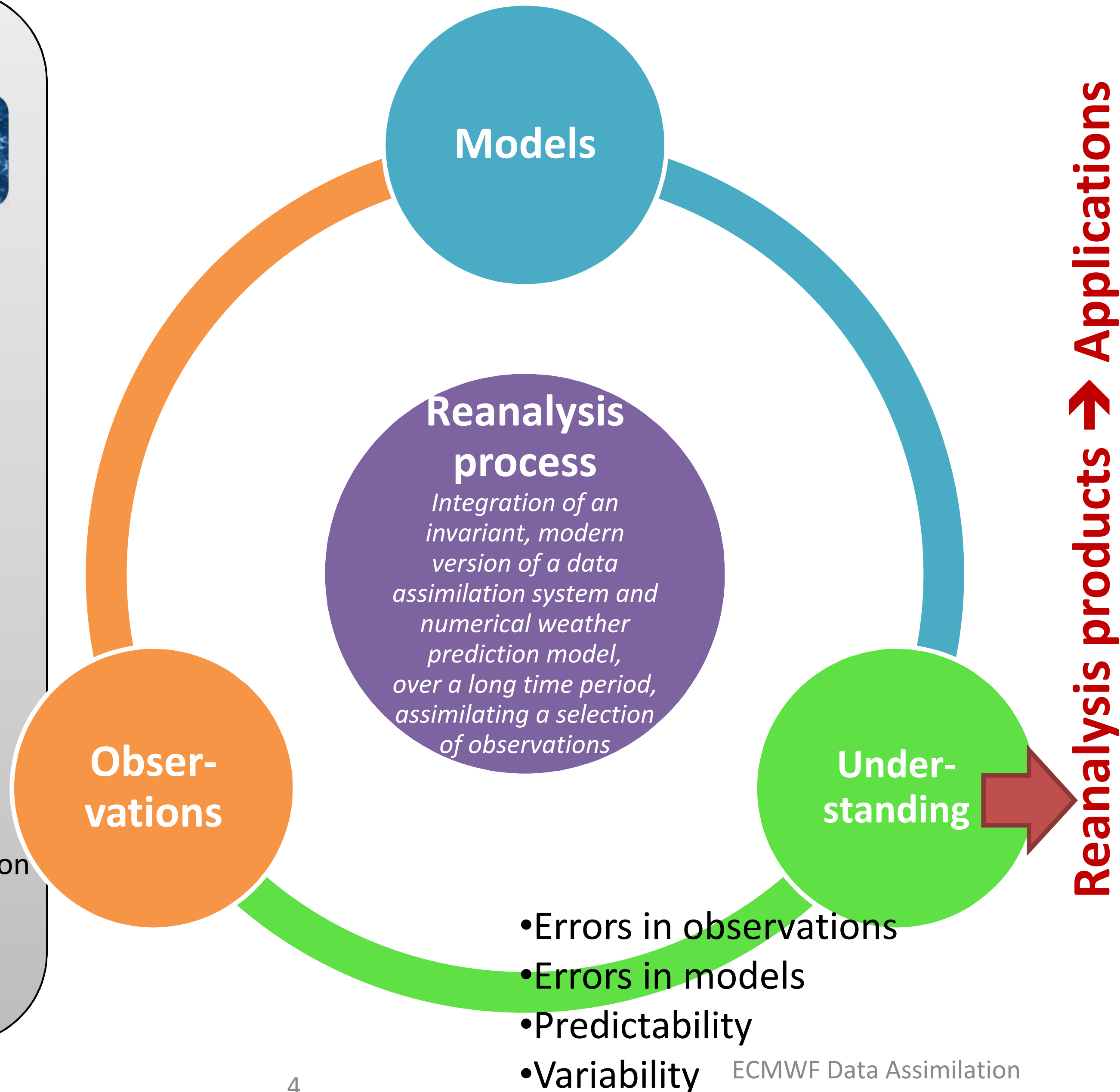
Reanalysis course outline

What is reanalysis?	• General concepts • Goals of reanalysis
How are reanalyses made?	• Observations • Model • Data assimilation

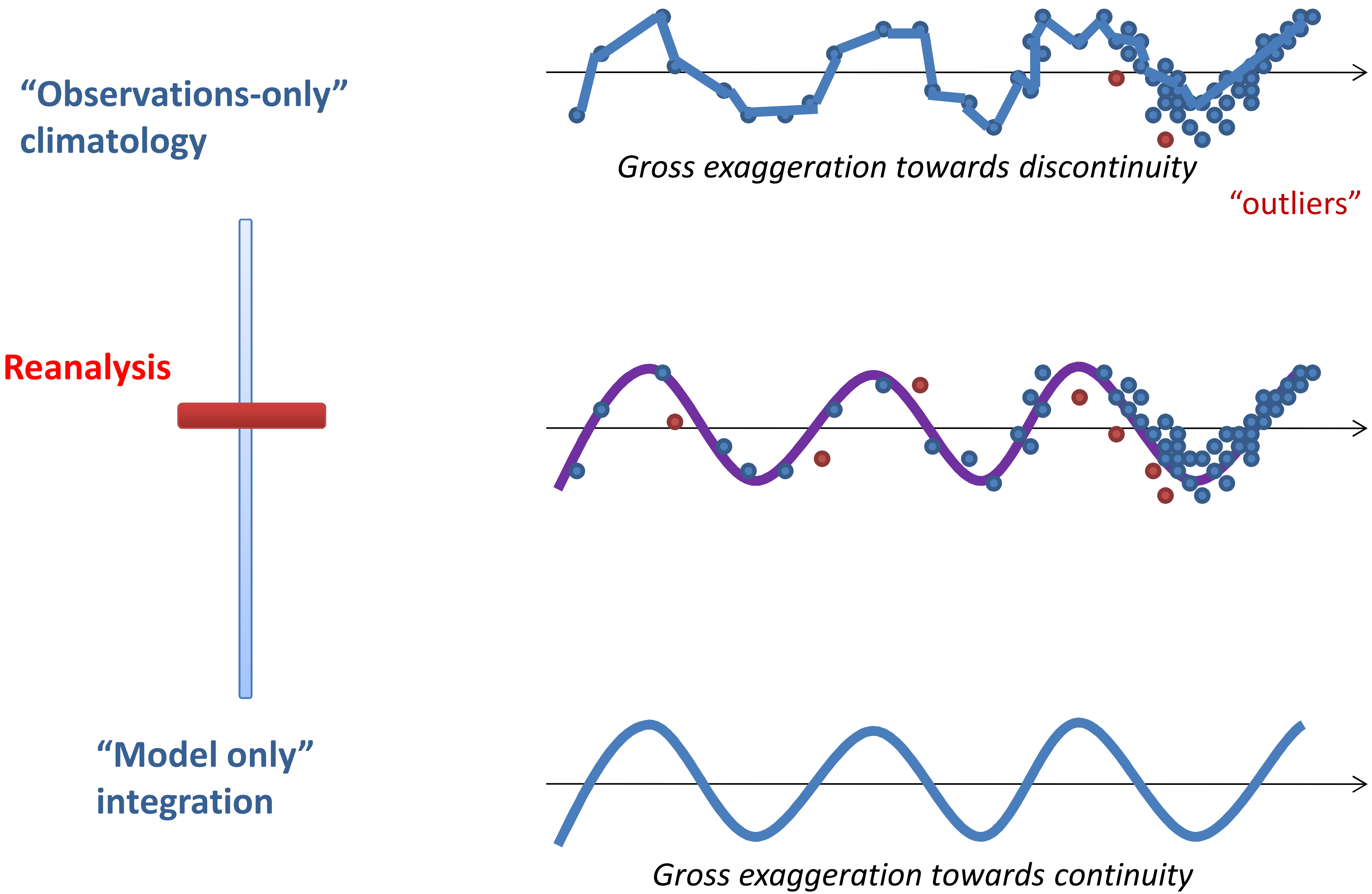
Background errors | Observation errors | Uncertainties

Reanalysis projects & applications	• Projects • Users • Applications
Conclusions	• Summary • Challenges ahead

The three pillars of geosciences



Objective: Reconstruct the past



Why not use simply observations?

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

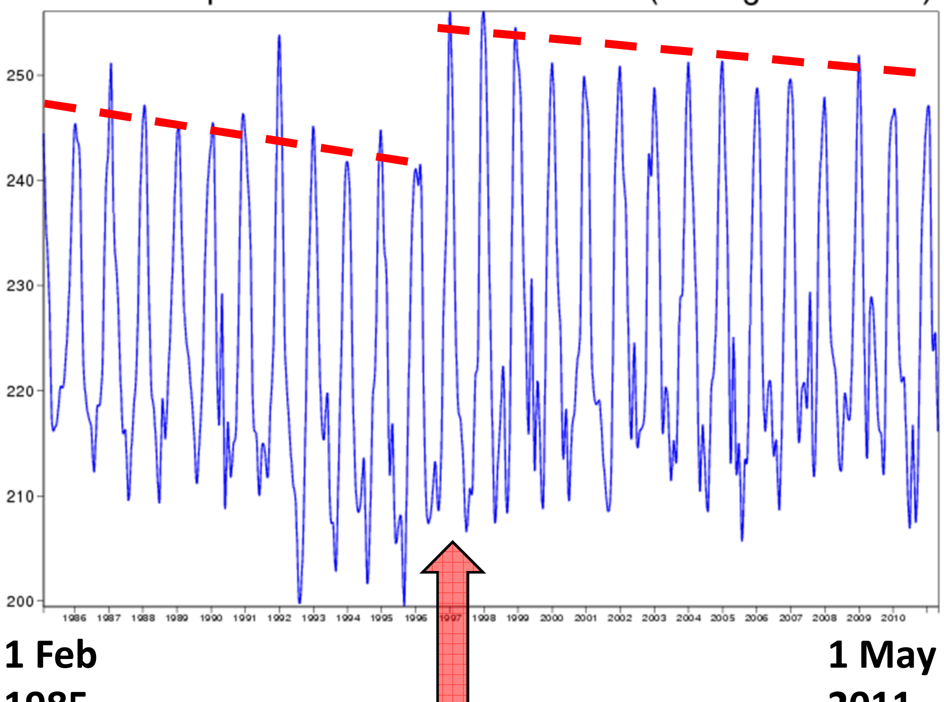
- LIKE ANY OTHER observations-based dataset.
- Reanalyses can adjust the observation influence to take account of how much information is already known (background errors). Example later with ERA-20C ensemble.

Observations-only datasets are the "observation limit" of reanalyses

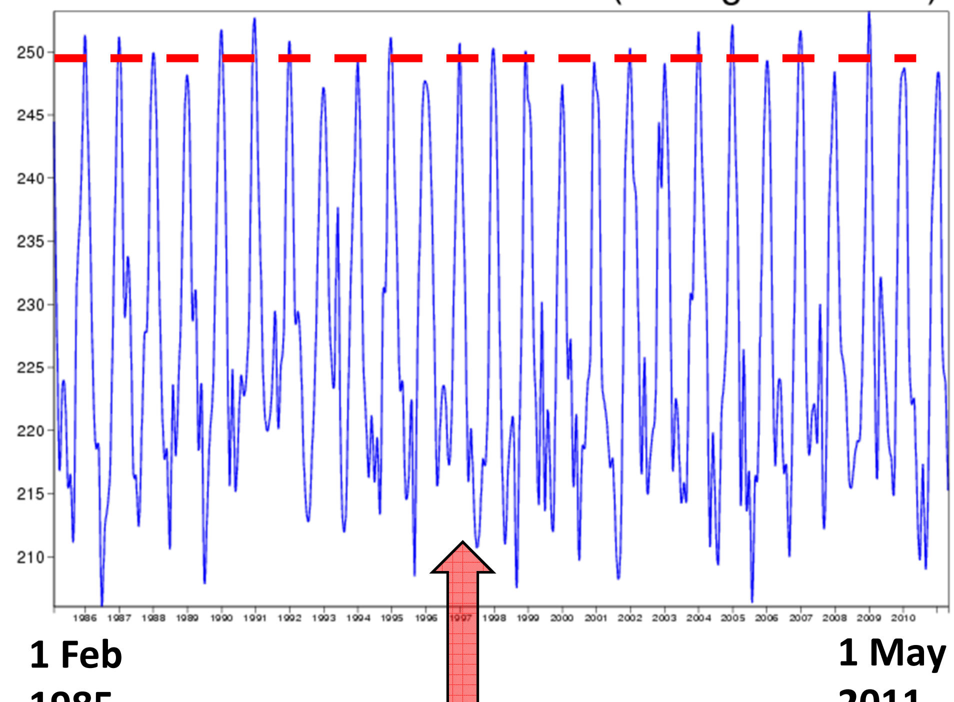
They also point out deficencies in reanalyses-- that further help improve understanding

Why not use simply operational NWP?

ECMWF Operations T2m at South Pole (average 88S-90S)



ERA-Interim T2m at South Pole (average 88S-90S)



1 Feb 1985 **1 May 2011** **1 Feb 1985** **1 May 2011**

Was there a sudden change in South Pole summer variability in 1997?

… probably not

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

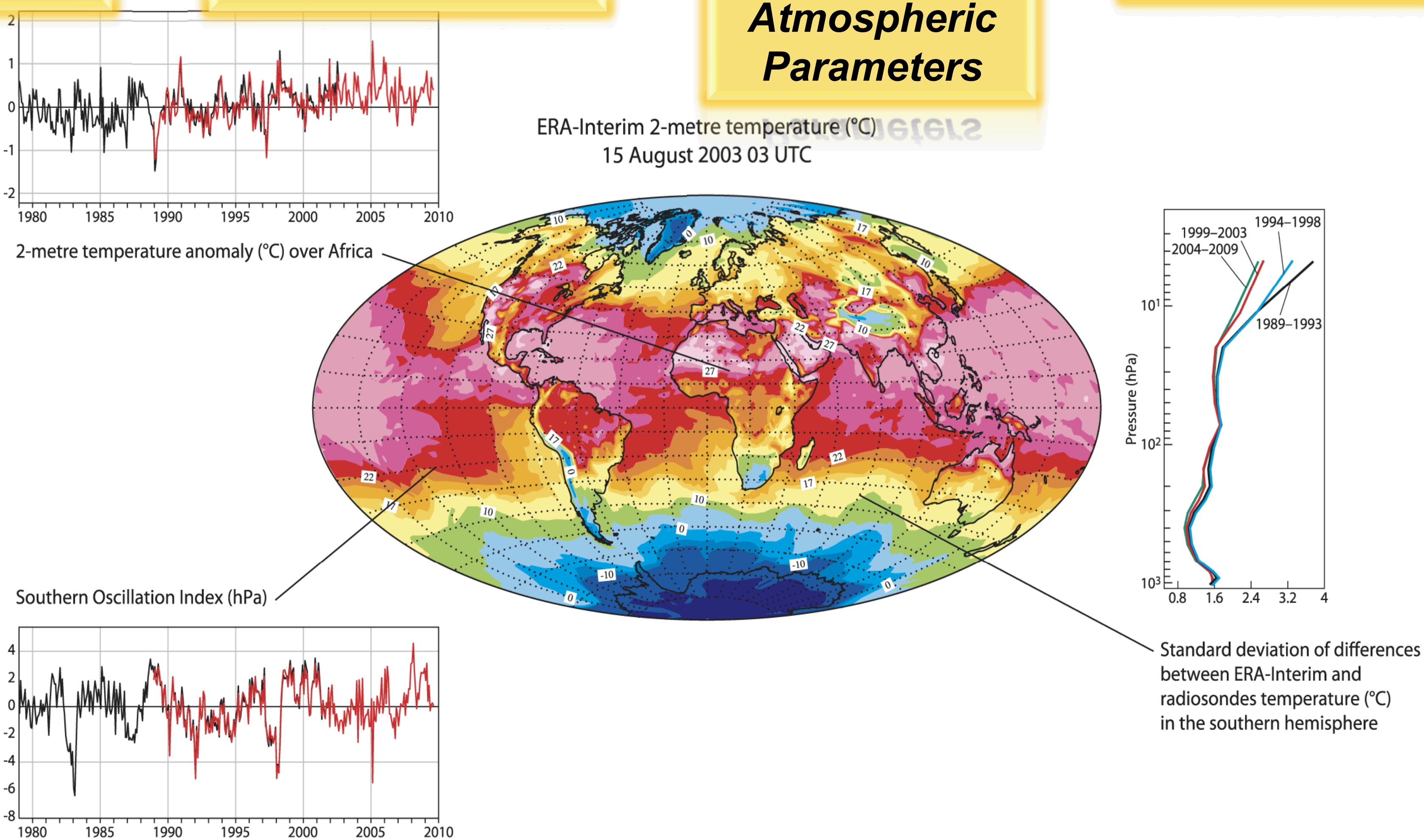
Summary of the goals: reanalysis products should be consistent …

…in Time

…in the Horizontal

…across Atmospheric Parameters

…in the Vertical



Reanalysis course outline

What is reanalysis?
- General concepts
- Goals of reanalysis

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

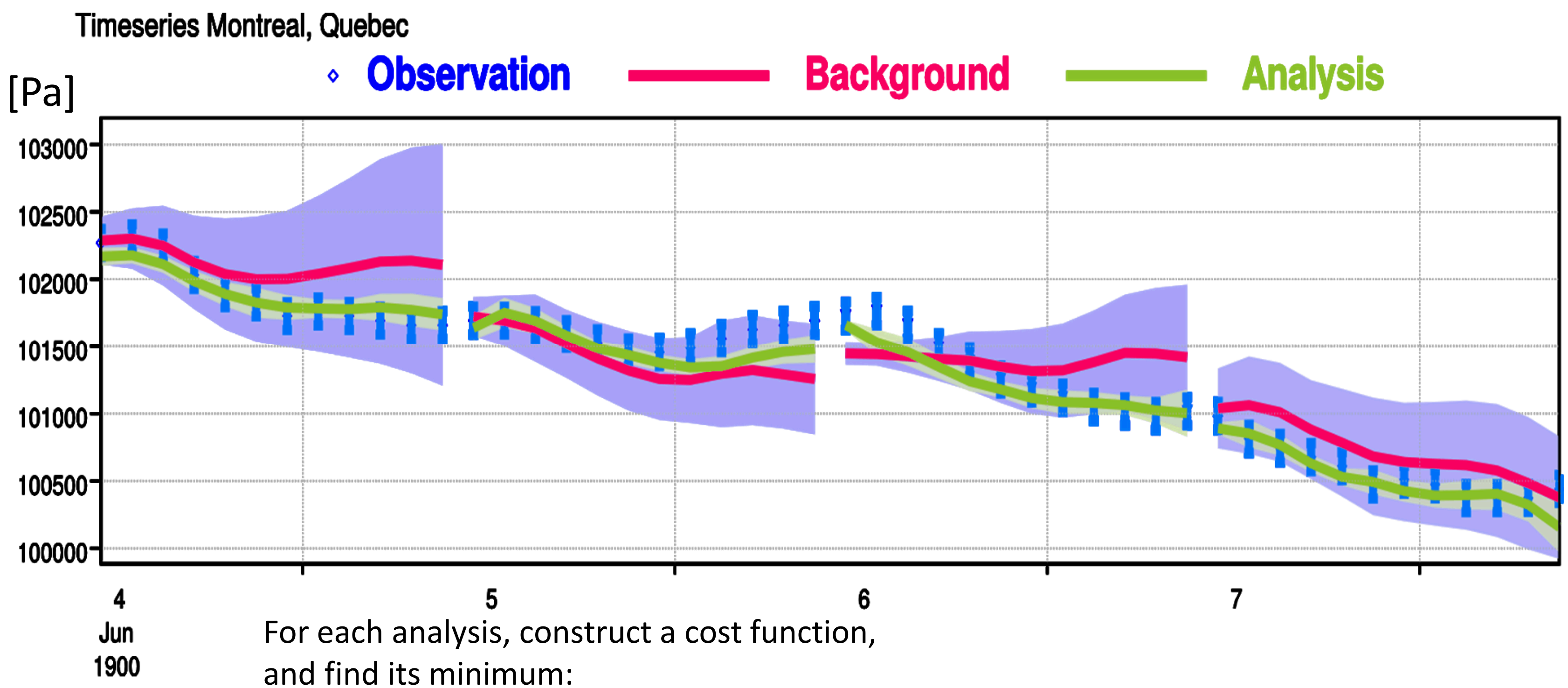
Background errors | Observation errors | Uncertainties

Reanalysis projects & applications
- Projects
- Users
- Applications

Conclusions
- Summary
- Challenges ahead

The ubiquitous data assimilation slide:

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



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

background constraint

observation constraint

$$J(\mathbf{x}) = (\mathbf{x}_b - \mathbf{x})^T \mathbf{B}^{-1} (\mathbf{x}_b - \mathbf{x}) + [\mathbf{y} - \mathbf{h}(\mathbf{x})]^T \mathbf{R}^{-1} [\mathbf{y} - \mathbf{h}(\mathbf{x})]$$

$$\mathbf{h}(\mathbf{x}) = \mathbf{h}[M(\mathbf{x})]$$

h simulates the observations

$$J(\mathbf{z}) = (\mathbf{z}_b - \mathbf{z})^T \mathbf{B}_z^{-1} (\mathbf{z}_b - \mathbf{z}) + [\mathbf{y} - \tilde{\mathbf{h}}(\mathbf{z})]^T \mathbf{R}^{-1} [\mathbf{y} - \tilde{\mathbf{h}}(\mathbf{z})]$$

$$\mathbf{z}^T = [\mathbf{x}^T \boldsymbol{\beta}^T]$$

$$\tilde{\mathbf{h}}(\mathbf{z}) = \mathbf{h}(\mathbf{x}) + \mathbf{b}(\mathbf{x}, \boldsymbol{\beta})$$

b simulates the observation biases

This produces the "most probable" atmospheric state *

* 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

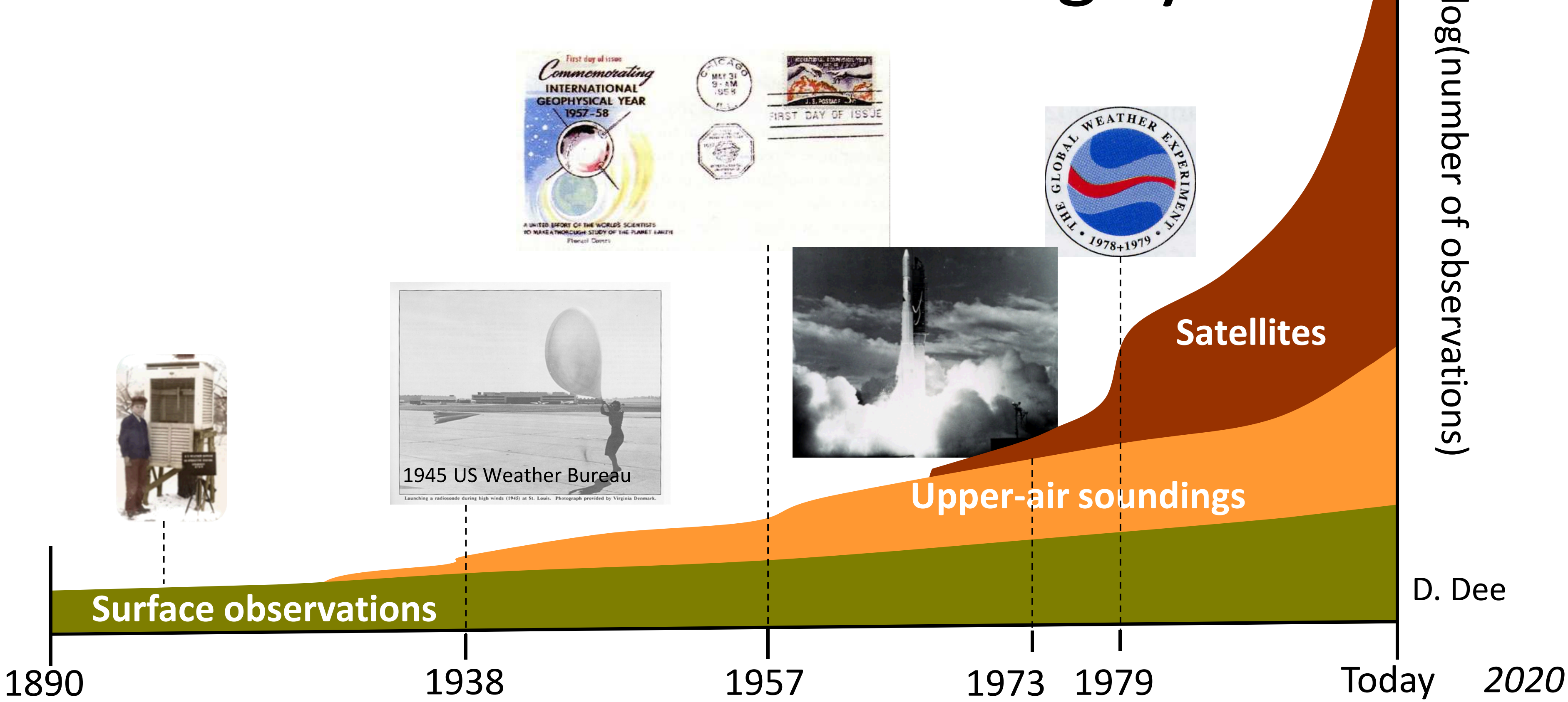
Reanalysis components
Part 1: Observations

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 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?

Evolution of the observing system



Manual stations, limited data exchange

First radiosonde networks, systematic soundings

International Geophysical Year: radiosonde network enhanced, especially in the Southern Hemisphere

First operational satellite soundings (NOAA-2)

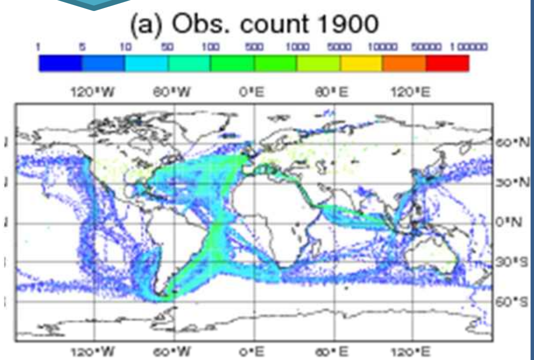
Improved sounding from polar orbiters; Winds from geostationary orbit; More data from commercial aircraft; First drifting buoys

More satellites, aircraft, buoys, ocean gliders and drifters. Fewer radiosondes, but probe higher. Better knowledge of instruments. More obs. per hour.

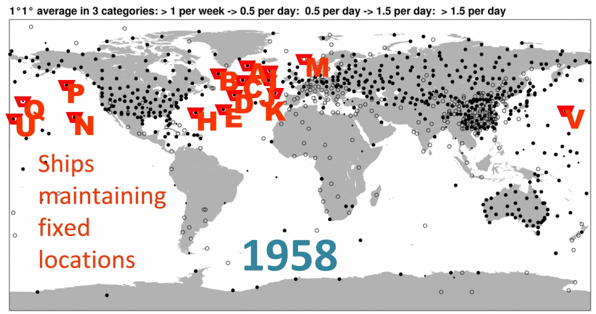
Opport-unity sensors : cell phones, UAVs, vehicles, rooftops, …

Evolution of the observation coverage

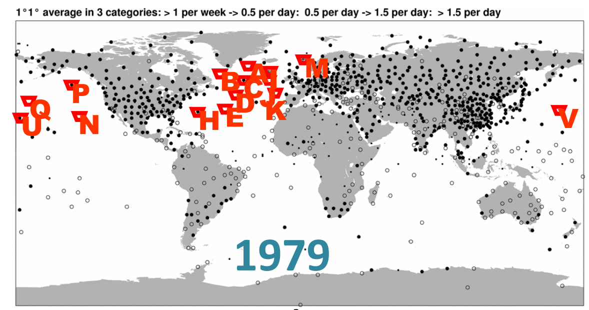
https://www.youtube.com/watch?v=NUfdFCHoxHM



Surface pressure network

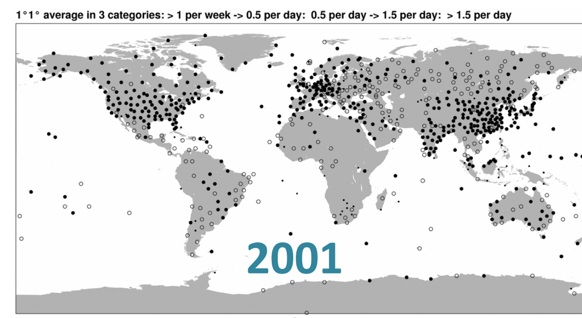


1609 soundings/day



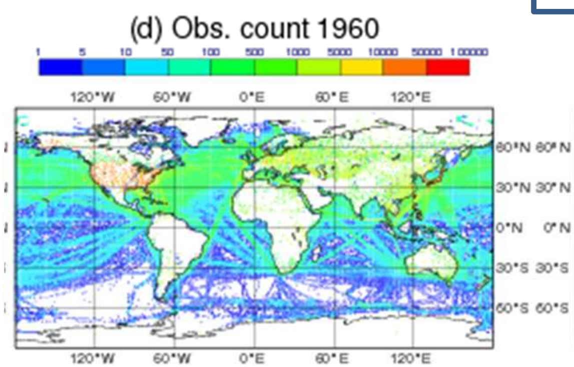
1626 soundings/day

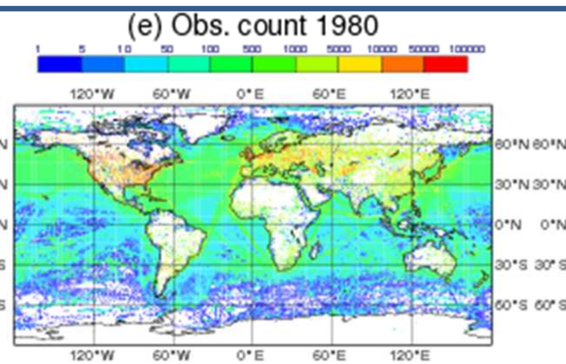
Radiosonde network

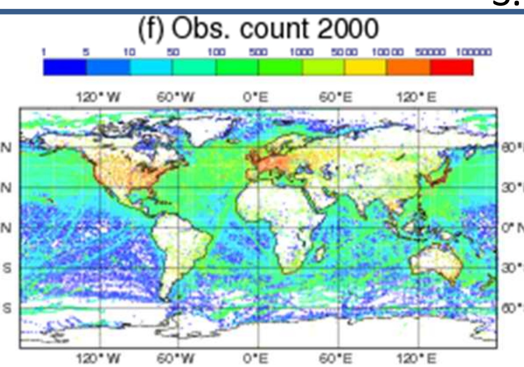


1189 soundings/day

S. Uppala

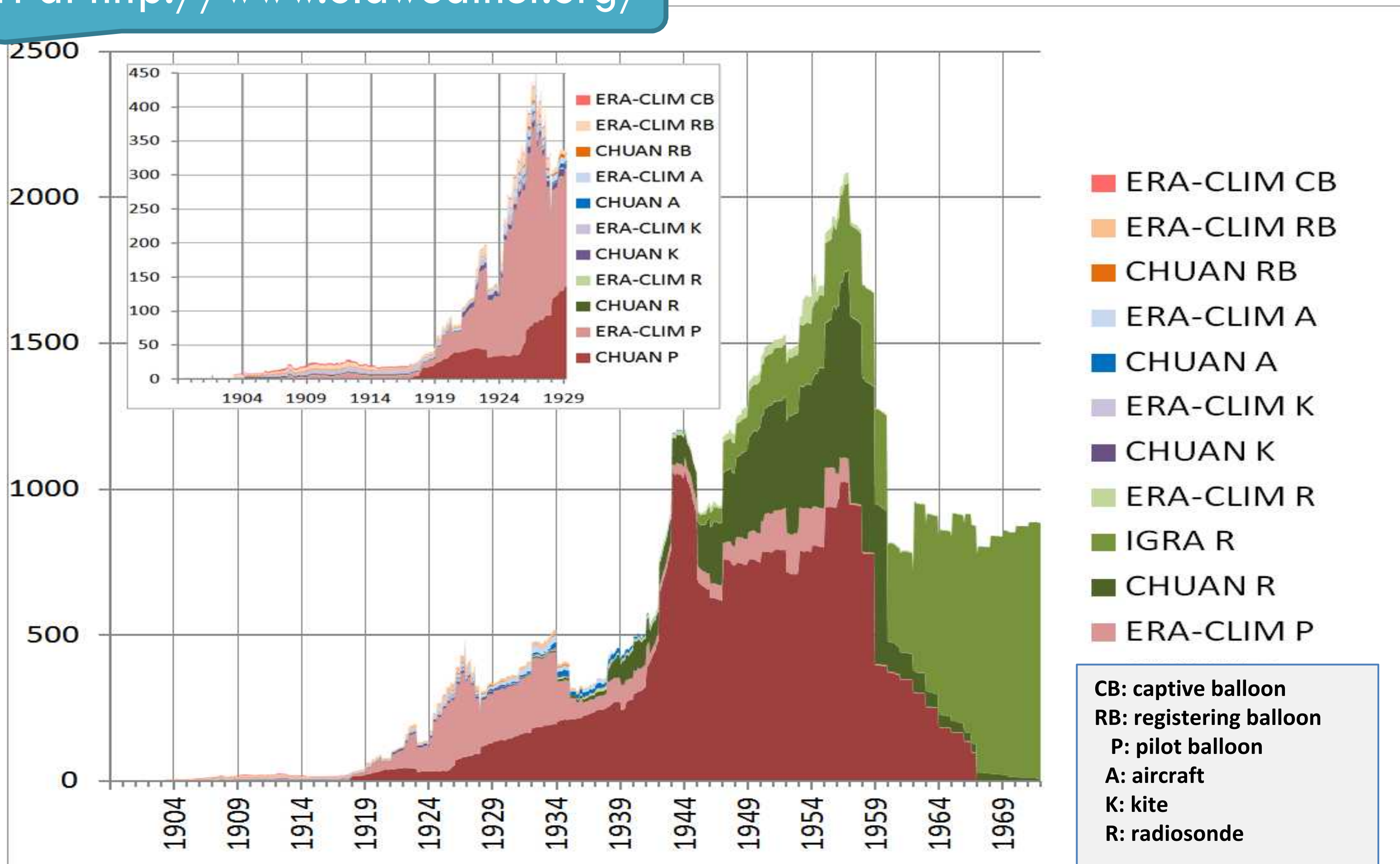






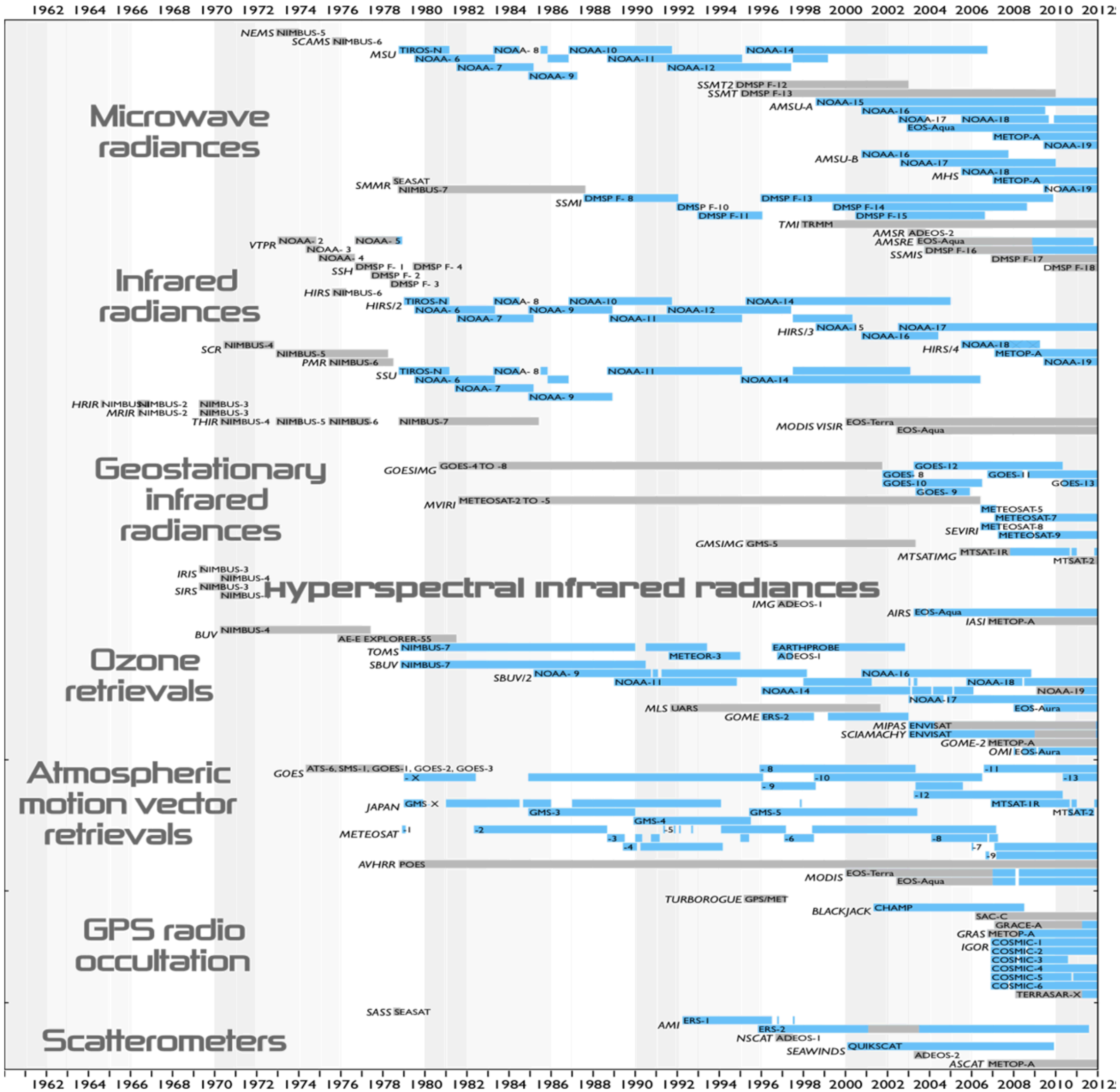
Efforts to improve the historical upper-air data record: “data rescue” a.k.a “data recovery”

Join the effort at http://www.oldweather.org/



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

Increased satellite observation diversity



Observation timeline (atmosphere)

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

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….

Model forcings for reanalysis

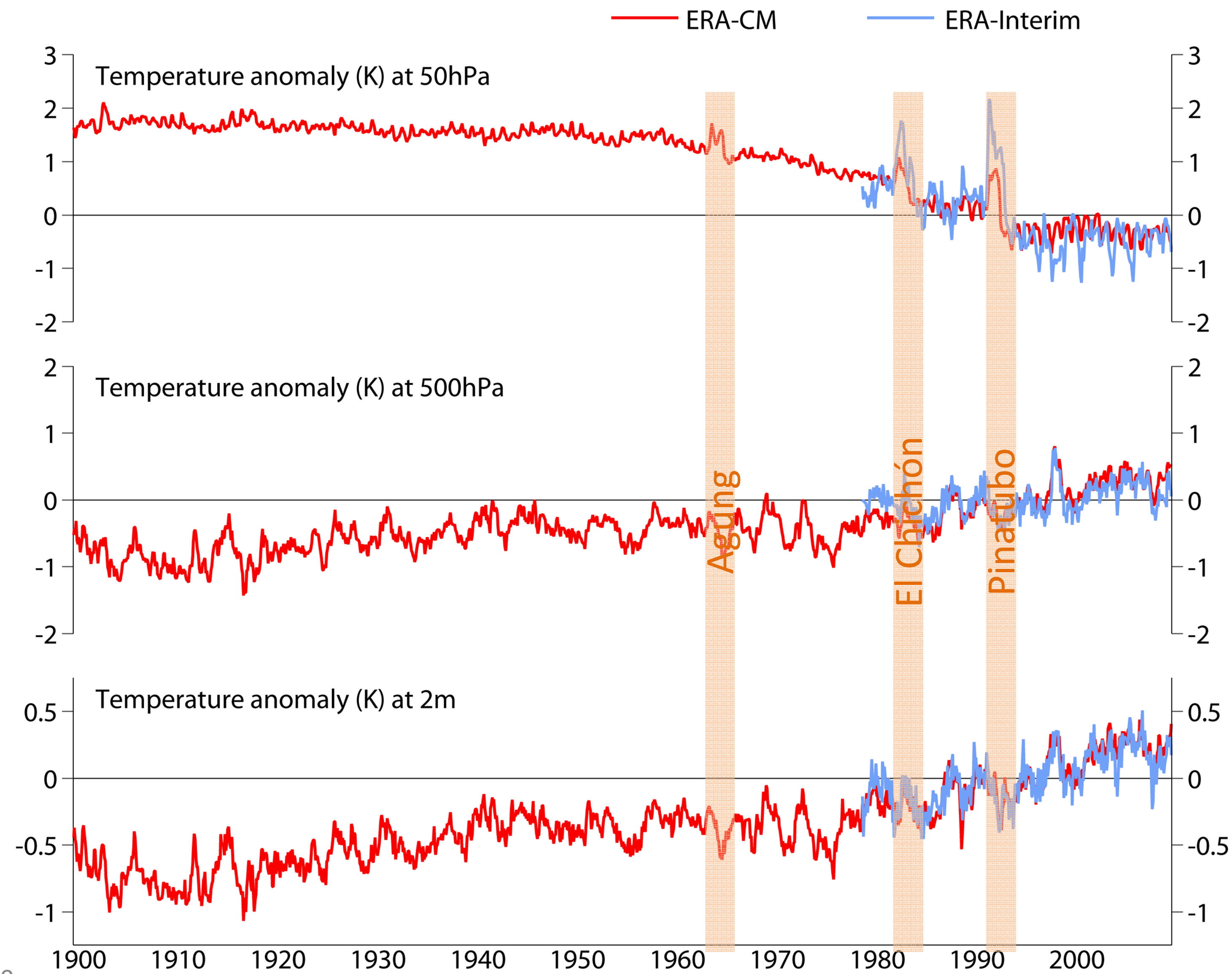
- ERA-20CM integrates the ECMWF model, but without data assimilation

- 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)

Model integration: ERA-20CM

http://onlinelibrary.wiley.com/doi/10.1002/qj.2528/pdf



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

What is reanalysis?

- General concepts
- Goals of reanalysis

How are reanalyses made?

- Observations
- Model
- Data assimilation

Background errors | Observation errors | Uncertainties

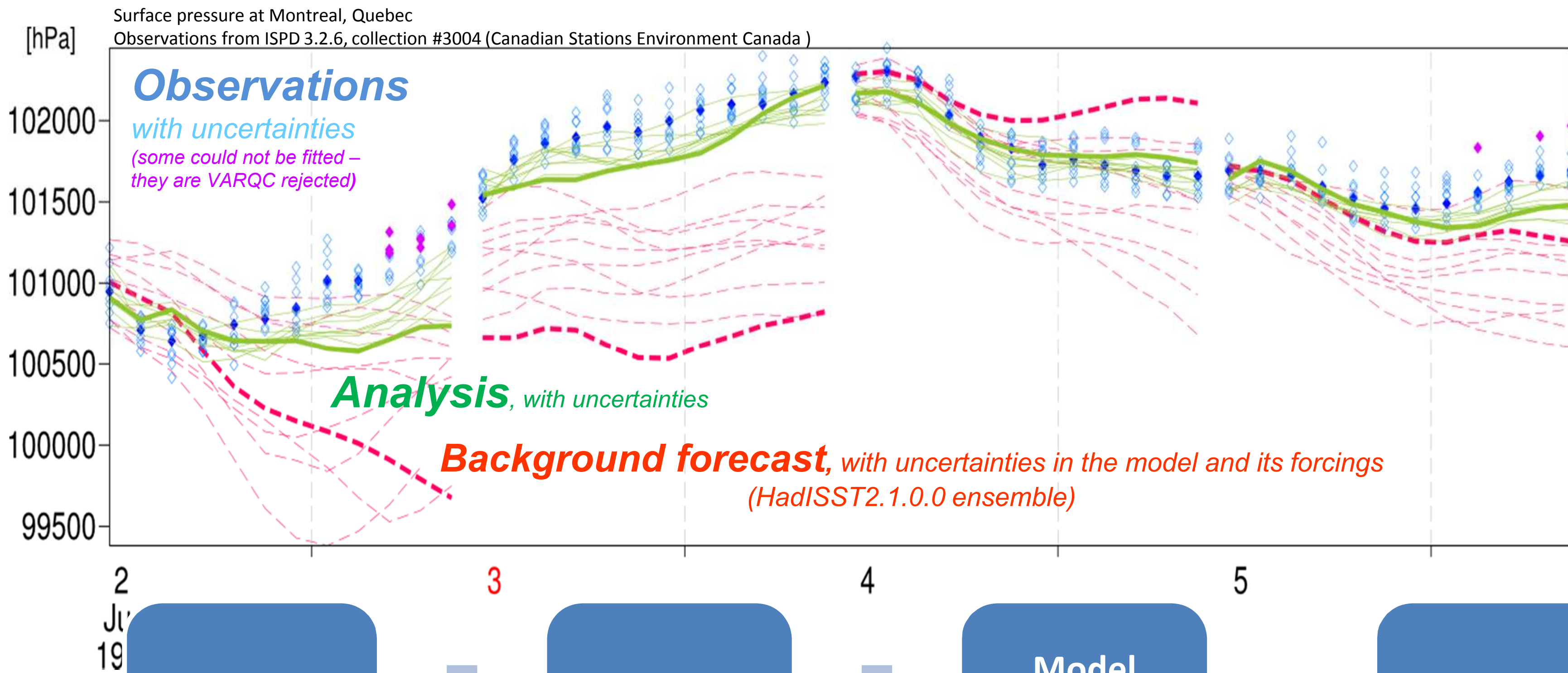
Reanalysis projects & applications

- Projects
- Users
- Applications

Conclusions

- Summary
- Challenges ahead

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

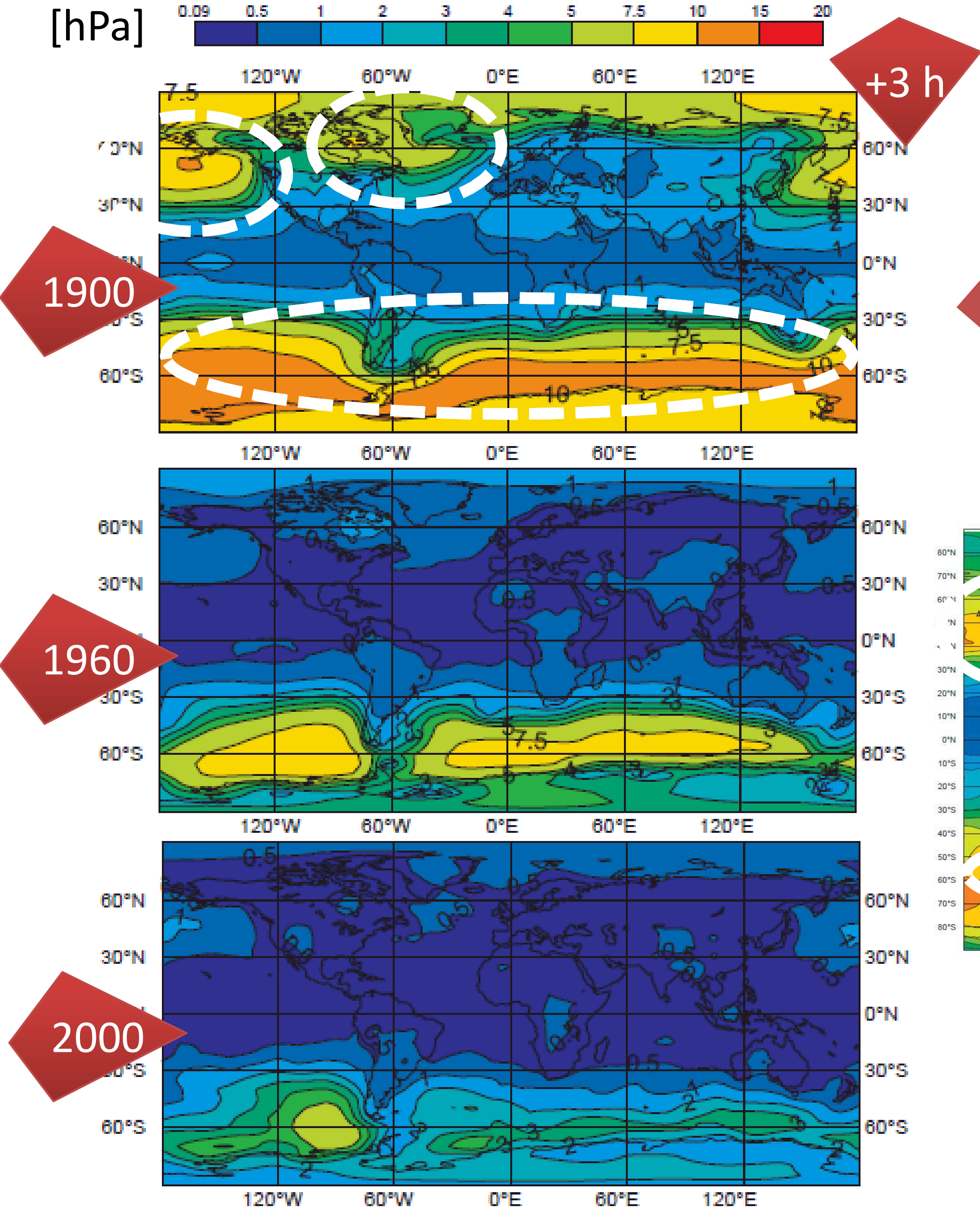


Observation uncertainties + Model uncertainties + Model forcing uncertainties = Reanalysis uncertainties

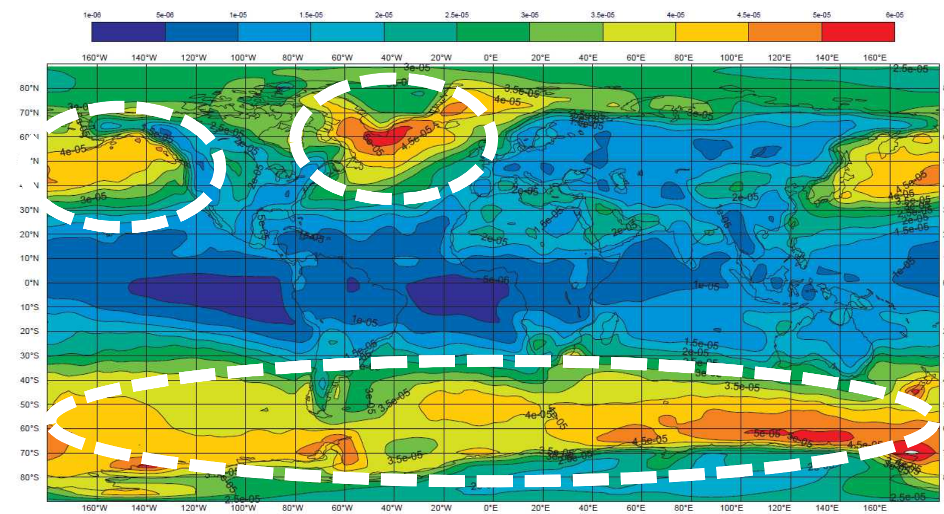
Aims:

1. Estimate automatically our background errors, and update them
2. Provide users with *some* uncertainties estimates

From ensemble spreads to background error variances

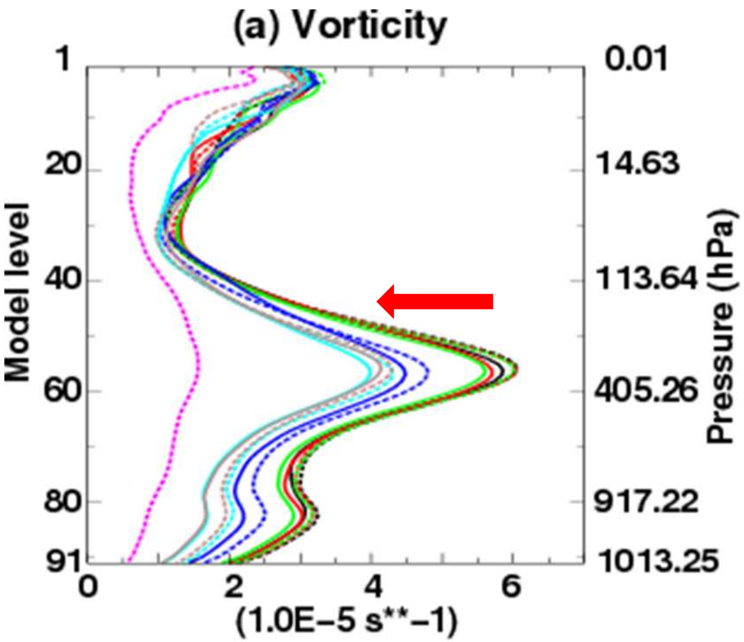


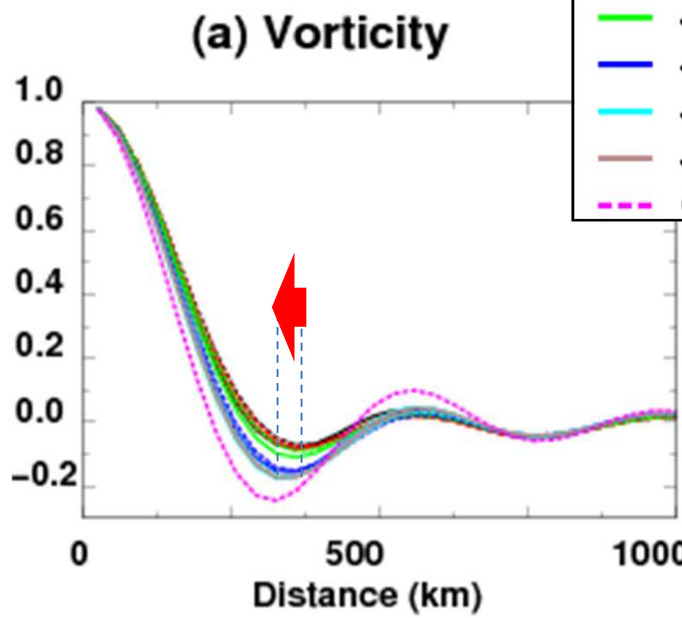
Estimate of bkg. error stdev. for vorticity at model level 89, for the year 1900 [s**-1]

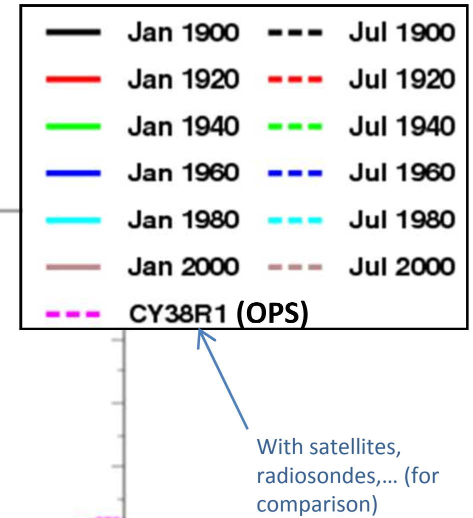


Self-updating background error covariances, throughout the century

(updated every 10 days, based on past 90 days)







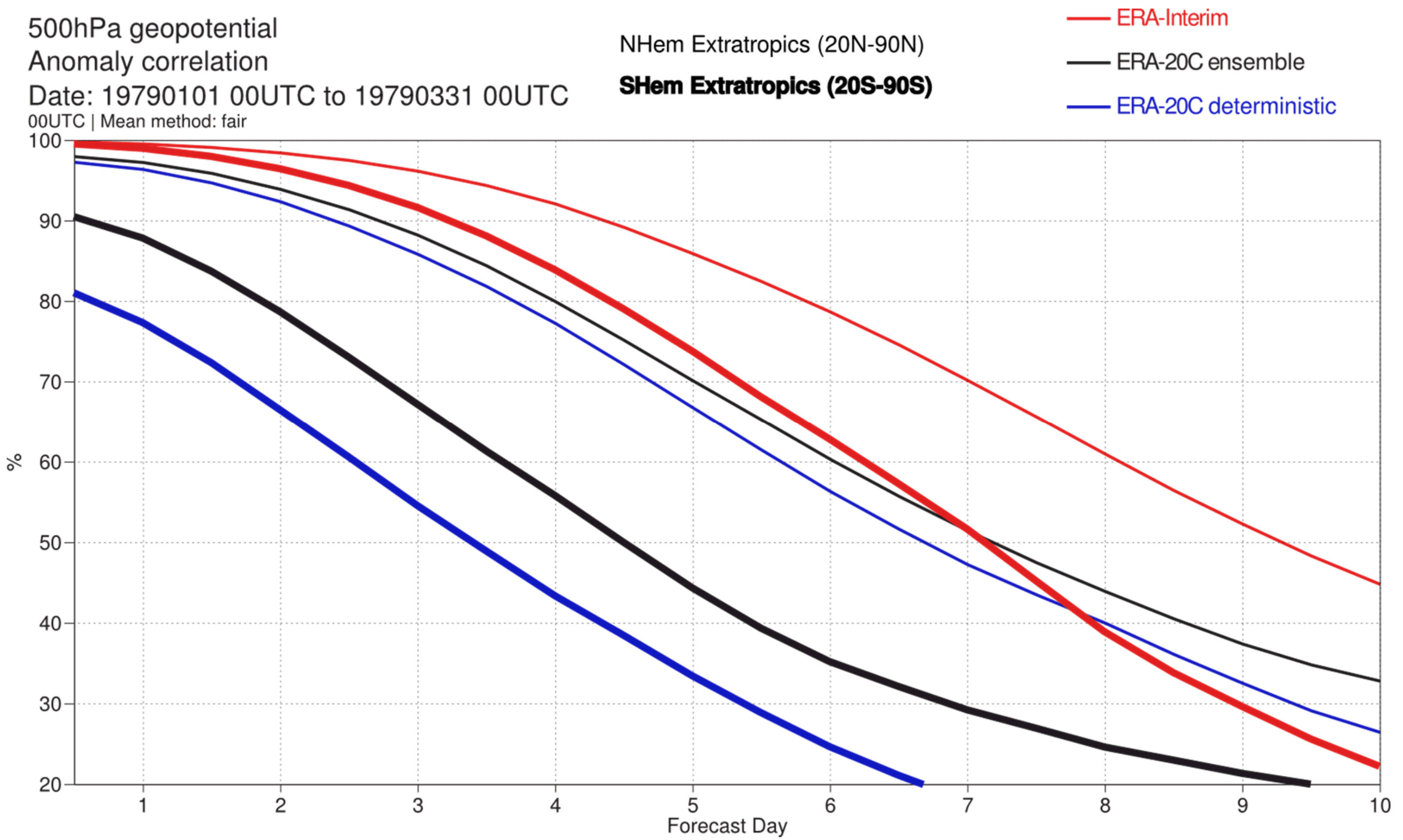
Over the course of the century, more observations result in…

➔**<u>Smaller</u>** background error variances, with **<u>sharper</u>** 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

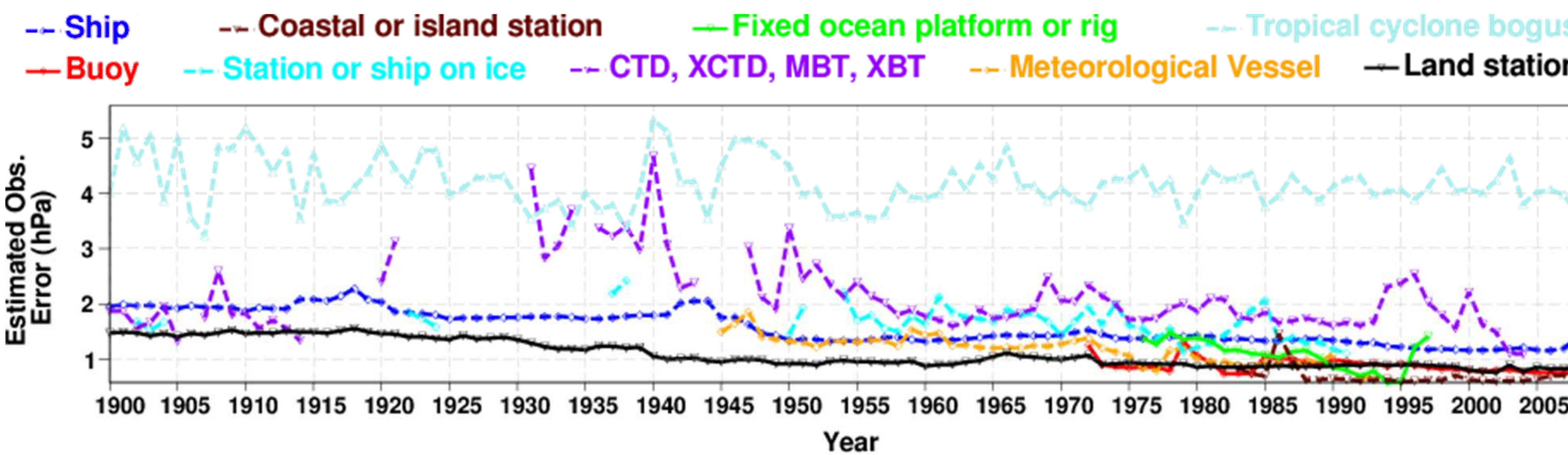


Background errors 4/4

Estimates of observation errors

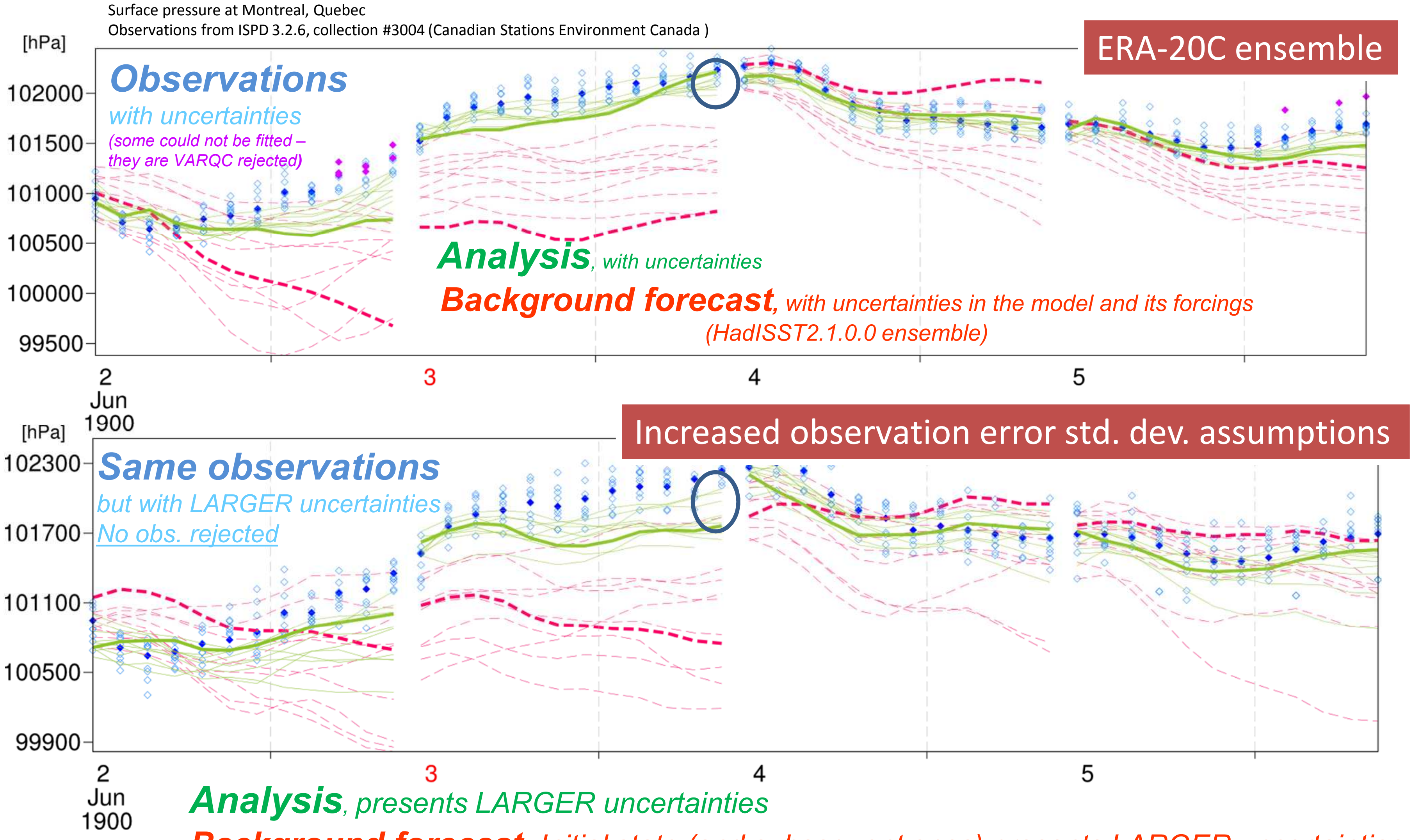
The following method can be used *a posteriori* to estimate observation error standard deviations: Desroziers *et al.* (2005), Diagnosis of observation, background and analysis-error statistics in observation space. Q.J.R. Meteorol. Soc., **131**: 3385–3396. doi: 10.1256/qj.05.108

The numbers that you obtain in such fashion are *for guidance only*. They are not absolute measures that you can plug in right away, but they are instructive to find out what observation types require attention, *e.g.* gross errors that would have slipped in, or changing errors over time etc… The example below is from ERA-20C ensemble assimilation statistics.



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

How useful are revised (larger) observation error standard deviation estimates?



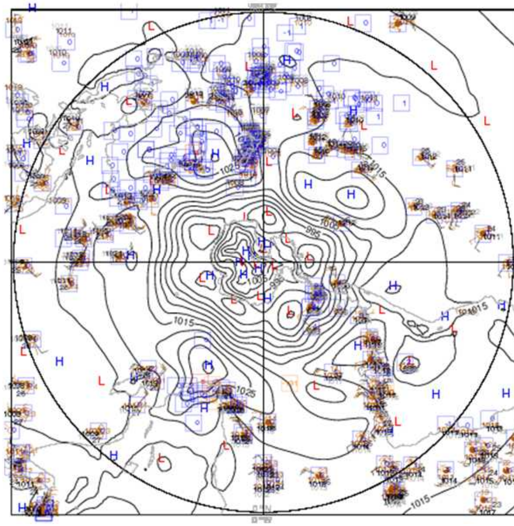
Analysis, *presents LARGER uncertainties*

Background forecast, *Initial state (and subsequent ones) presents LARGER uncertainties*

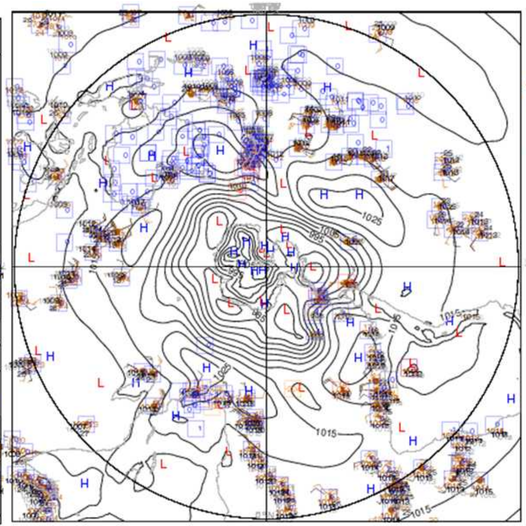
Impact of a single bad time-series

Southern Hemisphere

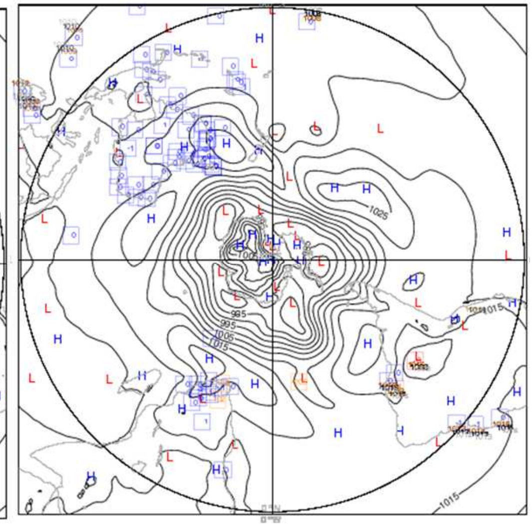
30 March 1954, 00 UTC



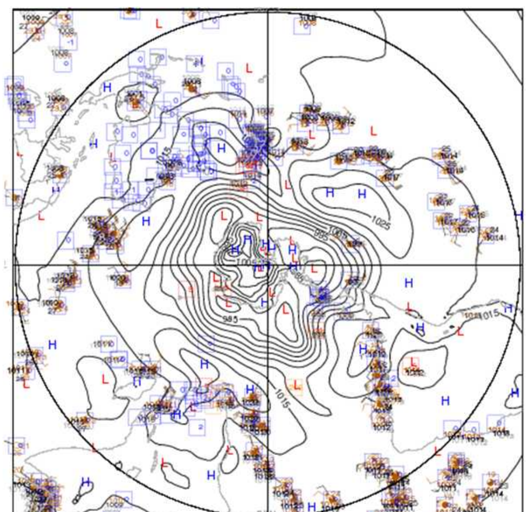
31 March 1954, 00 UTC



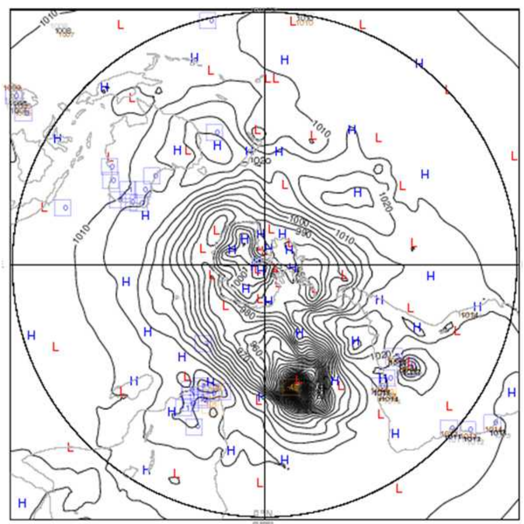
31 March 1954, 03 UTC



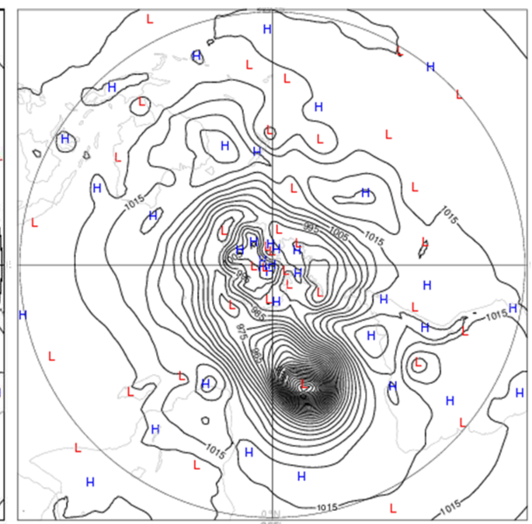
30 March 1954, 06 UTC



31 March 1954, 09 UTC

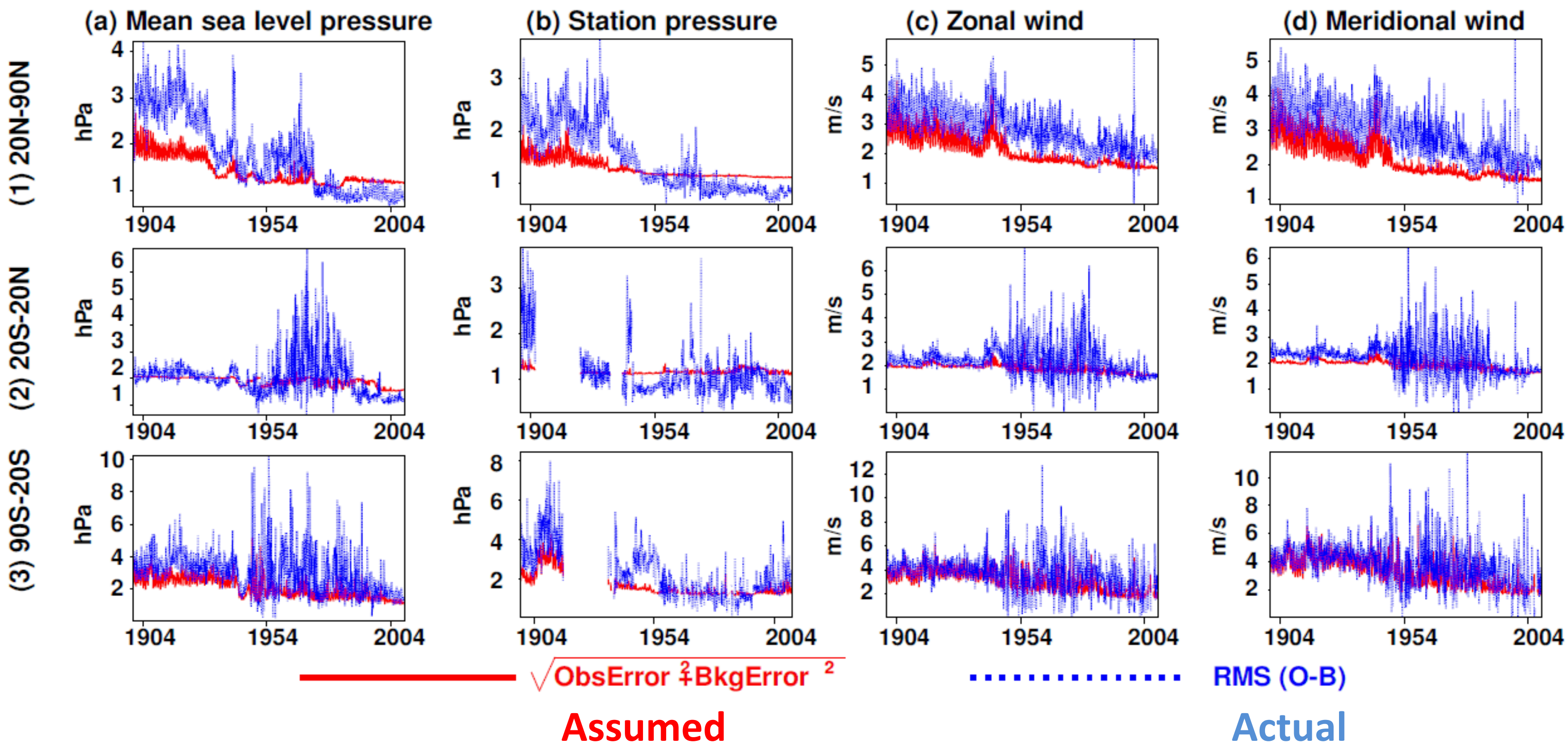


31 March 1954, 12 UTC



Observation errors 3/3

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

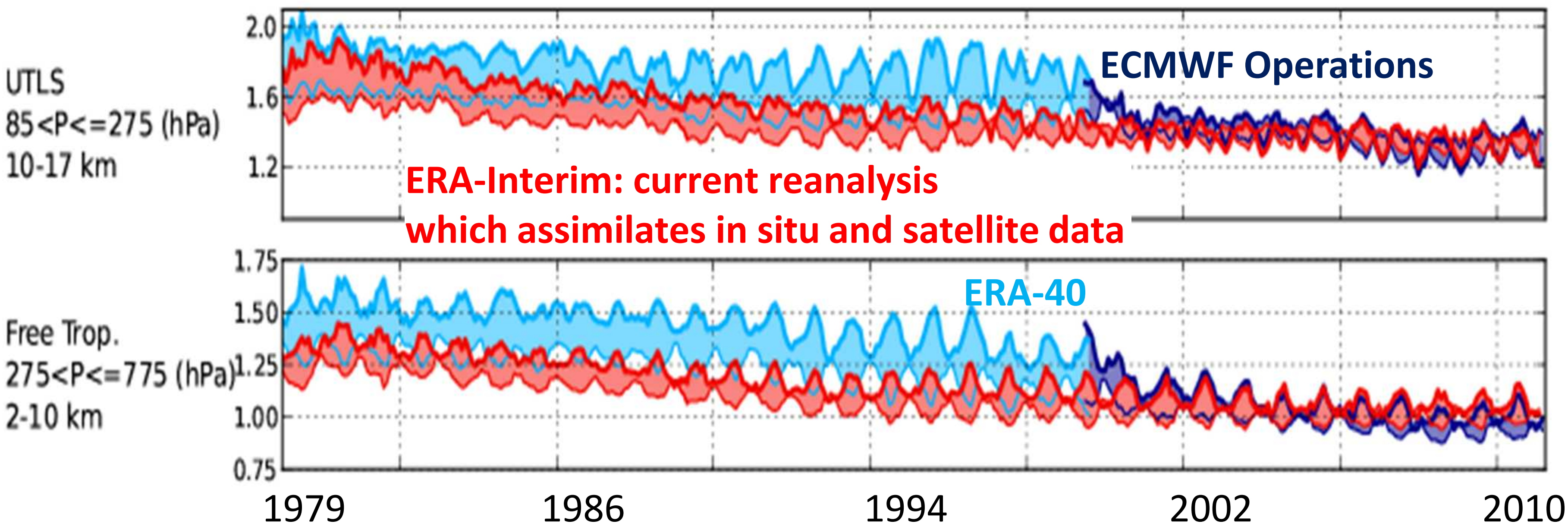


Showing only observations in the first 90 minutes of the 24-h window

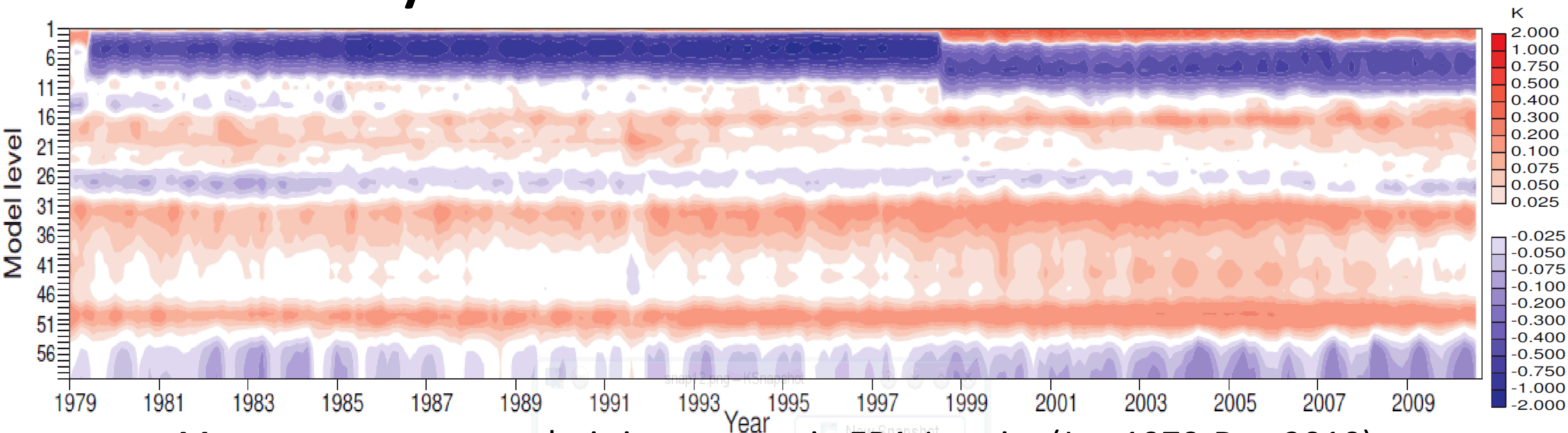
How to estimate uncertainty? Compare with observations

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

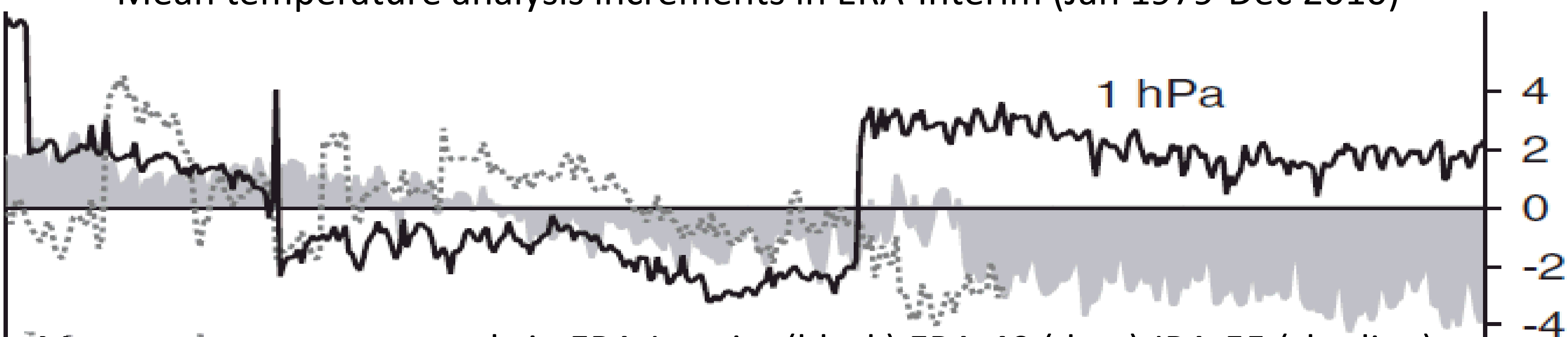
Thin line for Northern Hemisphere extratropics
Thick line for Southern Hemisphere, typically less well observed



Another look at uncertainties: Analysis increments and trends



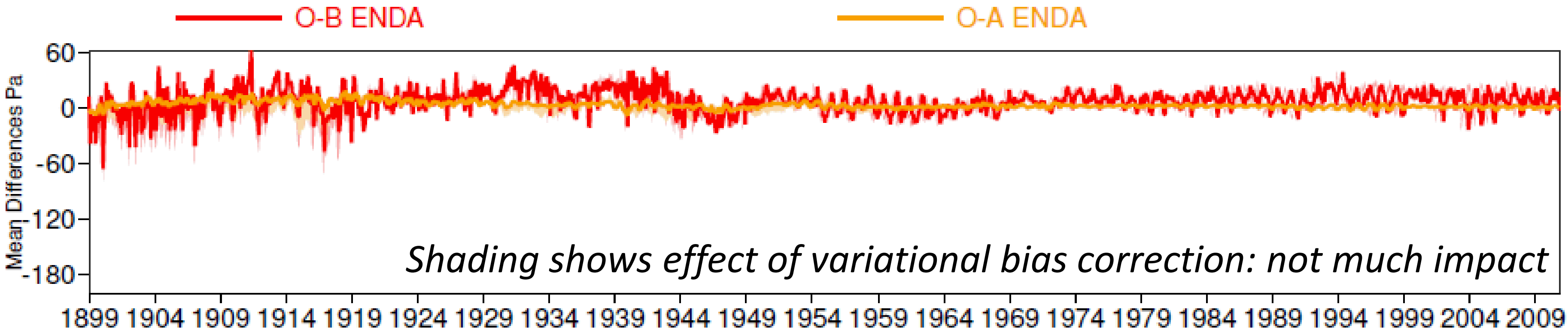
Mean temperature analysis increments in ERA-Interim (Jan 1979-Dec 2010)

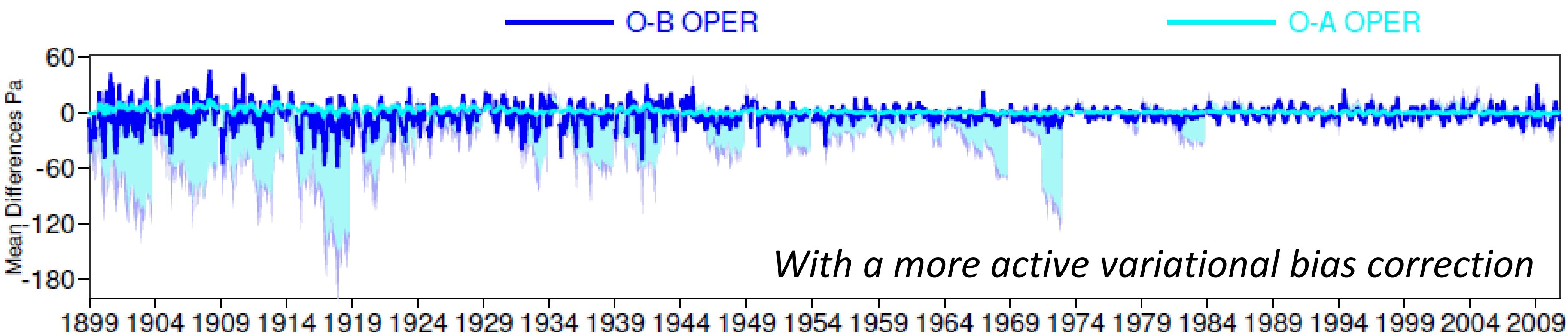


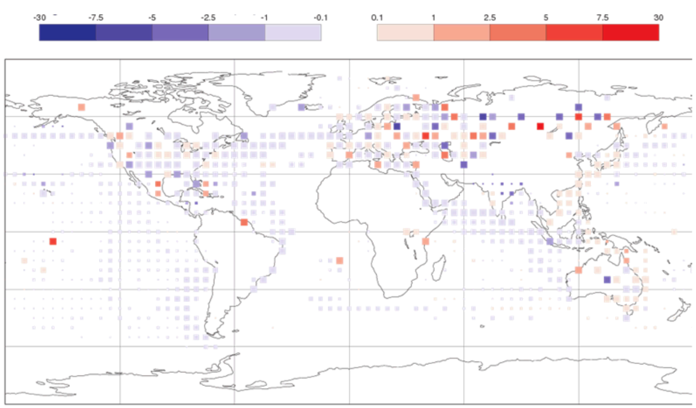
Mean temperature anomaly in ERA-Interim (black) ERA-40 (dots) JRA-55 (shading)

Simmons et al., 2014: Estimating low-frequency variability and trends in atmospheric temperature using ERA-Interim. Q.J.R. Meteorol. Soc., 140: 329–353. doi: 10.1002/qj.2317

Biases, the most difficult problem



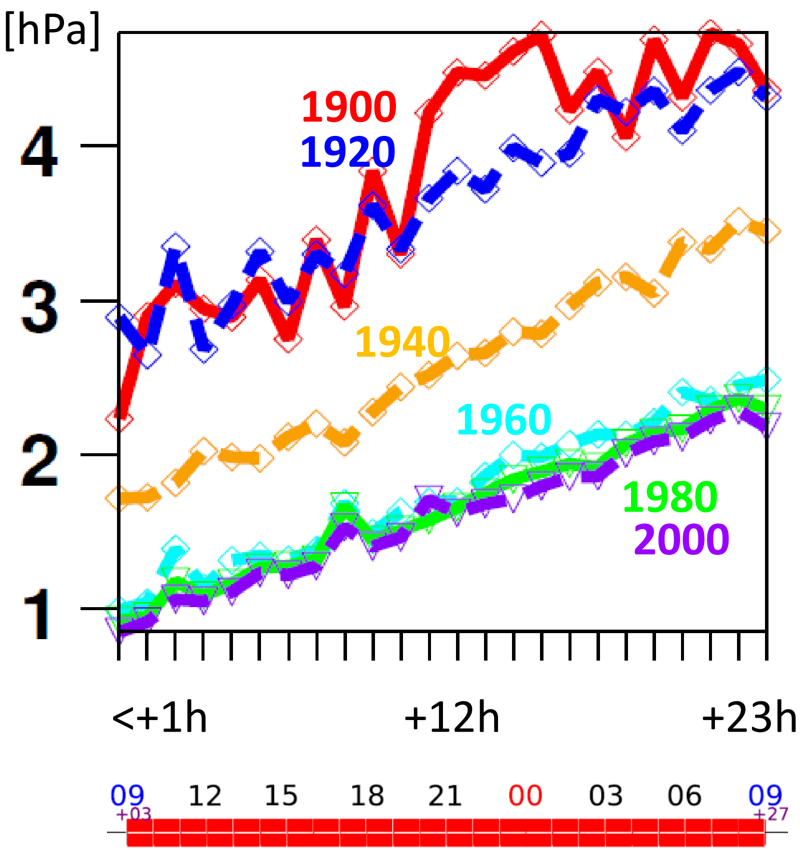




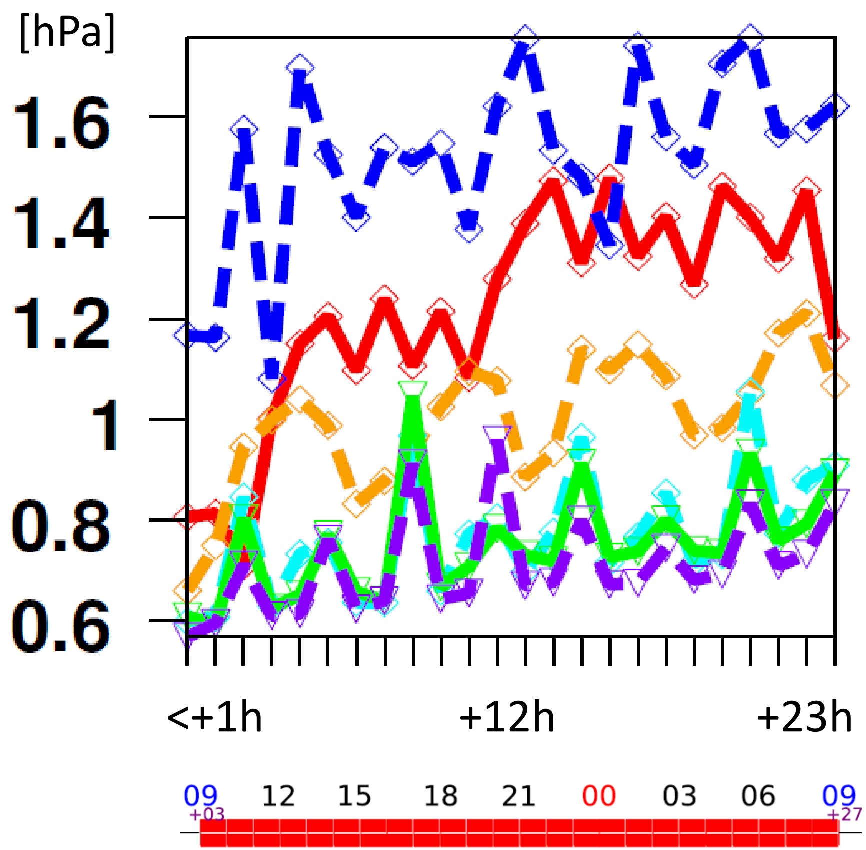
Map of surface pressure observation bias in 1906 estimated by the variational bias correction

What about error growth within the 24-hour window?

RMS (O-B)



RMS (O-A)



Uncertainties 5/5

Reanalysis course outline

What is reanalysis?	• General concepts • Goals of reanalysis
How are reanalyses made?	• Observations • Model • Data assimilation

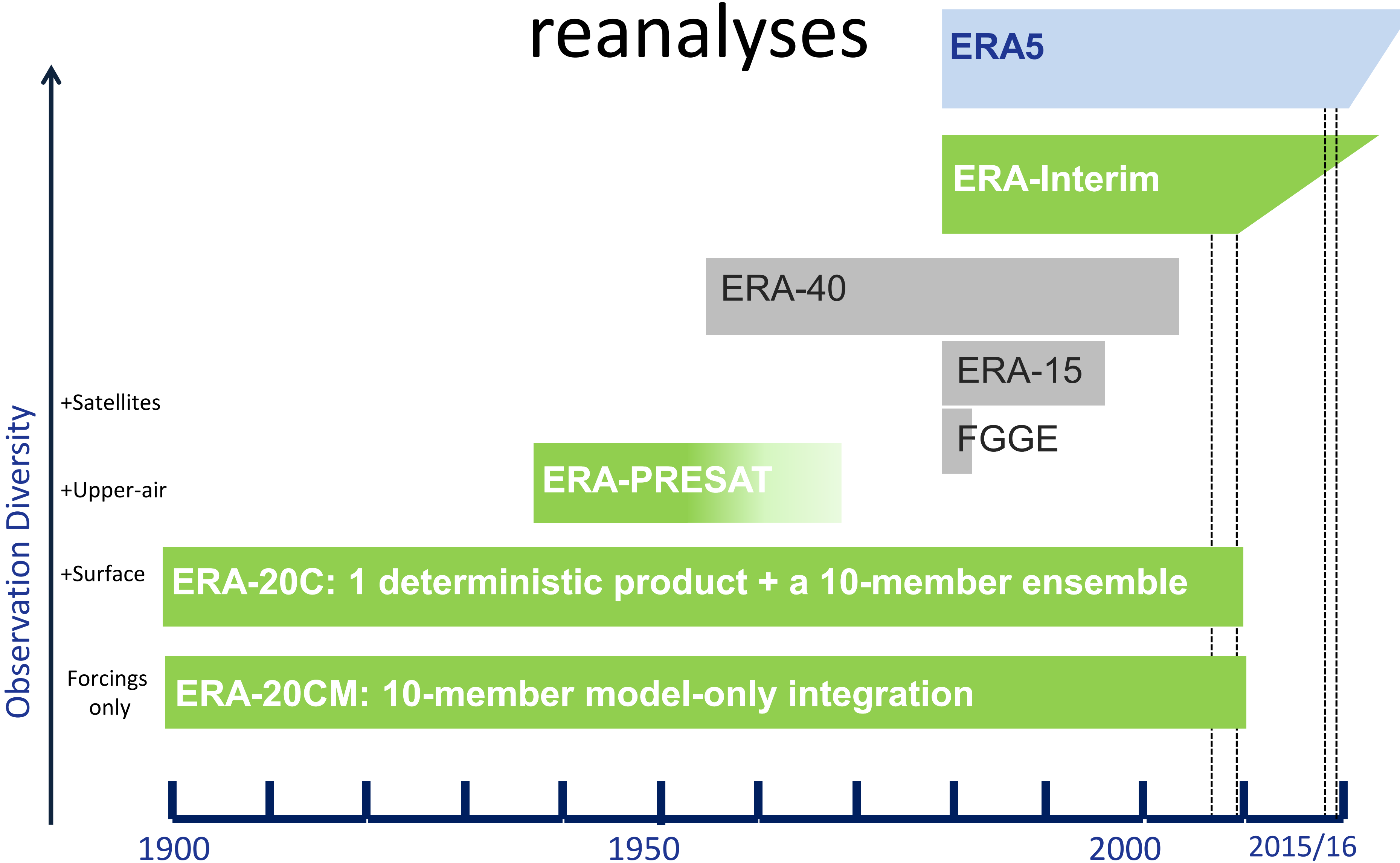
Background errors | Observation errors | Uncertainties

Reanalysis projects & applications	• Projects • Users • Applications
Conclusions	• Summary • Challenges ahead

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)

Overview of ECMWF atmospheric reanalyses



Datasets available from www.ecmwf.int

ERA-Interim: Comprehensive reanalysis of the recent times, extended monthly, assimilated 40×10^9 observations so far, serving >20,000 users

ECMWF | About | Forecasts | Computing | Research | Learning | Paul Poli | Search site | Go

Browse reanalysis datasets

- Data Assimilation
- Modelling and prediction
- Climate reanalysis
  - **Reanalysis datasets**
  - ERA-Interim
  - ERA-Interim/Land
  - ERA-20C
  - Coupled Earth-system reanalysis
  - Reanalysis for climate monitoring
  - Ocean reanalysis
- Projects
- Publications
- Special Projects

View published | New draft | Revisions | Access control

Dataset	Archive	Time period	Atmosphere	Atmospheric composition	Ocean waves	Ocean sub-surface	Land-surface	Observation Feedback Archive
ERA-Interim	Download	1979-present	✓		✓		✓	
ERA-Interim/Land	Download	1979-2010					✓	
ERA-20CM	Download	1900-2010	✓		✓		✓	
ERA-20C	Download	1900-2010	✓		✓		✓	Expected soon...
ERA-20CL	Expected soon...	1900-2010					✓	
ERA-40	Download	1957-2002	✓		✓		✓	
ERA-15	Download	1979-1993	✓				✓	

ERA-40 & ERA-15: Outdated now. Do not use to initiate new research.

ERA-20C family: ECMWF's first stab at the full 20th century

How (outside) users exploit reanalysis data

- Monitor the observing system
  - Feedback on observational quality, bias corrections
  - Basis for homogenization studies of long data records

ERA-Interim, ERA-40, ERA-20C: More than 20,000 users

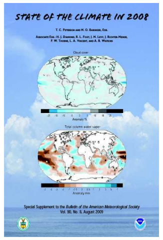
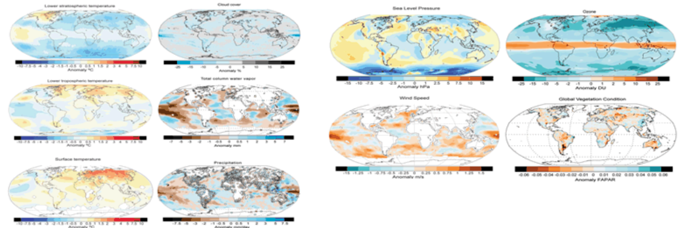
- Develop climate models
  - Use reanalysis products for verification, diagnosis, calibrating output,, …
- Drive users' models/applications
  - Use reanalysis as large-scale initial or boundary conditions for smaller-scale models (global→regional; regional→local), in various fields: wind energy, ocean 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

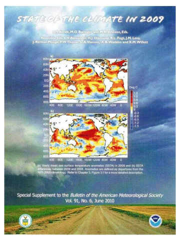
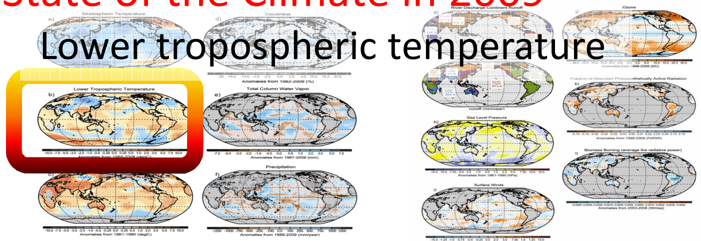
- Baseline to track NWP score improvements
- Calibration for seasonal forecasting system
- Reference to diagnose changes brought by model improvements

Growing recognition for climate application

BAMS State of the Climate in 2008

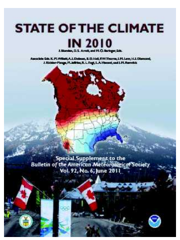
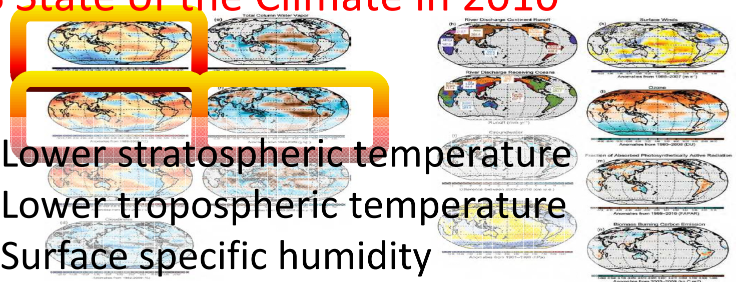



BAMS State of the Climate in 2009

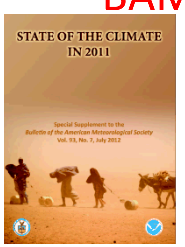
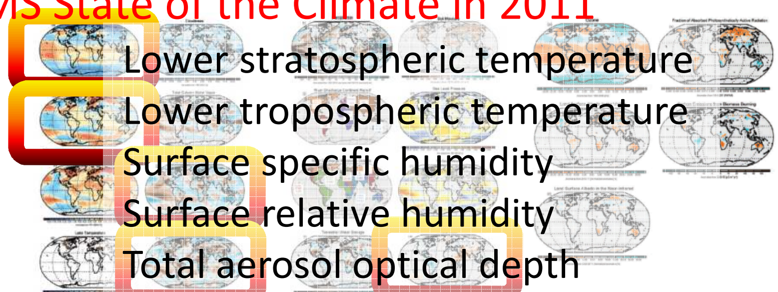
Lower tropospheric temperature

BAMS State of the Climate in 2010

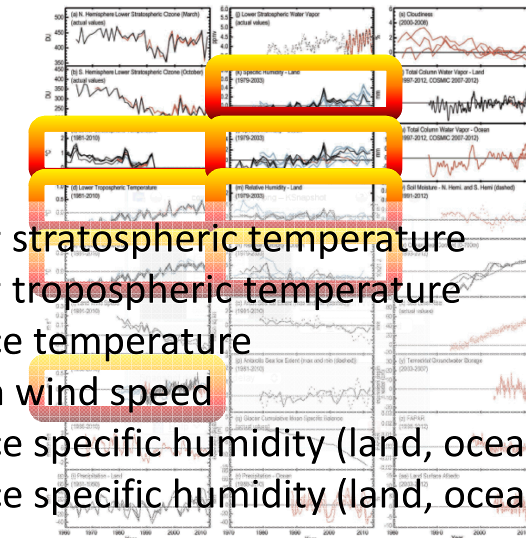
Lower stratospheric temperature
Lower tropospheric temperature
Surface specific humidity

BAMS State of the Climate in 2011

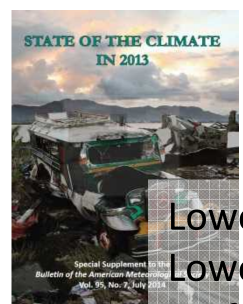
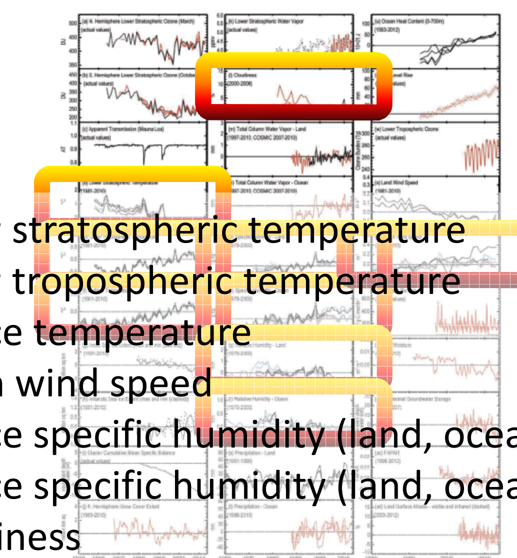
Lower stratospheric temperature
Lower tropospheric temperature
Surface specific humidity
Surface relative humidity
Total aerosol optical depth

BAMS State of the Climate in 2012

Lower stratospheric temperature
Lower tropospheric temperature
Surface temperature
Ocean wind speed
Surface specific humidity (land, ocean)
Surface specific humidity (land, ocean)

BAMS State of the Climate in 2013

Lower stratospheric temperature
Lower tropospheric temperature
Surface temperature
Ocean wind speed
Surface specific humidity (land, ocean)
Surface specific humidity (land, ocean)
Cloudiness

Reanalysis course outline

What is reanalysis?	• General concepts • Goals of reanalysis
How are reanalyses made?	• Observations • Model • Data assimilation

Background errors | Observation errors | Uncertainties

Reanalysis projects & applications	• Projects • Users • Applications
Conclusions	• Summary • Challenges ahead

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 <u>consistency in time, spanning several years</u>**
- **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**:
  - Publishing uncertainty estimates for the reanalysis products: how will they be used?
  - Bringing in additional or reprocessed observations
  - 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
  - Bridging the gap with climate models

Further reading and on-line material

(some mentioned already in this talk)

- 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 http://www.scidacreview.org/0801/pdf/climate.pdf with contributions from Compo and Whitaker
- Global and regional reanalyses twiki: http://www.reanalyses.org
- 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
- 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, http://www.ecmwf.int/publications/library/do/references/show?id=90833
- 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
- Hersbach *et al.* (2015), “ERA-20CM: a twentieth century atmospheric model ensemble”, *Q. J. R. Meteorol. Soc.* ***in press*** doi: 10.1002/qj.2528ERA-20CM (early on-line release at http://onlinelibrary.wiley.com/doi/10.1002/qj.2528/pdf)
- 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
- D-Day weather: http://www.ecmwf.int/en/about/media-centre/news/2014/ecmwf-revisits-meteorology-d-day-period
- Citizen science for the weather historian or the navy enthusiast: http://www.oldweather.org/
- Observation network movie: https://www.youtube.com/watch?v=NUfdFCHoxHM