

Reanalysis

Data assimilation training course 2019

Dinand Schepers,
Hans Hersbach, Joaquin Muñoz Sabater

Dinand.Schepers@ecmwf.int

A terrific storm at sea

3 February
1899

The New York Times
Published: February 16, 1899
Copyright © The New York Times

TERRIFIC STORMS AT SEA

Steamships from All Quarters Report Extremely Rough Voyages.

ALL MORE OR LESS BATTERED

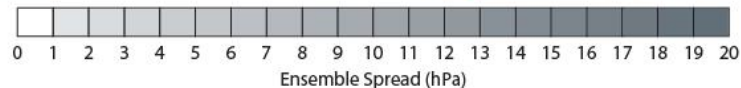
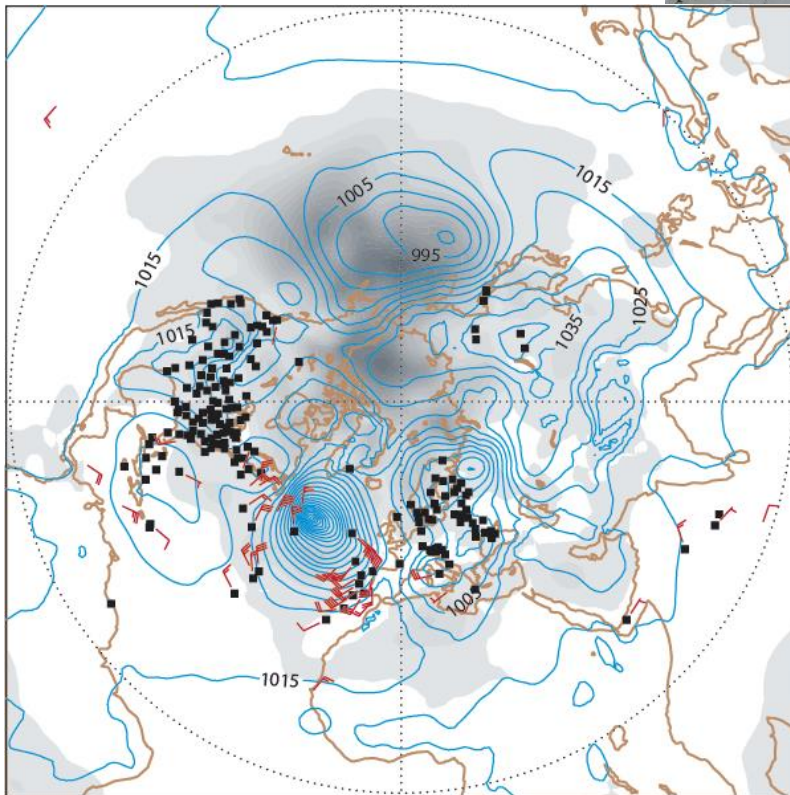
Vessels Sighted in Distress and Abandoned — Blinding Snow and Waves Like Mountains.

All the steamers that came in yesterday were coated with ice from the tops of the masts down to the water line, and all had passed through storms of blinding snow and mountainous waves. The British steamer *Ethelgonda*, from Bristol and Swansea, which left the latter port on Jan. 19, ran into a gale of hurricane force, and seas swept her decks repeatedly. So fierce was the wind that the boat drifted before the gales and was barely able to keep steerage way. She anchored outside the bar late Sunday afternoon. The cable parted and she lost her anchor, together with 100 fathoms of chain. Then the great snowstorm drove her 150 miles off the shore. She succeeded in getting back late on Tuesday night.

The French liner *La Bretagne*, from Havre, came in a little before noon yesterday, with 58 cabin and 225 steerage passengers. She left Havre on Feb. 4, and two days later she encountered a fierce gale, which soon raised tremendous seas. This lasted until Feb. 12, and during this time the vessel had to be slowed down. The following day heavy head seas were encountered, but the steamship proved herself a staunch craft in heavy weather. She had hard work in breaking the ice while getting into her slip.

The German steamer *Livorno* had a succession of heavy gales veering from the southwest to northwest, on Feb. 10, when she was 140 miles off Sandy Hook. She shipped heavy seas, while the snowstorm almost blinded all who were on deck. On Tuesday she sighted a British bark that had lost several sails. It was arriving very

“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.”



Outline

What is reanalysis?

- General concepts
- Goals of reanalysis

How are reanalyses made?

- Observations
- Model
- Data assimilation

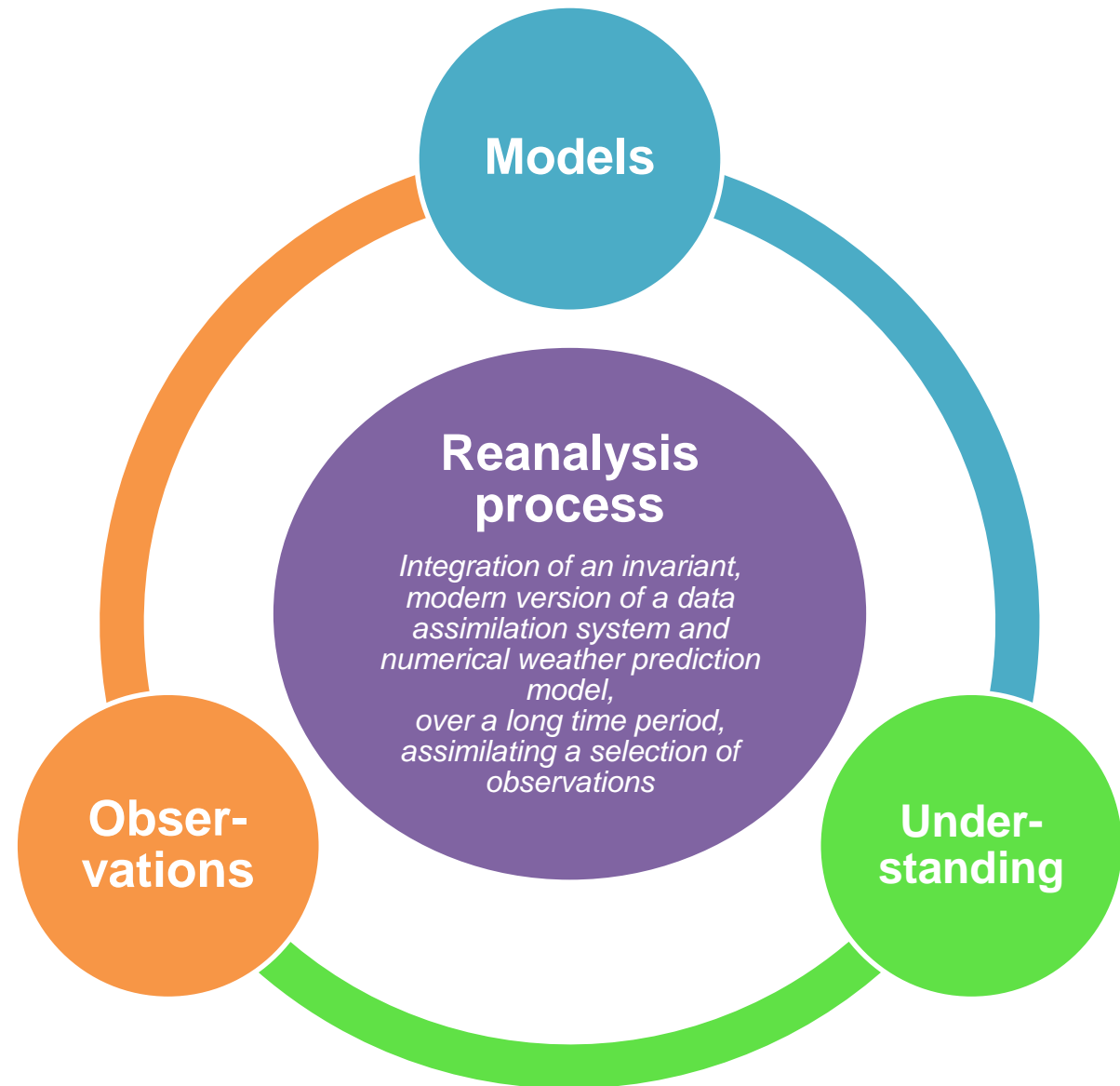
Reanalysis products & applications

- Types of reanalyses
- Applications

Conclusions

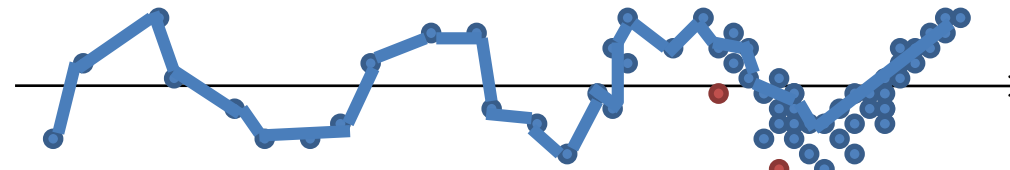
- Summary of concepts
- Challenges ahead

A broad view on reanalysis



Why not use simply observations?

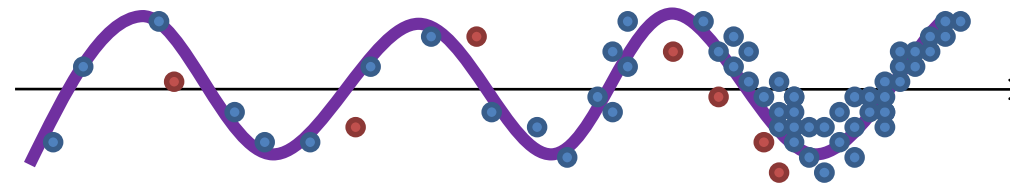
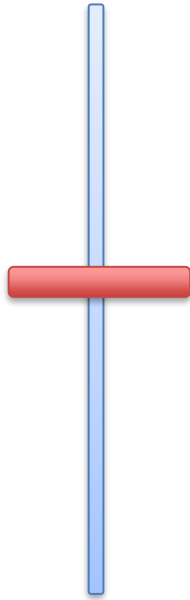
“Observations-only”
climatology



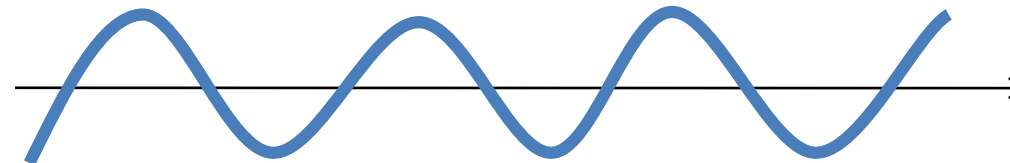
Gross exaggeration towards discontinuity

“outliers”

Reanalysis

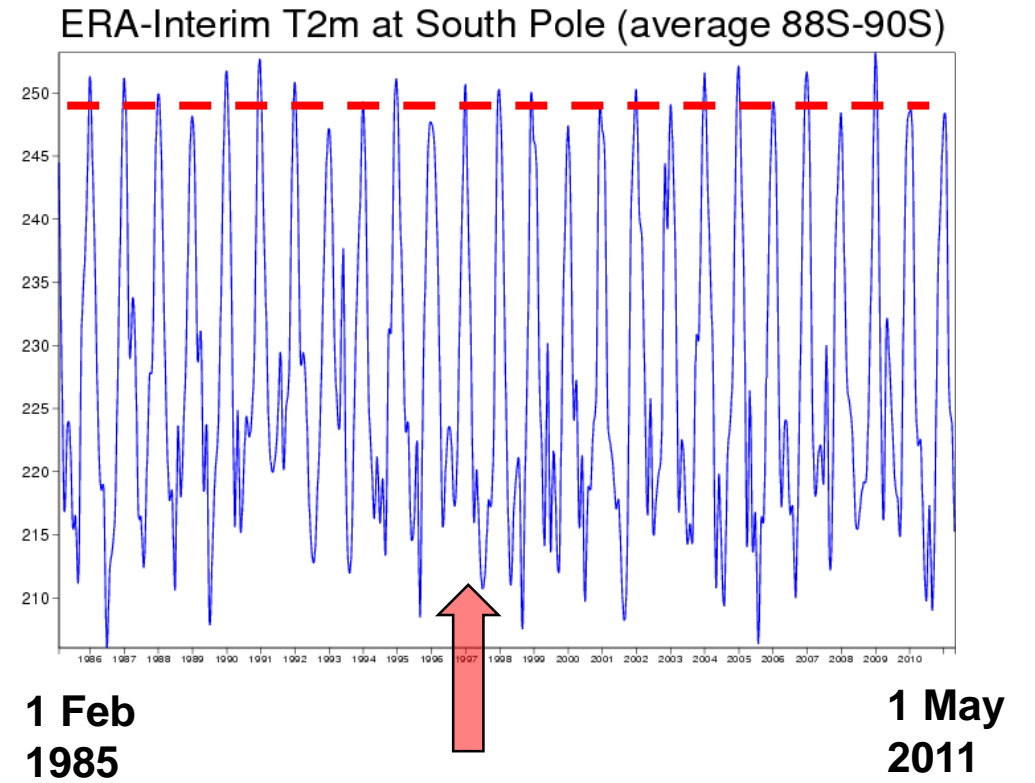
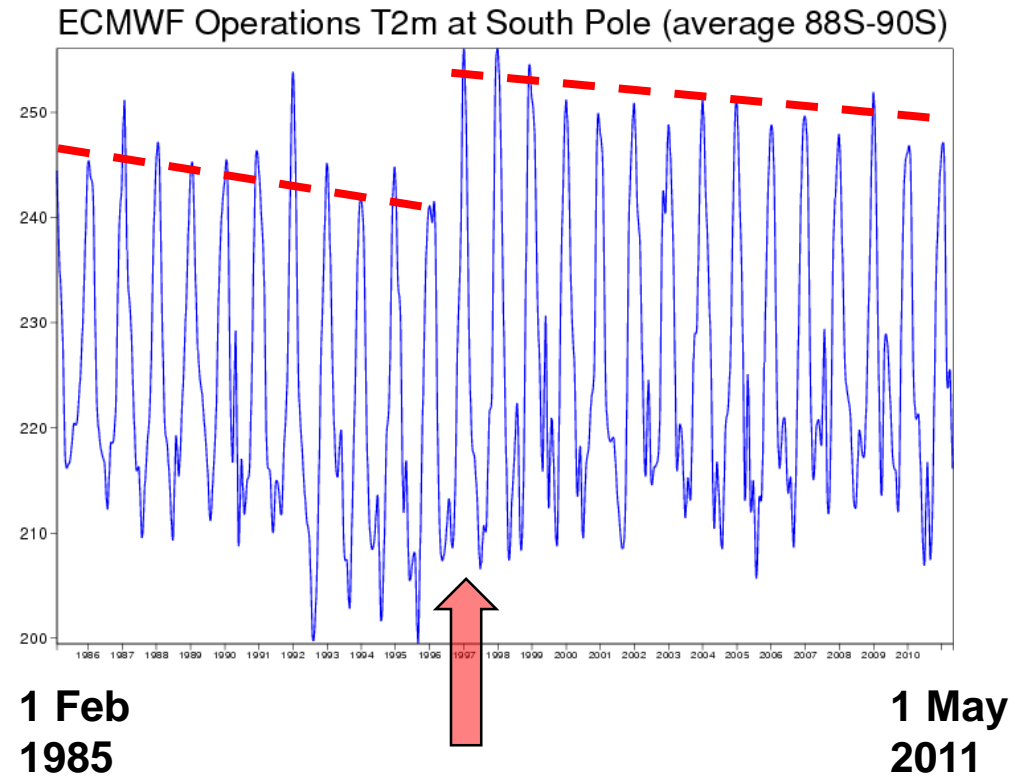


“Model only”
integration



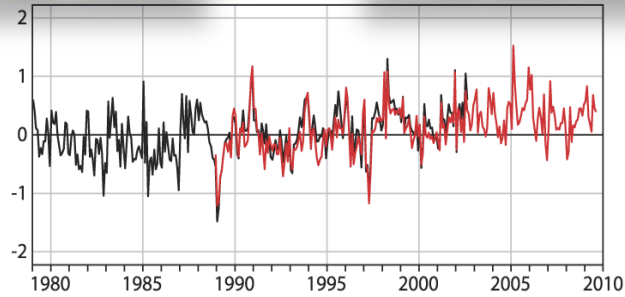
Gross exaggeration towards continuity

Why not use simply operational NWP?



A consistent and complete picture of the past atmosphere Earth system

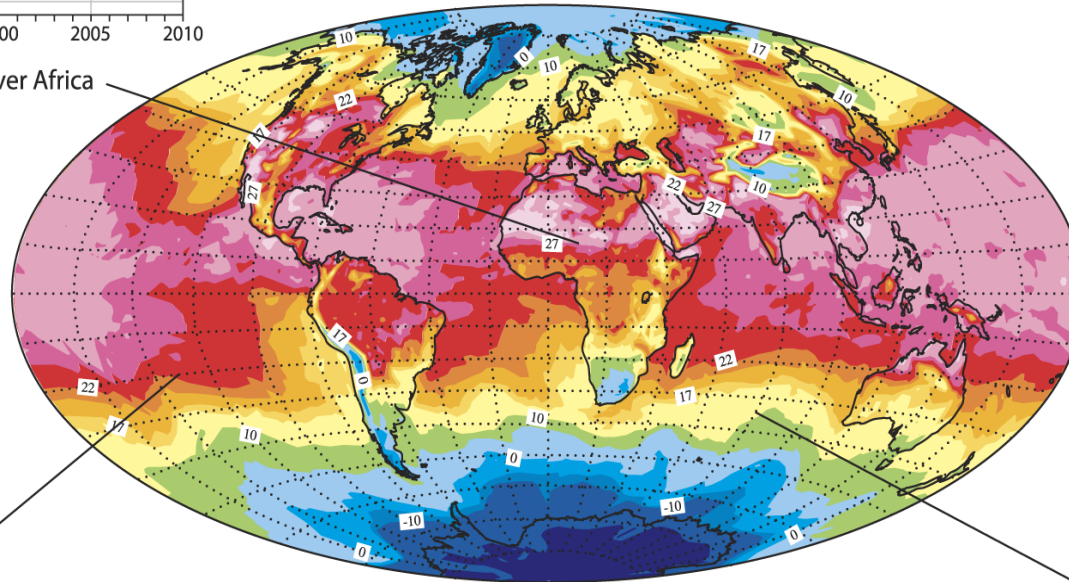
...in Time



2-metre temperature anomaly (°C) over Africa

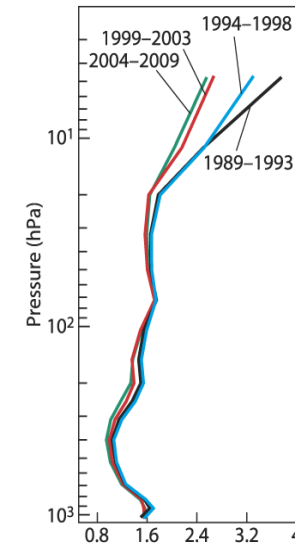
...in the Horizontal

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



**...across
Atmospheric
Parameters**

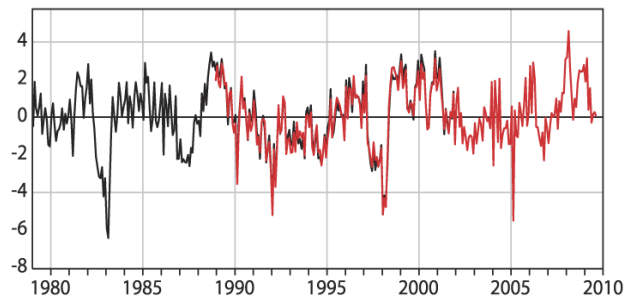
...in the Vertical



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

... across domains

Southern Oscillation Index (hPa)



Two classes of Reanalysis

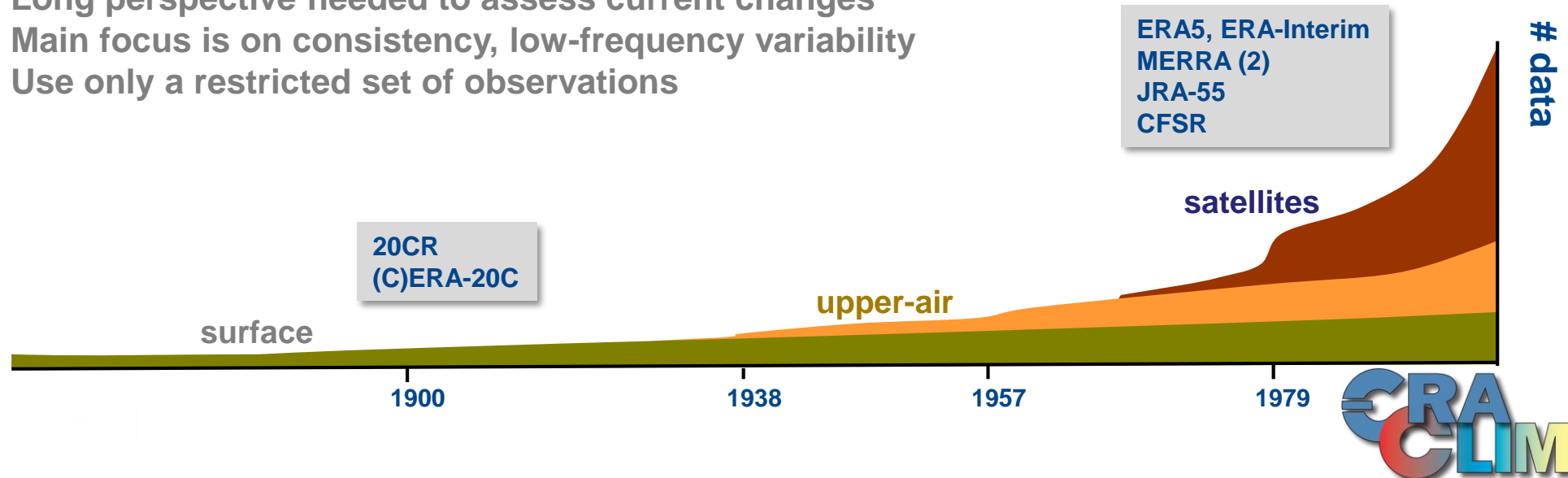
Reanalyses of the modern observing period (~30-50 years):

- 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

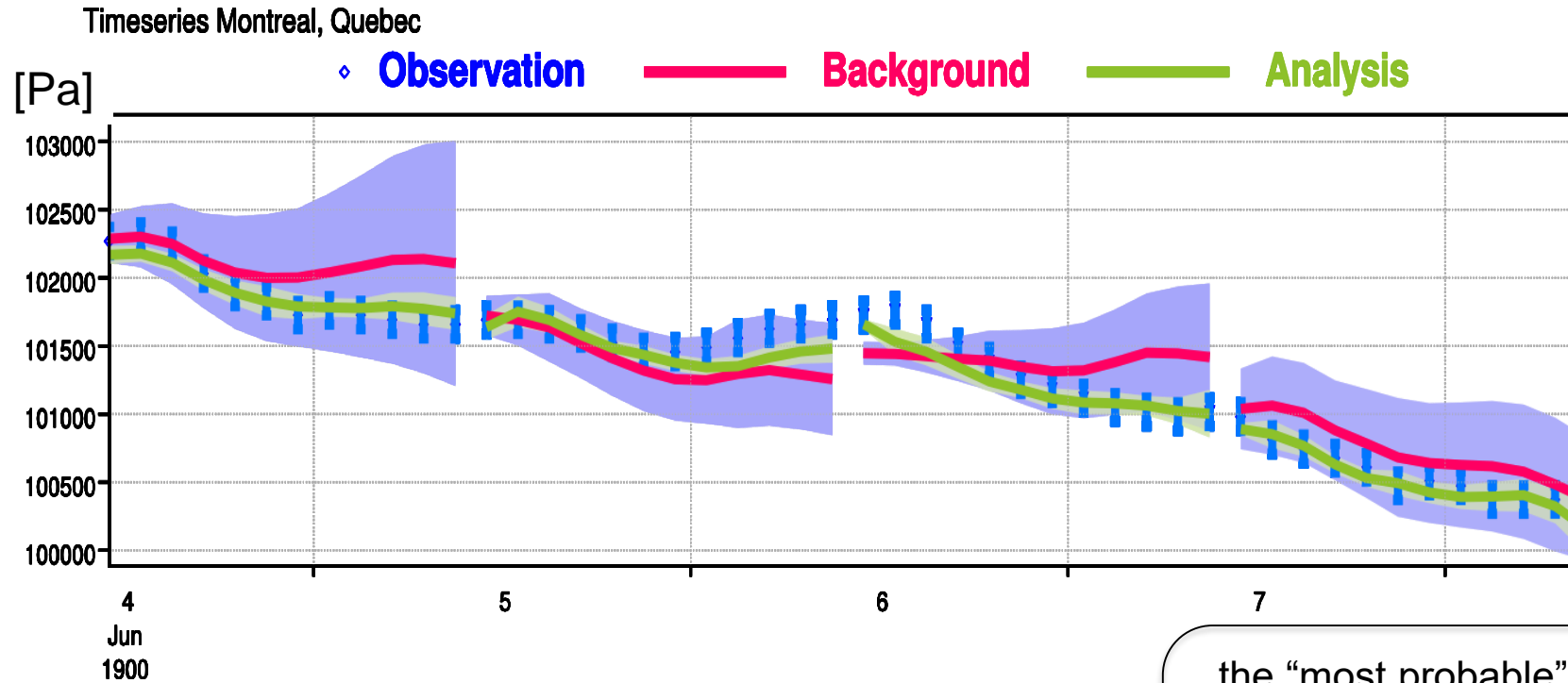
Reanalysis products & applications

- Users
- Applications

Conclusions

- Summary of concepts
- Challenges ahead

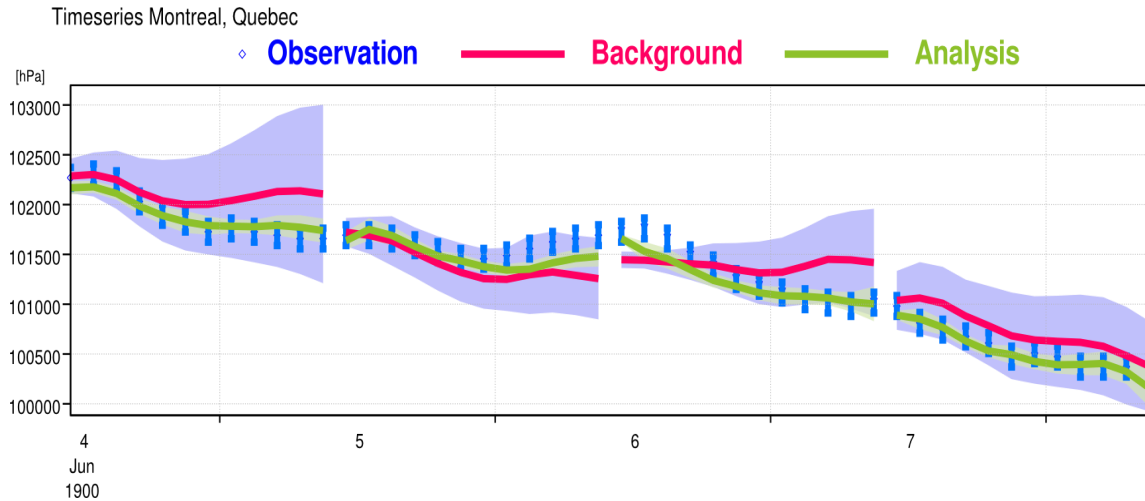
Constructing a history of the past with 4DVAR data assimilation



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

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

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

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

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

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

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

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

Bias correction schemes

Bias model b and coefficients may need expanding to tackle known biases in observations

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

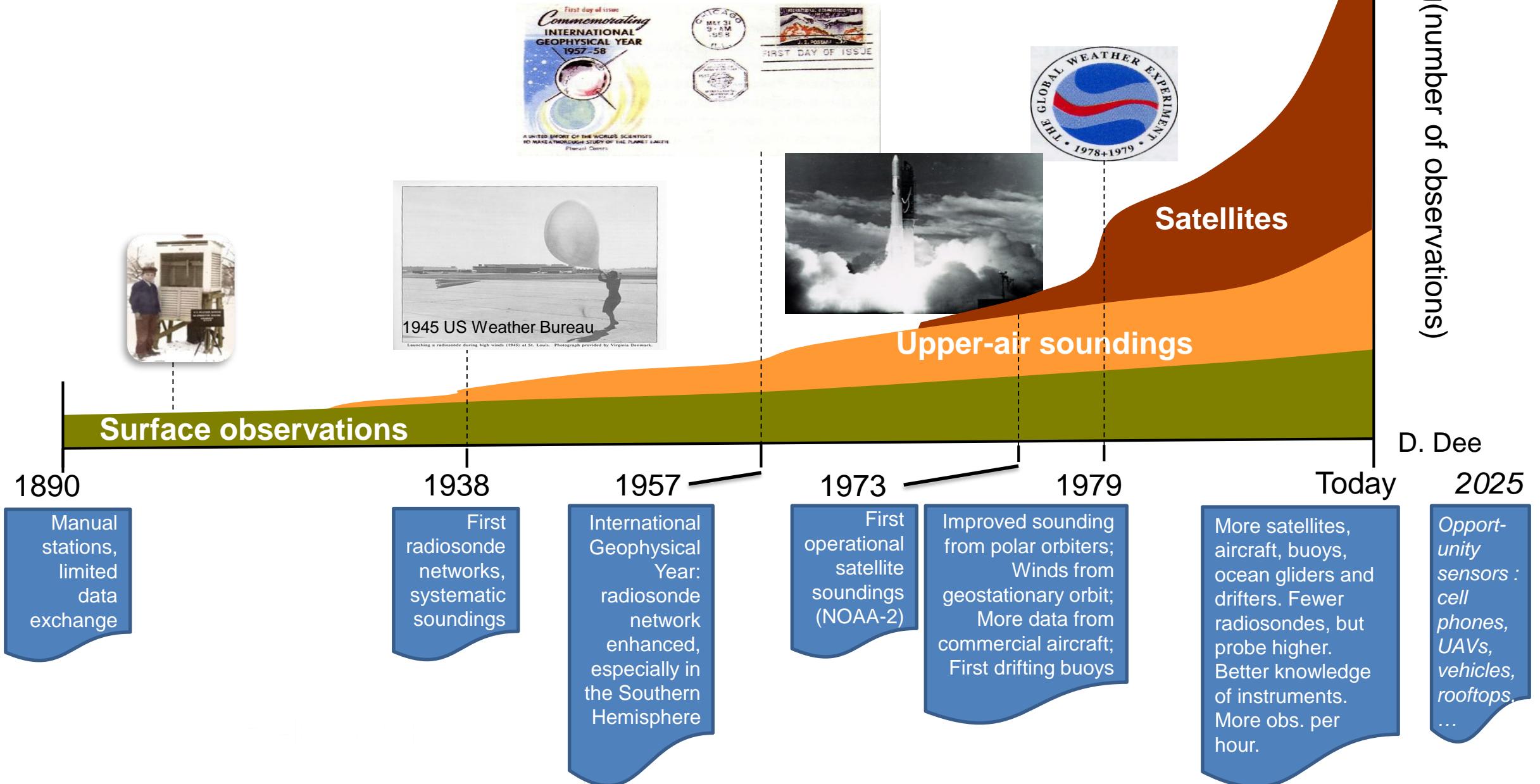
Keep track of what goes in/comes out

- 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



log(number of observations)

D. Dee

2025

Opportunity sensors: cell phones, UAVs, vehicles, rooftops, ...

Satellites

Upper-air soundings

Surface observations

1890

1938

1957

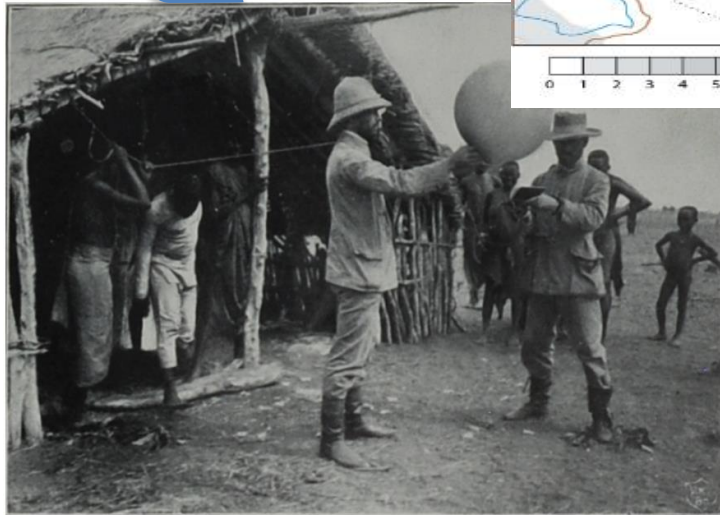
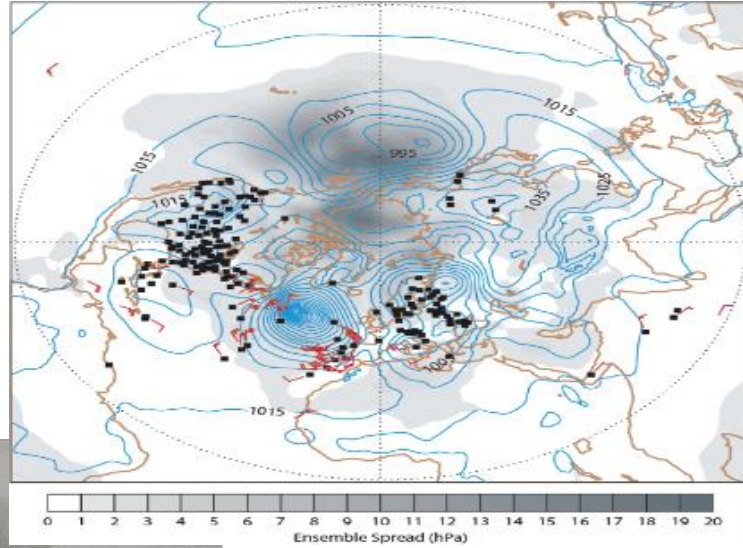
1973

1979

Today



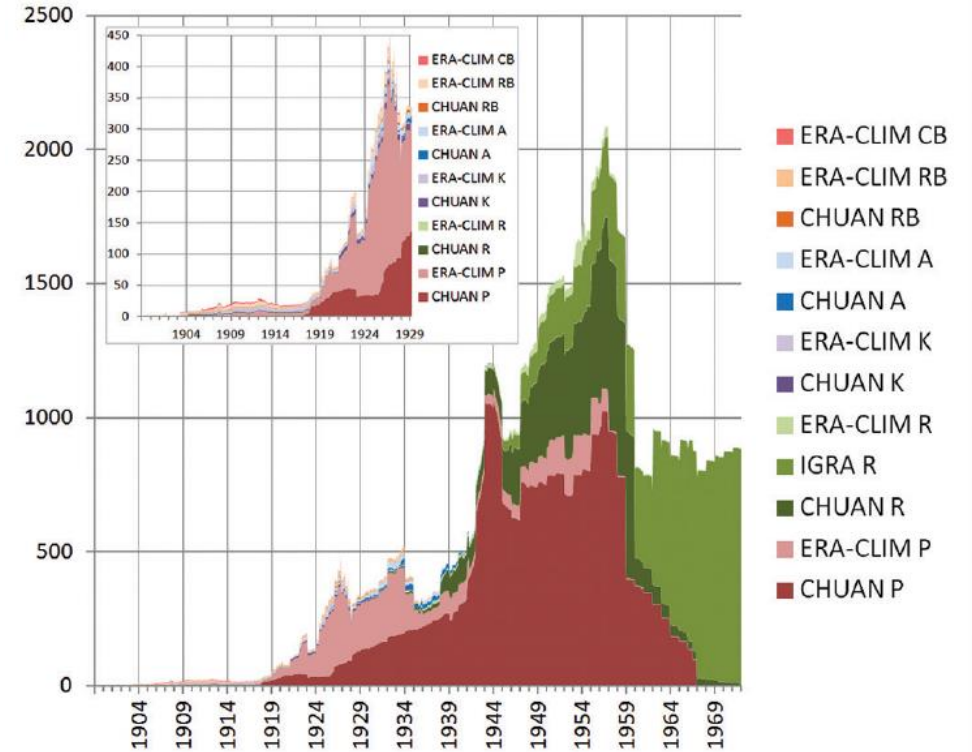
Data rescue



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



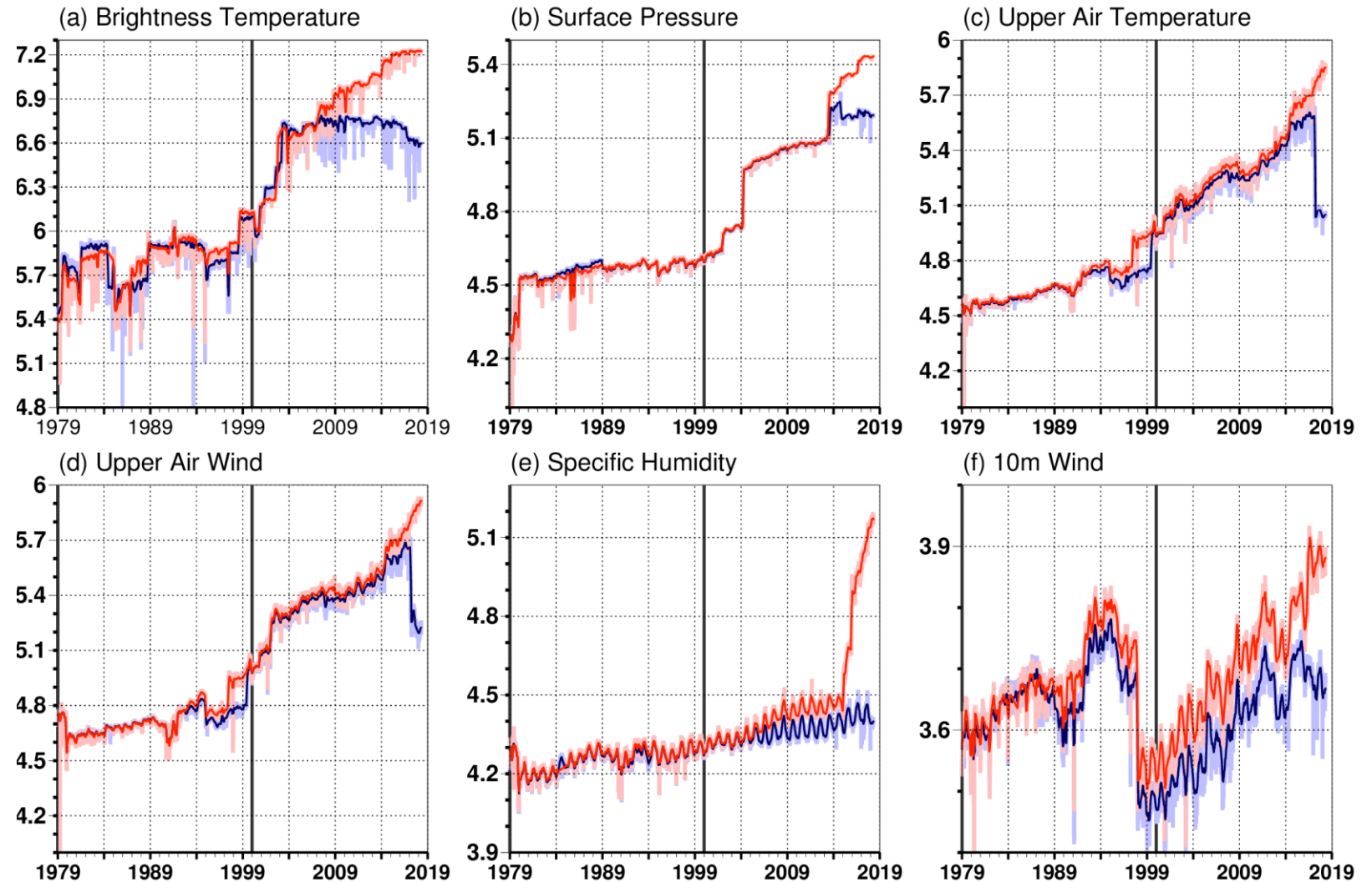
Stickler et al., 2014 : "ERA-CLIM: Historical Surface and Upper-Air Data for Future Reanalyses." *Bulletin of the American Meteorological Society*

ERA5 data usage compared to ERA-Interim

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)



Number of used observations per day (10log scale) for **ERA5** and **ERA-Interim**

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

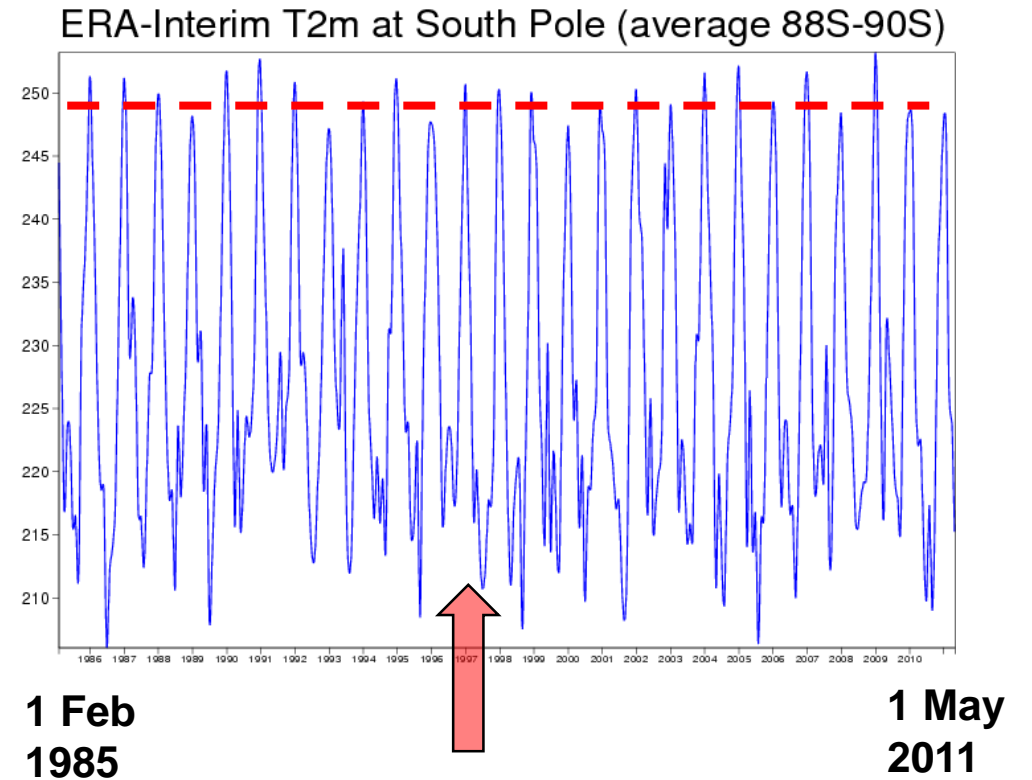
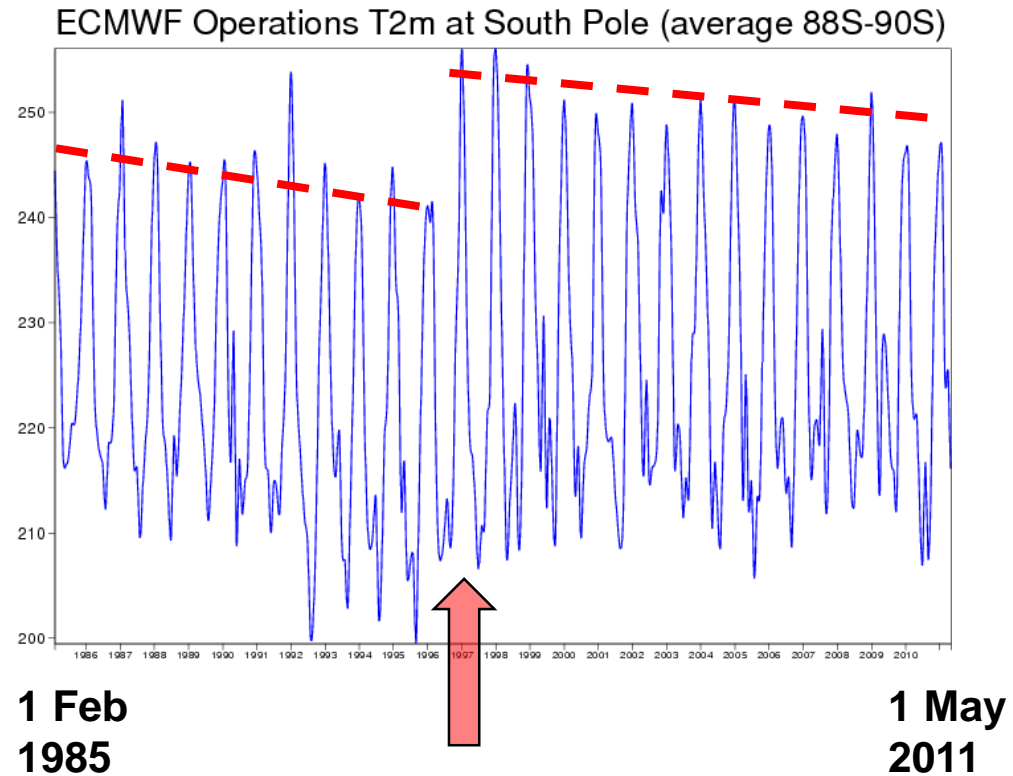
Shop around for forcing data

- 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?



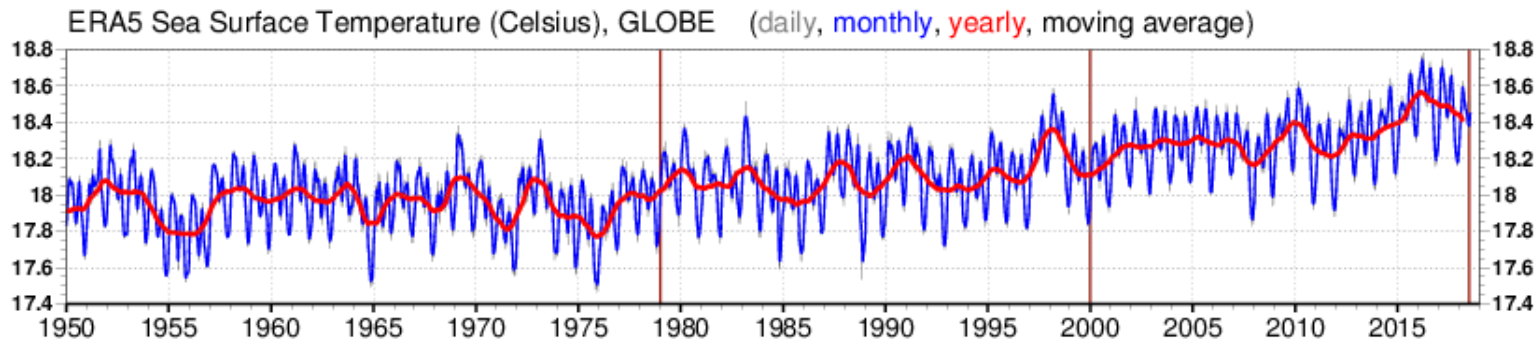
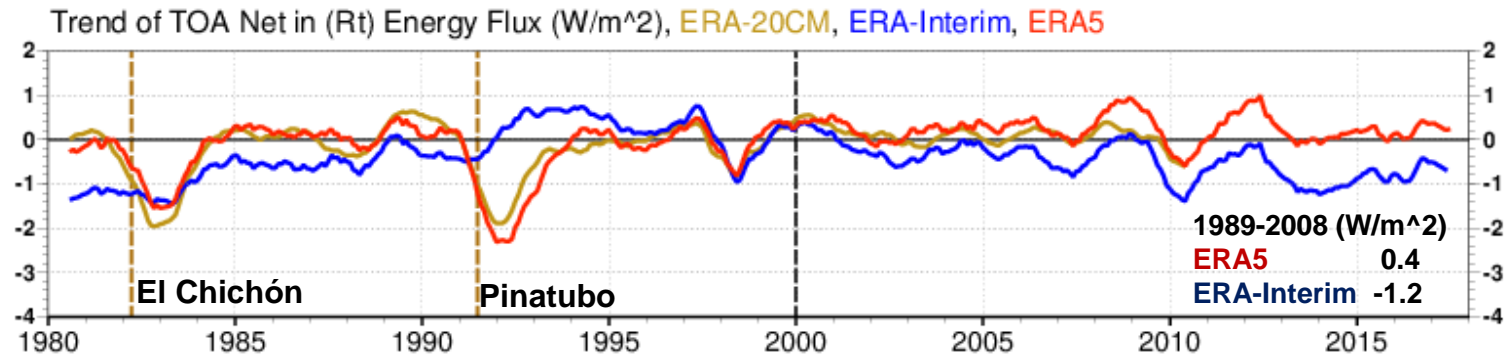
Models are essential tools to propagate the information and ensure consistency (over short) time scales between geophysical variables. The advances in models and in data assimilation help deliver improved products

Forcings appropriate for climate

CMIP5 recommended data sets

Total solar irradiance, greenhouse gases, ozone, aerosols (including volcanic)

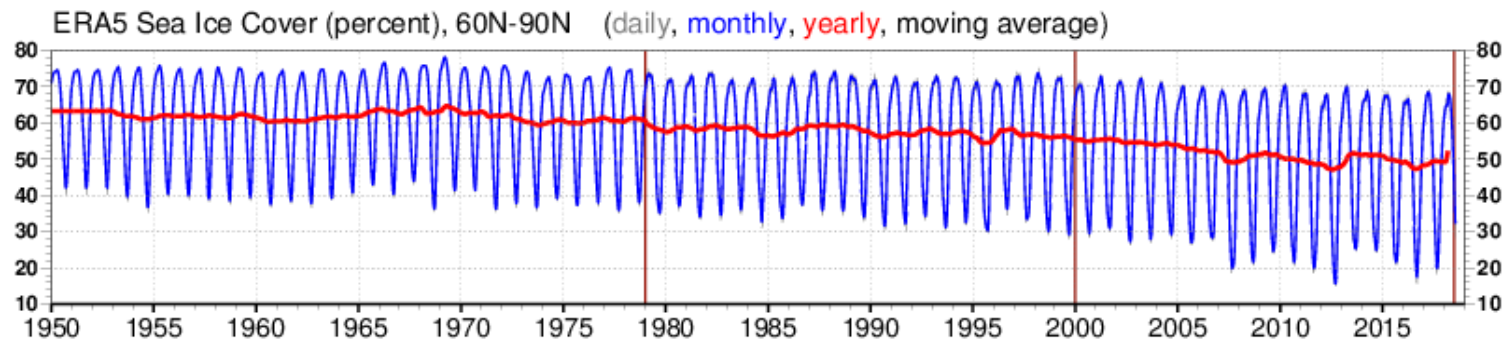
(Prepared in the ERA-CLIM project, *ERA-20CM*, *Hersbach et. al., 2015*)



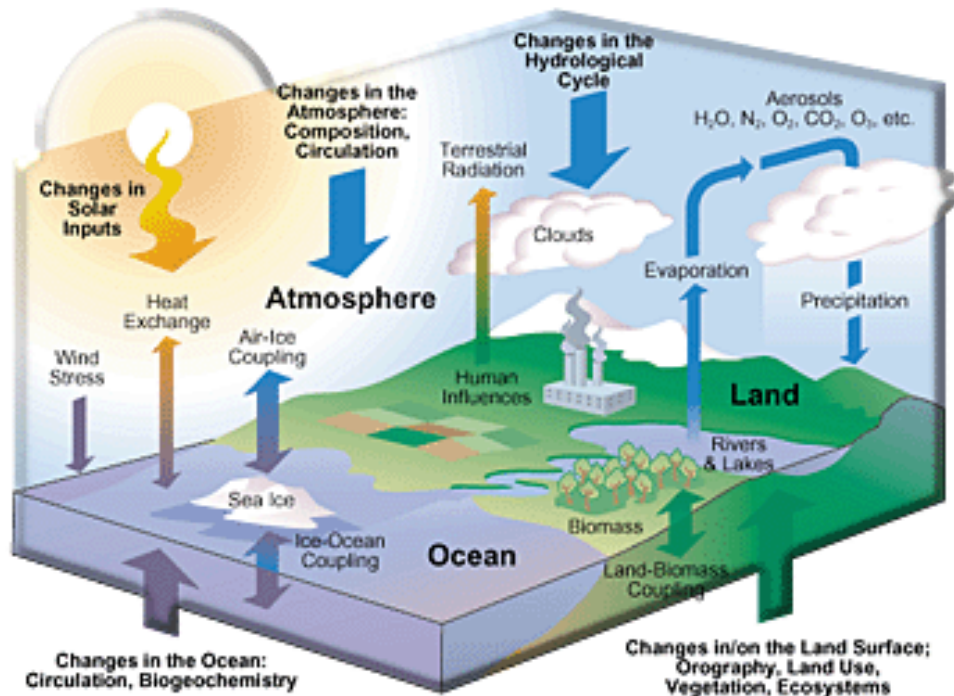
SST and sea ice cover

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

(*Hirahara et. al., 2016*)



Coupled processes



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**

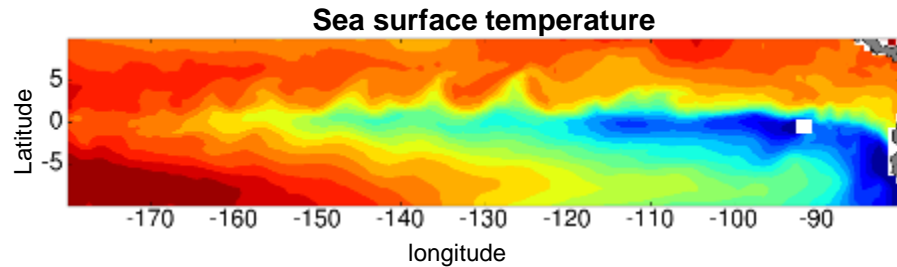
CERA-20C a 20th century reanalysis using an ocean-atmosphere coupled model and DA.

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)

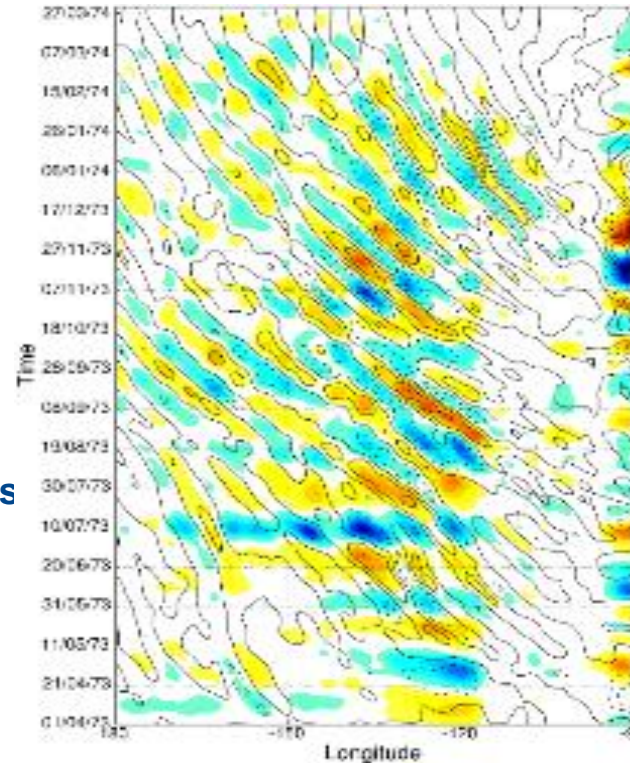
- represents TIWs thanks to the ocean dynamics
- atmosphere responds accordingly (surface wind stress is sensitive to the ocean TIW)

ERA20C (Forced reanalysis)

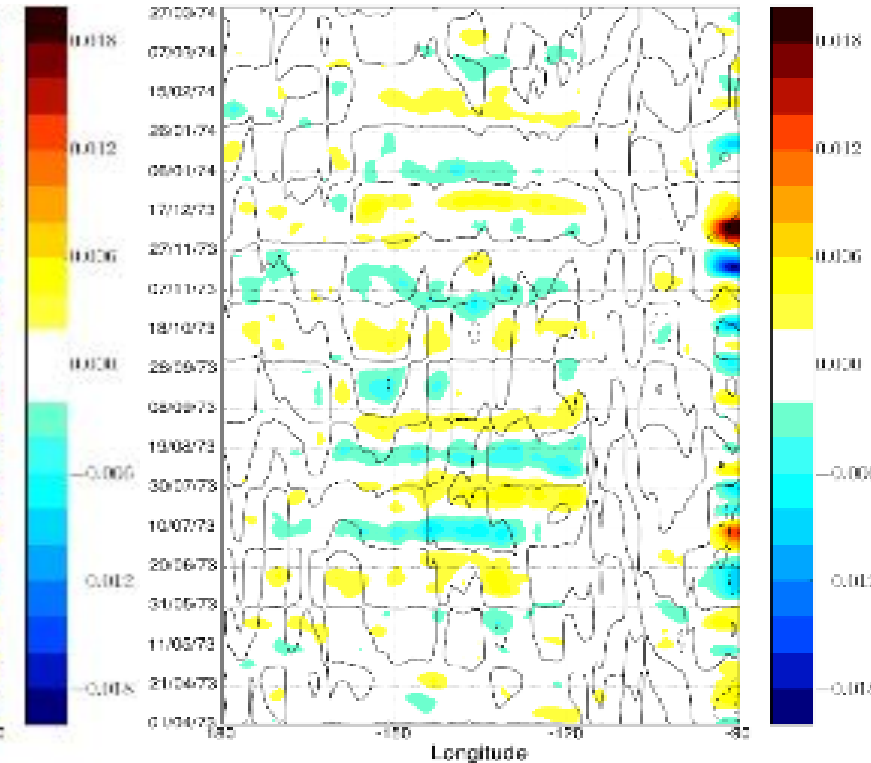
- no TIWs or wind stress signals (forced by monthly SST)

high-pass filtered SST (colour) and wind stress (contour)

CERA-20C



ERA-20C



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

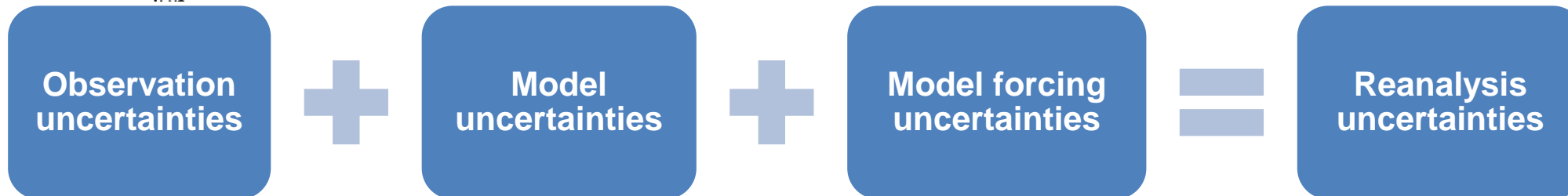
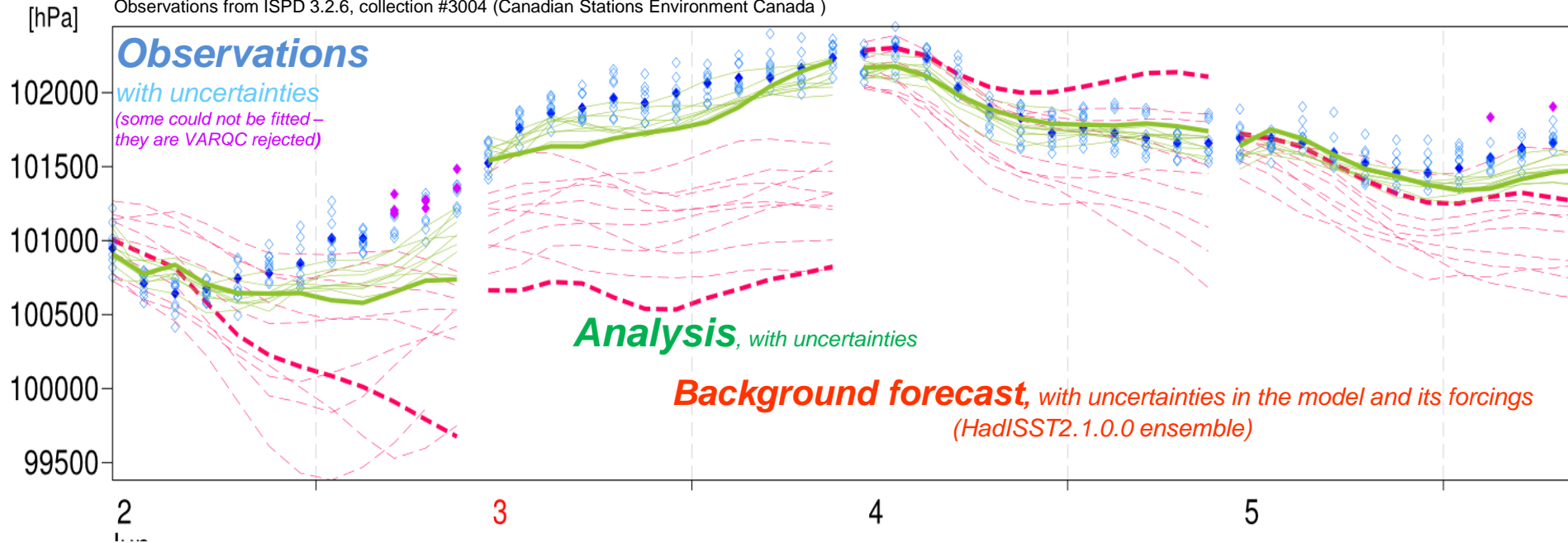
- Homework to find out Gross errors, Biases, and Random errors (std. dev. = specified as 'observation errors')

Keep that setup and monitor it

- Be extra careful during run-time etc...
- Implement automated monitoring for all the key steps of the assimilation

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

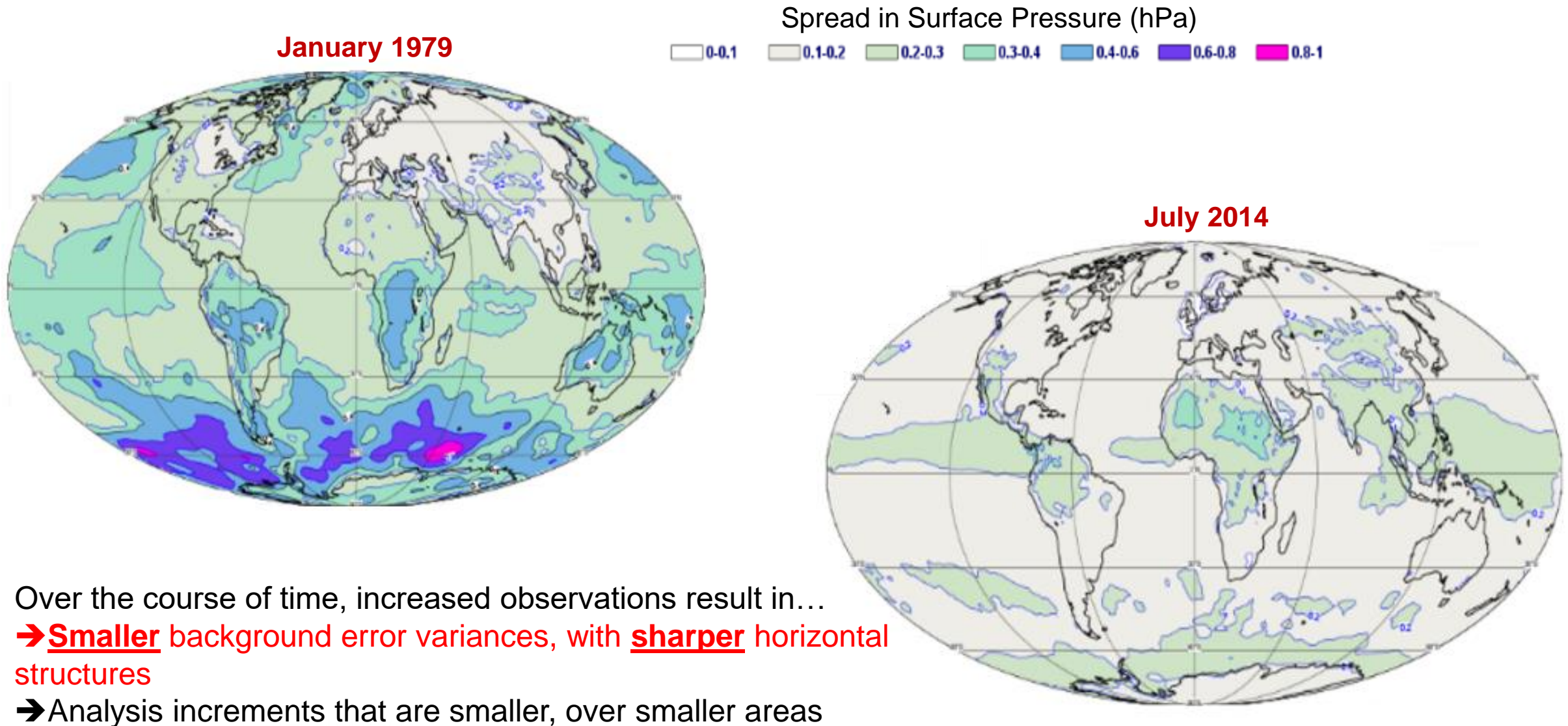
Surface pressure at Montreal, Quebec
Observations from ISPD 3.2.6, collection #3004 (Canadian Stations Environment Canada)



Aims:

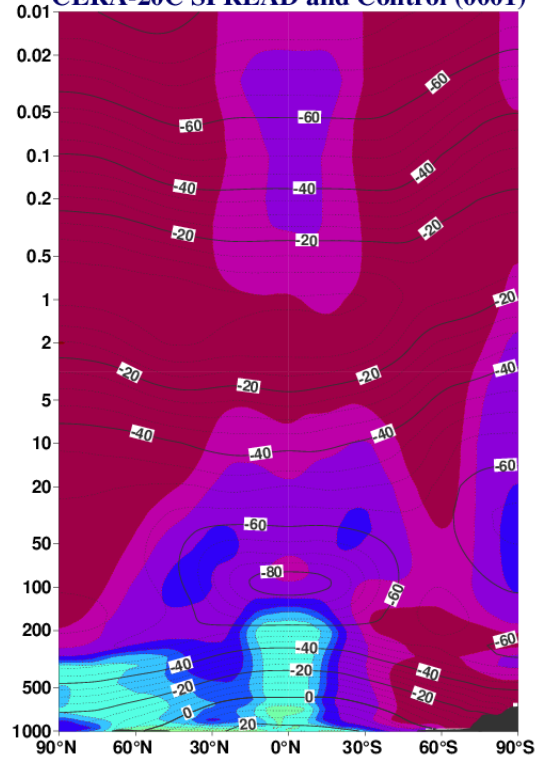
1. Estimate automatically our background errors, and update them
2. Provide users with *some* uncertainties estimates

ERA5 – Ensemble based uncertainty estimate



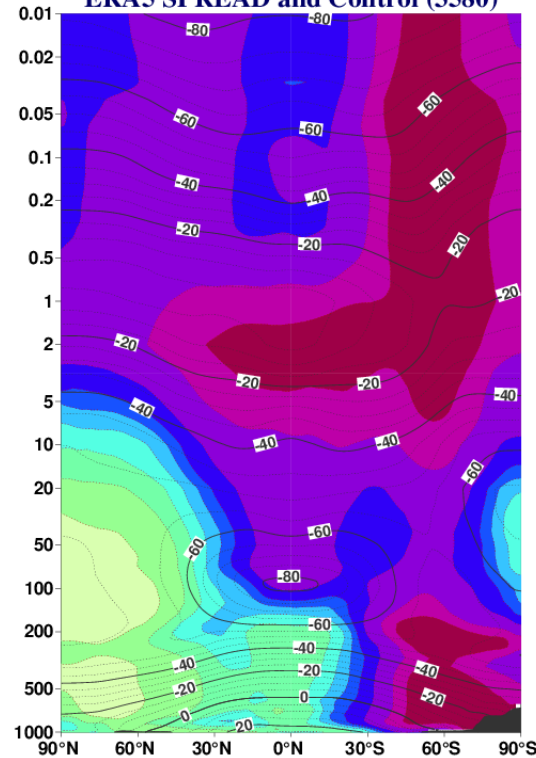
Ensemble spread as a proxy for the background error

Temperature (Celsius) in MAM 1971
CERA-20C SPREAD and Control (0001)



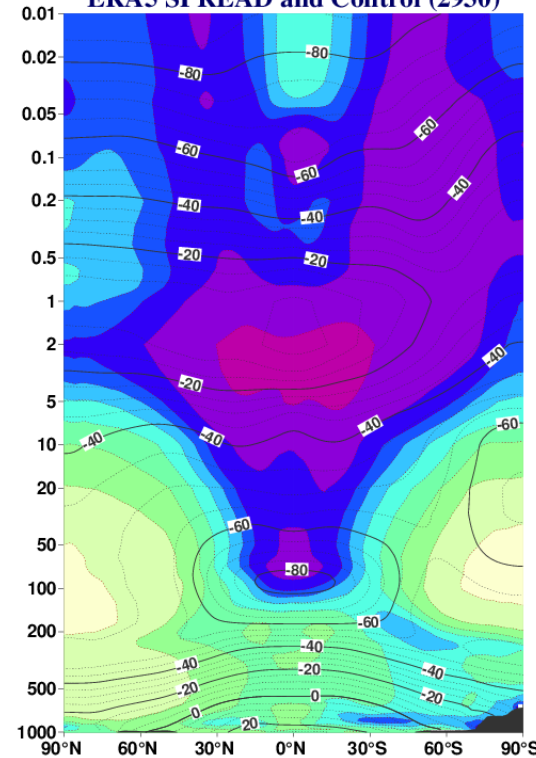
1971 CERA-20C:
Surface pressure,
marine wind, only

Temperature (Celsius) in MAM 1971
ERA5 SPREAD and Control (3580)



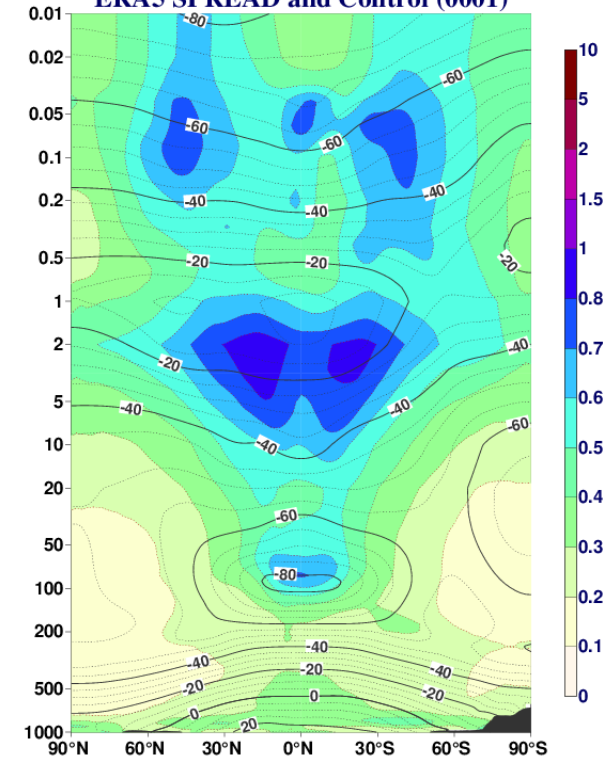
1971 ERA5:
Upper-air data

Temperature (Celsius) in MAM 1980
ERA5 SPREAD and Control (2930)



1980 ERA5:
Early-satellite era

Temperature (Celsius) in MAM 2018
ERA5 SPREAD and Control (0001)



2018 ERA5:
Current observing
system

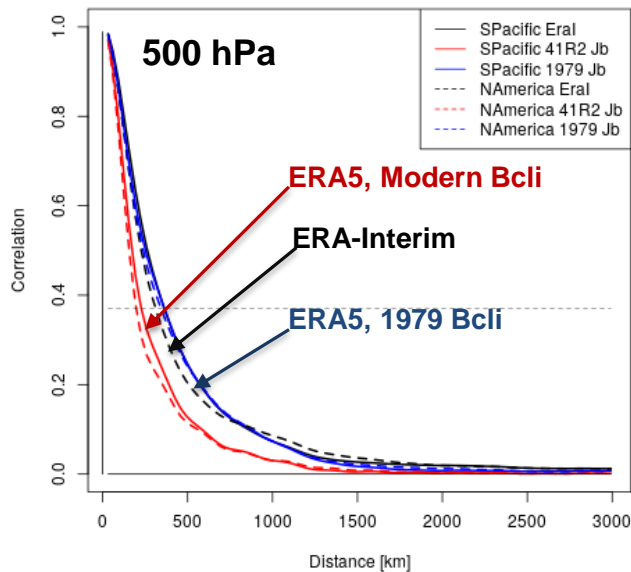
Long-term evolution of the background covariance matrix

For early decades correlation lengths for the static part of **B** appear too short.

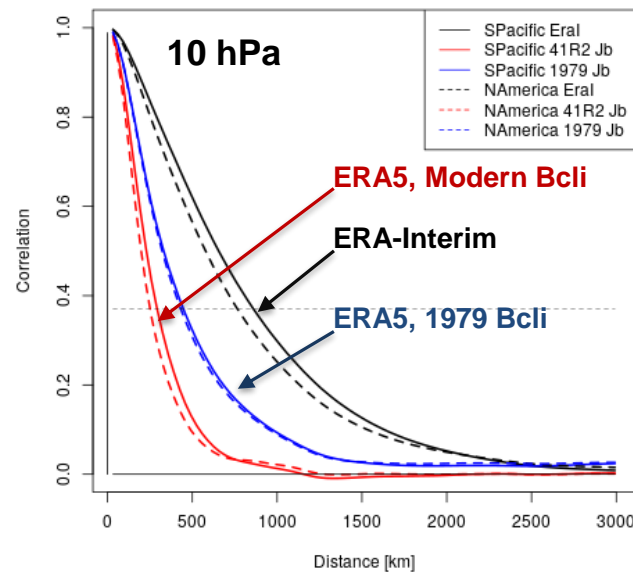
Solution: 1) build dedicated **B_{cli}** for distinct periods in time

However, this risks introducing discontinuities in the reanalysis

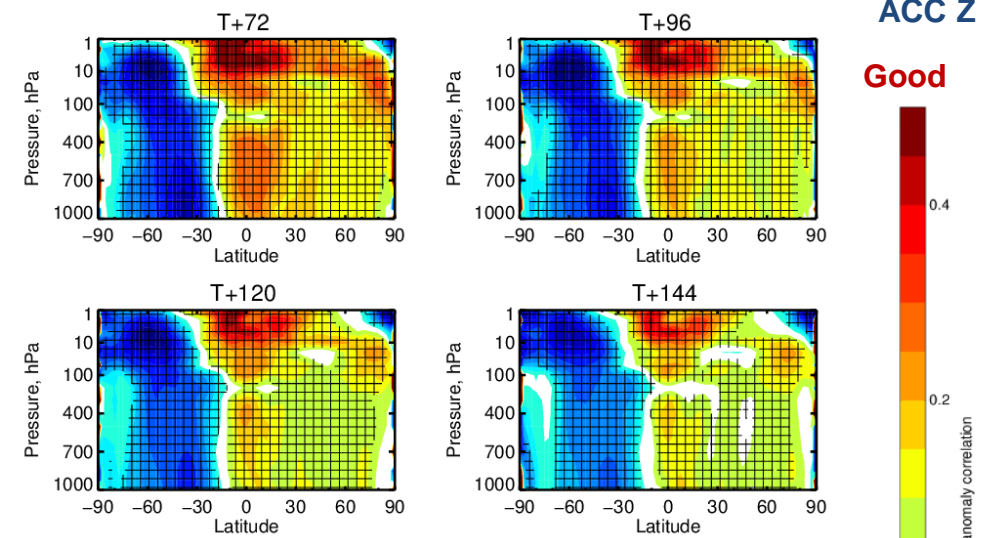
Horizontal Correlation of Temperature at 500hPa



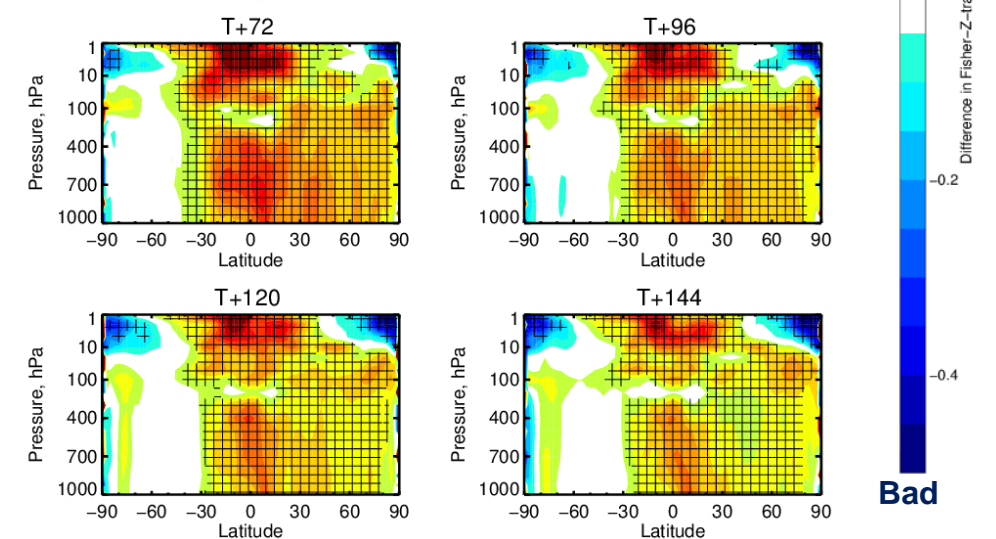
Horizontal Correlation of Temperature at 10hPa



ERA5 scout, Modern B_{cli} vs ERA-Interim for 1979

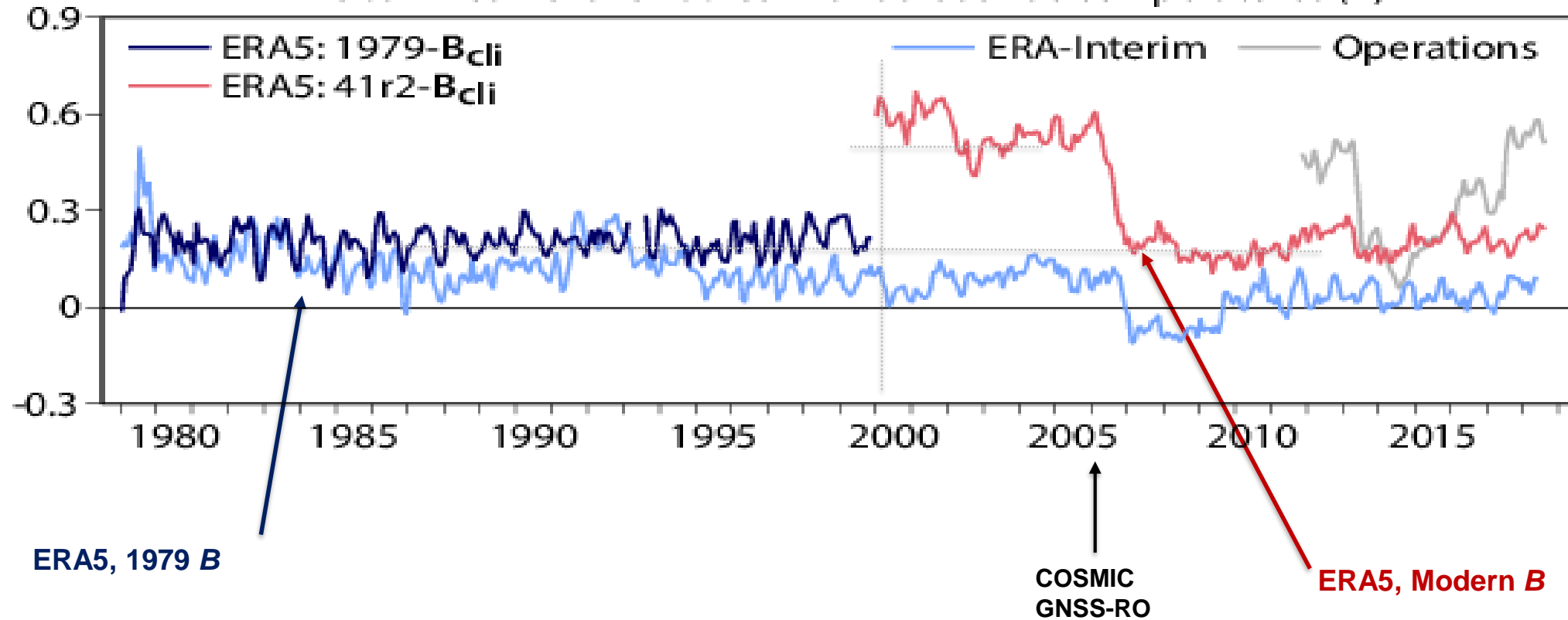


ERA5 scout, 1979 B_{cli} vs ERA-Interim for 1979



Stratospheric temperature (model) biases

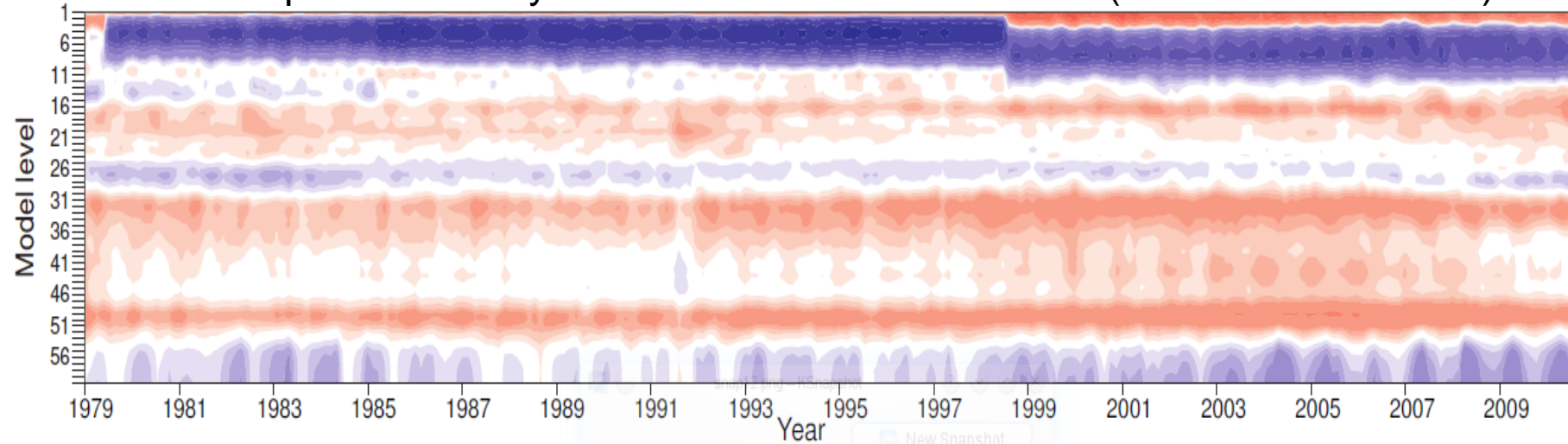
Global-mean o-b for 60-85hPa radiosonde temperatures (K)



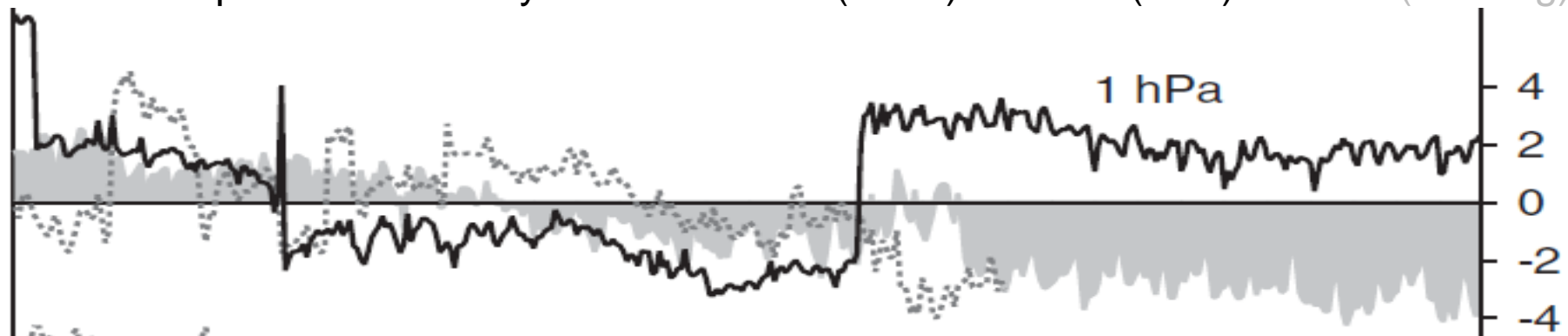
The transition from the **1979-B** the **ERA5 modern B** in 2000 introduces a **seam** in the stratosphere. Extended usage of the **1979-B** till the availability of **COSMIC GNSS-RO** could have alleviated this.

Another look at uncertainties: Analysis increments and trends

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



Mean temperature anomaly in ERA-Interim (black) ERA-40 (dots) JRA-55 (shading) at 1 hPa



Outline

What is reanalysis?

- General concepts
- Goals of reanalysis

How are reanalyses made?

- Observations
- Model
- Data assimilation

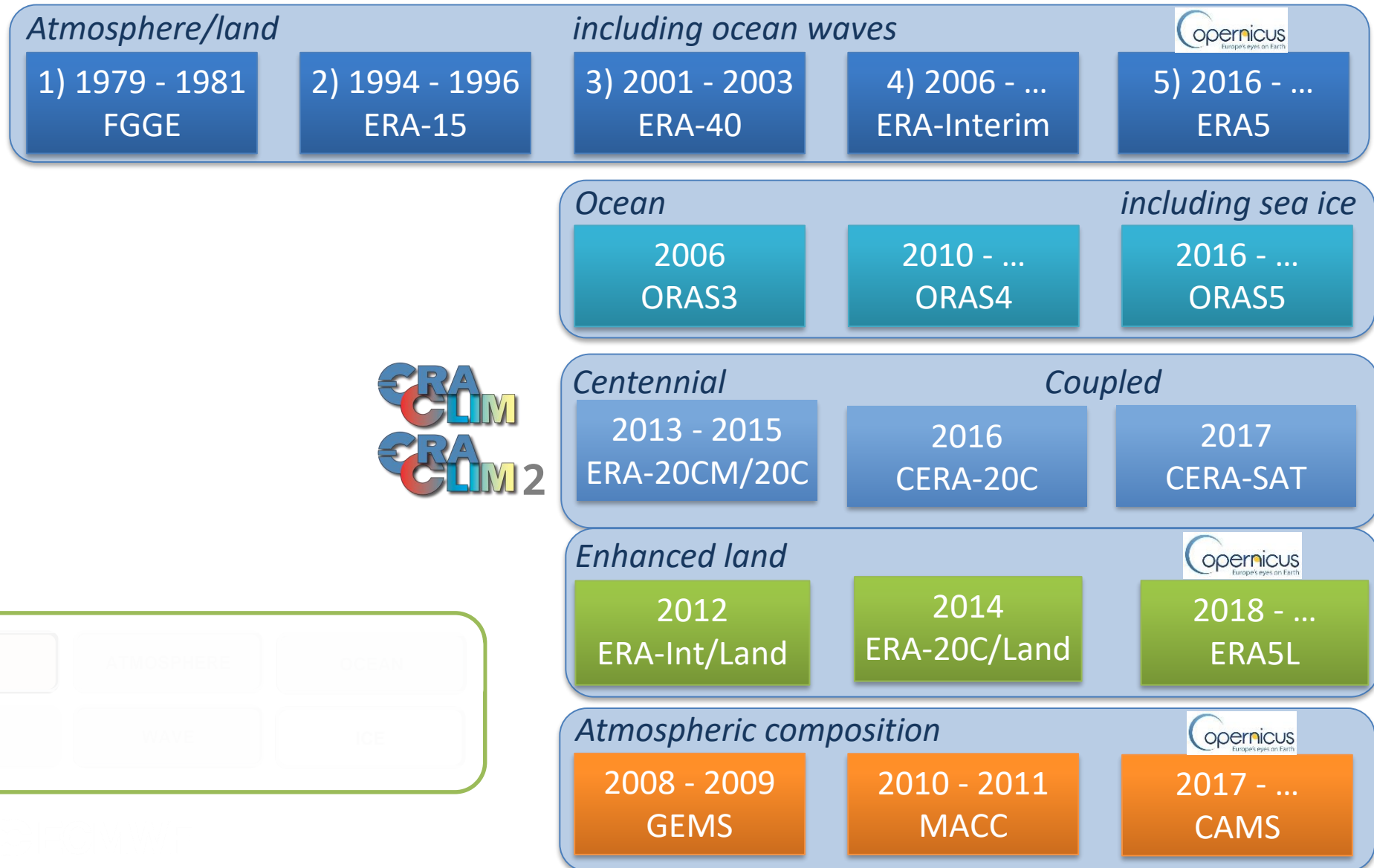
Reanalysis products & applications

- Users
- Applications

Conclusions

- Summary of concepts
- Challenges ahead

The ECMWF reanalysis landscape



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

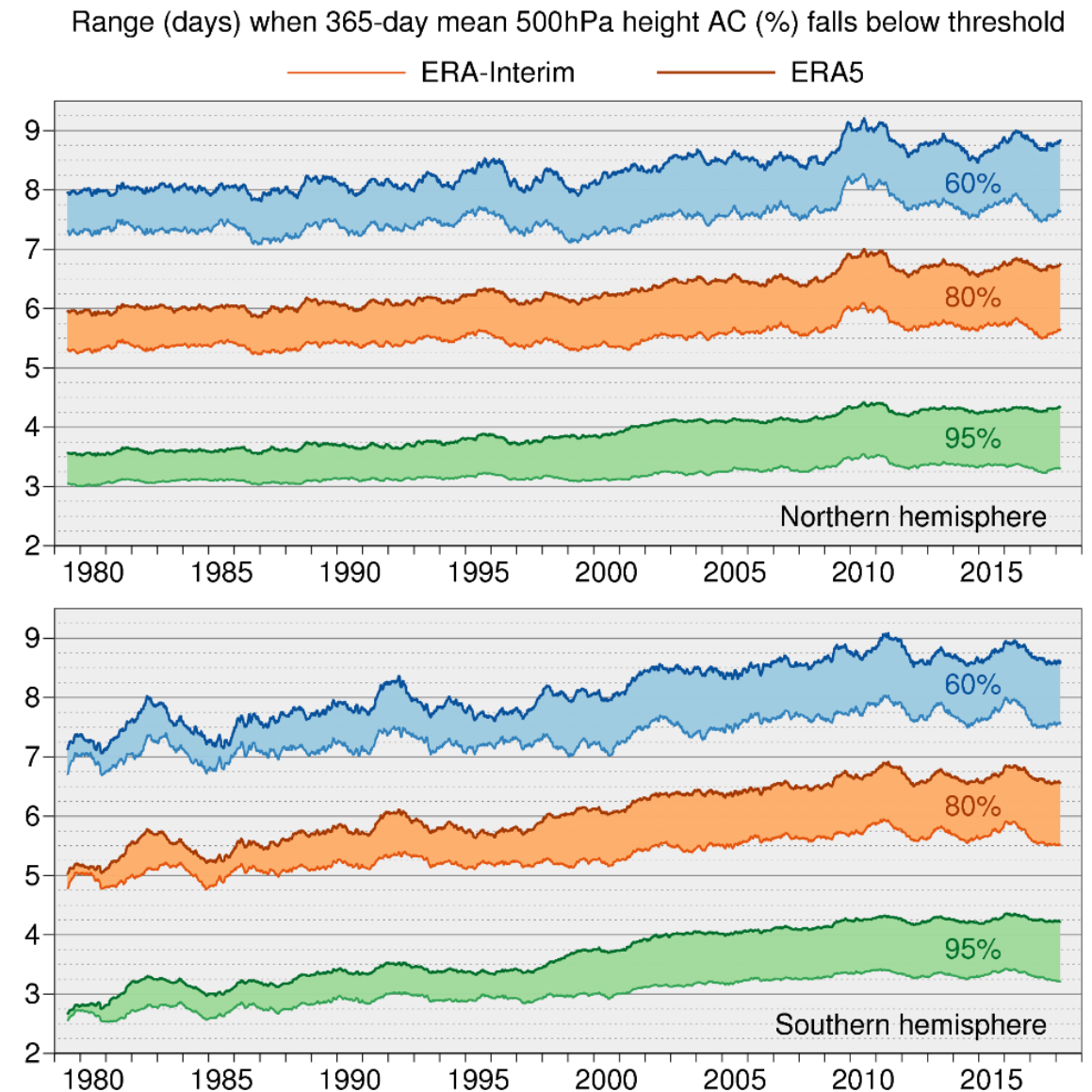
Phase 2: production of 1950-1978 has recently started

Publicly available in the C3S climate data store:

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

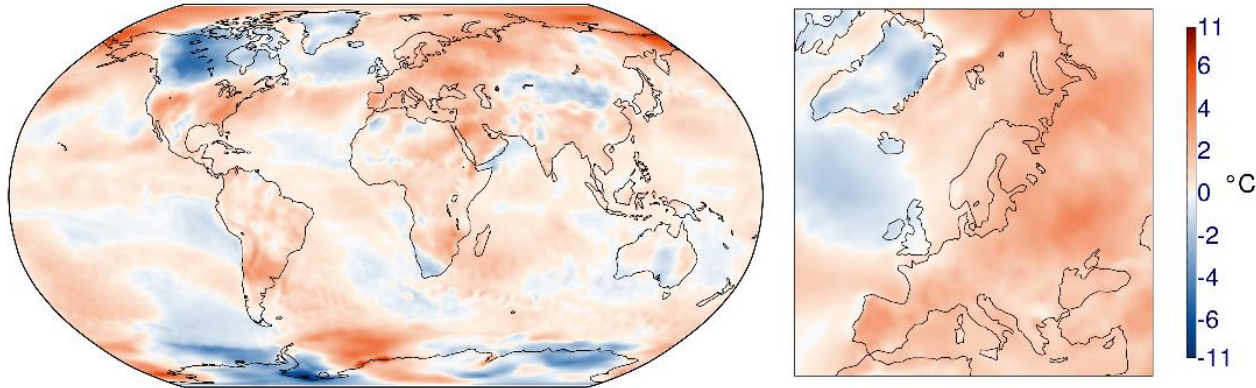
Soon: ERA5T, 2-5 days behind real time

Note: ERA-Interim production will not be supported after mid 2019



Usage of Reanalysis and its importance at ECMWF

Surface air temperature anomaly for September 2018 relative to 1981-2010

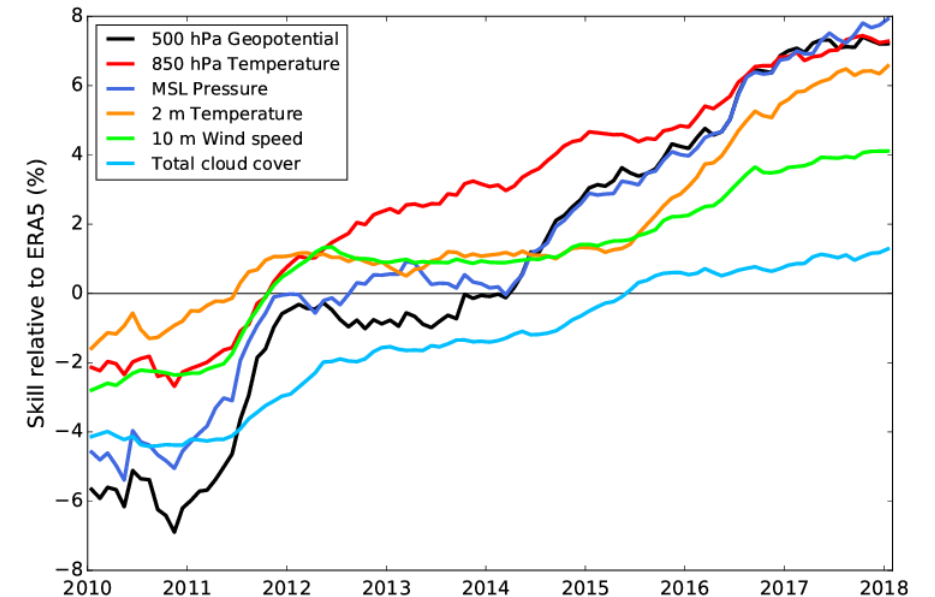


Reanalysis provides consistent “maps without gaps”.
Its role in climate monitoring is now widely recognized.

ERA-Interim has more than **30,000 users** worldwide

Importance at ECMWF:

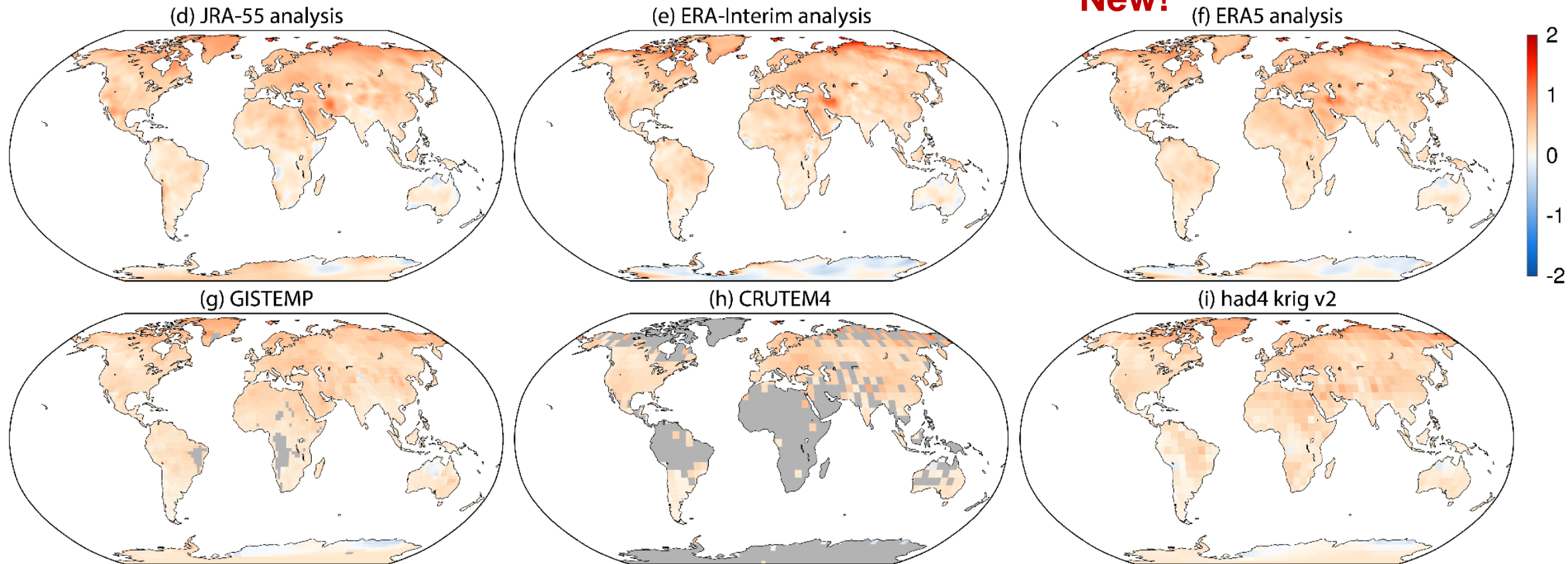
- Input to *timely* monthly C3S climate bulletins
- Required for re-forecasts
- Evaluation of progress in forecast skill
- Provision of climatologies:
 - EFI, ACC, probabilistic events
- Benchmark for developments in R&D
- Many more



Linear trend in surface air temperature over land

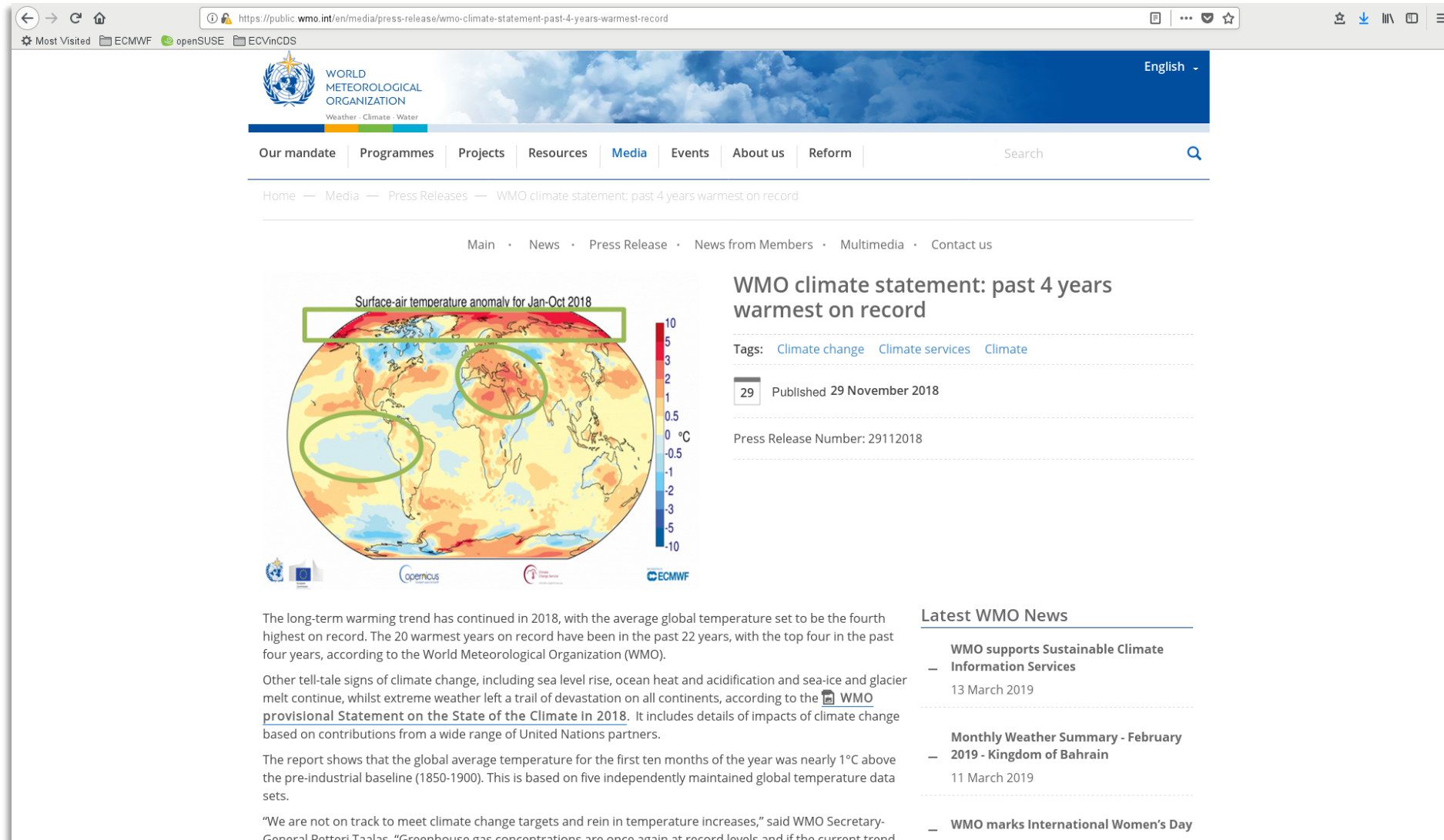
Kelvin/Decade for 1979-2017 (~0.18 globally)

New!



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

Climate monitoring and reporting



The screenshot shows the WMO website's press release page for the title "WMO climate statement: past 4 years warmest on record". The page features a navigation menu with categories like "Our mandate", "Programmes", "Projects", "Resources", "Media", "Events", "About us", and "Reform". A breadcrumb trail indicates the path: Home > Media > Press Releases > WMO climate statement: past 4 years warmest on record. A secondary navigation bar includes links for "Main", "News", "Press Release", "News from Members", "Multimedia", and "Contact us".

The main content area is dominated by a world map titled "Surface-air temperature anomaly for Jan-Oct 2018". The map uses a color scale from -10°C (dark blue) to 10°C (dark red). Significant red and orange areas are visible across the globe, particularly in the Northern Hemisphere, indicating above-average temperatures. A vertical color scale legend to the right of the map shows the temperature anomaly in degrees Celsius, ranging from -10 to 10. Logos for the WMO, European Commission, Copernicus, and ECMWF are displayed below the map.

WMO climate statement: past 4 years warmest on record

Tags: [Climate change](#) [Climate services](#) [Climate](#)

29 Published 29 November 2018

Press Release Number: 29112018

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 left 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.

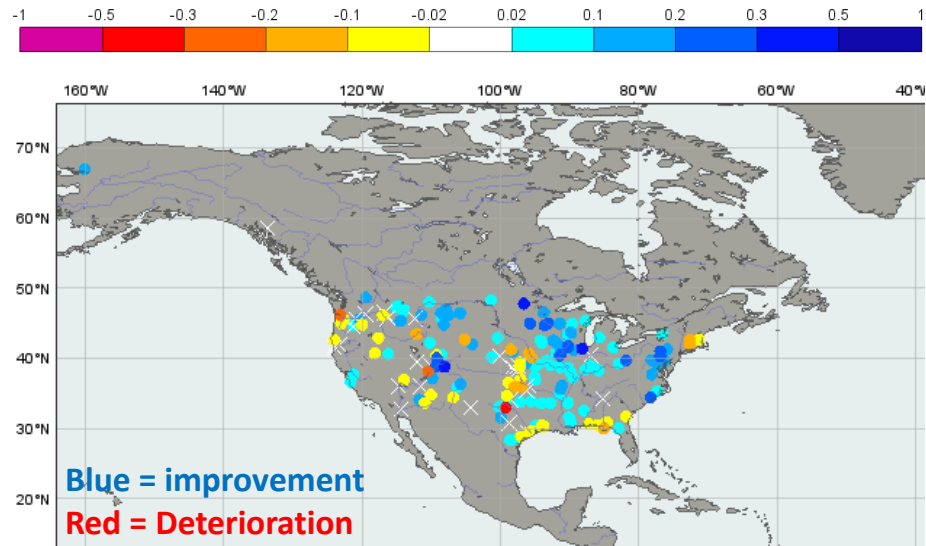
The report shows that the global average temperature for the first ten months of the year was nearly 1°C above the pre-industrial baseline (1850-1900). This is based on five independently maintained global temperature data sets.

"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 once again at record levels and if the current trend

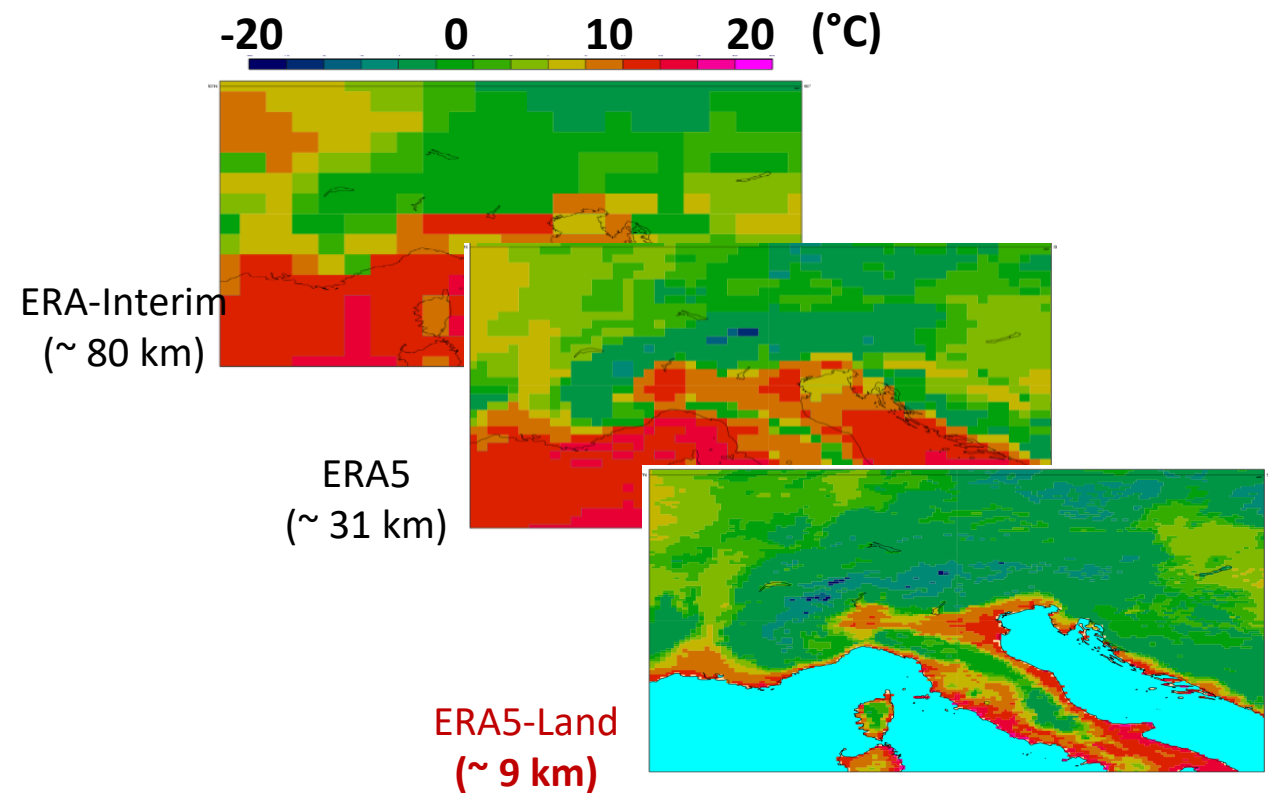
Latest WMO News

- WMO supports Sustainable Climate Information Services**
13 March 2019
- Monthly Weather Summary - February 2019 - Kingdom of Bahrain**
11 March 2019
- WMO marks International Women's Day**

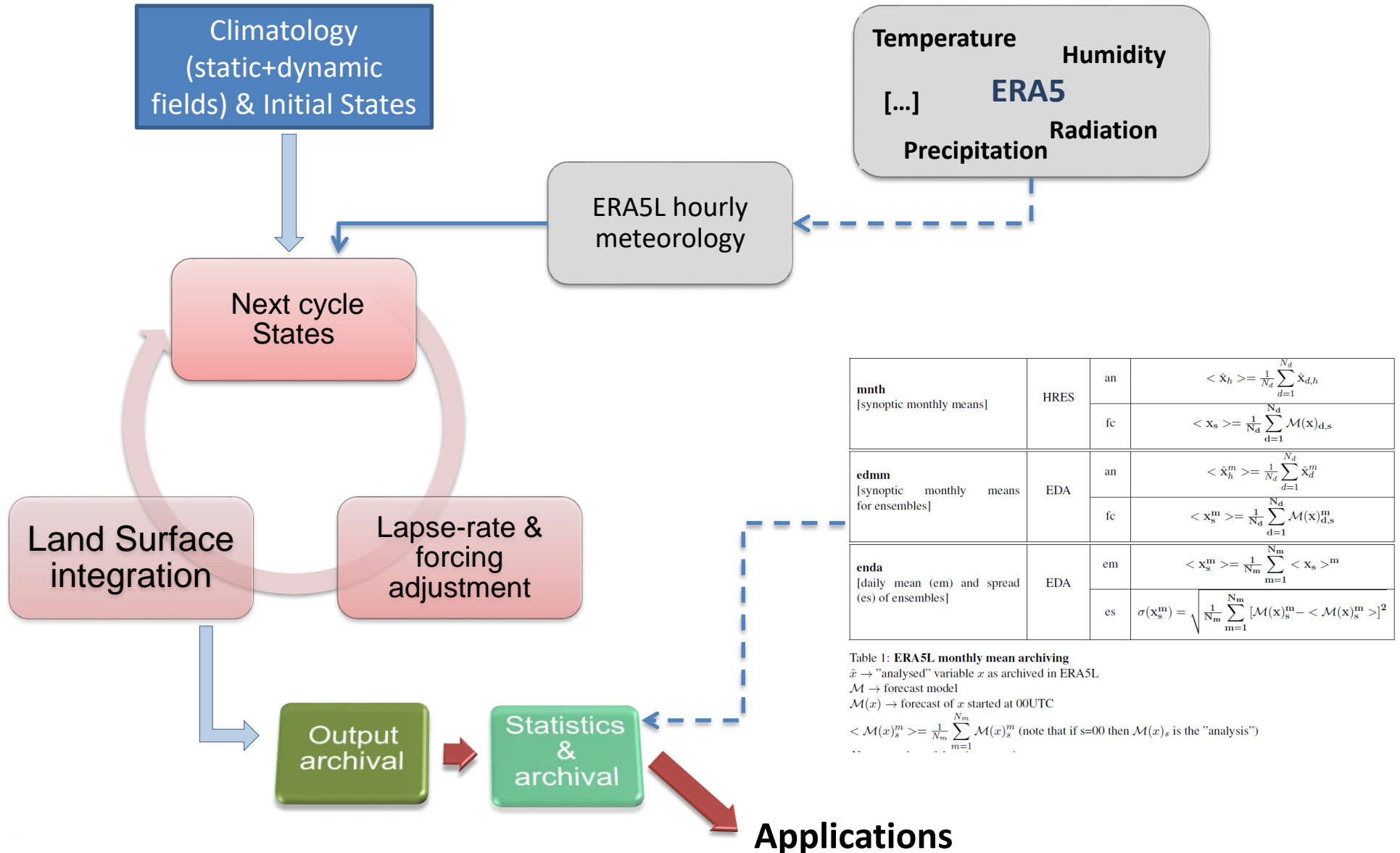
ERA5-Land, a high-resolution downscaling of the land-surface component



Discharge time series correlation difference ERA5-Land vs. ERA5



ERA5-Land in a simple diagram



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

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

Uppala *et al.* (2005), “The ERA-40 reanalysis”, *Q. J. R. Meteorol. Soc.* **131** (612), 2961-3012, doi:10.1256/qj.04.176

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.*, 141: 2350-2375. doi:10.1002/qj.2528

Laloyaux *et al.* (2018). CERA-20C: A coupled reanalysis of the twentieth century. *Journal of Advances in Modeling Earth Systems*, 10, 1172–1195. <https://doi.org/10.1029/2018MS001273>

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>