Reanalysis at ECMWF

Data assimilation training course 2024

Dinand Schepers,
Hans Hersbach, the Reanalysis Team and many others

Dinand.Schepers@ecmwf.int
Hans.Hersbach@ecmwf.int

Overview

- **Introduction reanalysis**
- How does reanalysis differ from NWP?
- The ERA5 reanalysis
  - Weather applications
  - Climate applications
- Coupled reanalysis
- Summary

What is reanalysis and how do we produce it

Reconstruction of the past climate using all observations we have:

- ✓ **Input:** non-gridded observations for a range of quantities (geophysical, radiances,..) + gridded boundary/forcing
- ✓ **Output:** consistent temporal evolution of 3D atmosphere (or ocean, atm. composition,..)

Methodology:

- ✓ Use a recent NWP data-assimilation system: ‘**re**do the **analysis** of old weather’
- ✓ For several decades or longer
- ✓ At lower resolution, such that we can afford it

- ✓ Made available to users in a convenient way
- ✓ Produced at several centres worldwide: ERA5; MERRA (2); JRA-3Q; CFSR; 20CR-v2,v3
- ✓ Regional and global, ingest full or selected observing system

Two main categories of applications:

- ✓ Study of **specific events** or phenomenae:
  - need accurate (3D) synoptic situation; i.e., **the weather of the day**
- ✓ **Climate** applications:
  - low-frequency variability of **the mean state**
  - Statistics, e.g., of extremes

A consistent and complete picture of the past ~~atmosphere~~ Earth system

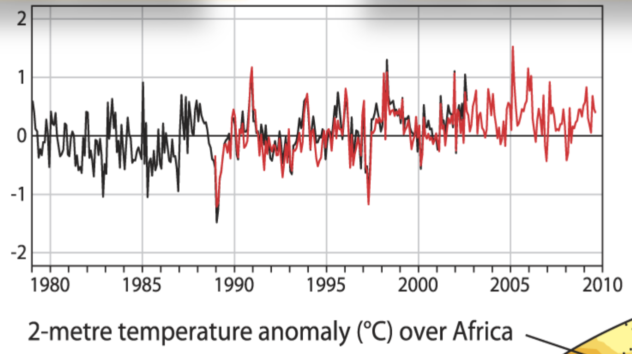
…in Time

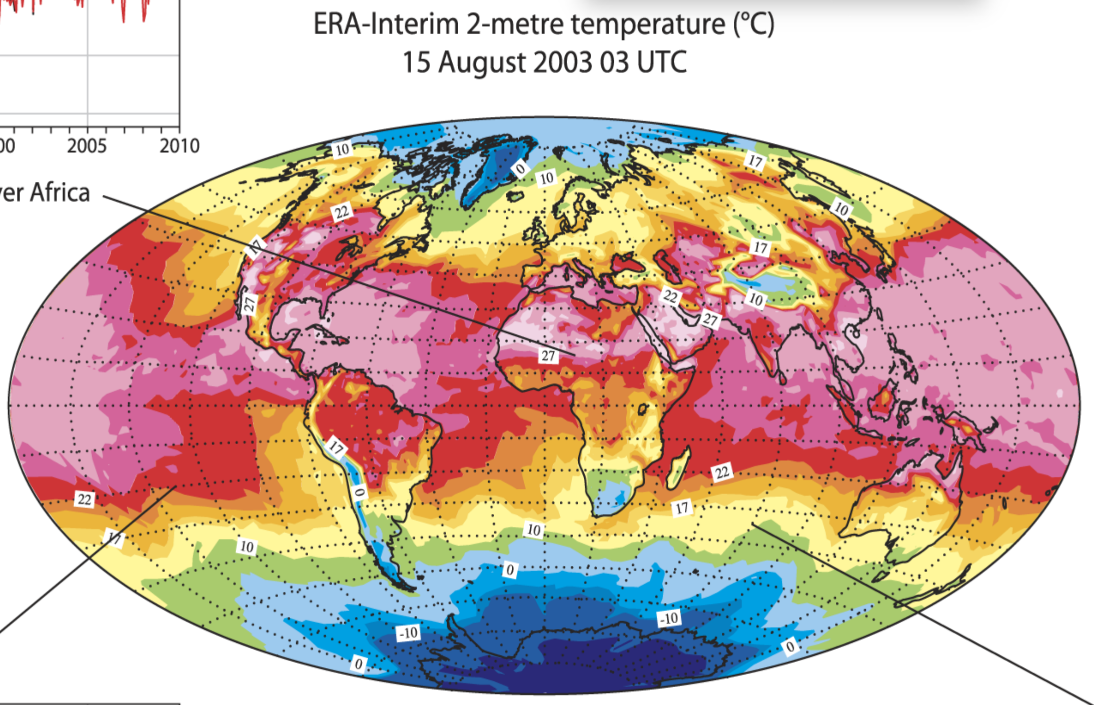
…in the Horizontal

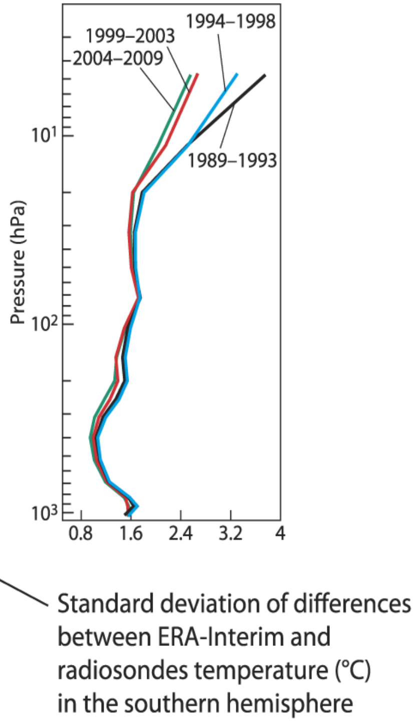
…across (Atmospheric) Parameters

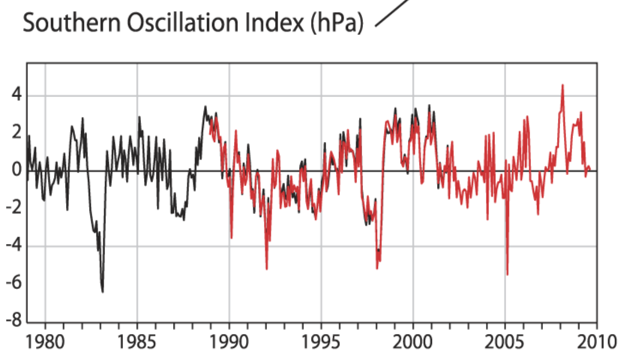
…in the Vertical

… across domains



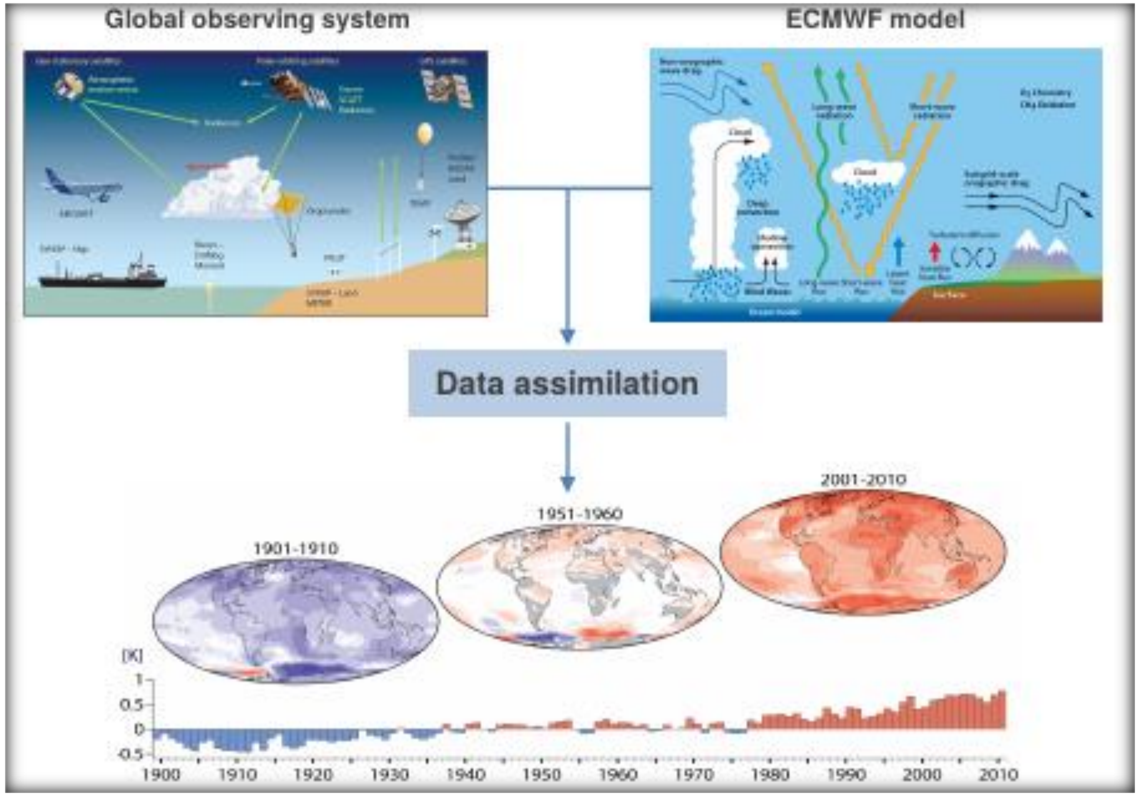






Reanalysis uses past observations with today's NWP DA system

The data from reanalysis are widely used



- ✓ Complete: combining vast amounts of observations into (global) fields
- ✓ Convenient: "maps without gaps", always available in the same way
- ✓ Consistent: use the same physical model and data assimilation system throughout

❖ *Observations are absolutely key!!*
- *fewer observations as we go back in time*
- *New, modern-day observations added where possible*

➢ provide an uncertainty estimate, e.g., from an ensemble.

ECMWF has a long experience with reanalysis

Atmosphere/land — *including ocean waves*

1) 1979 - 1981 FGGE	2) 1994 - 1996 ERA-15	3) 2001 - 2003 ERA-40	4) 2006 - 2019 ERA-Interim	5) 2016 - … ERA5

Ocean — *including sea ice*

2006 ORAS3	2010 - … ORAS4	2016 - … ORAS5

Centennial — *Coupled*

2013 - 2015 ERA-20CM/20C	2016 CERA-20C	2017 CERA-SAT

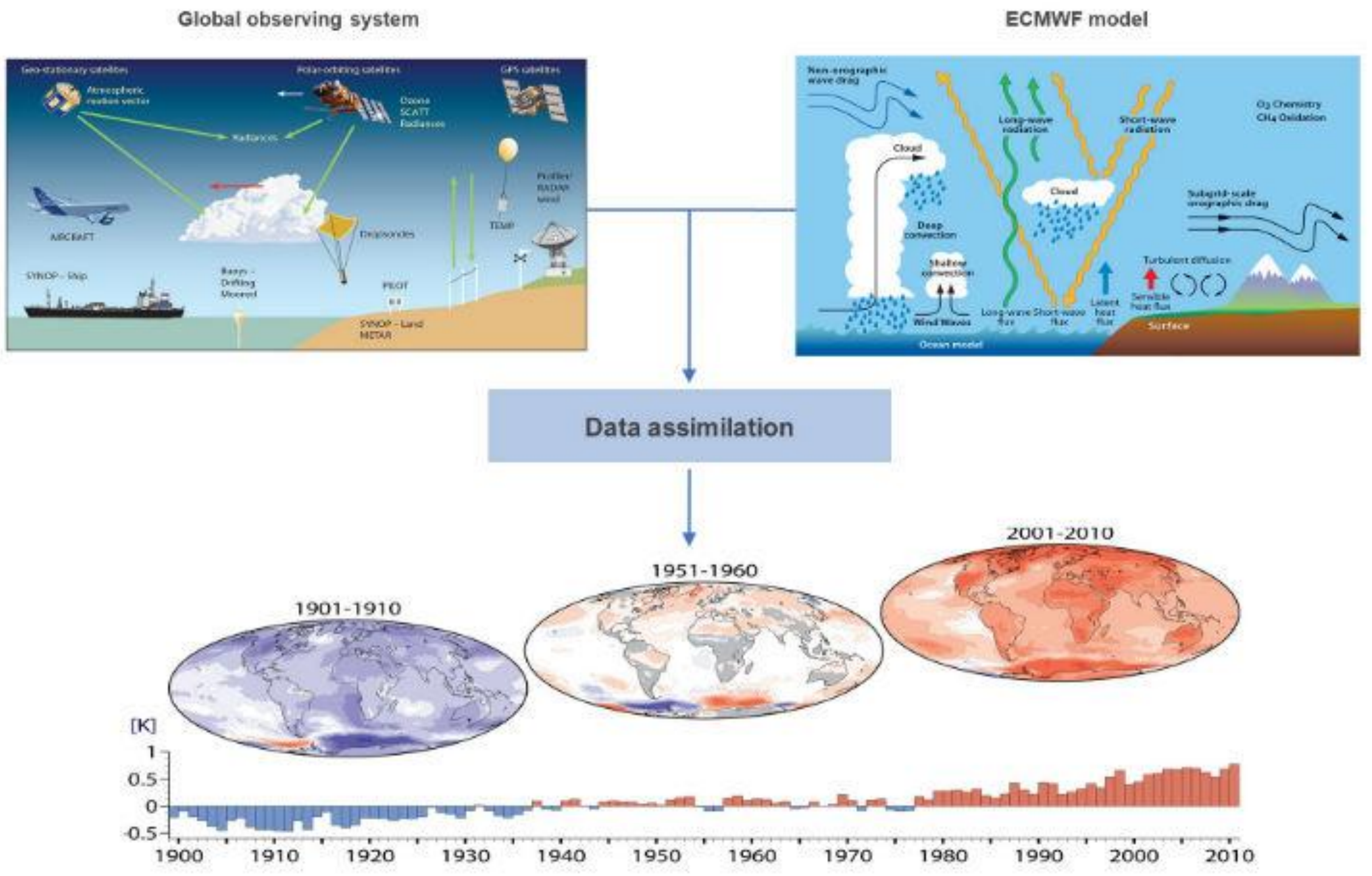
Enhanced land

2012 ERA-Int/Land	2014 ERA-20C/Land	2018 - … ERA5L

Atmospheric composition

2008 - 2009 GEMS	2010 - 2011 MACC	2017 - … CAMS

Data assimilation: combining the model and the observations

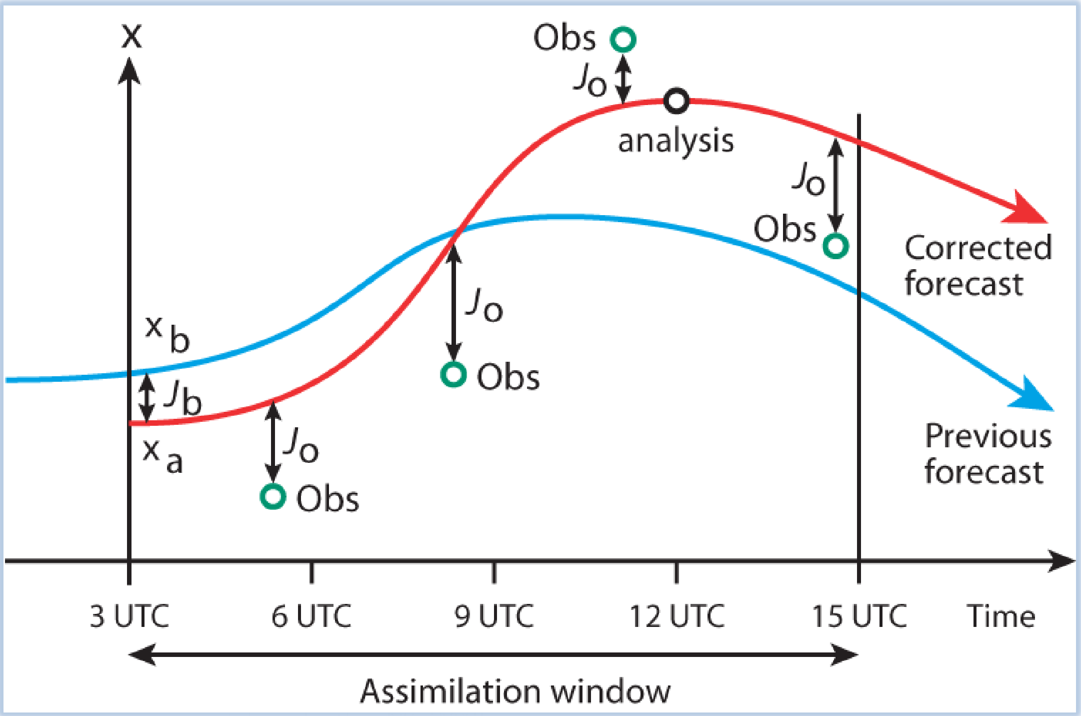


4D-Var data assimilation: using observations to correct the first guess

background constraint (J_b) observational constraint (J_o)

Minimise
$$J(z) = (z_b - z)^T B_z^{-1} (z_b - z) + \left[y - \tilde{h}(z)\right]^T R^{-1} \left[y - \tilde{h}(z)\right]$$

$$z^T = \left[x^T \beta^T\right] \qquad \tilde{h}(z) = h(x) + b(x, \beta)$$



Recipe:

- Start with a first guess $\mathbf{x}_b$, $\boldsymbol{\beta}_b$ from the previous analysis
- Compare with observations **y**
- Calculate the misfit: the cost **J**
- Change the first guess such that the fit is optimal

Result depends on:

- The confidence in your first guess: **B** matrix
- The confidence in your observations: **R** matrix
- How you compare the model to observations: **h(x)**
- The choice of the observation bias model **b(x,*β*)**

Result:

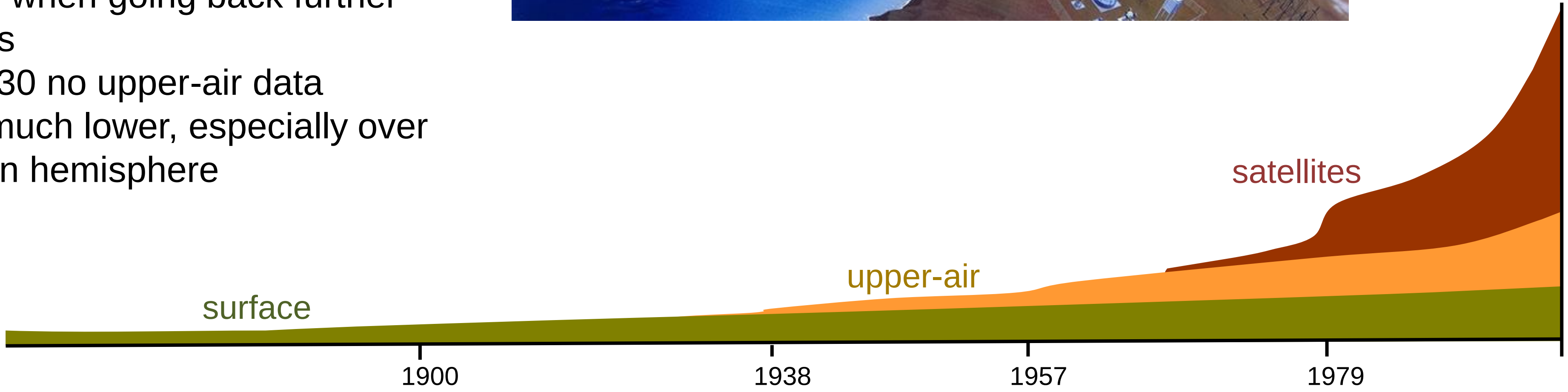
- The reanalysis $\mathbf{x}_a$, $\boldsymbol{\beta}_a$
- used to provide first guess for next analysis

The evolving observing system

Data sources:
- many satellites
- surface observations
- weather balloons, aircraft, etc.

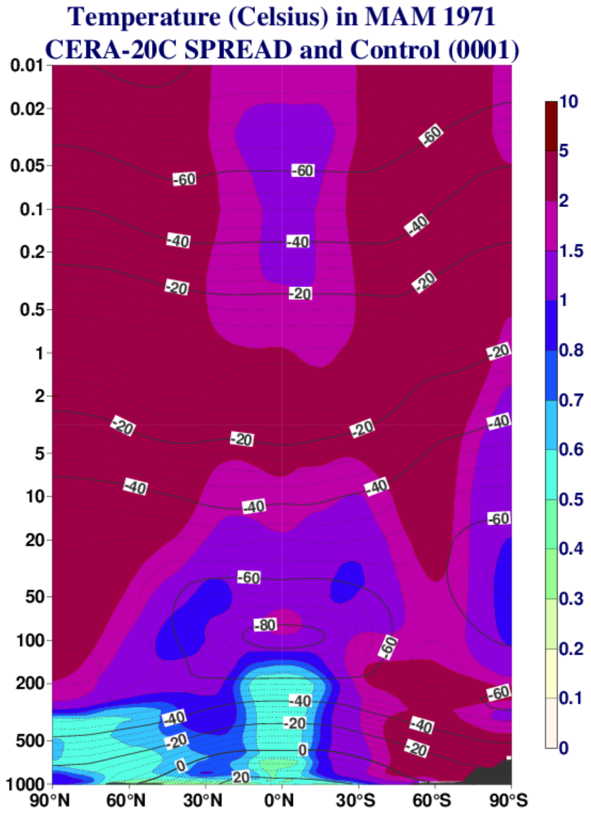
In the ERA5 reanalysis we daily use about:
- 53,000 observations in 1950
- and 26 million in 2021

- Amounts are continuing to grow,
- Less observations when going back further
    - No satellites
    - Before ~1930 no upper-air data
    - Coverage much lower, especially over the southern hemisphere

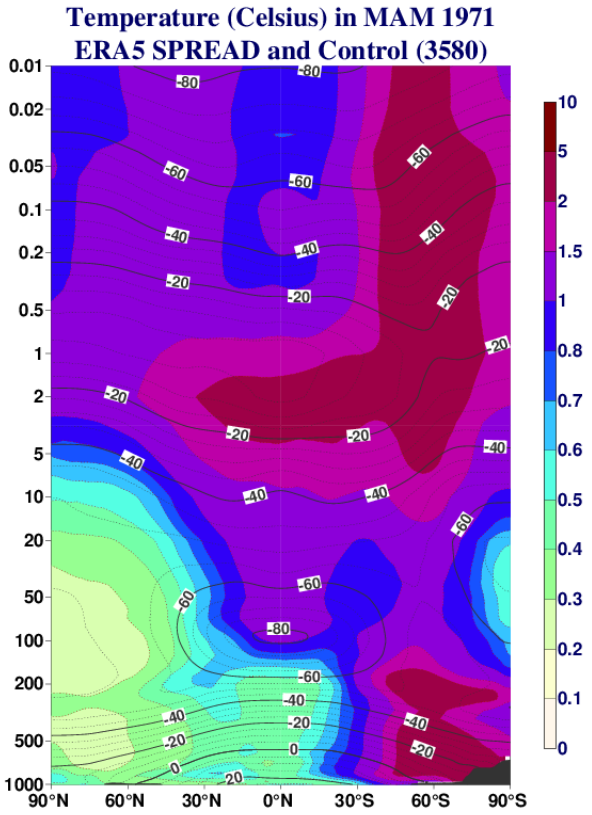


Courtesy: Paul Poli

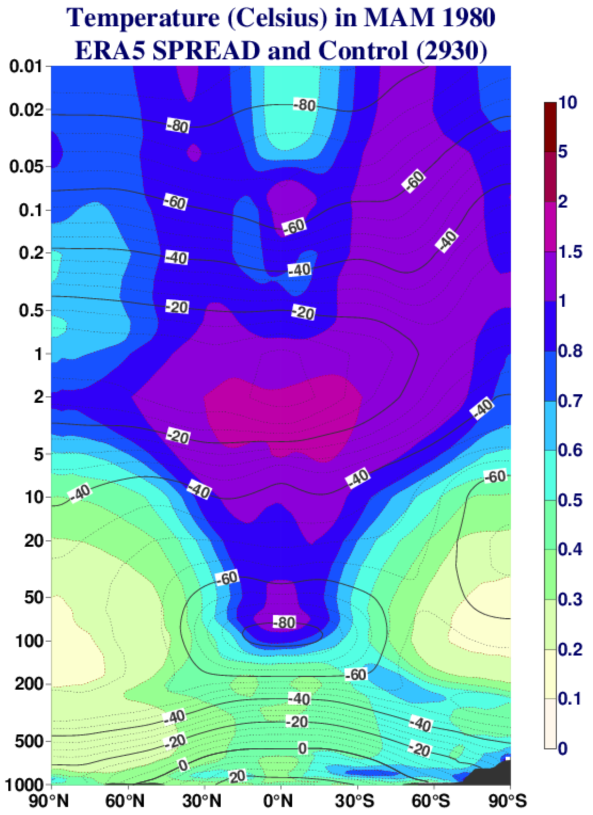
The evolution of ensemble spread; also proxy for synoptic uncertainty



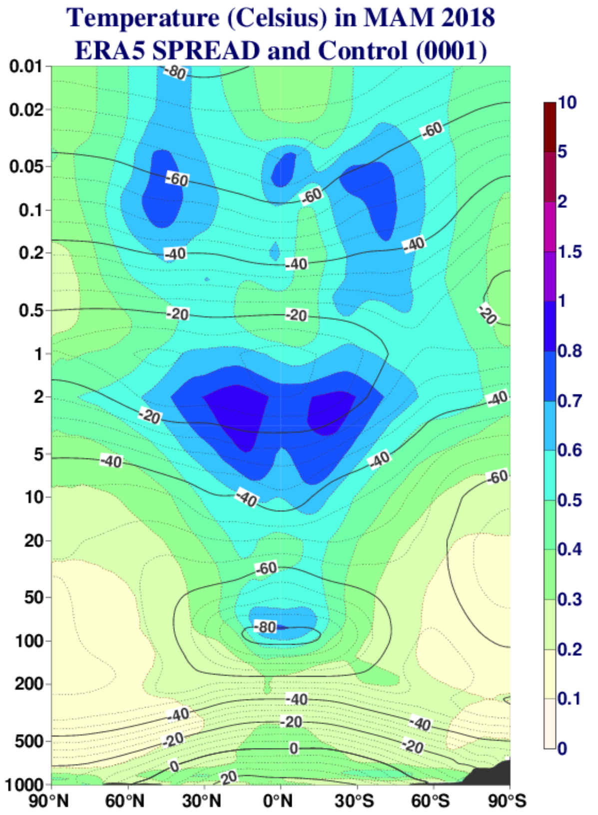
1971 CERA-20C:
Surface pressure, marine wind, only



1971 ERA5:
Upper-air data



1980 ERA5:
Early-satellite era



2018 ERA5:
Recent observing system

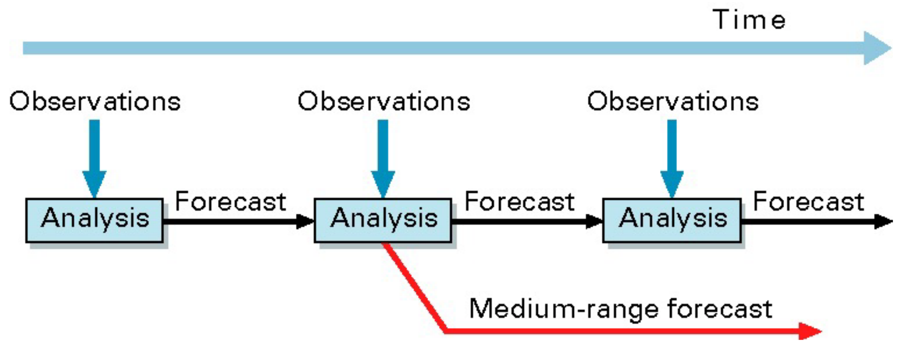
Overview

- Introduction reanalysis
- **How does reanalysis differ from NWP?**
- The ERA5 reanalysis
  - Weather applications
  - Climate applications
- Coupled reanalysis
- Summary

How does reanalysis differ from NWP?

It is good practice to base an operational reanalysis on a recent NWP system

✓ E.g., at ECMWF, ERA5 (2016) is based on Cy41r2



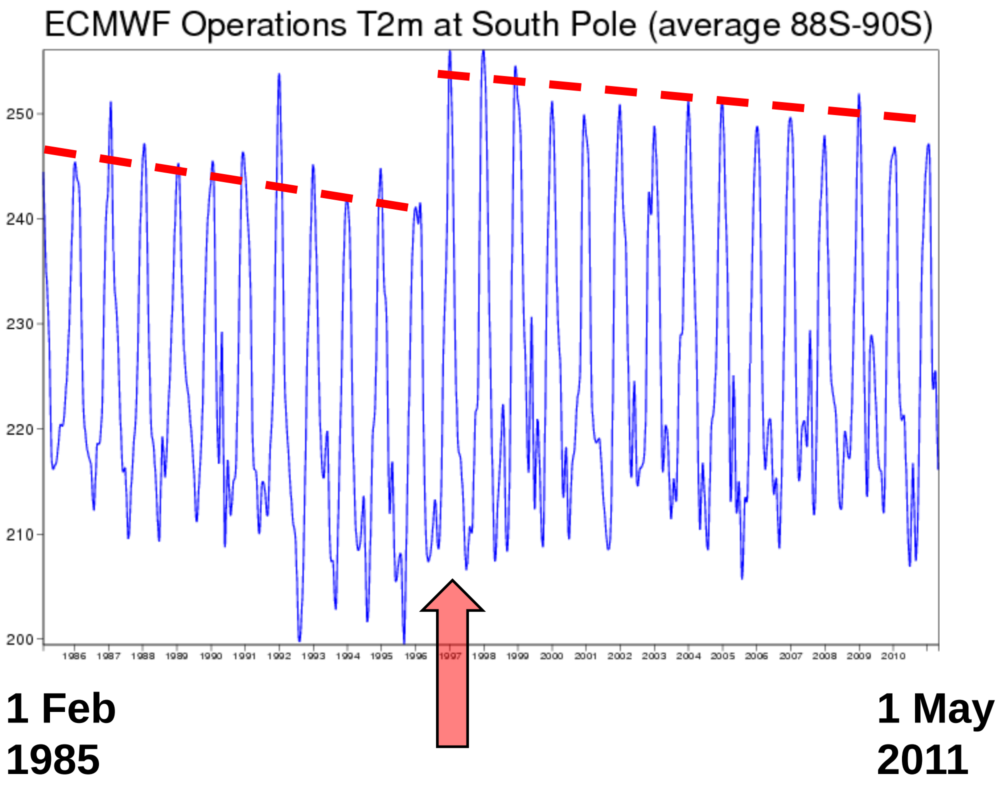
Differences:

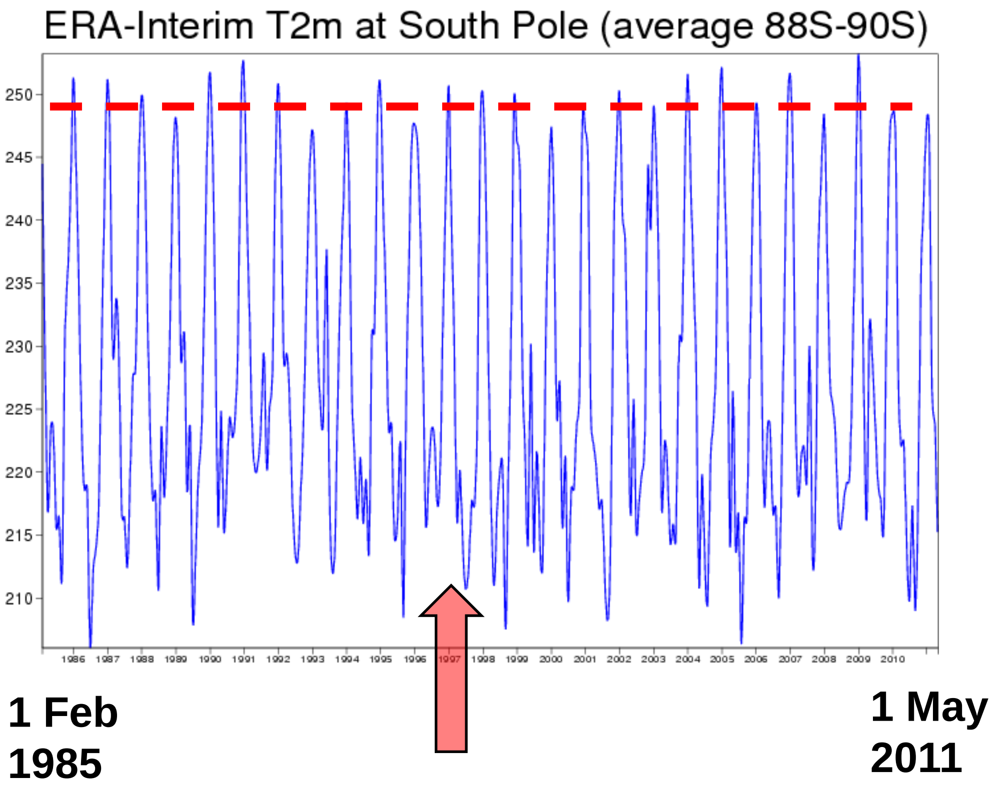
- ✓ The focus is on the quality and consistency of the analysis, not the forecast
  - ✓ Assimilation system is effectively frozen
- ✓ Need to ensure that you have good and as many as possible historical observations
  - ✓ Reprocessing and data rescue

❖ The NWP system is well-tuned for the recent data-rich era
Ensure that it also works well for the data-sparser past, e.g.:

- Appropriate forcing fields
- Background errors
- Observation errors
- Quality control
- Systematic model and observation errors

Why not use simply operational NWP?





The operational NWP system has evolved dramatically over time:

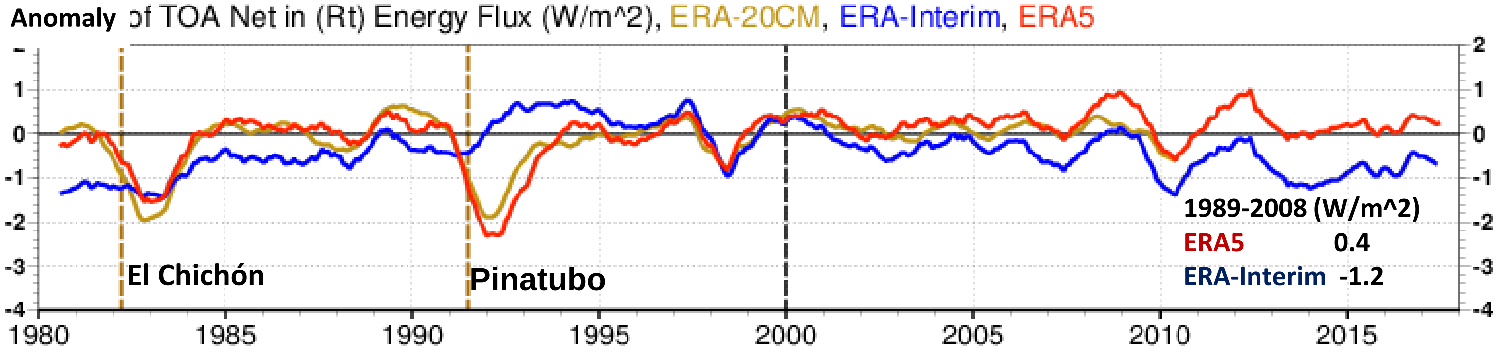
- Resolution
- Maturity of its NWP model and data-assimilation system

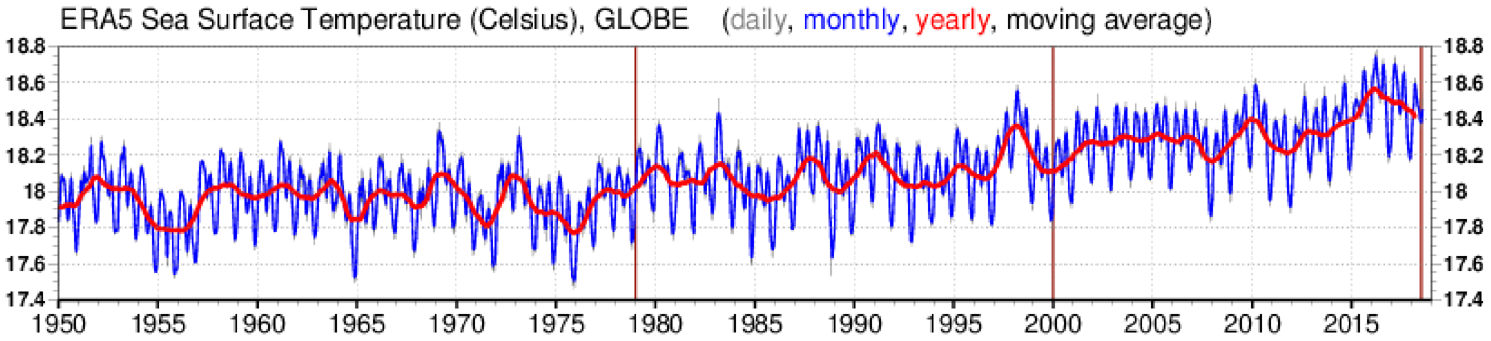
ERA5 forcing appropriate for climate; these are ingested ‘as is’

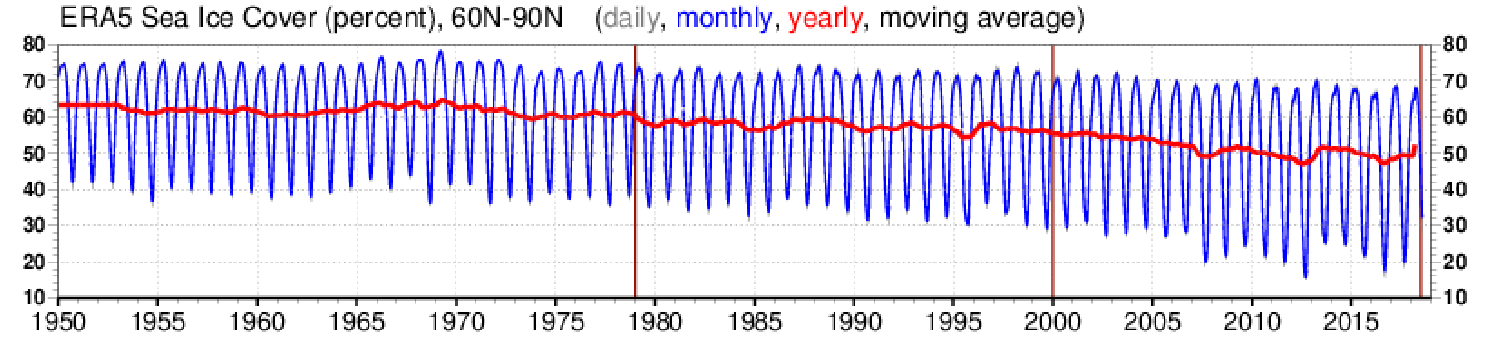
CMIP5 recommended data sets

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

(Prepared in the ERA-CLIM project, ERA-20CM)







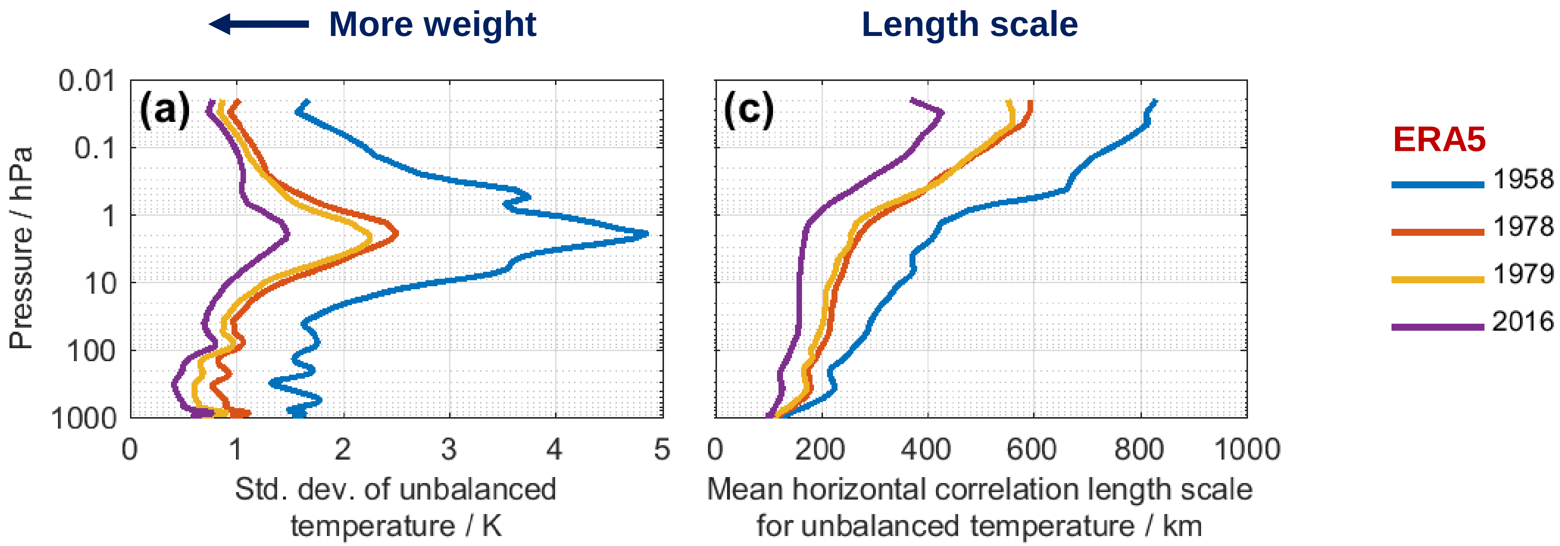
SST and sea ice cover

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

Different ensemble members use different SST realizations

(Hirahara et. al., 2016)

Long-term evolution of the background error covariance matrix



Lecture on **B**ackground error:

- For a single observation at a model grid point the analysis increment is proportional to a column of **B**
- The role of **B** is to spread out information from the observations

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

➔Smaller background error variances, with sharper structures

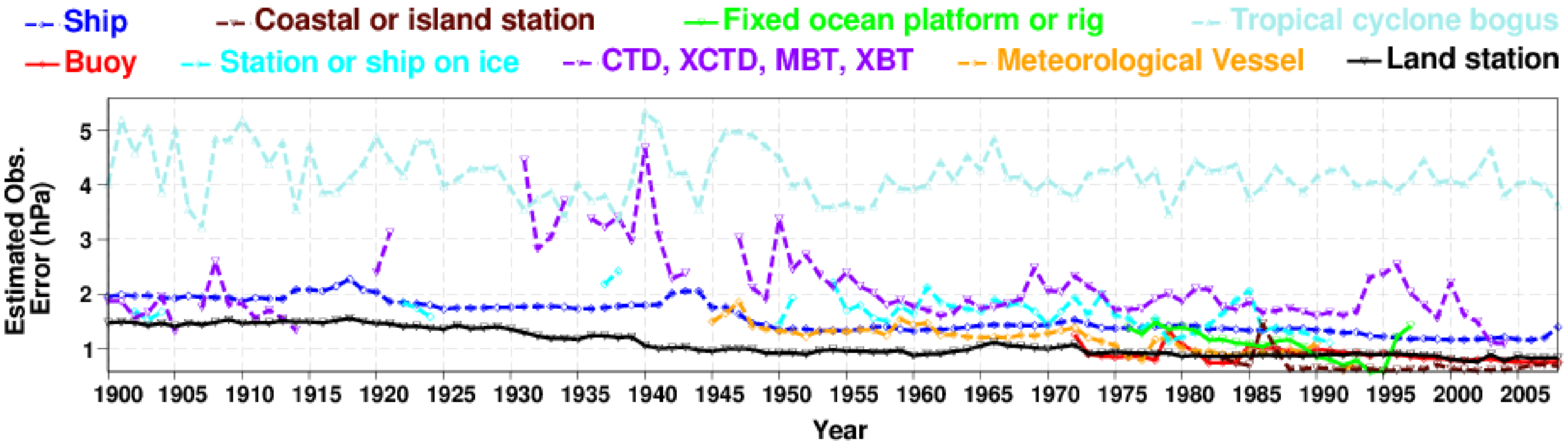
➔Analysis increments that are smaller, over smaller areas

Evolution of observation error

The quality of observations has evolved over time in line with changes in instrumentation. Therefore, the observation error should evolve accordingly

Methods exist that can be used *a posteriori* to estimate observation error standard deviations*

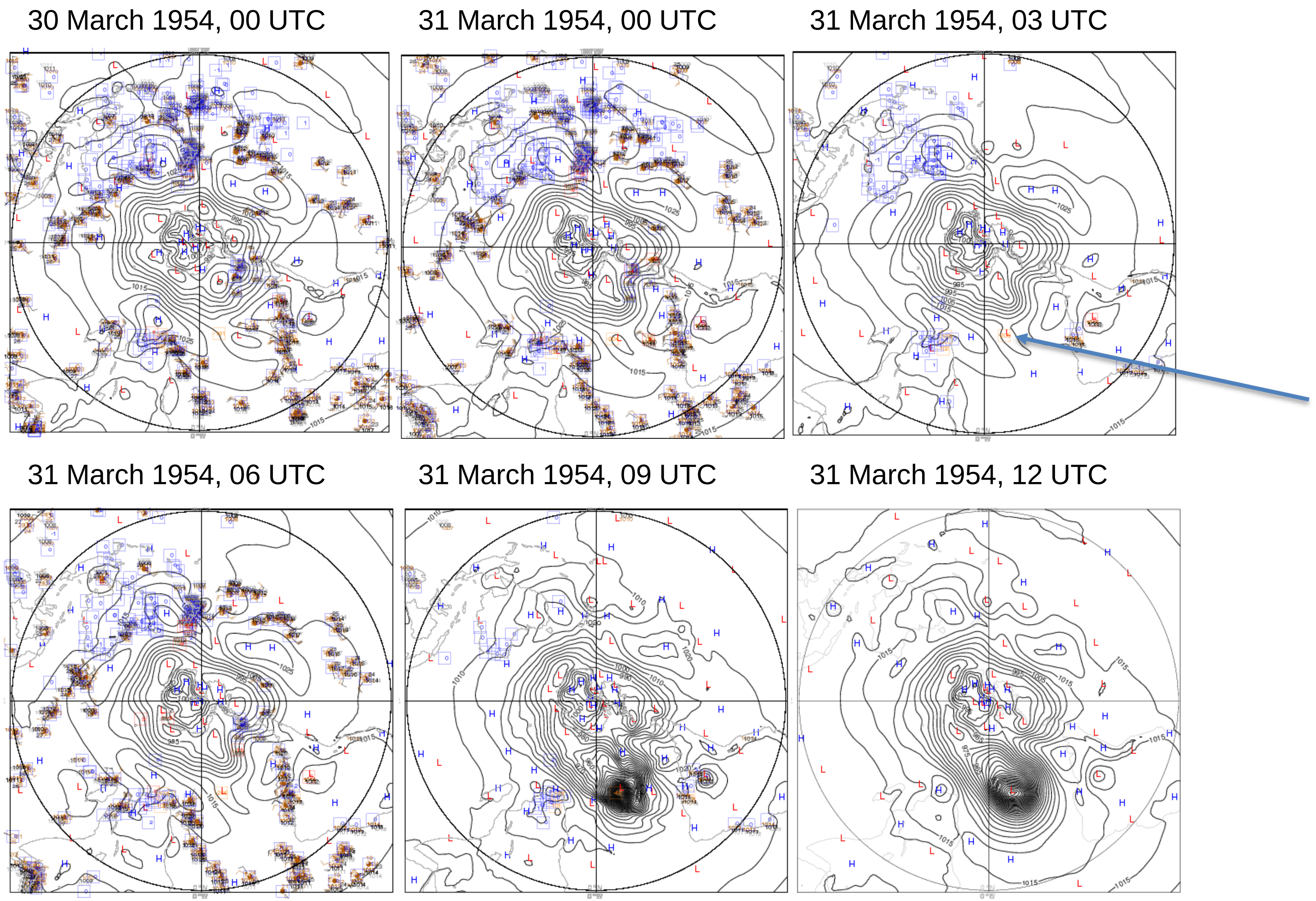
E.g., ERA-20C assumed time invariant observation errors. This does not seem to be the case… In CERA-20C these were evolved.



* 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

Quality control: Impact of a single bad time-series

Southern Hemisphere



Island station on Tristan da Cunha for years reported surface pressures around 800 hPa

Overview

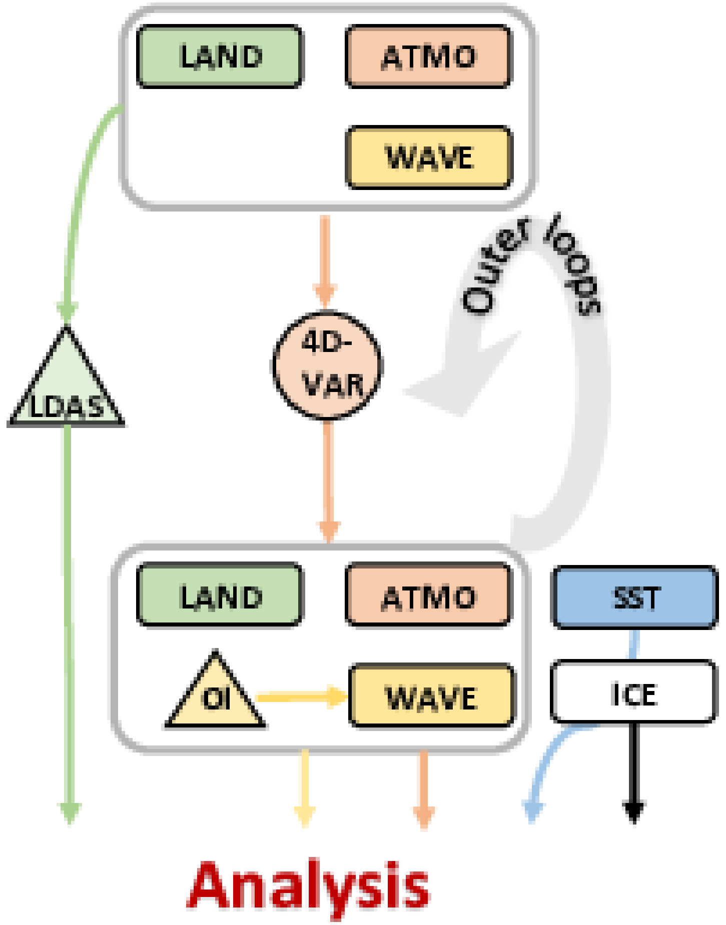
- Introduction reanalysis
- How does reanalysis differ from NWP?
- **The ERA5 reanalysis**
  - Weather applications
  - Climate applications
- Coupled reanalysis
- Summary

The ERA5 Global Reanalysis

ERA5: *A full-observing-system global reanalysis for the atmosphere, land surface and ocean waves*

- Produced at ECMWF, by the **Copernicus Climate Change Service**

- Produced in parallel streams to speed up production
- Daily updates **5 days behind real time from 1940 onwards**
- Hourly snapshots at 31km resolution up to about 80km height
- Uncertainty estimate from a 10-member ensemble at half resolution
- Total dataset about 12 petabyte



Observation usage:

- Around 100 billion so far
- Daily: 53,000 (1950), 0.5 million (1979), 26 million (2021)

Usage of external (gridded) products 'as is':

- SST and sea-ice cover
- GHGs, aerosols, Total Solar Irradiance , (diagnostic) ozone

The ERA5 observing system

0.75 (1979) – 26 Million (2021) obs per day
Over 200 types of reports

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, …
TAMDAR, MODE-S

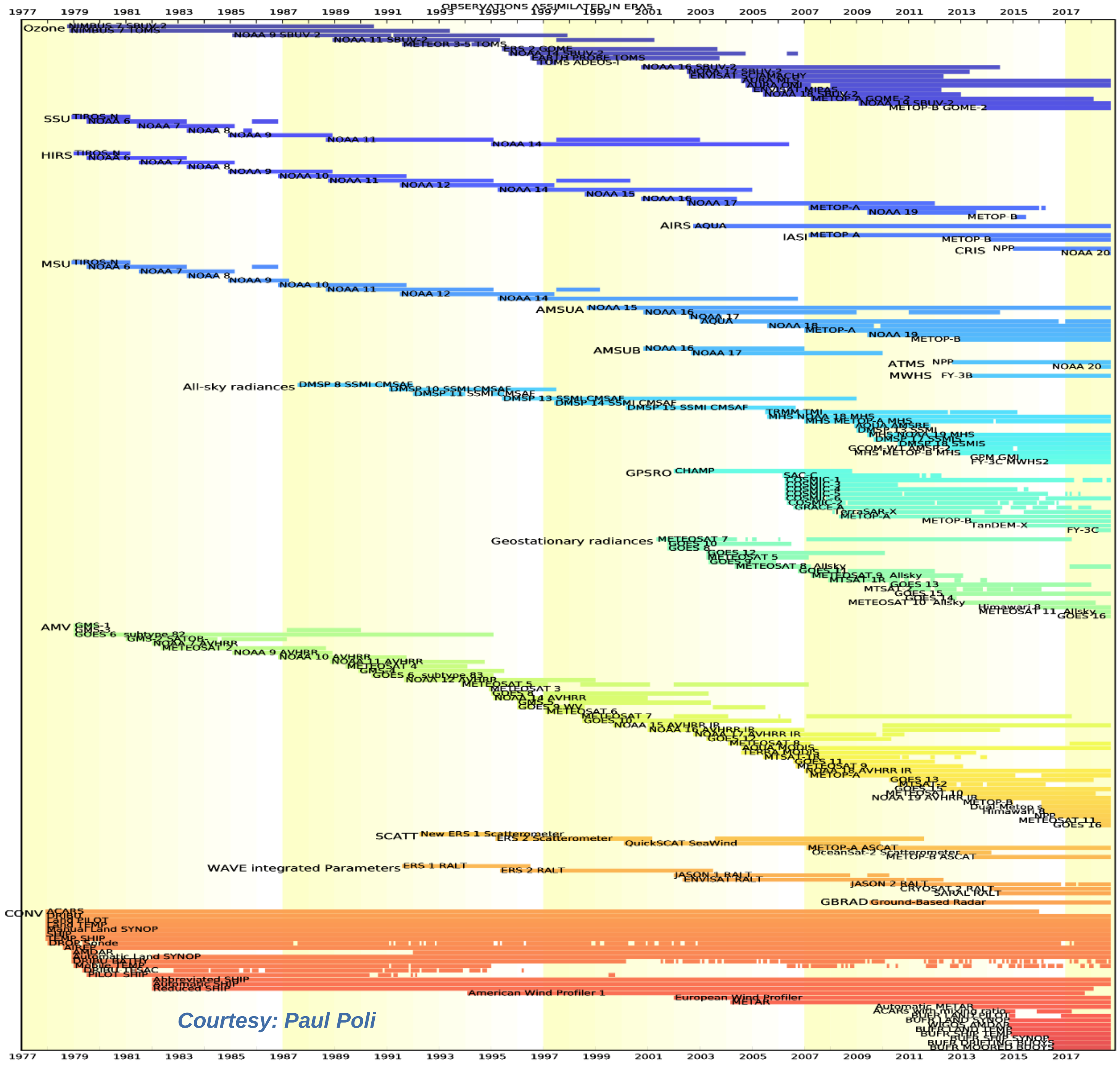
ERA5T vs ECMWF NWP operations:

ERA5 only: AURA MLS
ERA5 not possible: Saphir, Aeolus, TEMP descent, SPIRE, …

Improved data usage compared to ERA-I

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

Needs to be monitored all during production!



Courtesy: Paul Poli

ERA5 uses an Ensemble of Data Assimilations (EDA)

Concept (see presentation from Massimo):

- Perturb observations (including SST and sea ice)
- Perturb model in short forecasts linking analyses
- Do not Perturb VarBC

From this estimate a flow-dependent B matrix:

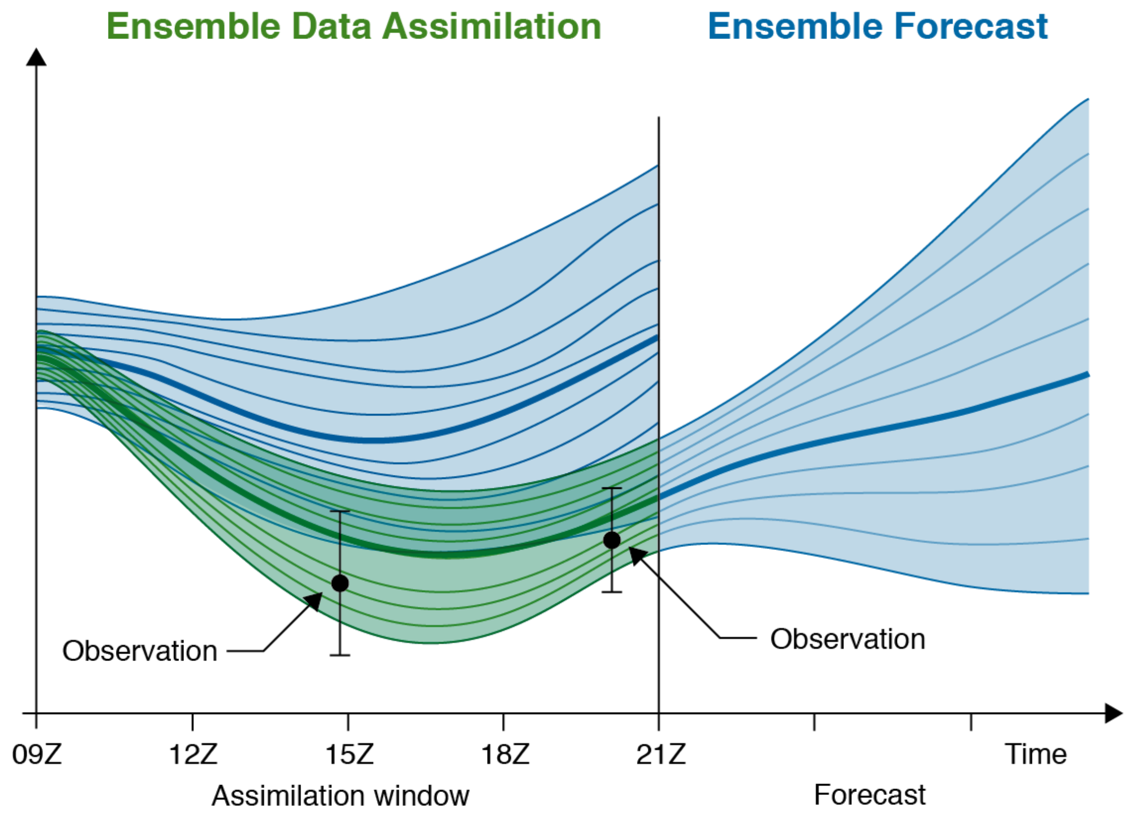
$$B(t) = (1-\alpha)B_{cli} + \alpha B_{EDA}(t)$$

ERA5:

10 members, rather than 50 in NWP

Much reduced resolution

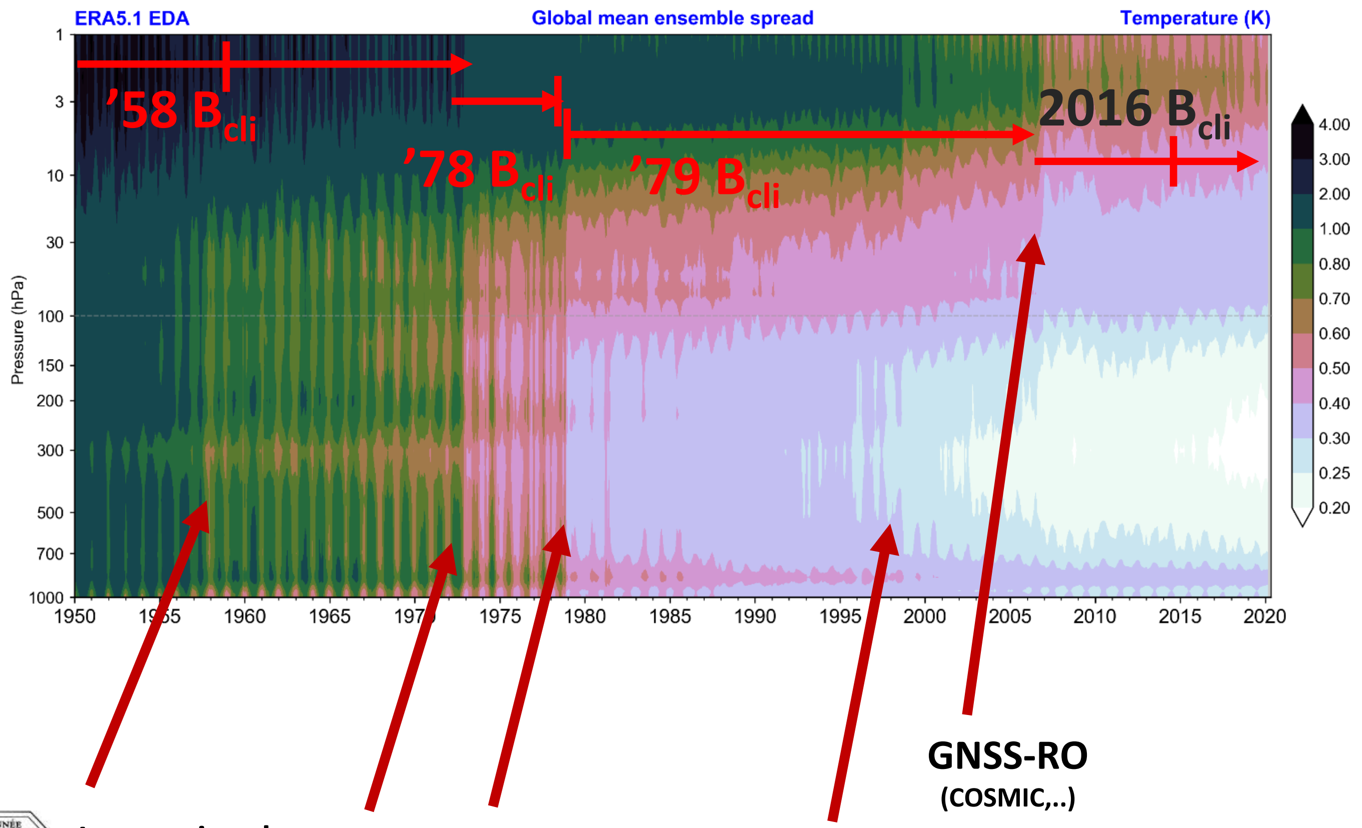
Smaller mixing α



EDA Ensemble spread as a measure for the synoptic ERA5 uncertainty

Spread decreases over time when more and more observations become available

Major changes in the observing system are clearly visible



International Geophysical Year (1957/58)

VTPR

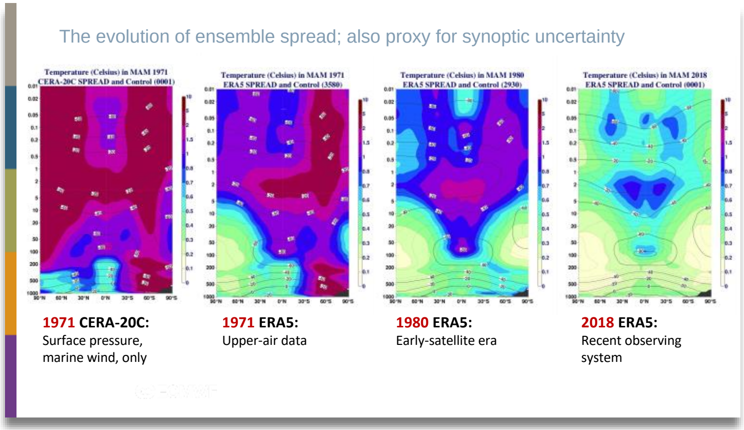
TOVS (HIRS, MSU, SSU)

ATOVS (AMSU-A)

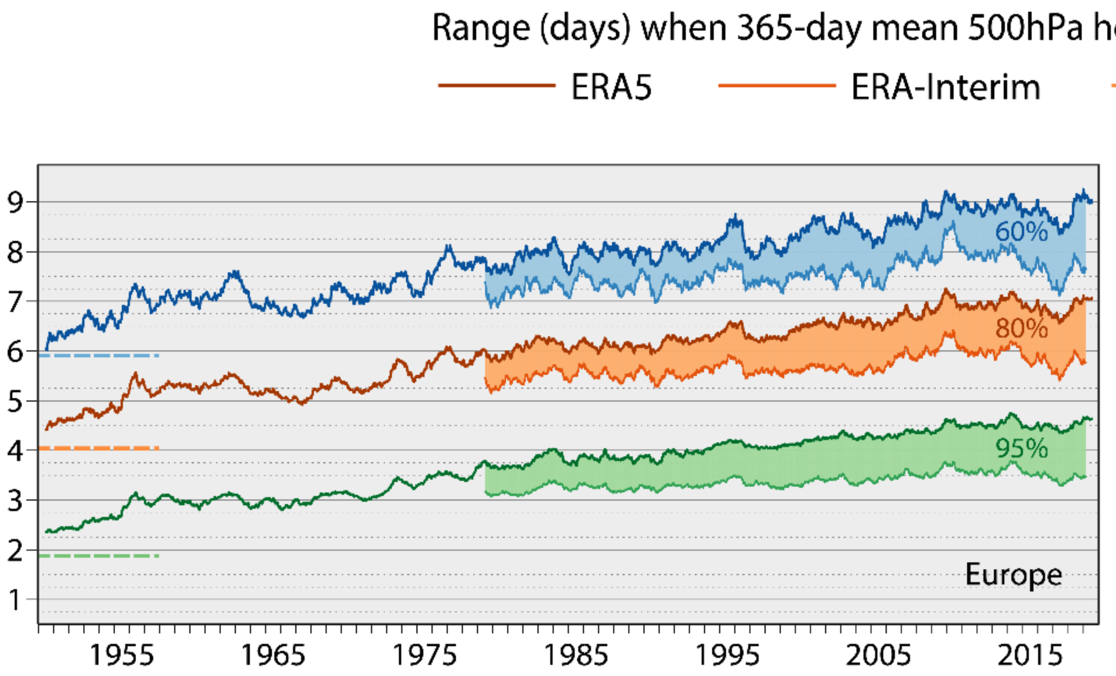
GNSS-RO (COSMIC,..)

Flow-dependent B matrix:

$$B(t) = (1-\alpha)B_{cli} + \alpha B_{EDA}(t)$$



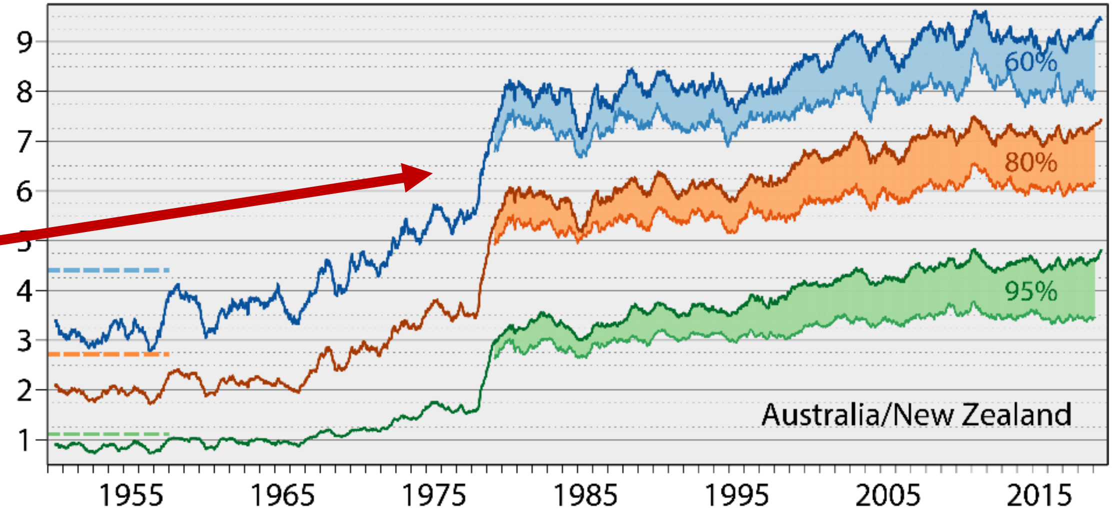
the quality of re-forecasts issued from reanalysis evolves accordingly



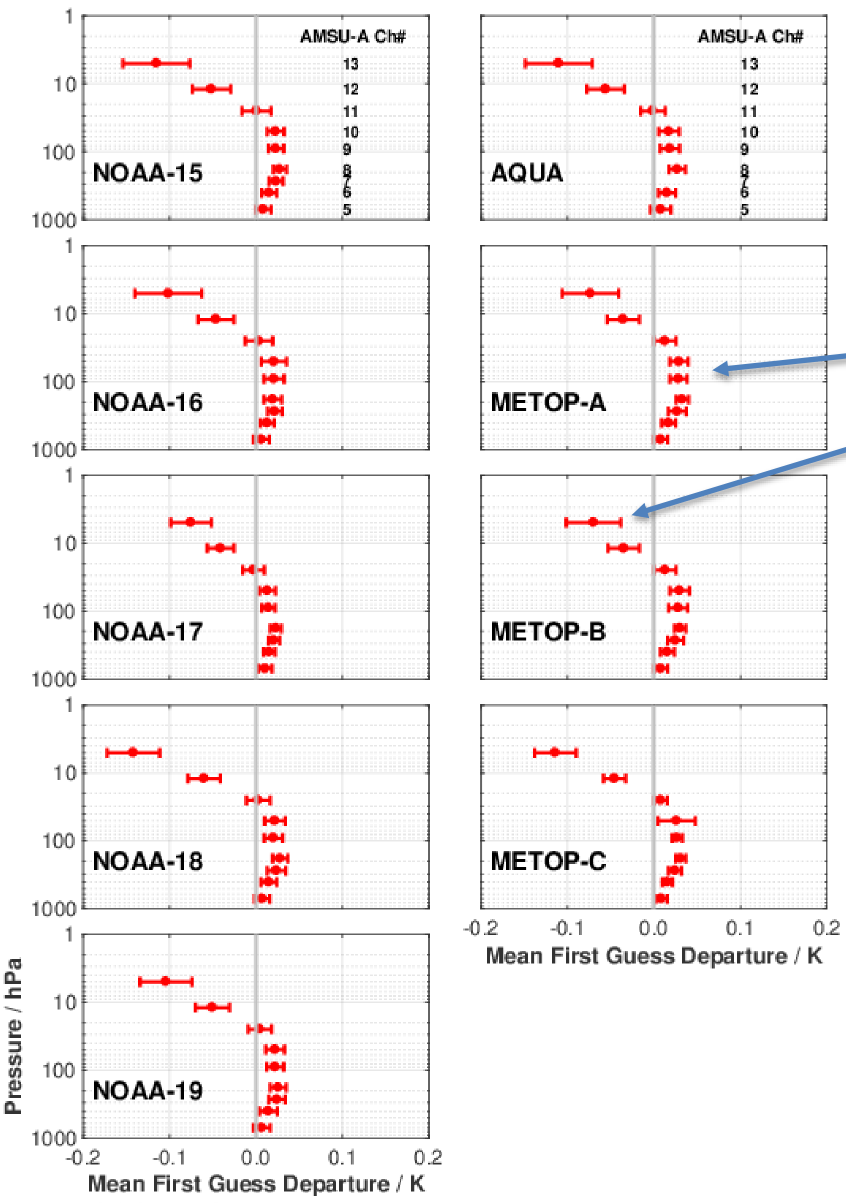
ERA5 back extension:

NHEM (especially Europe) skill is promising, but lower prior to 1957-1958

Over SHEM there is a dramatic improvement following the introduction of TOVS satellite data in late 1978.



Model Error: diagnosed from AMSU-A Mean FG_DEPS in ERA5



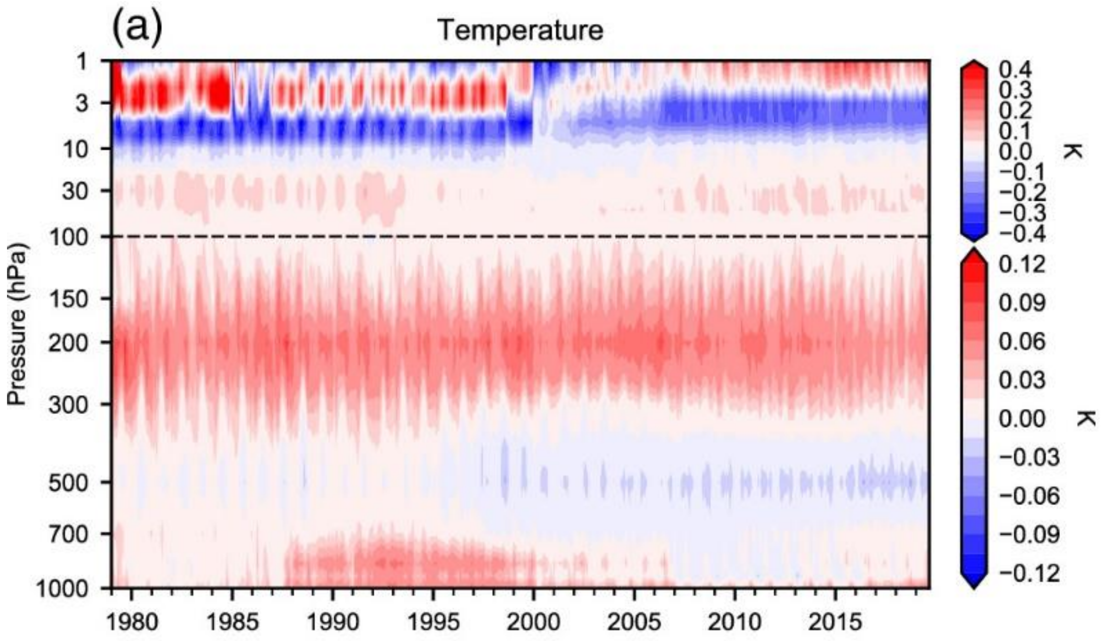
ERA5 mean first guess departures shown for AMSU-A

Error bars represent (±1σ) spread over the lifetime of each sensor

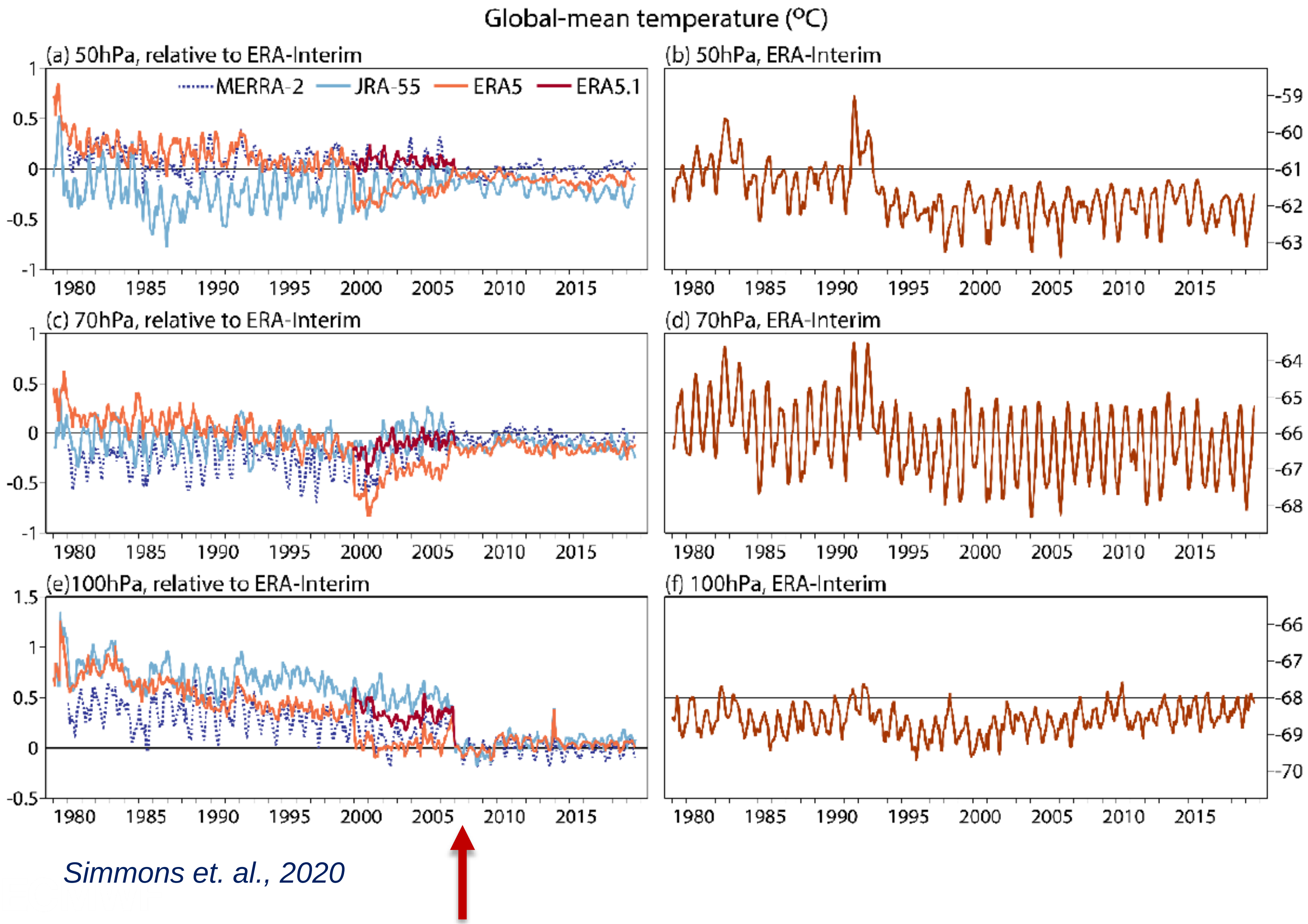
Consistent picture of :

- a cold model bias mid-trop to mid-stratosphere
- a (larger) warm model bias above 10 hPa

Broadly consistent with analysis increments in ERA5



The importance of anchor observations: COSMIC RO from 2006



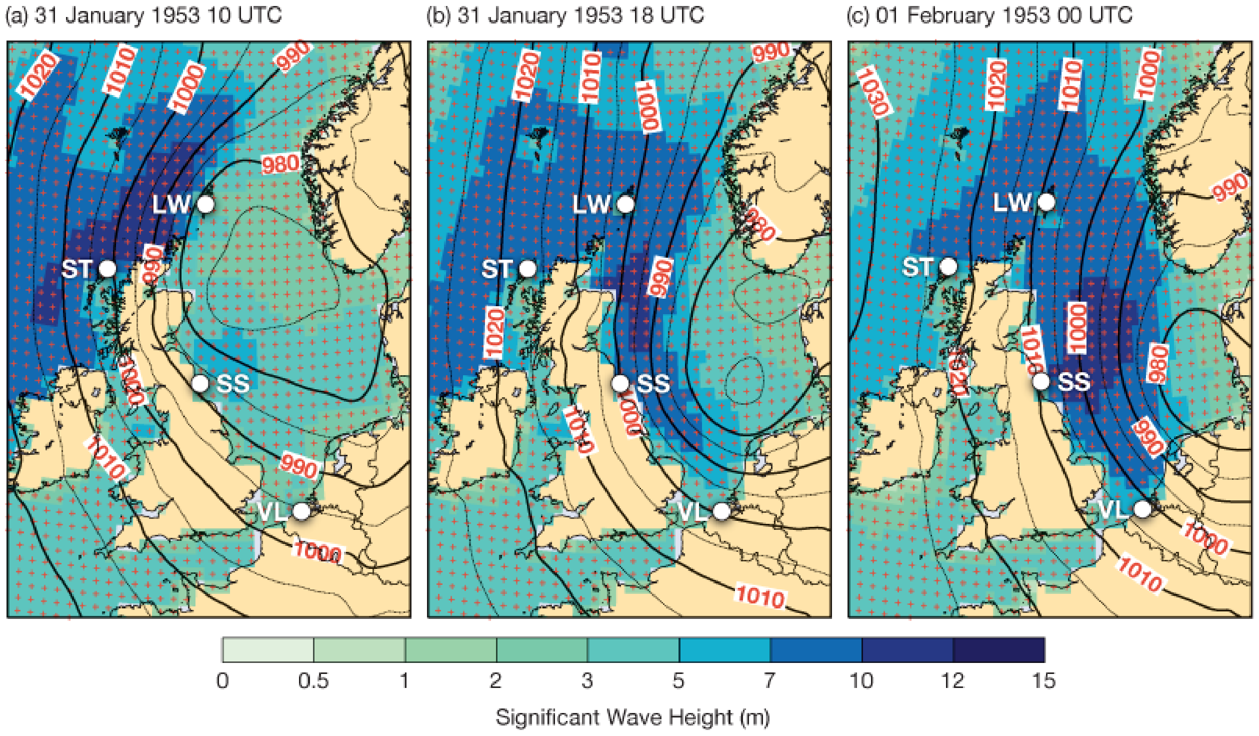
Simmons et. al., 2020

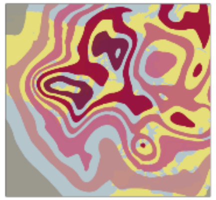
Overview

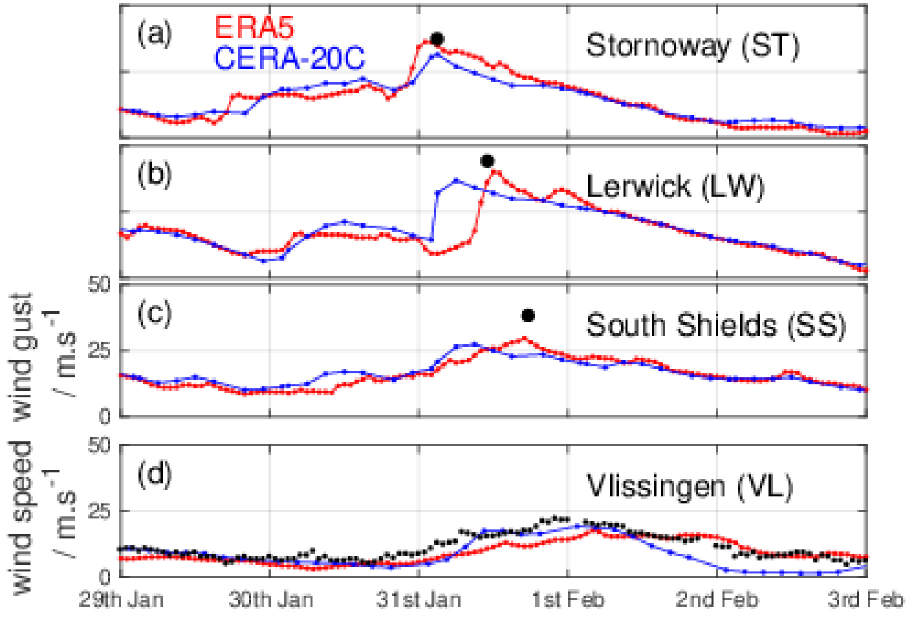
- Introduction reanalysis
- How does reanalysis differ from NWP?
- The ERA5 reanalysis
  - **Weather applications**
  - Climate applications
- Coupled reanalysis
- Summary

The North Sea Storm of February 1953

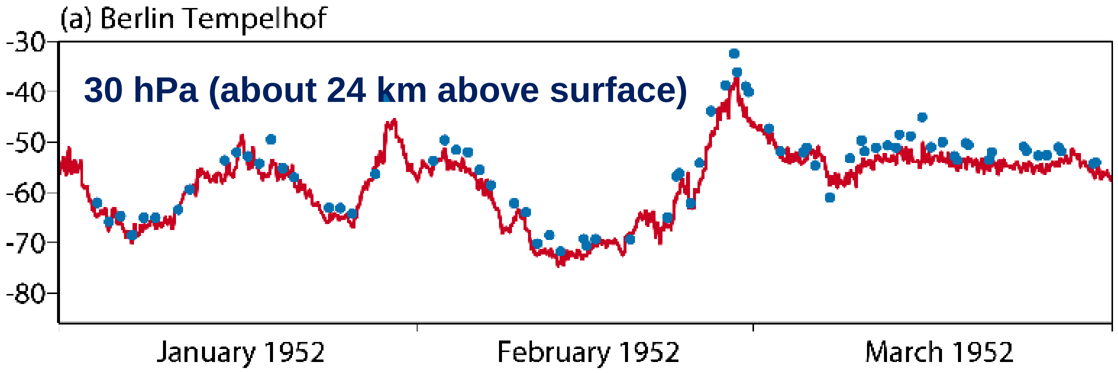
An intense storm in combination with spring tide caused widespread breaches of sea defenses resulting into about 2,500 deaths.
Subsequently, the UK and the Netherlands fortified their defences (Thames flood barrier and Delta Works).





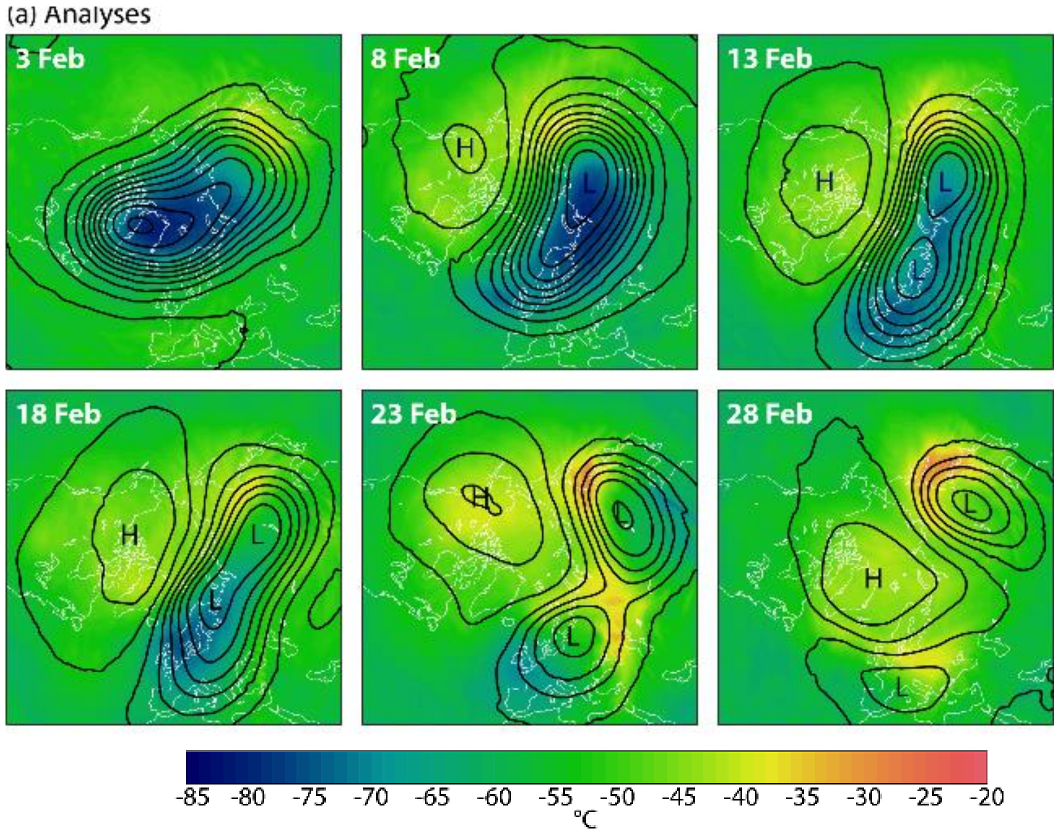


Stratospheric Sudden Warming, February 1952



The discovery of the stratospheric sudden warming phenomenon, was made by Scherhag (1952) by studying radiosonde ascents from Tempelhof Airport, Berlin, many of which were assimilated by ERA5.

In addition, ERA5 shows the full three-dimensional picture of the related split of the stratospheric polar vortex.



Data-driven NWP

Pangu-Weather: A 3D High-Resolution System for Fast and Accurate Global Weather Forecast

Kaifeng Bi, Lingxi Xie, Hengheng Zhang, Xin Chen, Xiaotao Gu, and Qi Tian✉, *Fellow, IEEE*

[physics.ao-ph] 3 Nov 2022

Abstract—In this paper, we present Pangu-Weather, a deep learning based
this purpose, we establish a data-driven environment by downloading 43 ye
ECMWF reanalysis (ERA5) data and train a few deep neural networks with
of forecast is $0.25° \times 0.25°$, comparable to the ECMWF Integrated Forecas
AI-based method outperforms state-of-the-art numerical weather prediction
RMSE and ACC) of all factors (*e.g.*, geopotential, specific humidity, wind sp
hour to one week). There are two key strategies to improve the prediction a
(3DEST) architecture that formulates the height (pressure level) information
aggregation algorithm to alleviate cumulative forecast errors. In deterministi
short to medium-range forecast (*i.e.*, forecast time ranges from one hour to
downstream forecast scenarios, including extreme weather forecast (*e.g.*, tr
forecast in real-time. Pangu-Weather not only ends the debate on whether A
but also reveals novel directions for improving deep learning weather foreca

Index Terms—Numerical Weather Prediction, Deep Learning, Medium-ran

1 Introduction

Weather forecast is one of the most important scenarios of scientific computing. It offers the ability of predicting future weather changes, especially the occurrence of extreme weather events (*e.g.*, floods, droughts, hurricanes, etc.) which has large values to the society (*e.g.*, daily activ-

learn
to ca
data)
cializ
ods

GraphCast: Learning skillful medium-range global weather forecasting

Remi Lam[*,1], Alvaro Sanchez-Gonzalez[*,1], Matthew Willson[*,1], Peter Wirnsberger[*,1], Meire Fortunato[*,1], Alexander Pritzel[*,1], Suman Ravuri[1], Timo Ewalds[1], Ferran Alet[1], Zach Eaton-Rosen[1], Weihua Hu[1], Alexander Merose[2], Stephan Hoyer[2], George Holland[1], Jacklynn Stott[1], Oriol Vinyals[1], Shakir Mohamed[1] and Peter Battaglia[1]

[*]equal contribution, [1]DeepMind, [2]Google

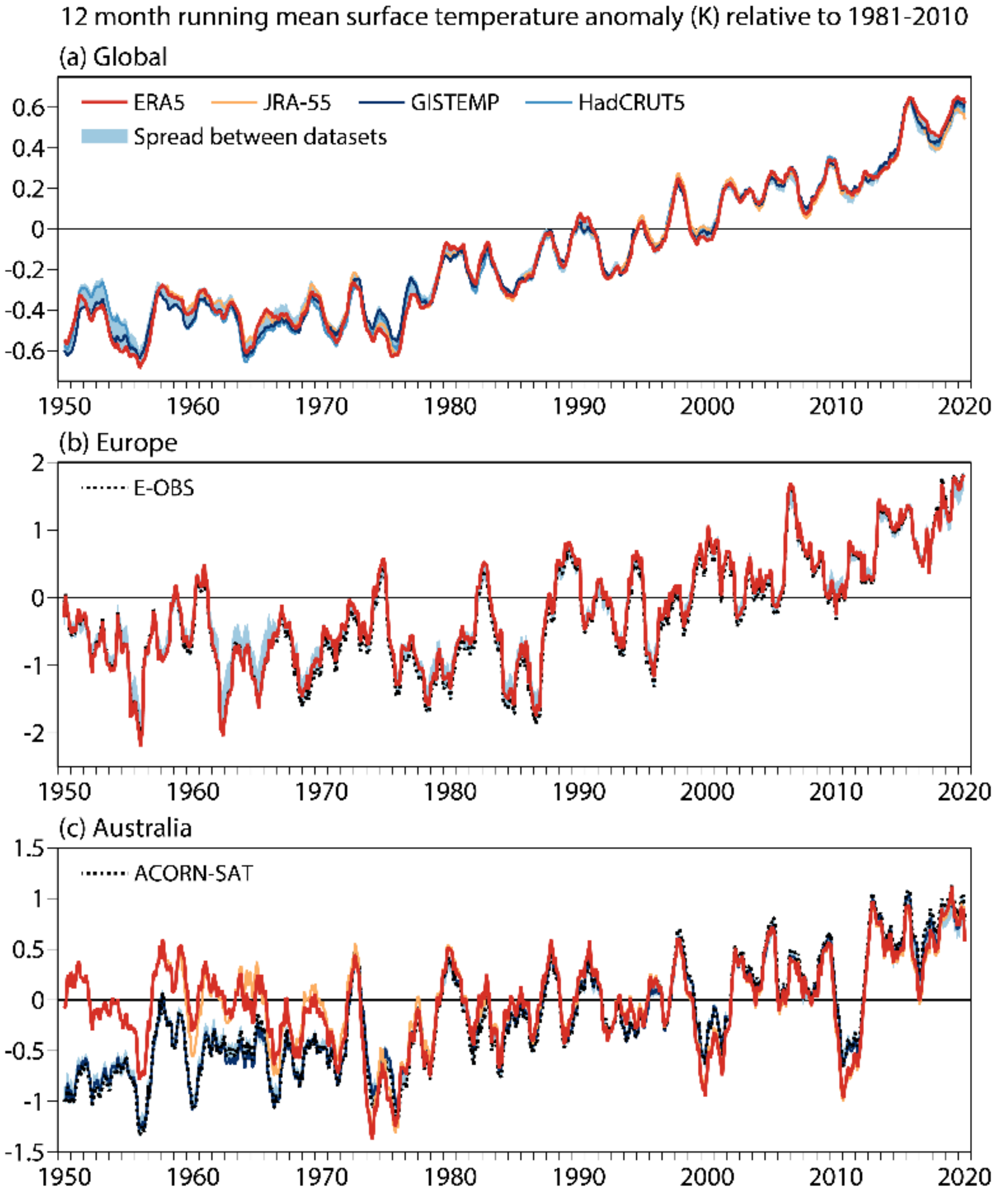
24 Dec 2022

We introduce a machine-learning (ML)-based weather simulator—called "GraphCast"—which outperforms the most accurate deterministic operational medium-range weather forecasting system in the world, as well as all previous ML baselines. GraphCast is an autoregressive model, based on graph neural networks and a novel high-resolution multi-scale mesh representation, which we trained on historical weather data from the European Centre for Medium-Range Weather Forecasts (ECMWF)'s ERA5 reanalysis archive. It can make 10-day forecasts, at 6-hour time intervals, of five surface variables and six atmospheric variables, each at 37 vertical pressure levels, on a 0.25° latitude-longitude grid, which corresponds to roughly 25×25 kilometer resolution at the equator. Our results show GraphCast is more accurate than ECMWF's deterministic operational forecasting system, HRES, on 90.0% of the 2760 variable and lead time combinations we evaluated. GraphCast also outperforms the most accurate previous

Overview

- Introduction reanalysis
- How does reanalysis differ from NWP?
- The ERA5 reanalysis
  - Weather applications
  - **Climate applications**
- Coupled reanalysis
- Summary

Climate change: evolution of 2m temperature and comparison with other datasets



Temperature trends:

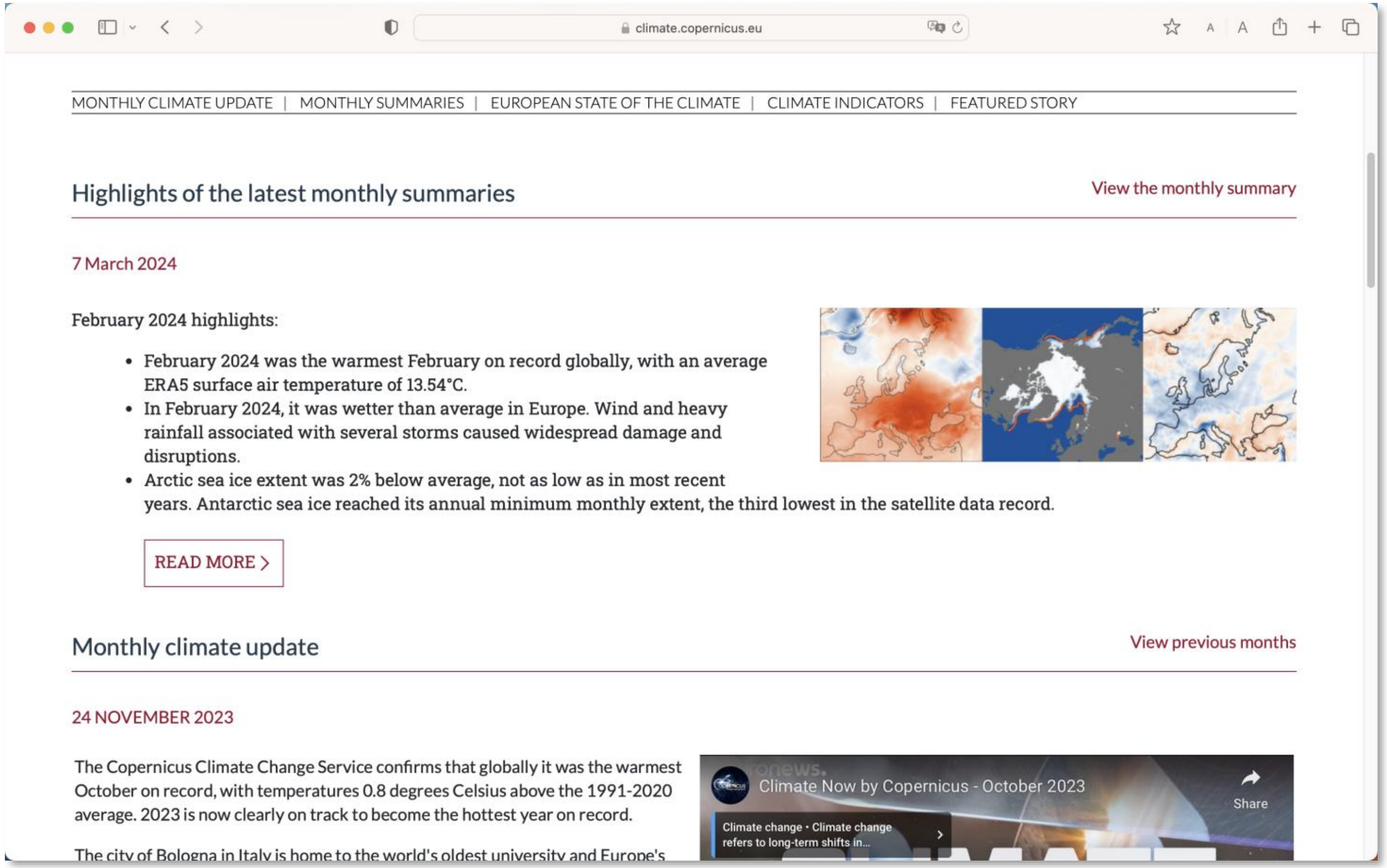
- The global mean temperature shows little trend from 1950 to the mid 1970s.
- After that global warming becomes clearly visible with a global trend of around
  - 0.18 K/decade for 1981-2010
  - 0.24 K/decade for 1991-2020

Consistency between datasets:

- reanalyses and more direct observation-based datasets.

- In general, quite good and reassuring,
  - especially over Europe,
- However, there are some decrepancies,
  - especially over Australia

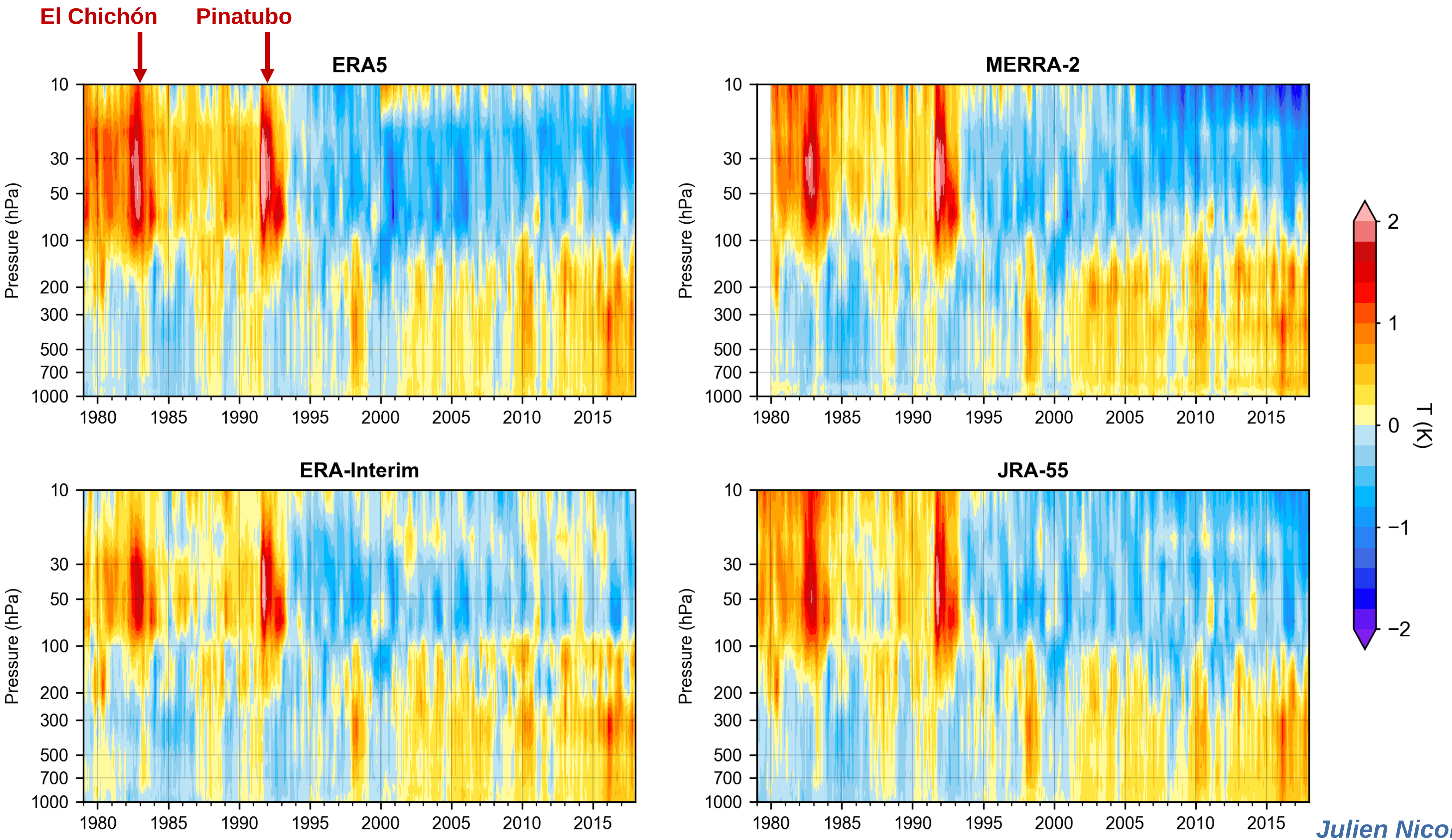
Regular reporting on state of the climate



- **Monthly climate bulletins**
  present the current condition of the climate using monthly maps of key climate change indicators

- **Yearly 'European Ttate Of The Climate' report**
  Provides global context through Climate Indicators for which data are available for the majority of the year.

Intercomparison with reanalysis products from other major centres

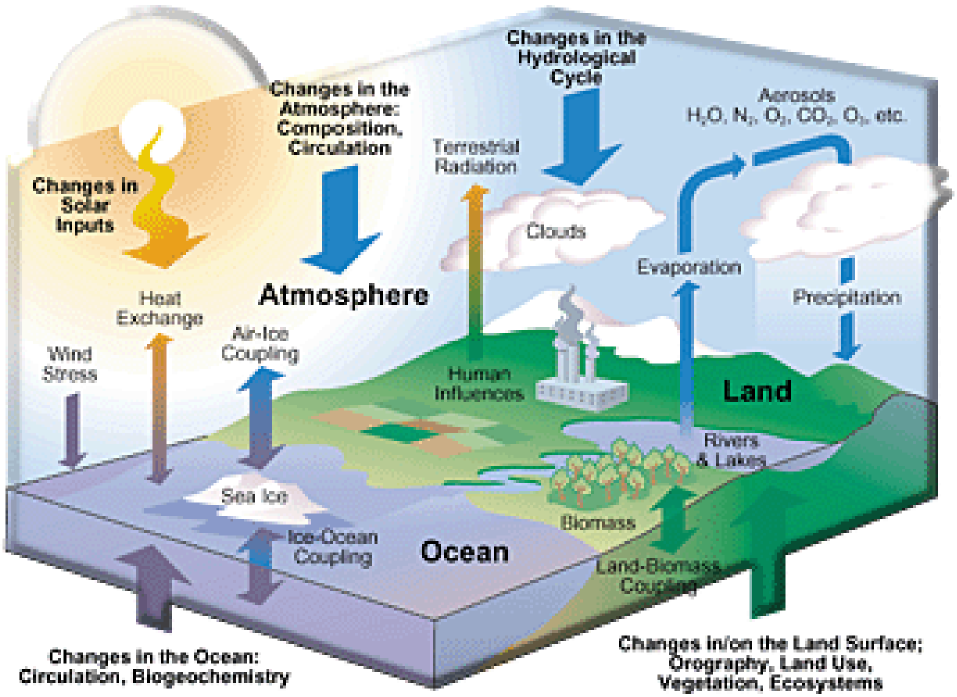


Julien Nicolas

Overview

- Introduction reanalysis
- How does reanalysis differ from NWP?
- The ERA5 reanalysis
    - Weather applications
    - Climate applications
- **Coupled reanalysis**
- Summary

Coupled processes



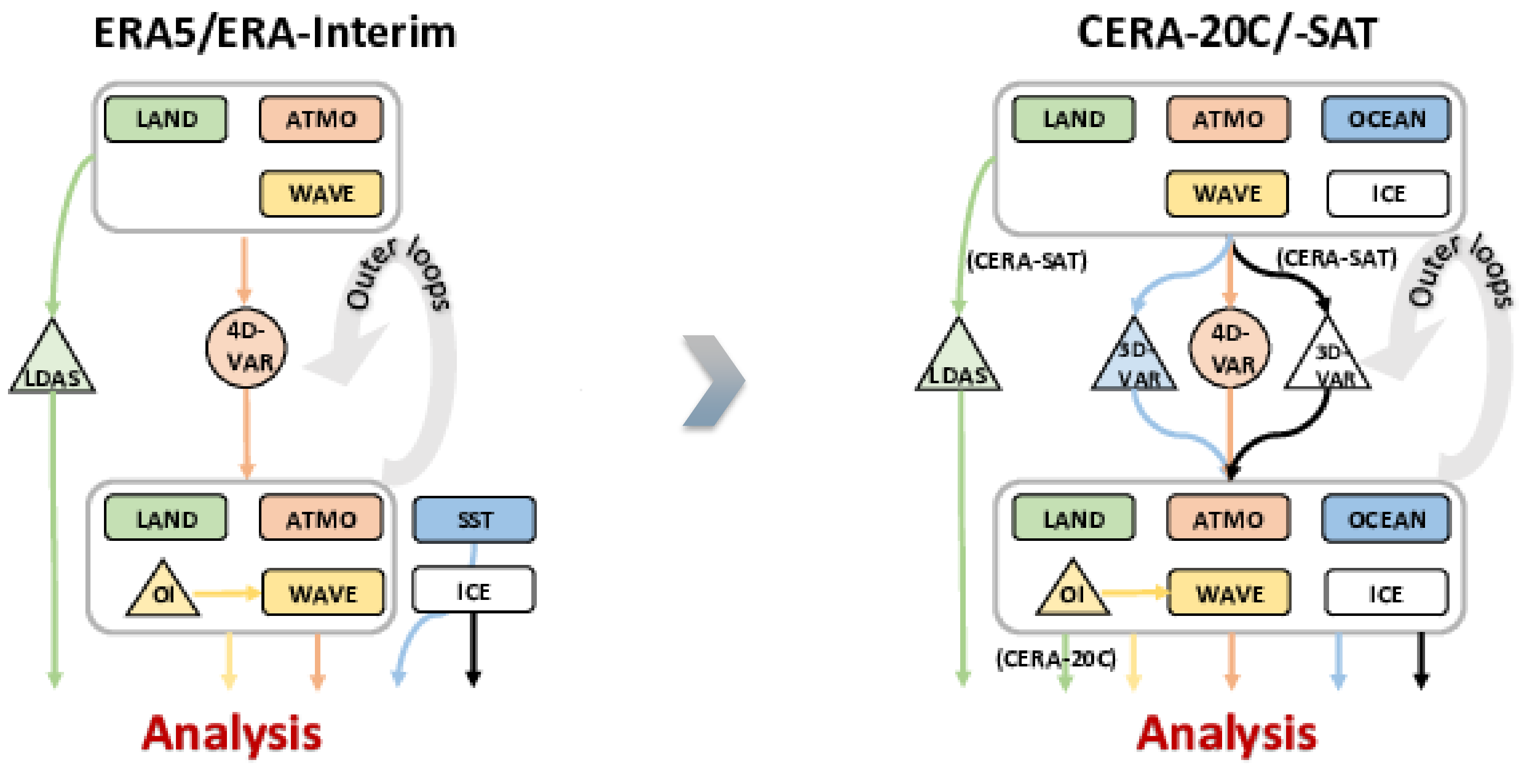
The FP7 ERA-CLIM2 project (2014-2017)

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.*

ERA-CLIM2: towards coupling with the ocean and sea ice



The **ERA-CLIM2** project pioneered the development of an **outer-loop coupled** data-assimilation in climate reanalysis

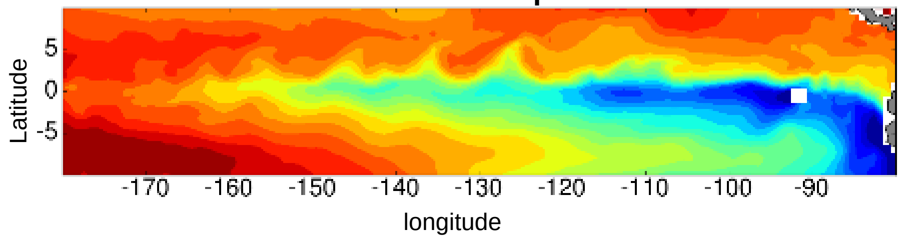
- **CERA-20C:** centennial reanalysis using surface observations only
- **CERA-SAT:** proof of concept for a recent 9-year period using the full observing system at the ERA5 EDA resolution
- Land data assimilation (LDAS) remains **weakly-coupled**

Coupled processes: Tropical instability waves

Tropical instability waves (TIW)

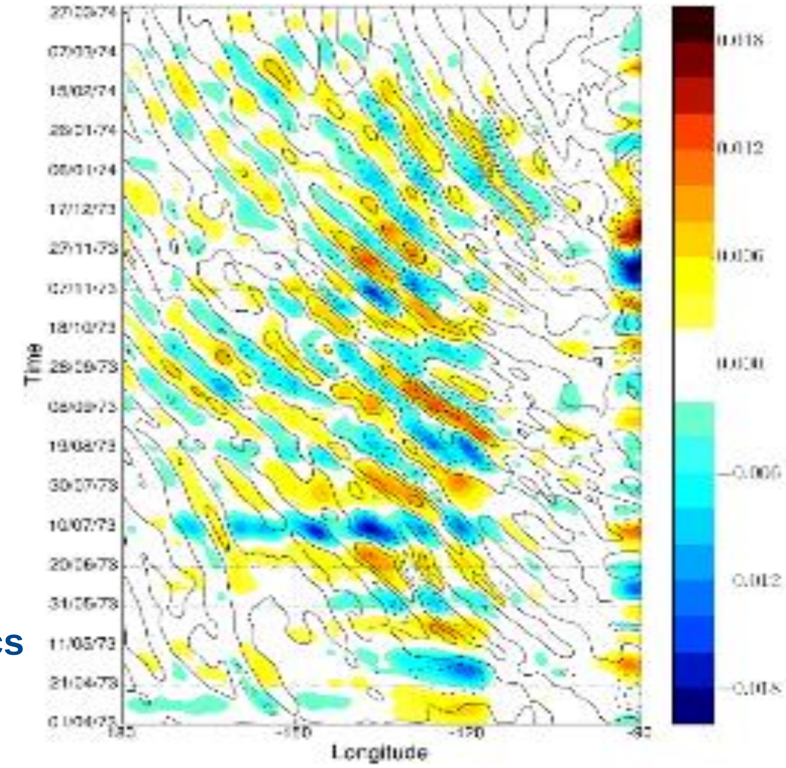
westward-propagating waves near the equator

Sea surface temperature

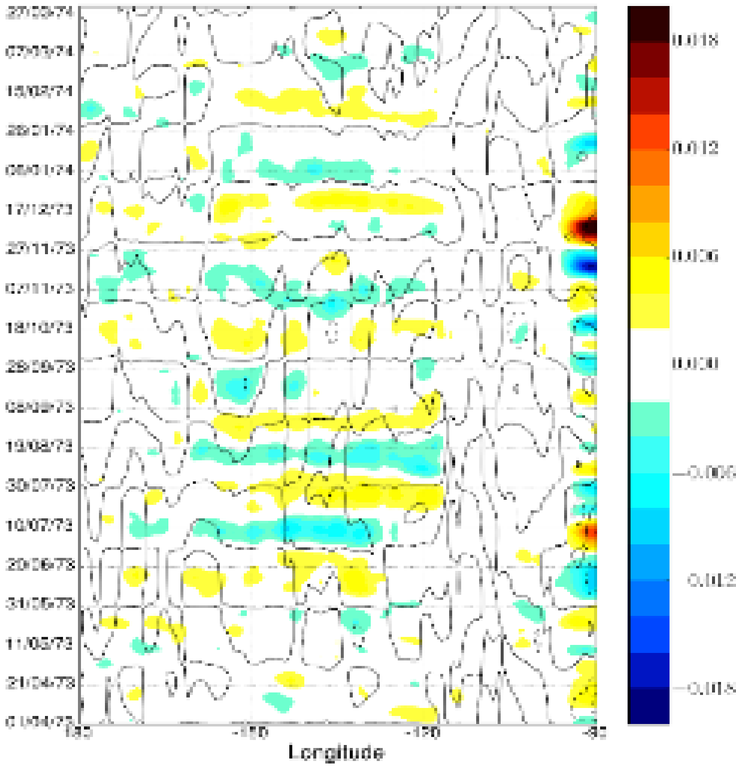


high-pass filtered SST (colour) and wind stress (contour)

CERA-20C



ERA-20C



ERA20C (Forced reanalysis)

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

CERA-20C (Coupled reanalysis)

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

Overview

- Introduction to reanalysis
- How does reanalysis differ from NWP?
- The ERA5 reanalysis
  - Weather applications
  - Climate applications
- Coupled reanalysis
- **Summary**

Summary of important concepts

Reanalyses are extremely popular datasets:

- convenient: transform non-gridded historical observations into
- consistent and accurate 'maps without gaps'
- Many types of users (academic, commercial, policy makers) and applications (weather, climate)
- produced at several centres around the globe

Reanalysis is typically based on a recent NWP data-assimilation system

- for several decades or longer and at lower resolution to keep affordable
- focus is on the analysis, not the forecast (although of course better analyses produce better forecasts)

The long time dimension needs attention:

- Appropriate forcing fields for e.g., radiative forcing, SST, sea ice
- Reprocessing and data rescue of historical observations are important

The data assimilation system is more challenged in the data-sparser past:

- Need to evolve background (and observation) errors accordingly
- Systematic model error has more chance to grow
    - Can affect the mean state and changes as the observing system evolves
    - The method of weak constraint 4D-Var could alleviate this situation

An uncertainty estimate is important to reflect the increasing confidence when more observations become available

- Ensemble spread provides a proxy for the synoptic uncertainty, not the uncertainty of the mean state

In line with the developments in NWP, reanalysis is **progressing towards the Earth system approach**

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.* (2020), “*Global stratospheric temperature bias and other stratospheric aspects of ERA5 and ERA5. 1*”, European Centre for Medium Range Weather Forecasts.

Simmons *et al.* (2021), “on ERA5 surface temperature and humidity”, European Centre for Medium Range Weather Forecasts.

Hersbach, *et al.* *(2020),* "The ERA5 global reanalysis." *Quarterly Journal of the Royal Meteorological Society* 146.730: 1999-2049.

Bell, et al. (2021), “The ERA5 global reanalysis: Preliminary extension to 1950”. Quarterly Journal of the Royal Meteorological Society, https://doi.org/10.1002/qj.4174.

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