Data Assimilation Training

Reanalysis methods

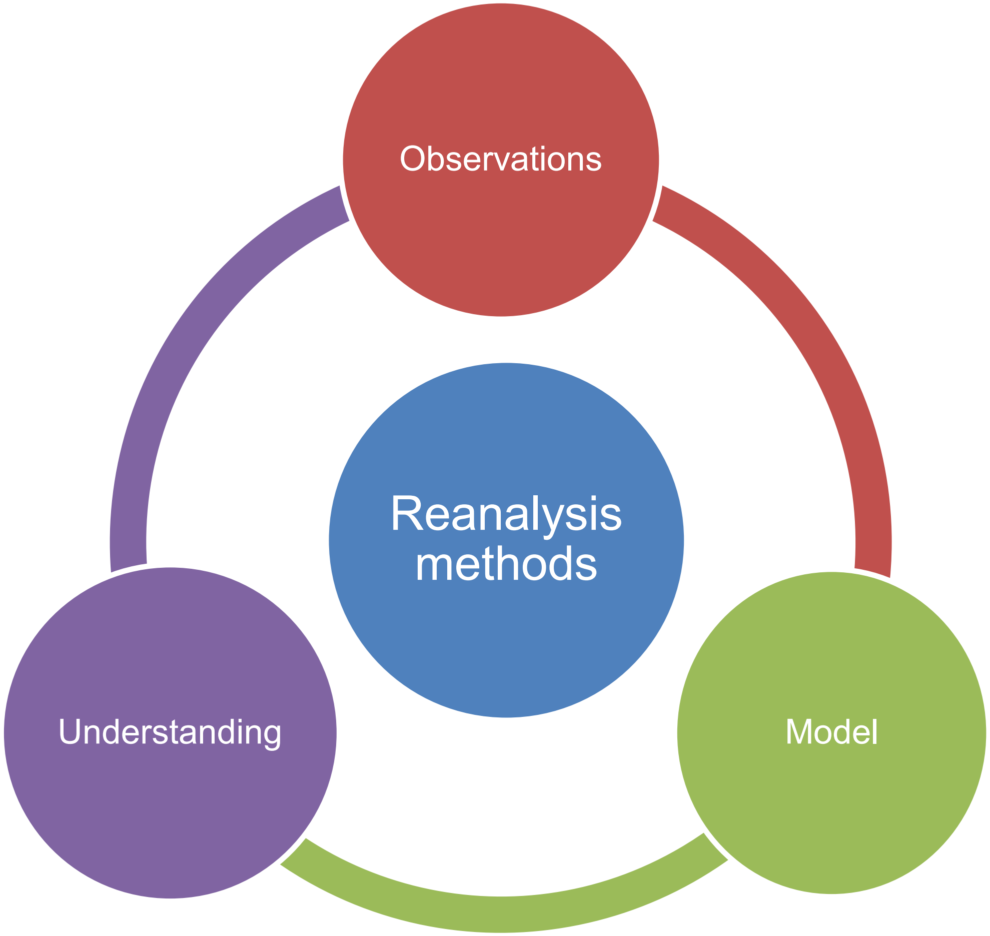
Patrick Laloyaux - Earth System Assimilation Section

patrick.laloyaux@ecmwf.int

Acknowledgement: Paul Poli, Dick Dee, Hans Hersbach, Adrian Simmons

ECMWF

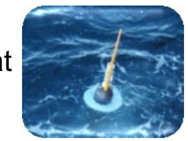
Earth Science has three pillars



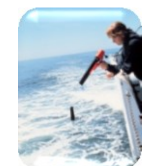
Earth Observations

Measurements from many platforms with different types of sensors


Polar-orbiting Satellite

Argo Float



Geostationary Satellite

Bathythermograph



Aircraft

Buoy



Balloon, Radiosonde

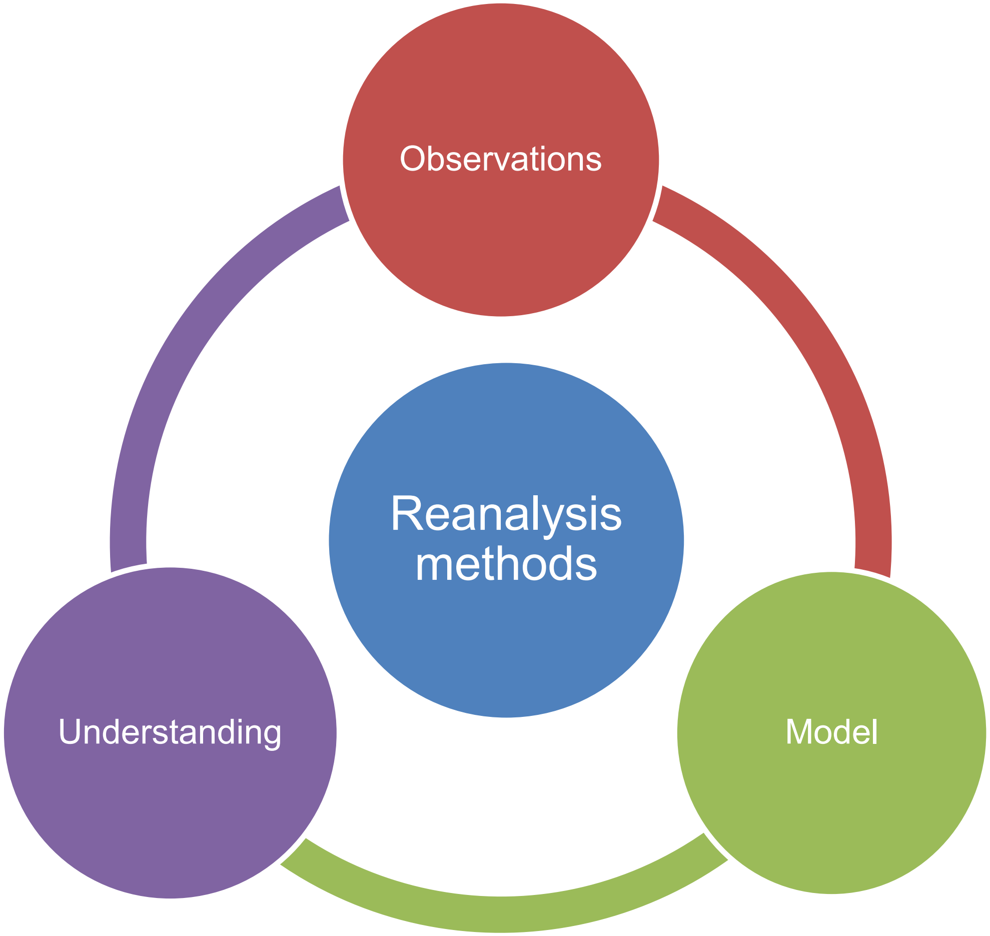
Ship



(Semi-) Automatic Station

Observer,
with instruments

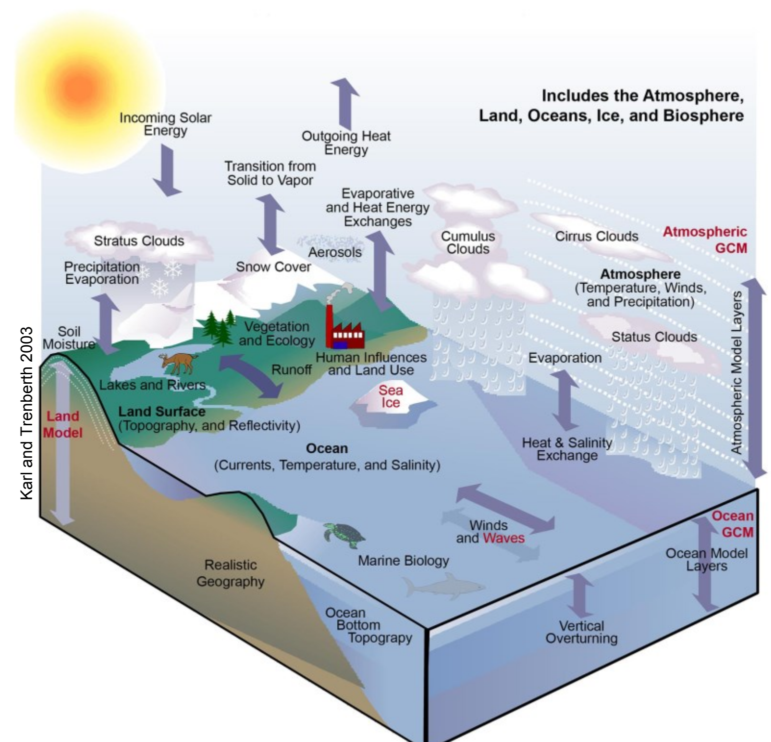

Earth Science has three pillars



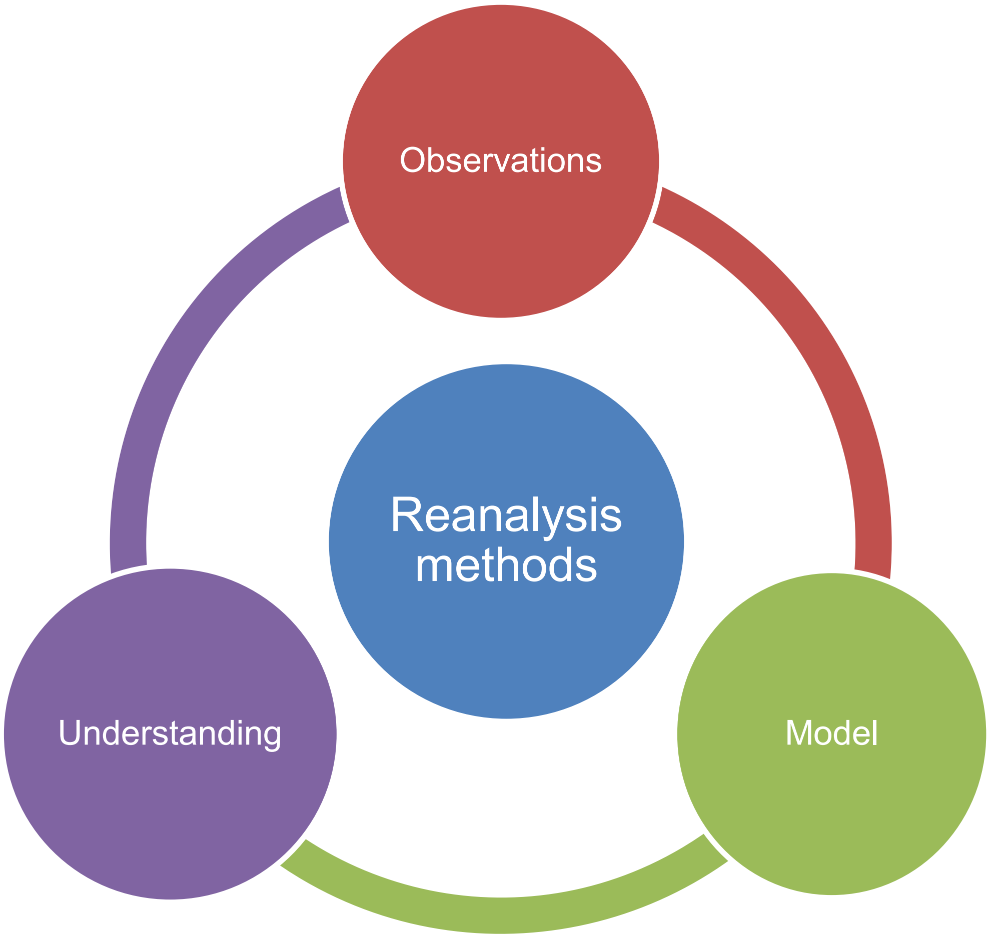
Earth Modelling

State-of-the-art numerical models based on physical laws

- tie geophysical variables
- enforce balance
- ensure mass conservation



Earth Science has three pillars



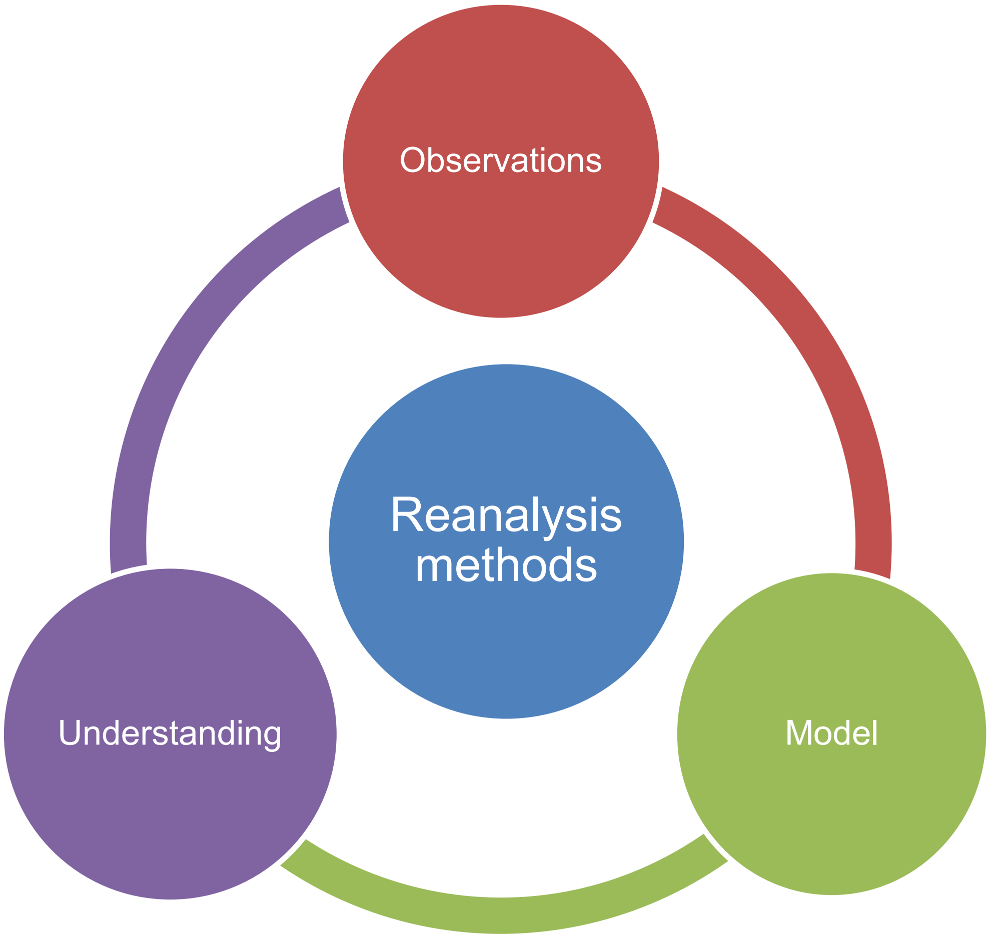
Earth Understanding

Confronting the models with the observations to identify limitations

- imperfections in the observations
- mistakes in model concepts
- systematic errors (bias)

From there, we can improve the instruments, and refine the models (infinite loop)

Earth Science has three pillars



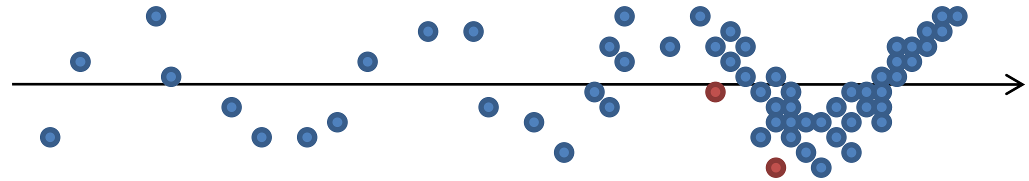
Reanalysis Methods

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

Target:
Reconstruct the past climate and/or weather ensuring consistency
- horizontal and vertical dimensions
- time
- across geophysical variables

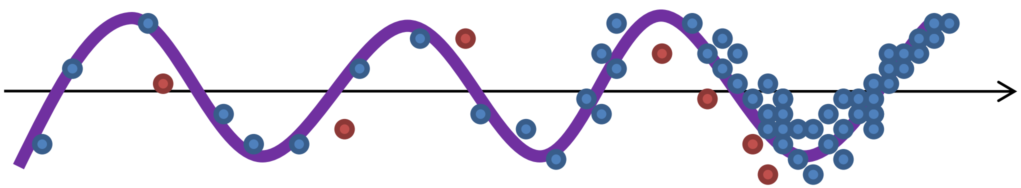
Different methods to reconstruct the past climate and/or weather

“Observations-only” climatology



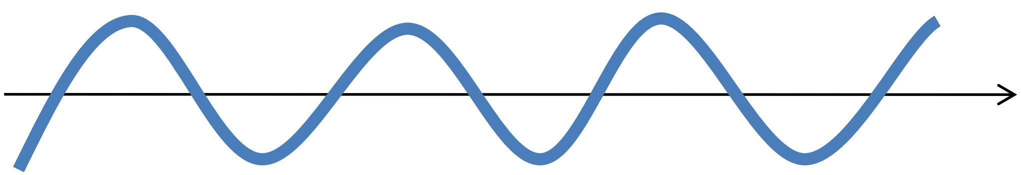
Reconstruction based on observations, little use of model

Reanalysis



Balance between use of observations and model

“Model only” integration

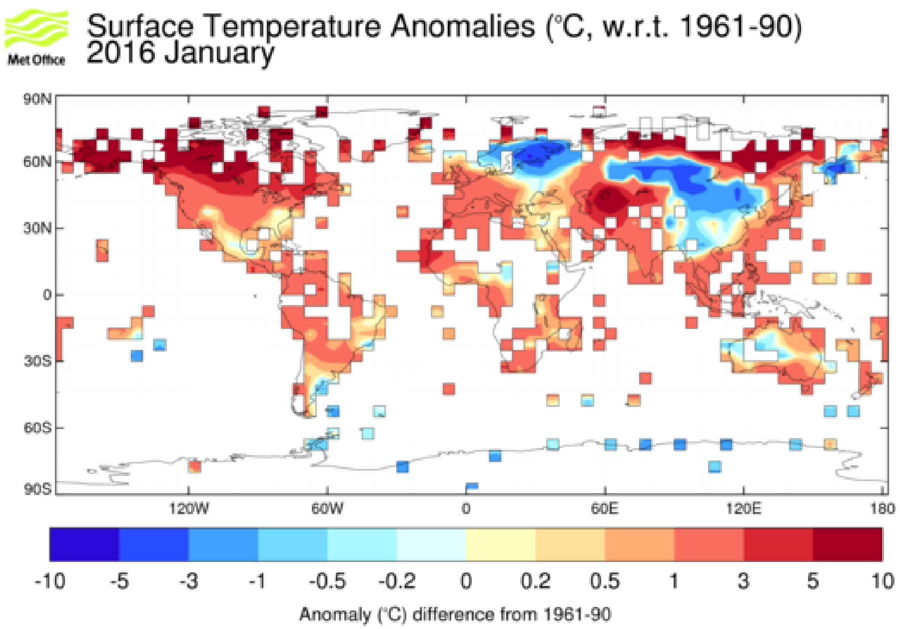


Reconstruction based on a model, little use of observations

"Observations-only" climatology - CRUTEM4

Observations: archive of monthly mean temperatures provided by 5500 weather stations distributed around the world

Method: each station temperature is converted to an anomaly from the 1961-90 average temperature for that station. Each grid-box value is the mean of all the station anomalies within that grid box.

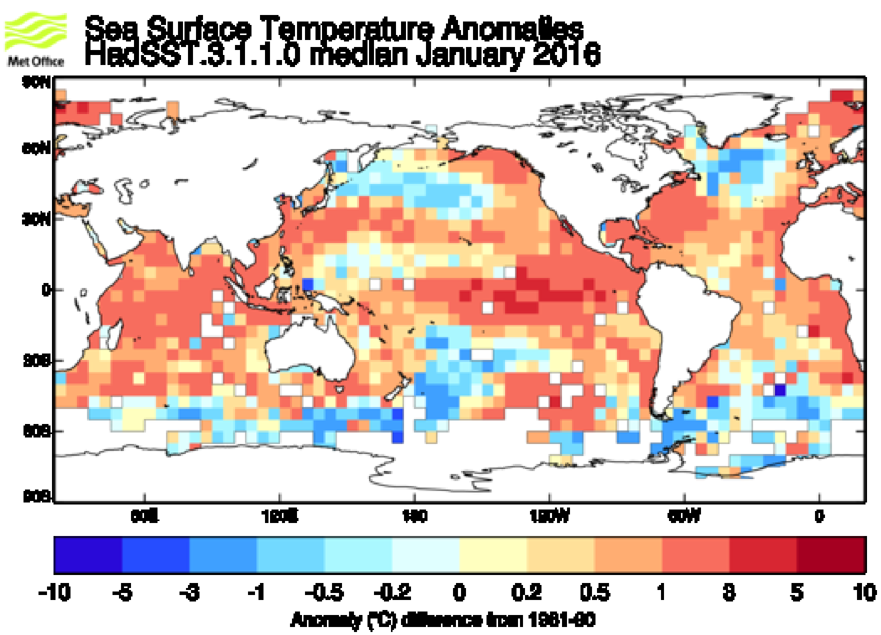


A gridded dataset (5 degree grid) of historical near-surface air temperature anomalies over land available for each month from January 1850 to present

“Observations-only” climatology - HadSST3

Observations: in-situ measurements of Sea Surface Temperature (SST) from ships and buoys coming from ICOADS and GTS archives

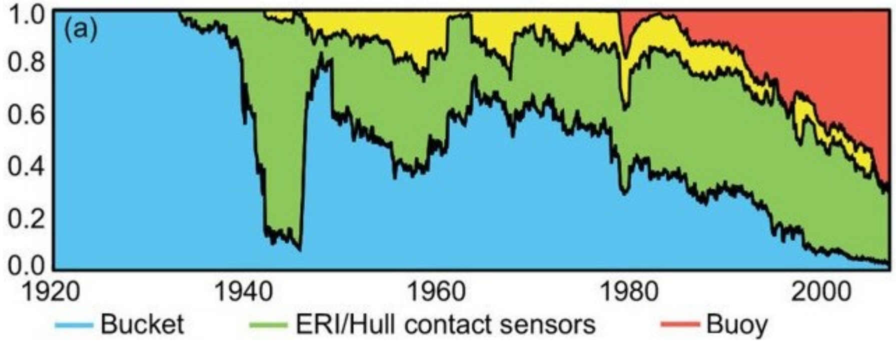
Method: the measurements are converted to anomalies. Bias adjustments to reduce the effects of spurious trends caused by changes in SST measuring practices



A gridded dataset (5 degree grid) of historical SST anomalies available for each month from January 1850 to present

Bias adjustment of Sea Surface Temperature in HadSST3

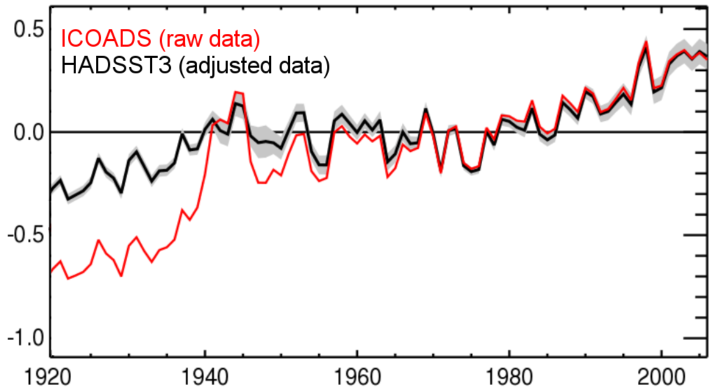
Evolution of instruments for SST measurements



Different instruments and sampling methods lead to different observation biases

- buckets have cold biases
- ERI have small warm biases

Annual SST anomalies (relative to 1961-1990)



Spurious climate signal in the raw data due to a change in the observing system

Raw data are bias corrected to produce **HADSST3**

Adjustments suffer from

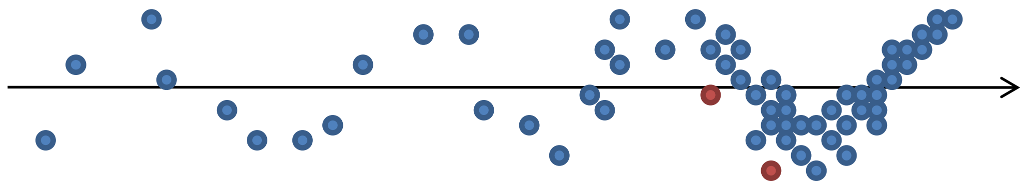
- poor documentation of sampling characteristics
- no proper overlap for intercomparison

“Observations-only” climatology:

- long record (extending back to 1850), based only on observations
- spatial and temporal discontinuities
- no use of a NWP model, but average operator, bias correction, QC, interpolation

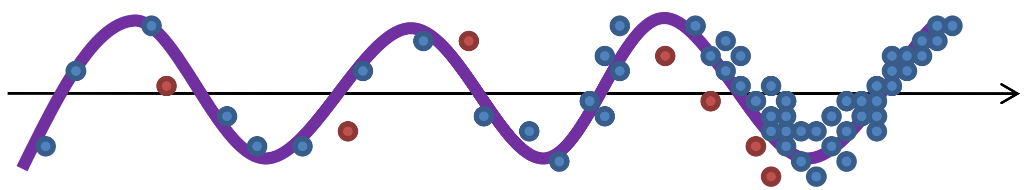
Different methods to reconstruct the past climate and/or weather

“Observations-only” climatology



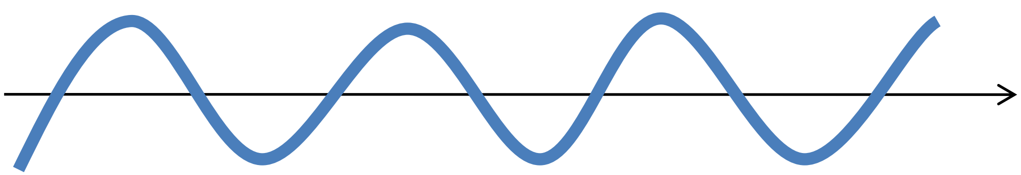
Reconstruction based on observations, little use of model

Reanalysis



Balance between use of observations and model

“Model only” integration



Reconstruction based on model, little use of observations

"Model only" integration - ERA-20CM

Model: the IFS atmospheric model developed for NWP at low resolution (125 km)

Method: the model is integrated from 1900 to 2010. Observations are not assimilated but the model is constrained by atmospheric forcings

CMIP5 atmospheric forcing are used:

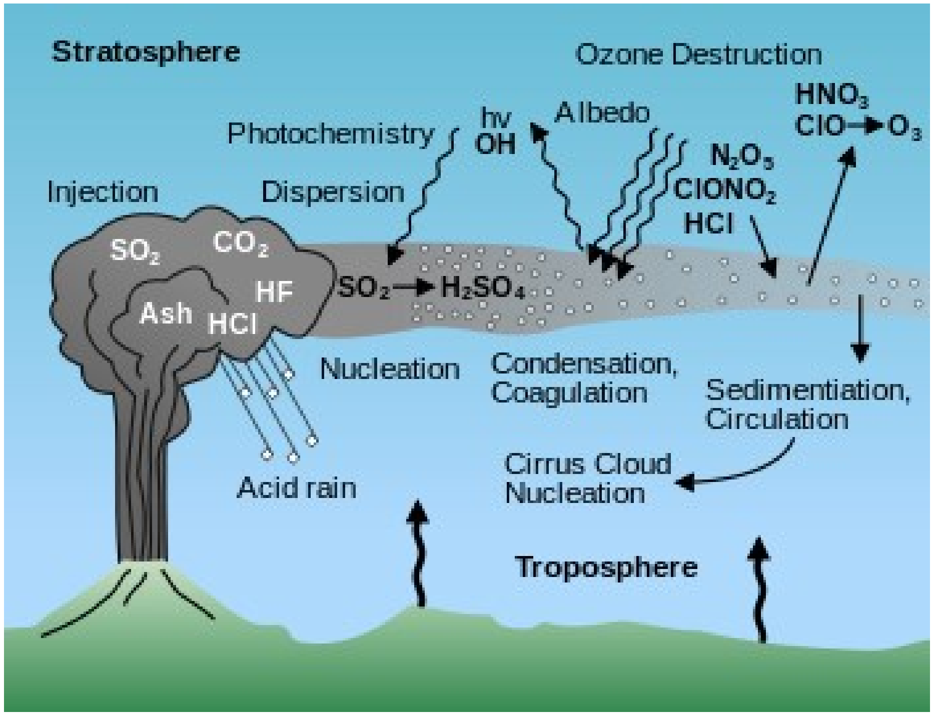
- Solar irradiance (CMIP5)
- Greenhouse gases (CMIP5)
- Ozone for radiation (CMIP5)
- Tropospheric aerosols (CMIP5)
- Stratospheric aerosols (CMIP5)
- Sea-surface temperature and sea-ice cover (Hadley Centre)

These forcings are based indirectly on observations

The example of stratospheric aerosols

Stratospheric aerosols (optical depth) mainly have a volcanic origin

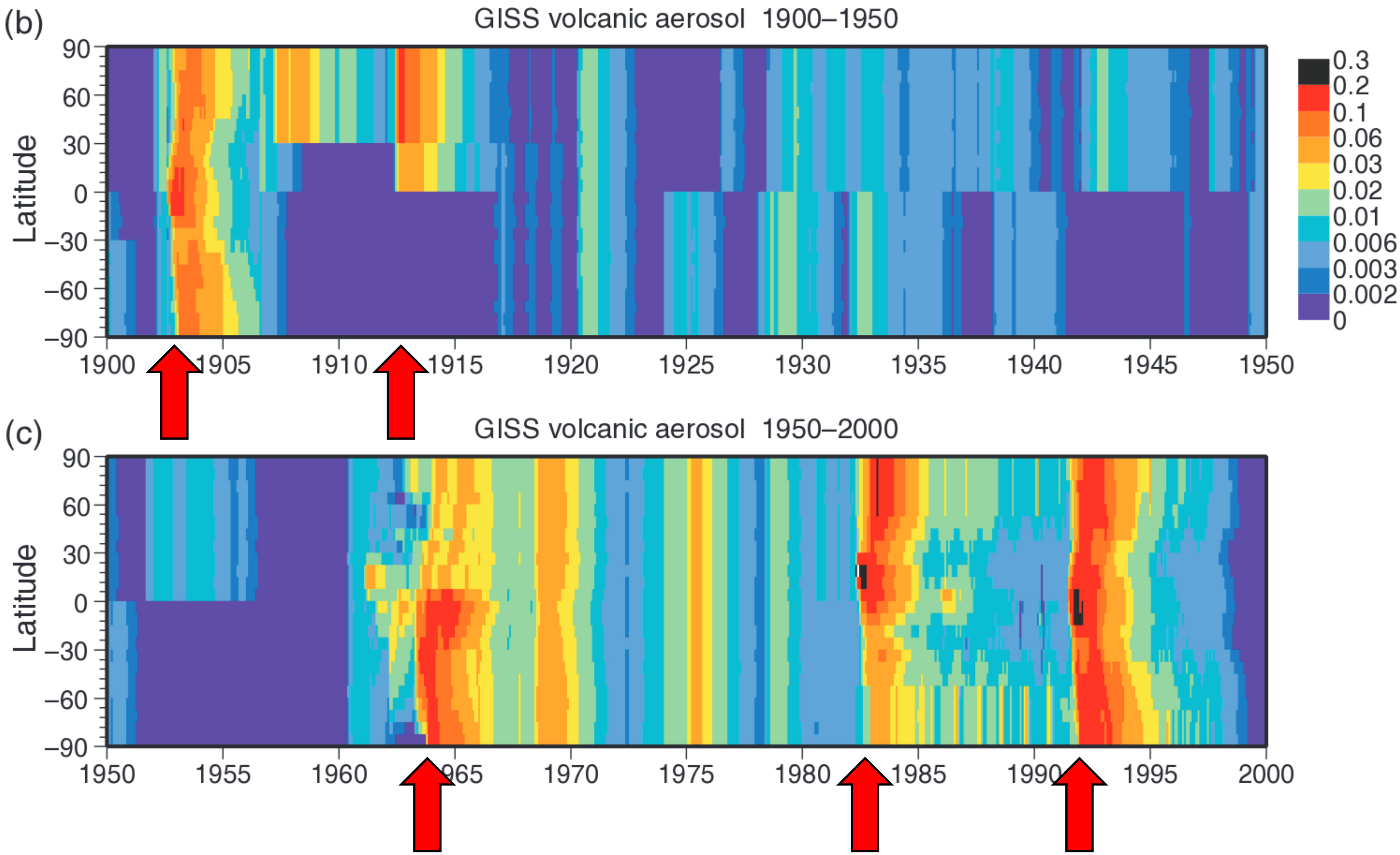
Volcanic sulphate can remain in the stratosphere for many months, where it mixes within large predominantly zonal bands, increasing atmospheric opacity



In the IFS model used in operation, volcanic sulphate is assumed to be constant (evenly distributed over the stratosphere, assuming a constant volume-mixing ratio)

Stratospheric aerosols in ERA-20CM

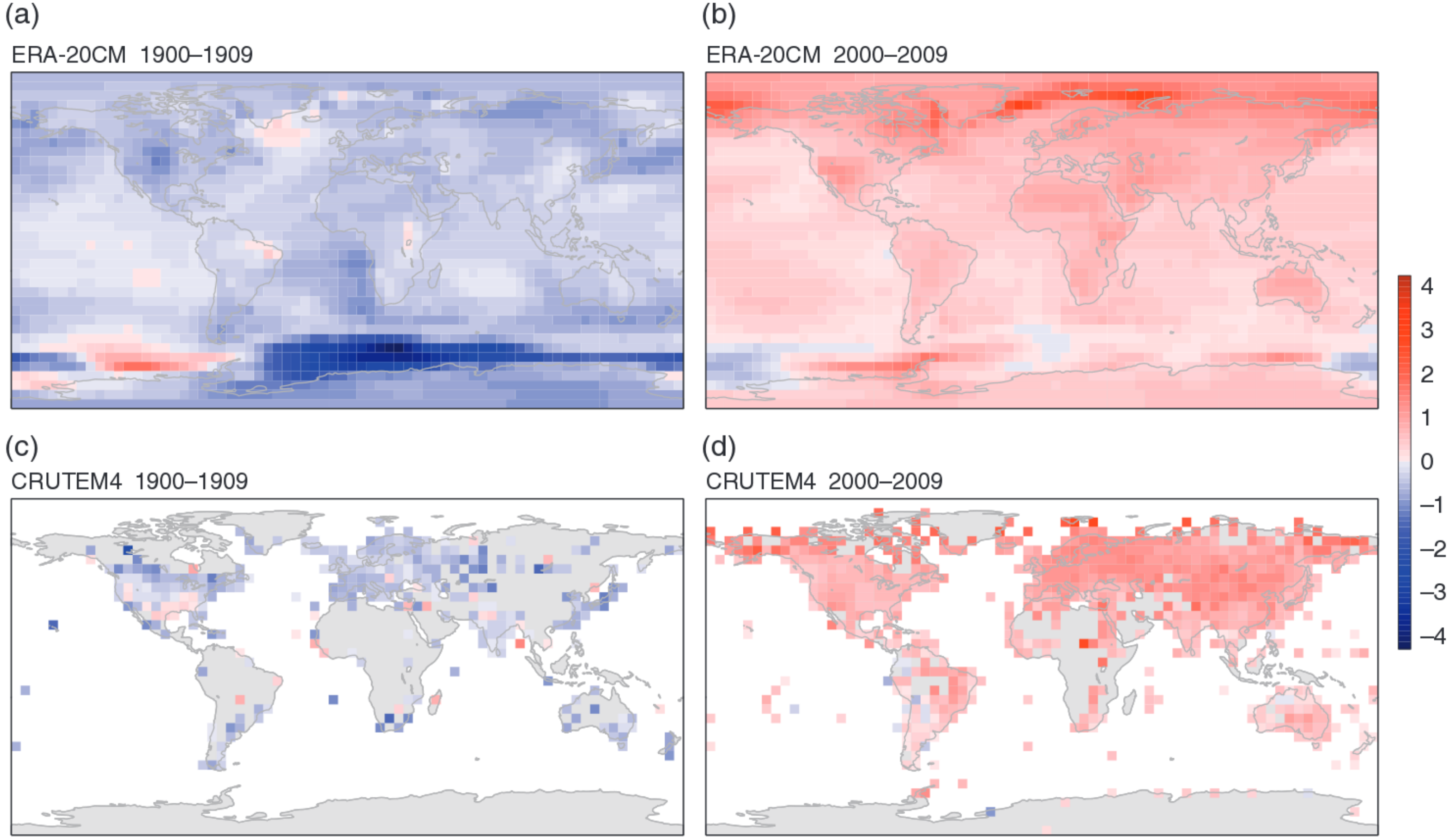
CMIP5 dataset reconstructs the evolution of volcanic sulphate (1850 to present)



Major eruptions are clearly visible: Santa Maria (1902), Novarupta (1912), Agung (1963), Fernandina (1968), El Chichon (1982) and Pinatubo (1991)

Comparison between ERA-20CM with CRUTEM4

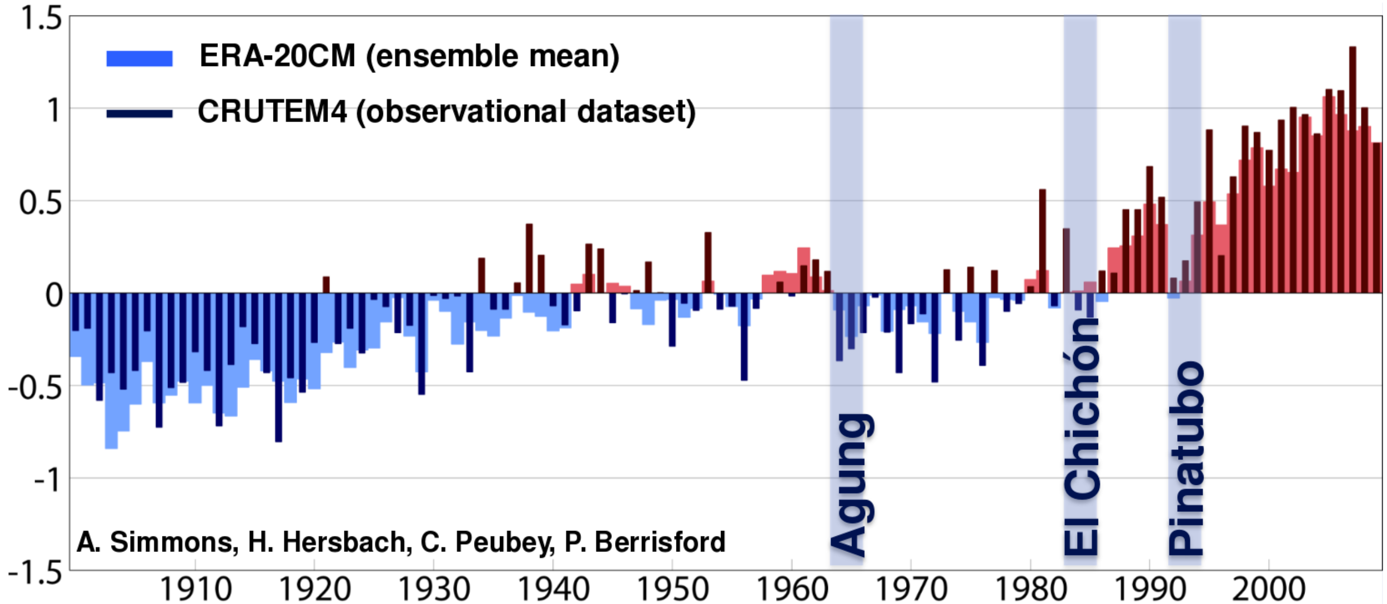
Temperature anomalies for 1900-1909 (left) and 2000-2009 (right)



Similar global warming in the “model-only” and the “observation-only” reconstructions
Differences in Southern United States for 1900-1909

Comparison between ERA-20CM with CRUTEM4

Annual mean anomalies for ERA-20CM (light) and CRUTEM4 (dark)



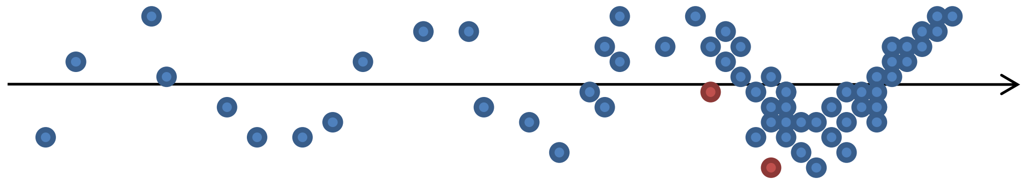
ERA-20CM reproduces the long term variation and capture interannual variability after volcanic eruptions

“Model only” integration:

- long record (extending back to 1900), based on a forced NWP model
- space and time consistency
- capture interannual variability, not expected to reproduce actual synoptic weather

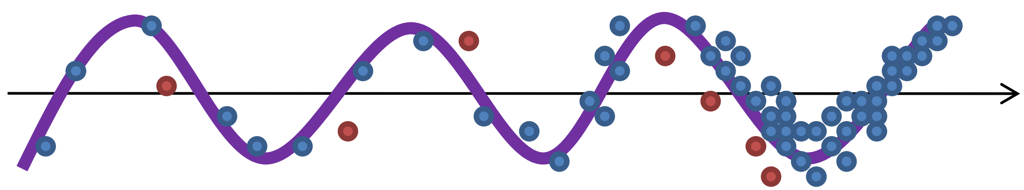
Different methods to reconstruct the past climate and/or weather

“Observations-only” climatology



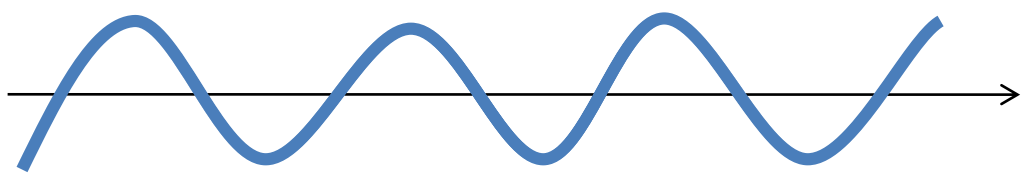
Reconstruction based on observations, little use of model

Reanalysis



Balance between use of observations and model

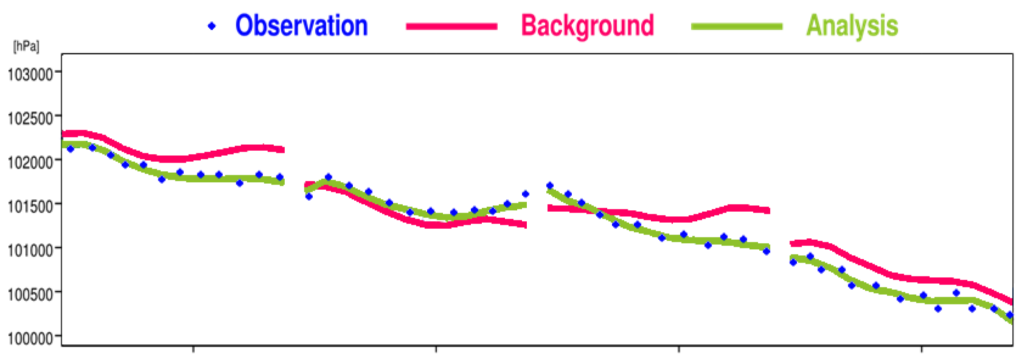
“Model only” integration



Reconstruction based on model, little use of observations

Reanalyses - ERA-Interim

- use an invariant version of a data assimilation system and NWP model
- use as many observations as possible, including from satellites
- run over a long time period (1979 to present)



Model and assimilation system from cycle 31R2 (released in December 2006, deterministic 4D-Var)

prior state constraints

prior parameter constraints

$$J(\mathbf{x}, \beta) = (\mathbf{x}_b - \mathbf{x})^T \mathbf{B}_x^{-1} (\mathbf{x}_b - \mathbf{x}) + (\beta_b - \beta)^T \mathbf{B}_\beta^{-1} (\beta_b - \beta) + [\mathbf{y} - \mathbf{b}_o(\mathbf{x}, \beta) - \mathbf{h}(\mathbf{x})]^T \mathbf{R}^{-1} [\mathbf{y} - \mathbf{b}_o(\mathbf{x}, \beta) - \mathbf{h}(\mathbf{x})]$$

observational constraints

For each 12-hour assimilation window, 4D-Var computes an estimate for all the model outputs and prognostic variables at any given time

Reanalyses

Reanalysis uses a NWP model (physical laws) which allows to:

- deal with missing data
  (the model is used to "fill in the blanks", from past and neighbouring information)
- ensure consistency in horizontal and vertical dimensions
- ensure consistency across geophysical variables

Reanalysis uses and evaluates all observations in a consistent way:

- observation accuracy explicitly taken into account
- quality control (QC) procedures apply across all observation types

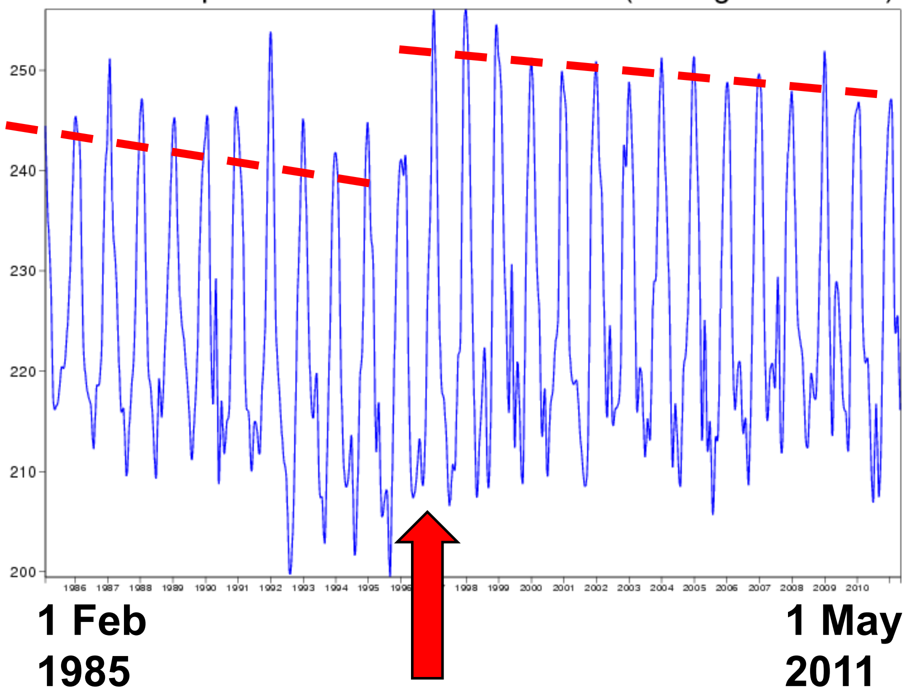
Reanalysis uses the widest variety of observations (40,000 millions in ERA-Interim)

- reduce the model biases
- reproduce actual synoptic situation

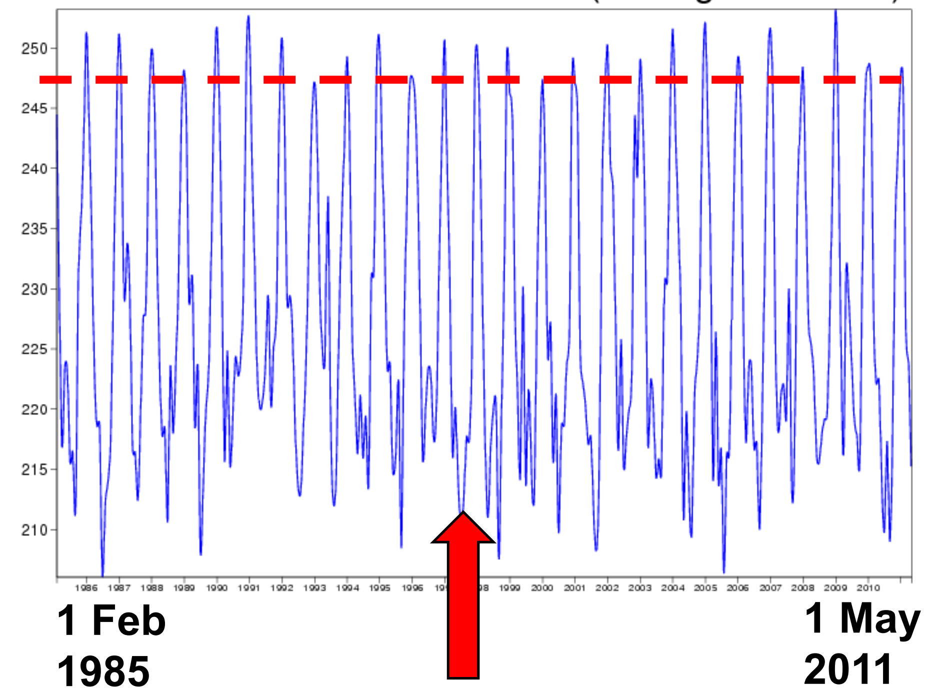
Why not use simply operational NWP analysis?

The models and data assimilation methods have improved a lot over time, so analysis timeseries feature spurious changes.

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



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

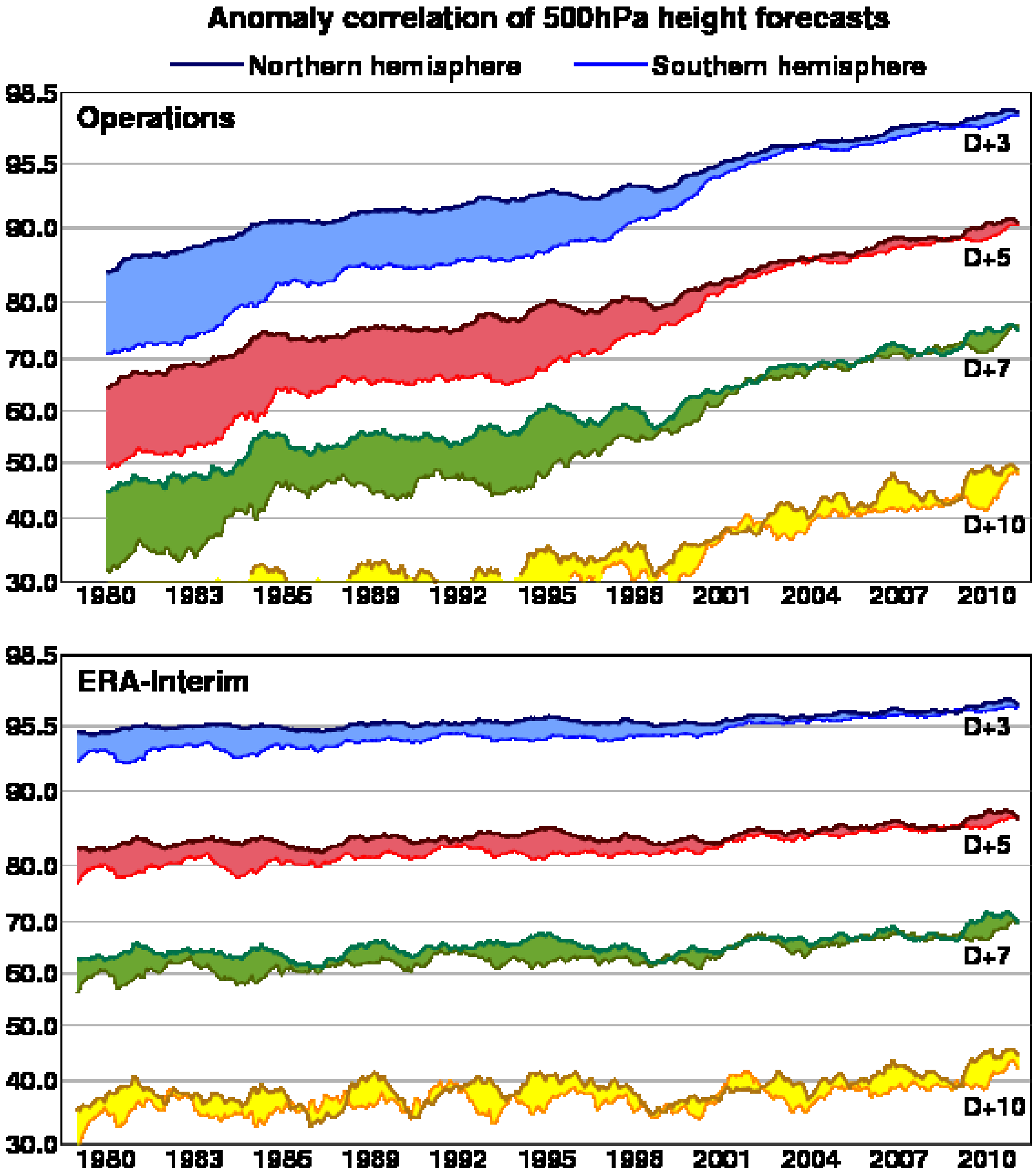


Was there a sudden change in South Pole summer variability in 1997?

No.

To remove these spurious sources of variability, model and data assimilation systems are frozen and rerun to produce a reanalysis dataset

Reanalysis supports NWP development and evaluation



Rate of improvements in operations:

- model
- data assimilation
- observing system

Rate of improvement in reanalysis:

- observing system

The comparison shows that most of the improvements in operational forecasts comes from a better model and data assimilation system

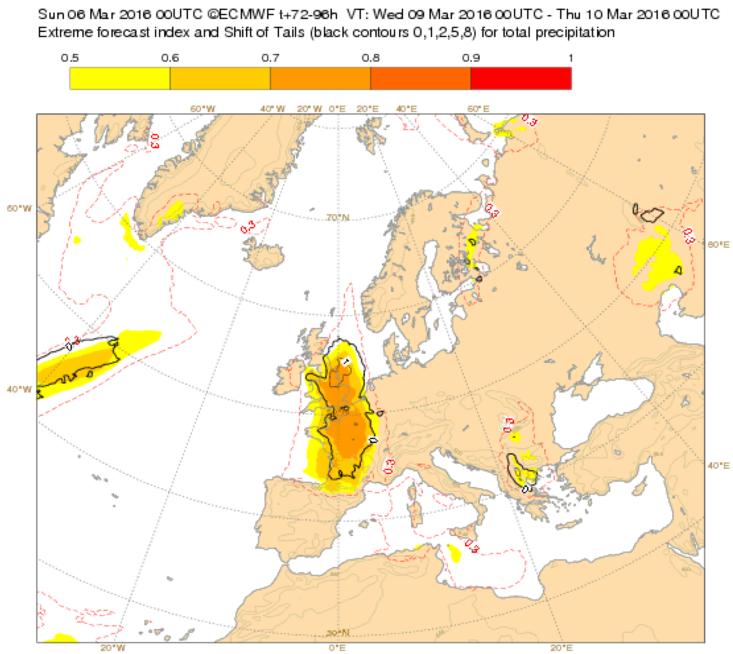
These improvements benefits from better observations

Reanalysis supports the computation of operational products

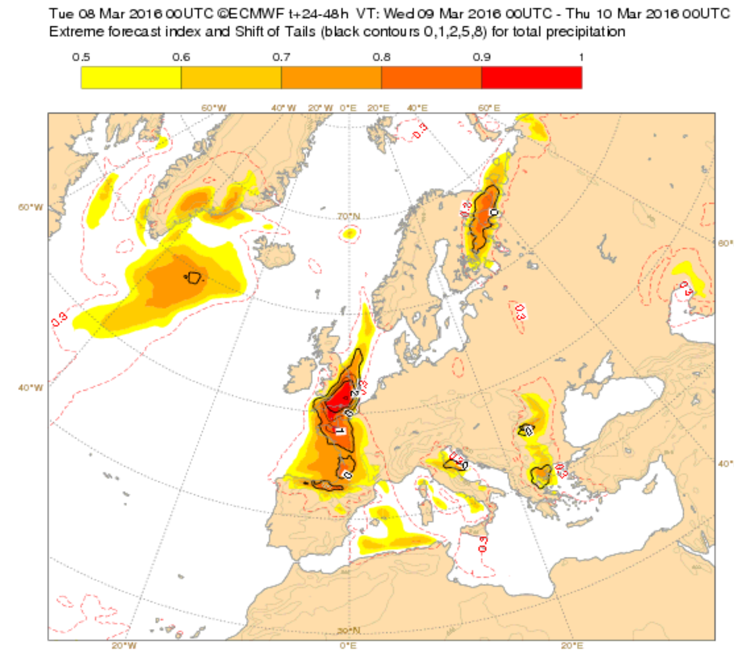
Extreme Forecast Index (EFI) detects extreme events in a given ensemble forecast.

Difference between the ensemble forecast distribution and a reference (M-climate)

- reference distribution is an ensemble of re-forecast for the most recent 20 years
- initial conditions taken from ERA-Interim (atmosphere) and ORAS4 (ocean)



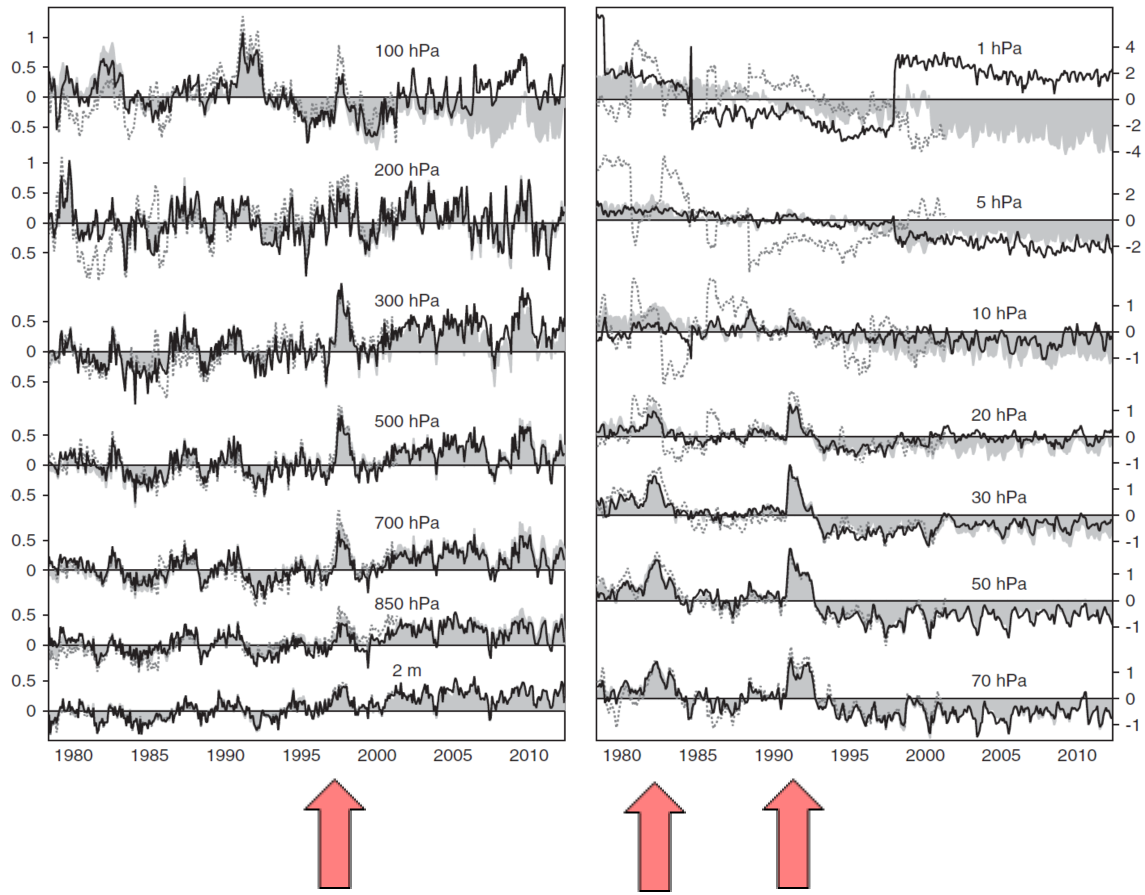
EFI for precipitation for last Wednesday (issued last Sunday, warning 4-days in advance)



EFI for precipitation for last Wednesday (issued last Tuesday, warning 1-days in advance)

Climate signals in reanalysis

Solid line: ERA-Interim temperature anomalies relative to 1979–2001 (monthly and globally averaged)



Good trends for surface temperature

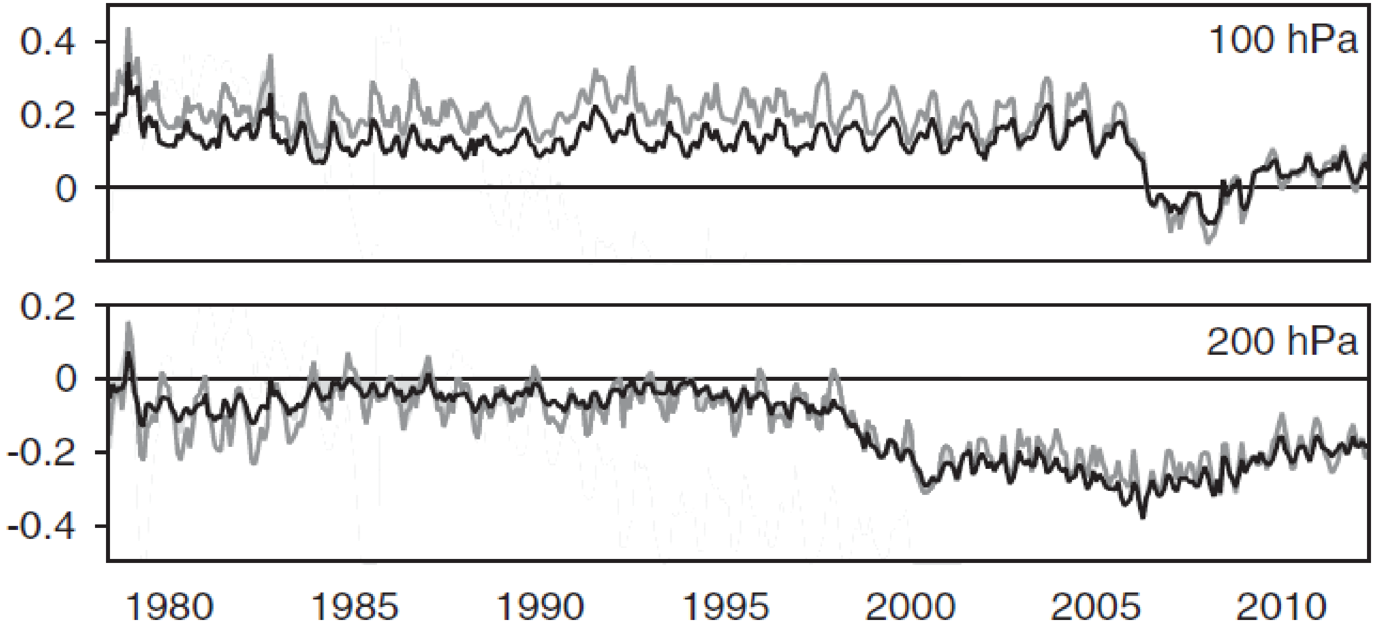
El-Nino, El Chichon and Pinatubo events

Issue at 1hPa with the introduction of a new satellite (AMSU-A)
→ Improve the use of these observations to improve climate signal (e.g. Bias correction)

Some climate signals may be affected by changes in the observing system

Use departure statistics to detect biases

ERA-Interim observation-minus-analysis (black lines) and observation-minus-background (grey lines) differences for radiosonde temperatures (K)

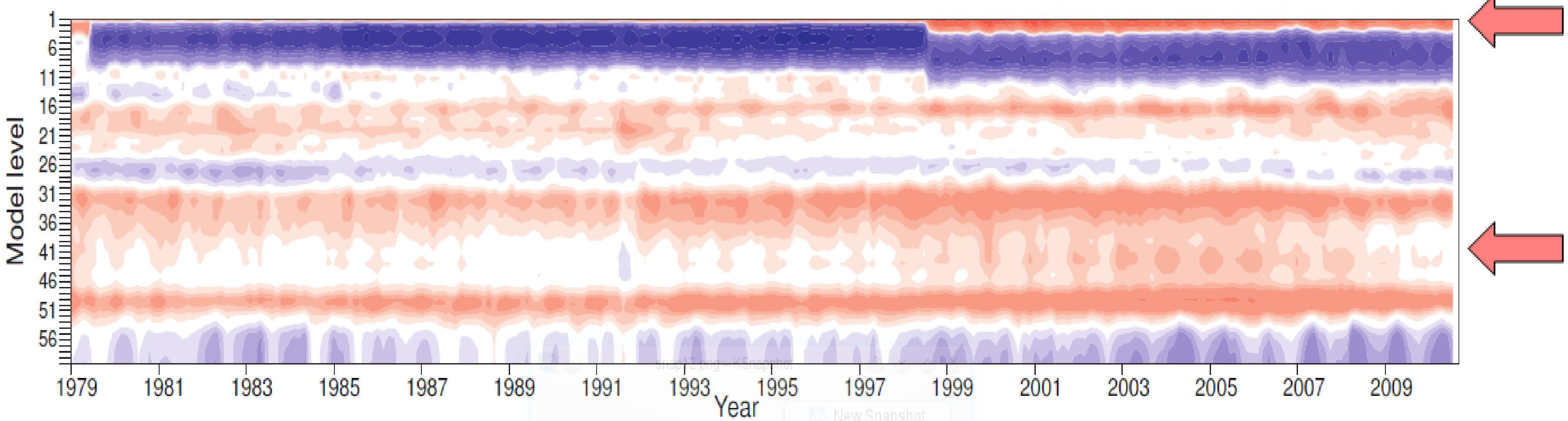


100hPa: before 2006, cold bias in the analysis relative to the assimilated radiosondes. Assimilation of GPSRO data from late 2006 onwards brings the analysis into much closer agreement with radiosondes

200hPa: in the late 1990s, warm bias in the analysis relative to the assimilated radiosondes. This is associated with the assimilation of an increasing amounts of temperature data from commercial aircrafts (warm-biased)

Use analysis increments to detect biases

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



Warm increments produced by the assimilation of AMSUA measurements in the stratosphere and by the assimilation of aircraft measurements in the troposhere

We can explain most of the features by now (putting together many time-series of increments, observation mean departures and bias corrections).
In practice, it is about near to impossible to know how mean increments will turn out in advance in a reanalysis.

ERA-Interim will be replaced by ERA-5

Reanalysis is worth repeating as all ingredients continue to evolve:

- model, data assimilation, observation reprocessing
- with each new reanalysis, understanding of model/observations biases is improved

ERA-Interim: model and assimilation from cycle 31R2 (released in December 2006)
ERA-5: model and assimilation from cycle 41R2 (released in March 2016)

ERA-5 took on board 10 years of research and development in NWP model and data assimilation methods:

- EDA technique
- surface analysis (SEKF),
- better bias correction
- improved model physics
- higher resolution,…

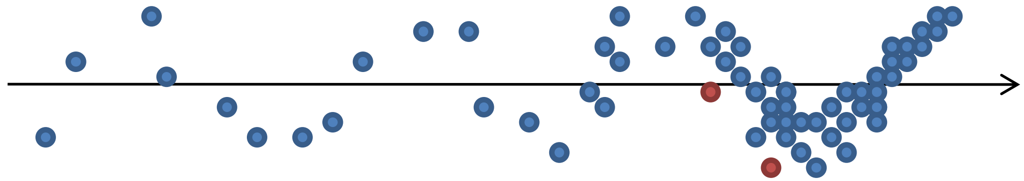
Improved observations: **34** observation data records either reprocessed or updated

SSMI, *CM SAF*	ASMRE, *JAXA*	TMI, *JAXA*	AVHRR POES AMV, *CIMSS*
AVHRR METOP AMV, *EUMETSAT*	MFG CSR, *EUMETSAT*	MSG ASR, *EUMETSAT*	GMS & GOES-9 & MTSAT CSR, *JMA*
GOES post-1995 AMV, *CIMSS*	MFG AMV, *EUMETSAT*	MSG AMV, *EUMETSAT*	GMS & GOES-9 & MTSAT AMV, *JMA*
ASCAT METOP sigma0, *EUMETSAT*	QuikSCAT sigma0, *KNMI* or *OSI-SAF?*	GPSRO METOP, *EUMETSAT*	GPSRO COSMIC, *UCAR*
GPSRO CHAMP, *UCAR*	GPSRO GRACE, *UCAR*	GPSRO SAC-C, *UCAR*	GPSRO TSX, *UCAR*
SWH ERS-1, *ESA*	SWH ERS-2, *ESA*	SWH ENVISAT v2.1, *ESA*	TOMS v8.0 Ozone total column, *NASA*
GOME Ozone profile, *ESA CCI*	GOME-2 Ozone profile, *ESA CCI*	MIPAS Ozone profile, *ESA CCI*	MLS Ozone profile, *NASA*
OMI Ozone total column, *KNMI*	SBUV v8.6 Ozone profile, *NASA*	SCIAMACHY Ozone total column, *ESA CCI*	ASCAT soil moisture, *H-SAF*
Upper-air RS & pilot balloons, *NCAR*	Ship and buoys, *NOAA*		

Brightness Temp. (6)	**Sigma0 (2)**	**Bending Angle (6)**	**Soil moisture (1)**	**Ozone (8)**	**Signif. wave height (3)**	**Atmos. mot. vector (6)**	In-situ T/Q/P/U/V (2)

Different methods to reconstruct the past climate and/or weather

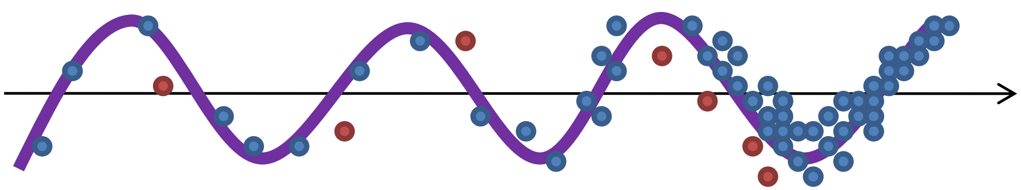
“Observations-only” climatology



Reconstruction based on observations, little use of model

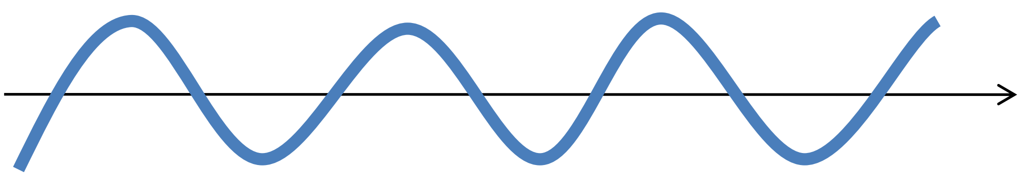
Reanalysis

Extended climate reanalysis



Balance between use of observations and model

“Model only” integration



Reconstruction based on model, little use of observations

Another type of reanalysis: extended climate reanalysis

ERA-Interim (1979 to present)

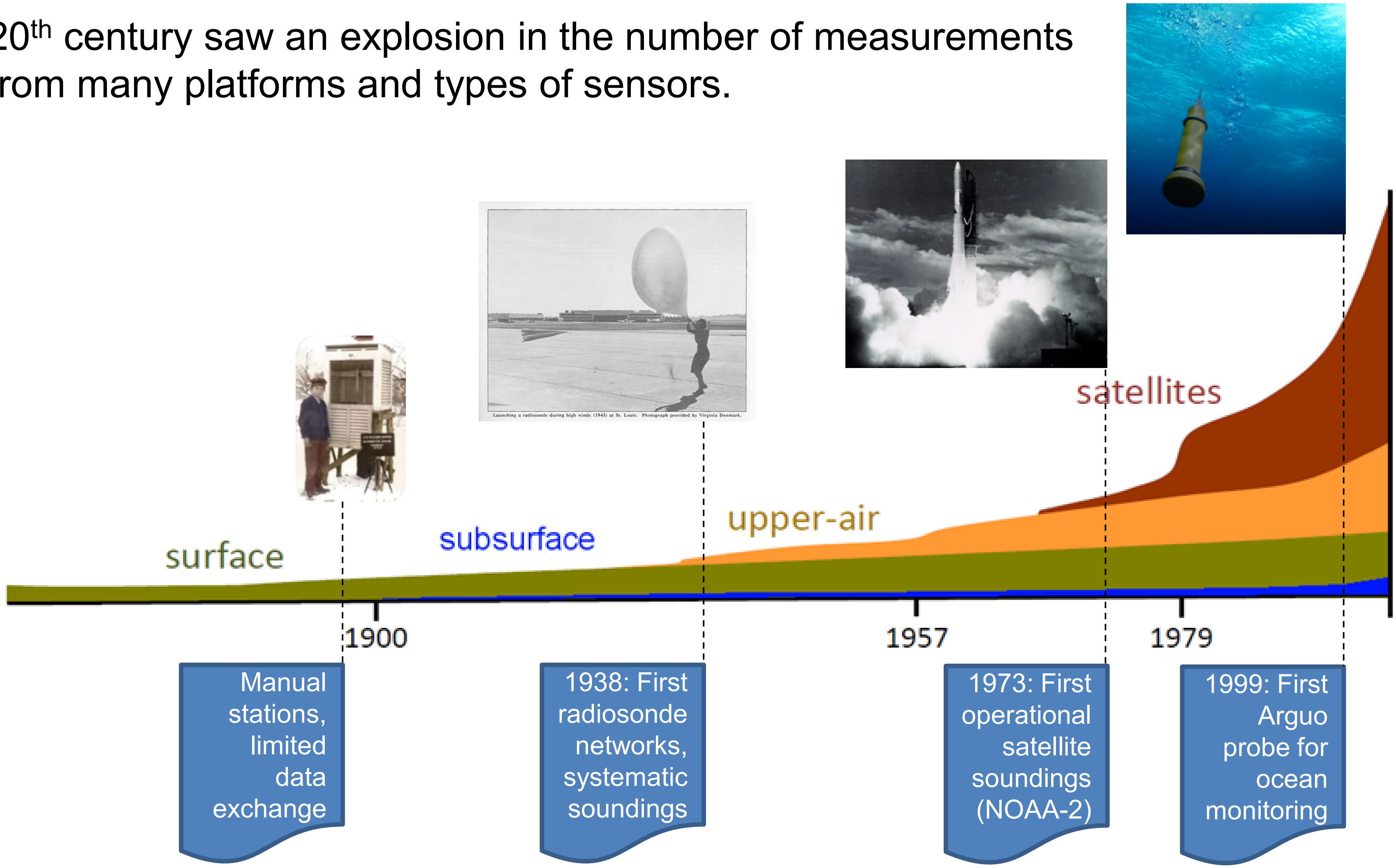
- use an invariant version of a data assimilation system and NWP model
- use as many observations as possible, including from satellites
- produce the best state estimate at any given time

ERA-20C (1900 to 2010)

- use an invariant version of a data assimilation system and NWP model
- use only a restricted set of observations
- focus on consistency and low-frequency climate variability

Evolution of the observing system

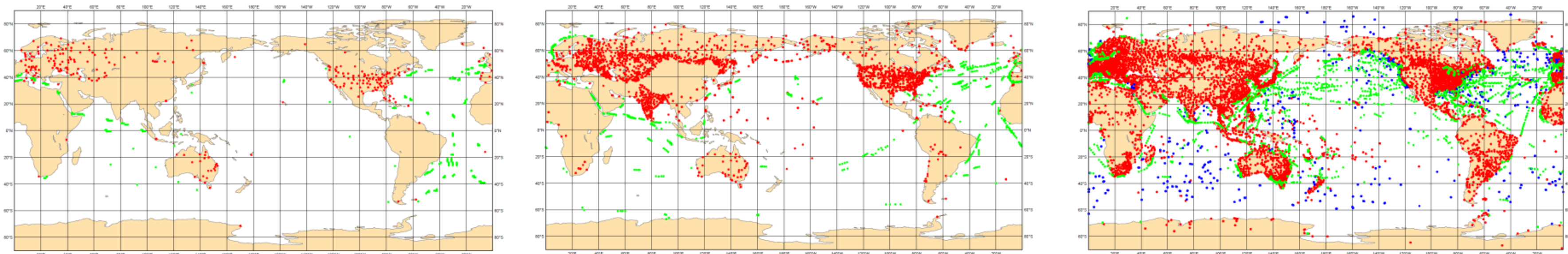
20th century saw an explosion in the number of measurements from many platforms and types of sensors.



ERA-20C assimilates only surface pressure and ocean surface winds from conventional instruments (targeting a consistent observing system)

Observing system still evolves

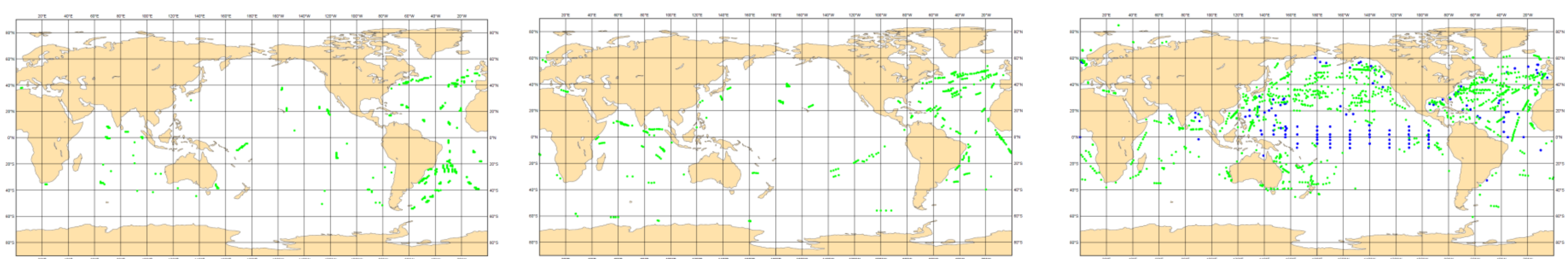
Surface pressure observations (observations per day) Stations Ships Buoys



1900 1950 2010

- observations initially concentrated in the northern hemisphere
- global coverage increases with time
- few observations for the poles even for the recent period

Surface marine wind observations (observations per day) Ships Buoys

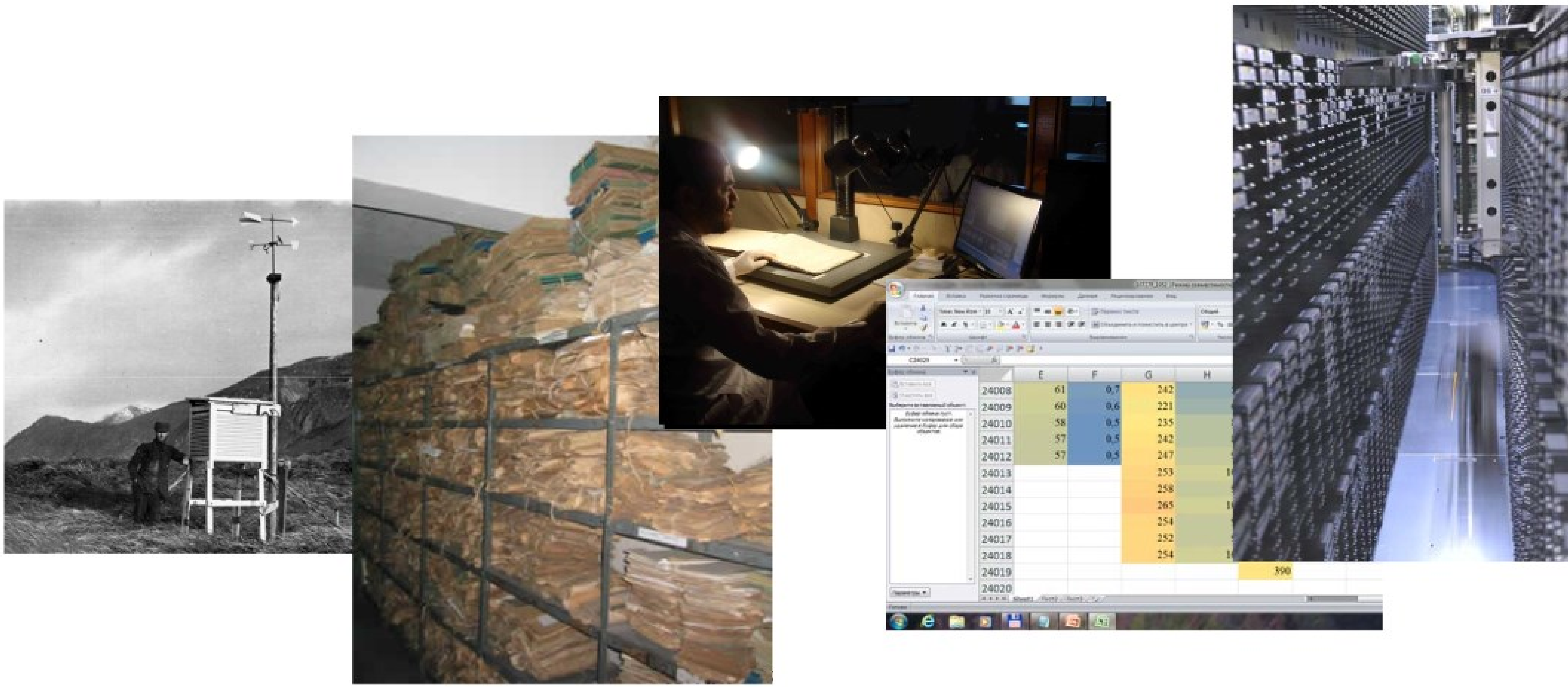


1900 1950 2010

Historical data record has to be improved

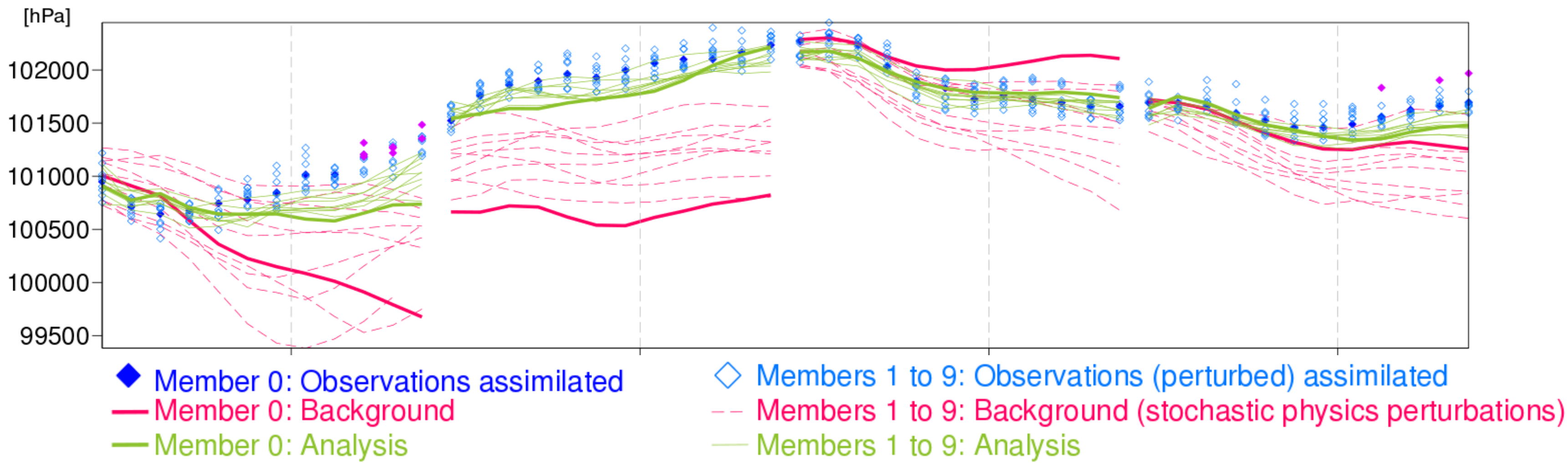
Data rescue activities:

- inventory past measurements
- produce high resolution image
- digitalize (manual keying or automatic recognition software)
- reformat to ASCII files
- import to our MARS system



ERA-20C: Extended climate reanalyses (1900-2010)

10-member EDA system to assimilate only surface pressure and ocean surface winds from conventional instruments

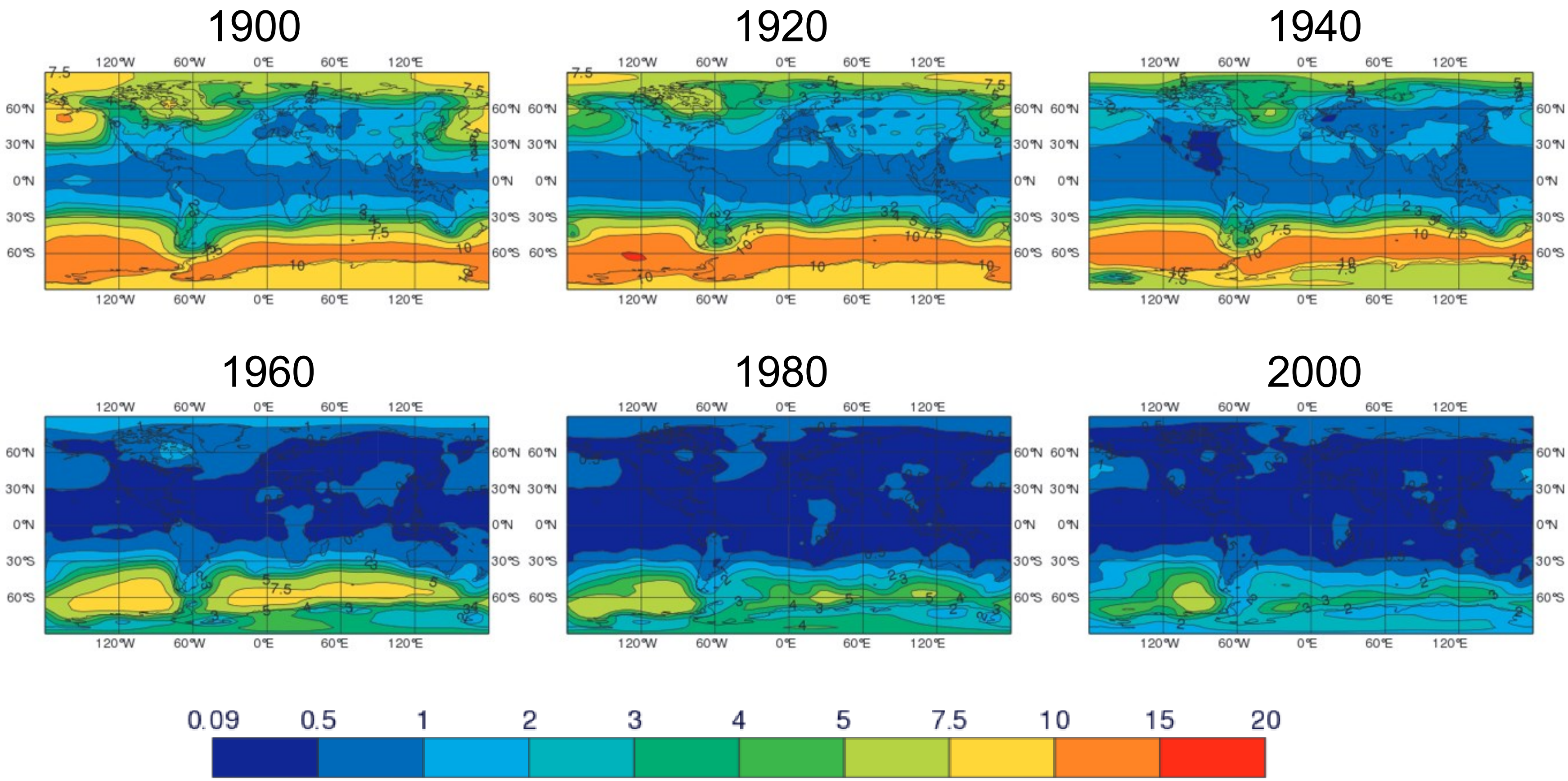


EDA system can:

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

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

Evolution of the background error for mean sea-level pressure

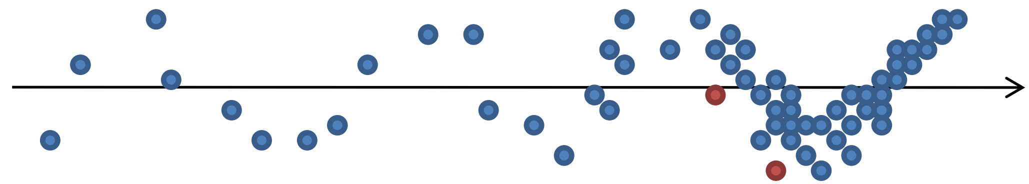


At the beginning of the century, the background error from 1 hPa to 15 hPa
At the end of the century, the background error from 1 hPa to 3 hPa

As the reanalysis system ingests more observations, it is learning from them: the backgrounds become more accurate.

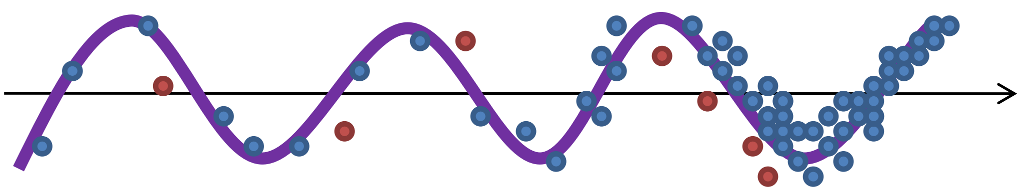
Different methods to reconstruct the past climate and/or weather

“Observations-only” climatology



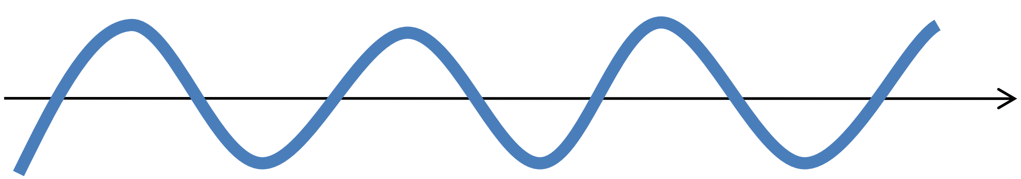
Reconstruction based on observations, little use of model

Reanalysis
Extended climate reanalysis
Coupled climate reanalysis



Balance between use of observations and model

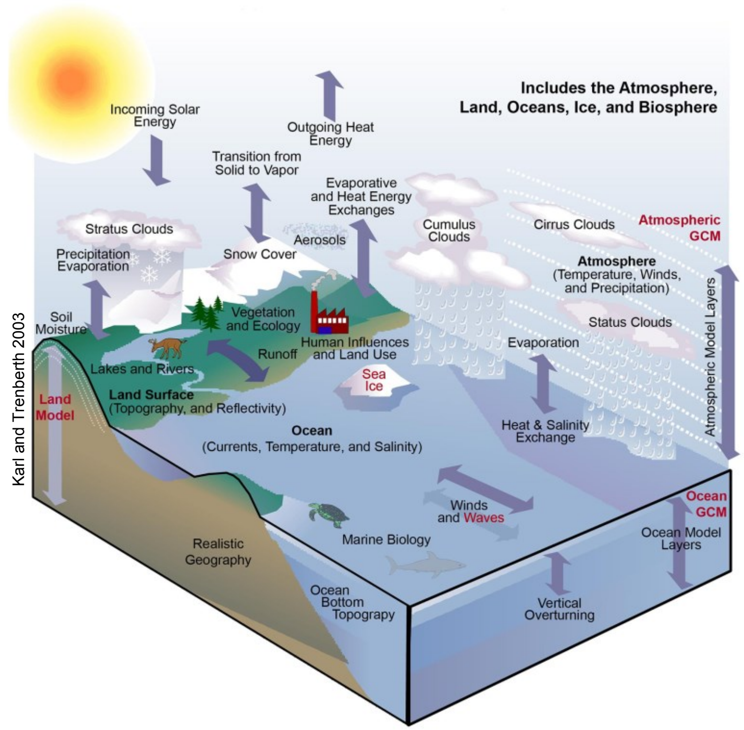
“Model only” integration



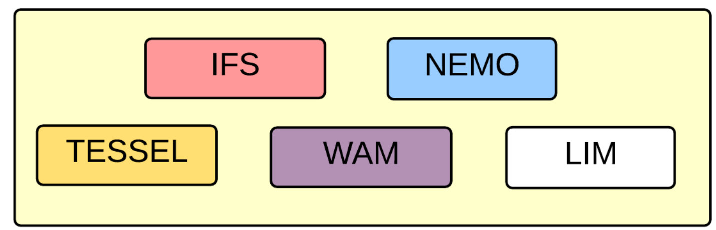
Reconstruction based on model, little use of observations

Extended climate reanalyses for the coupled earth model

ECMWF coupled Earth model for medium-range weather forecasting



ECMWF coupled Earth model

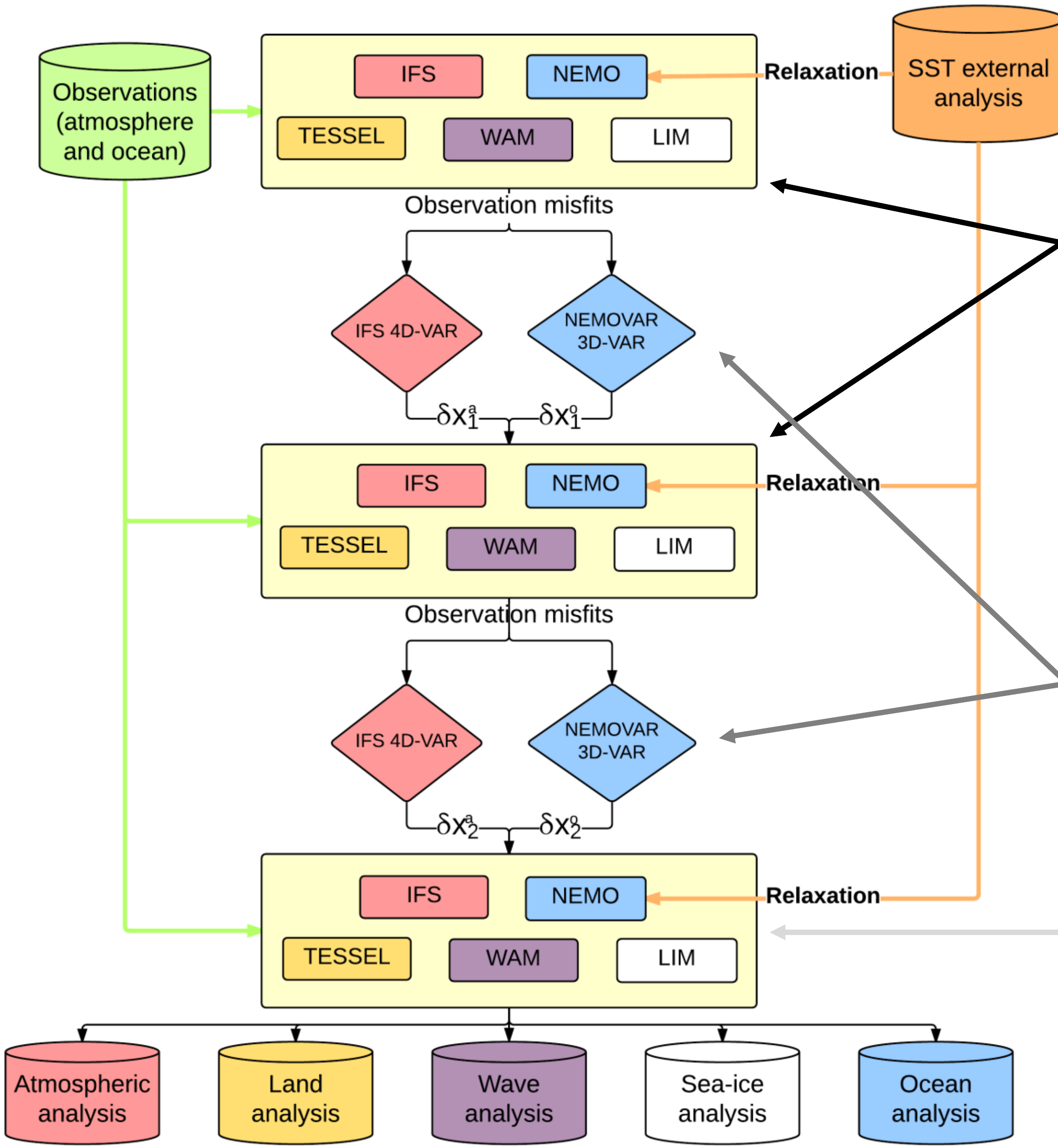


New coupled assimilation system (CERA) for the coupled Earth model:

- atmospheric and ocean observations assimilated simultaneously
- ocean observations can impact atmospheric estimate and conversely
- CERA-20C reanalysis in production (1900-2010)

Coupled assimilation system (CERA)

EDA variational approach with a 24-hour window that assimilates simultaneously atmospheric and ocean observations



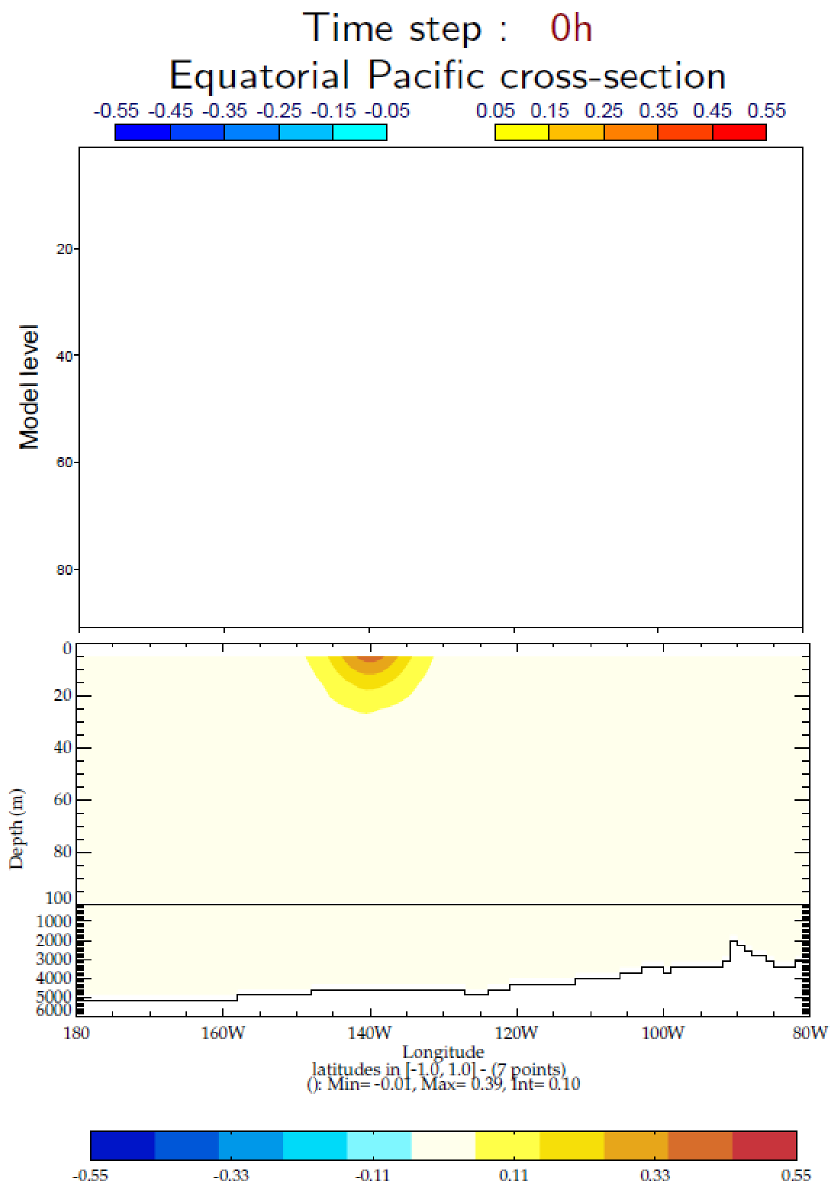
coupled model computes observation misfits in each outer iteration

SST computed in NEMO and constrained by relaxation

atmospheric and ocean increments are computed in parallel to correct the initial state

analysis dynamically consistent with respect to the coupled model

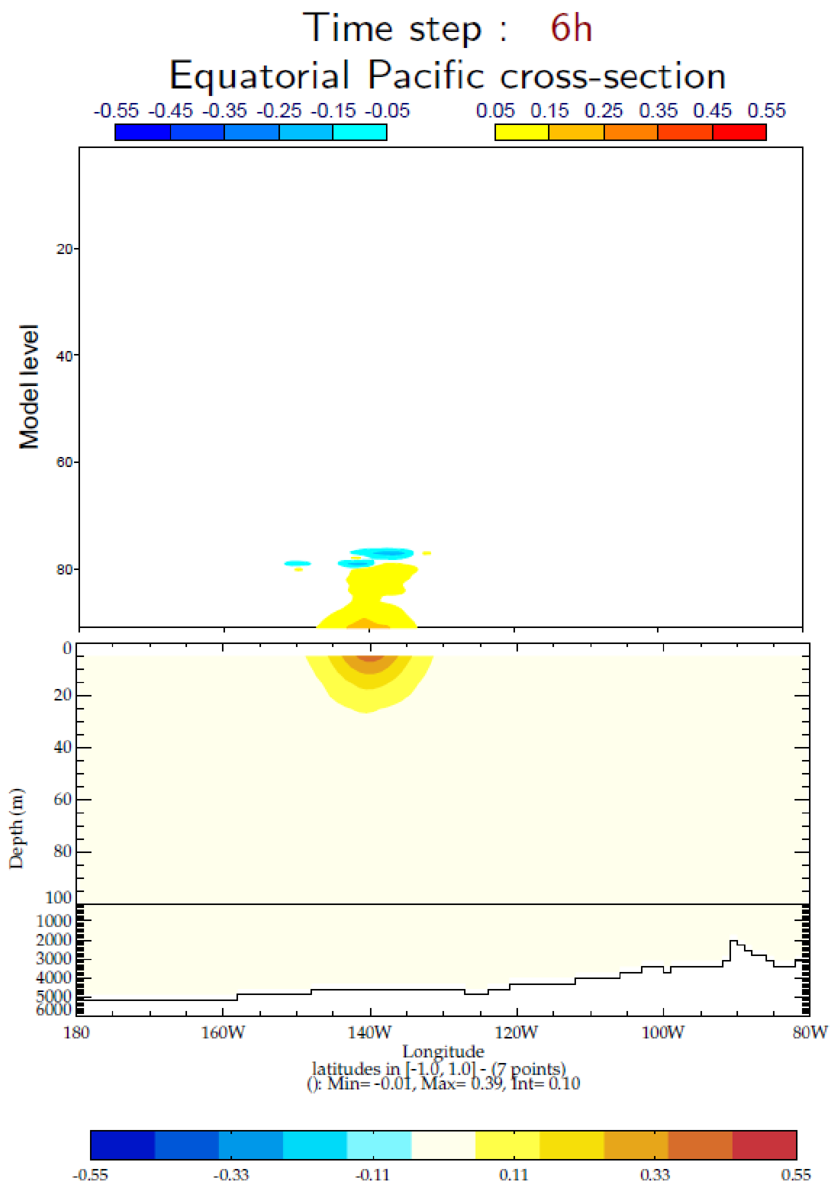
Information exchange in a coupled assimilation system



Atmosphere-ocean temperature cross-section

Ocean increment (assimilation of one temperature observation at 5-meter depth) spreads in the atmosphere during the assimilation process

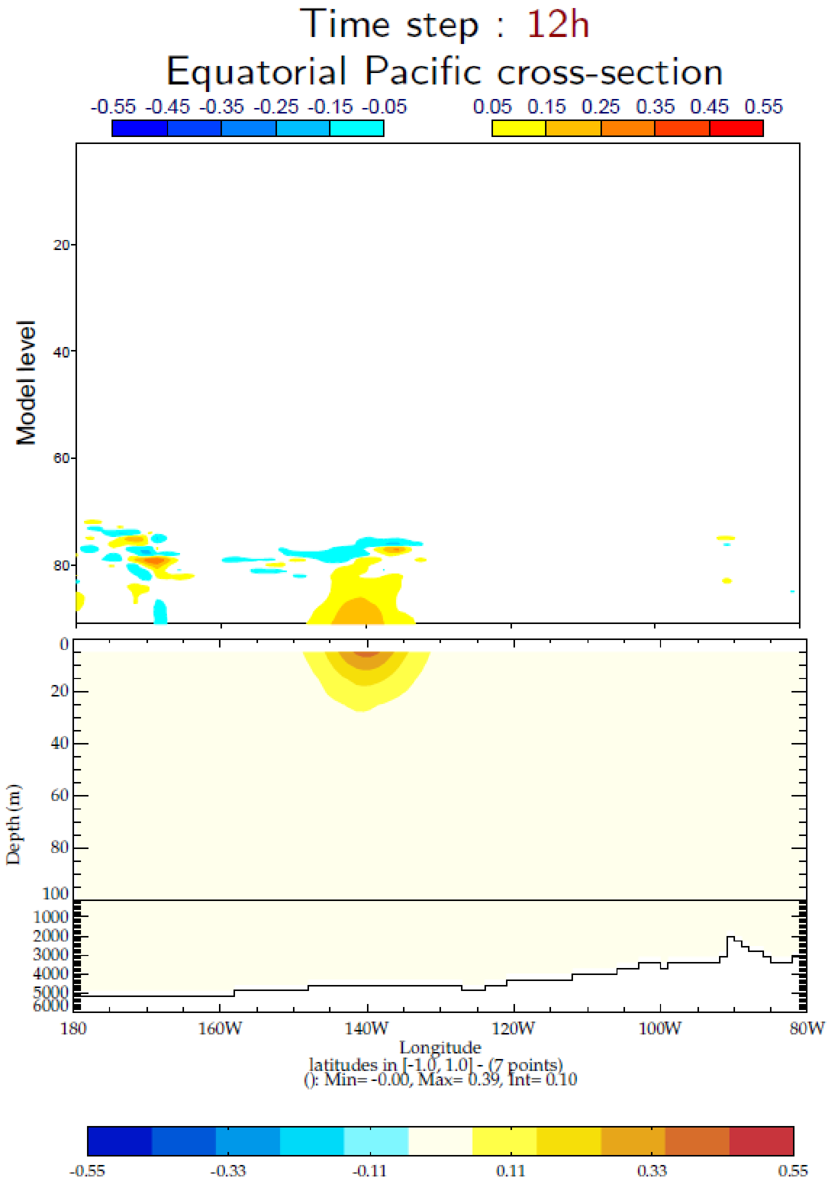
Information exchange in a coupled assimilation system



Atmosphere-ocean temperature cross-section

Ocean increment (assimilation of one temperature observation at 5-meter depth) spreads in the atmosphere during the assimilation process

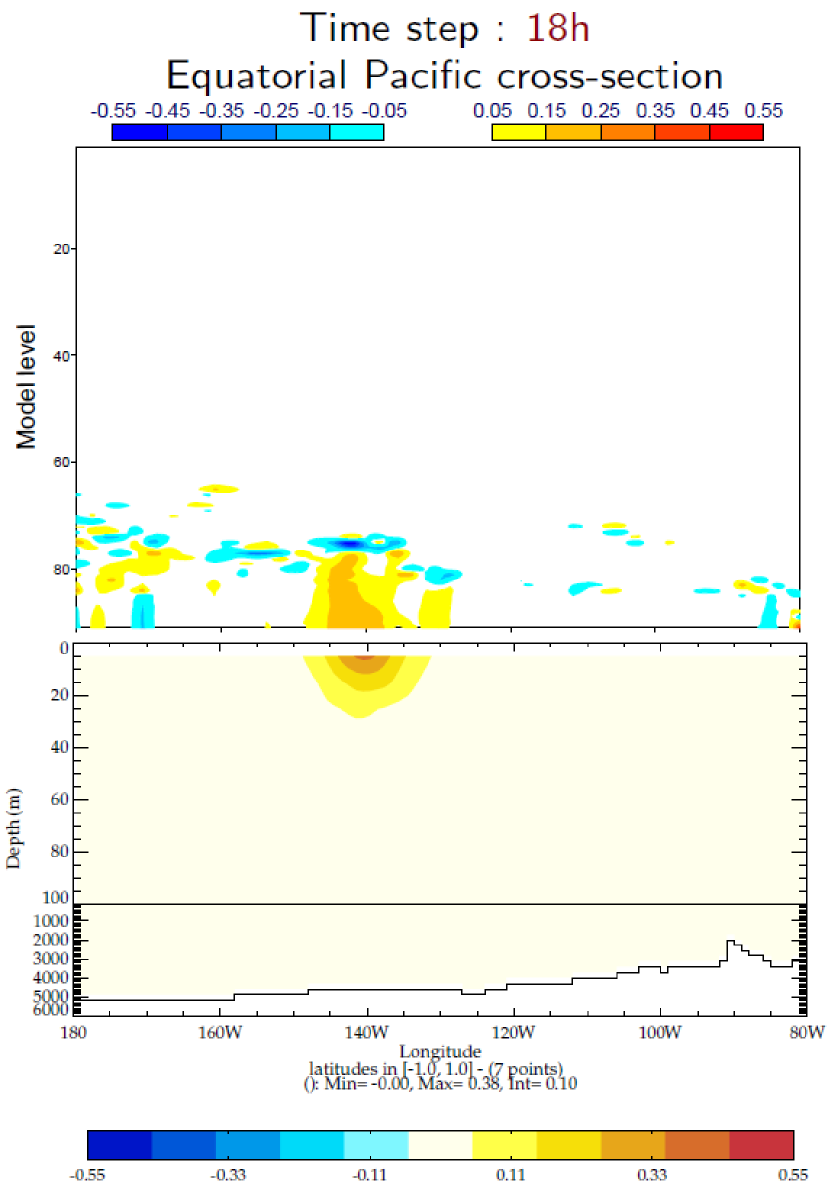
Information exchange in a coupled assimilation system



Atmosphere-ocean temperature cross-section

Ocean increment (assimilation of one temperature observation at 5-meter depth) spreads in the atmosphere during the assimilation process

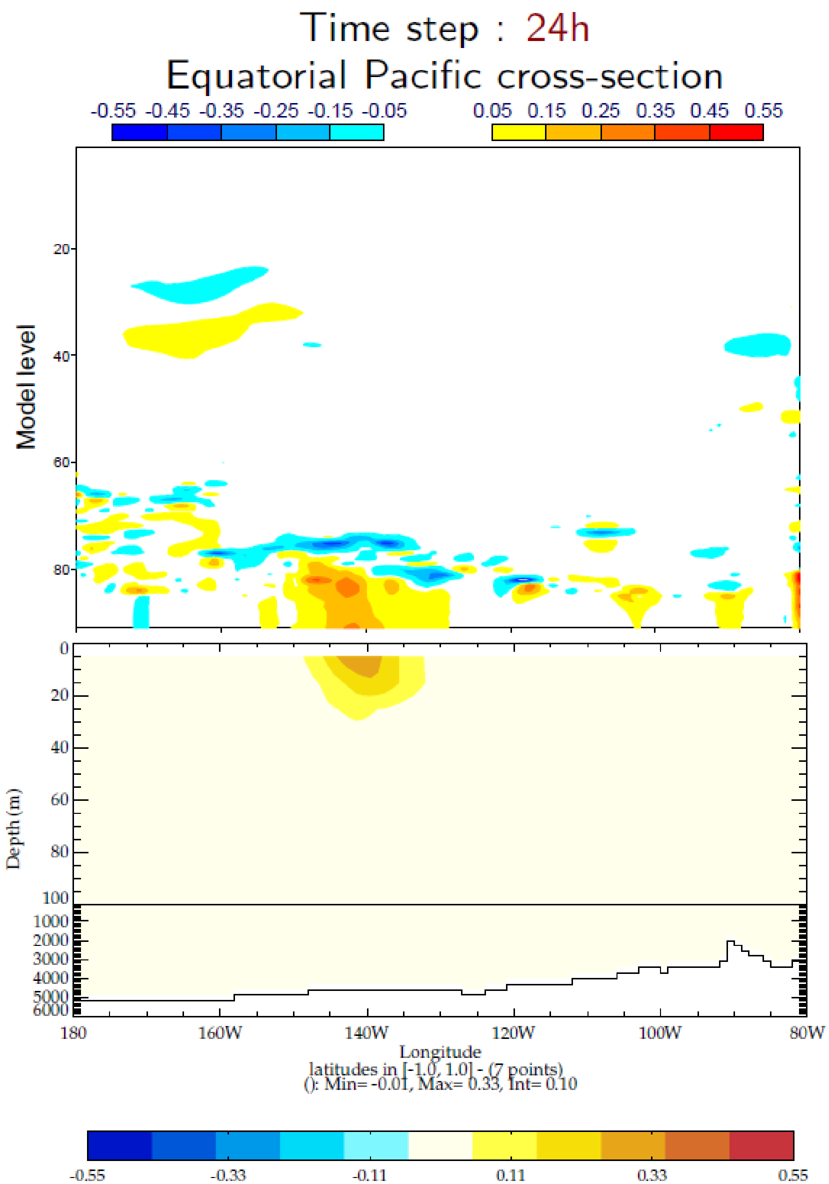
Information exchange in a coupled assimilation system



Atmosphere-ocean temperature cross-section

Ocean increment (assimilation of one temperature observation at 5-meter depth) spreads in the atmosphere during the assimilation process

Information exchange in a coupled assimilation system



Atmosphere-ocean temperature cross-section

Ocean increment (assimilation of one temperature observation at 5-meter depth) spreads in the atmosphere during the assimilation process

Production of a coupled analysis which should be better balanced and consistent with respect to the coupled model

Summary of important concepts

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

Reanalysis neither produces "gridded observations" nor "model data"

- extract information from observations using the model to propagate the information in space and time, and across variables

Unlike NWP, a very important concern in reanalysis is the consistency in time, spanning several years

Reanalysis is worth repeating as all ingredients continue to evolve

- models, data assimilation, observation reprocessing and data rescue
- with each new reanalysis, understanding of model/observations biases is improved

Summary of important concepts

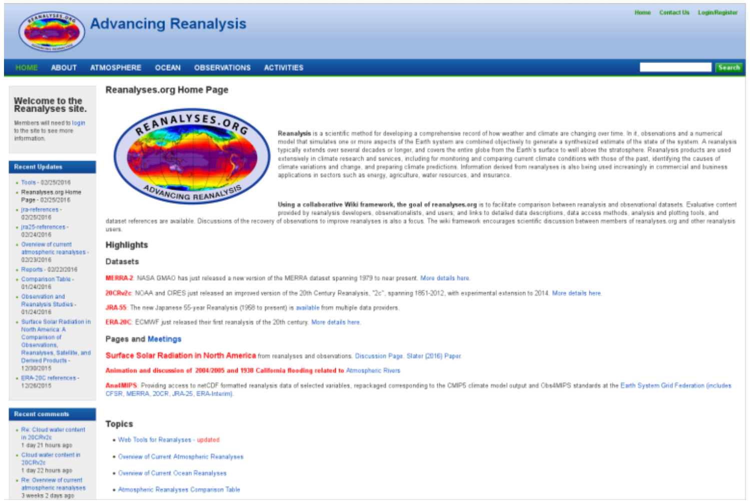
Key aspects that require particular attention in reanalysis

- external forcing fields for the NWP model
- biases in the model and observations
- changes in the observing system
- specification of background and observation errors

More challenges for comprehensive reanalyses

- publishing uncertainty estimates for reanalysis: how will they be used?
- coupling with ocean and land surface
- making observations used in reanalysis more accessible to users
- bridging the gap with climate models

Websites

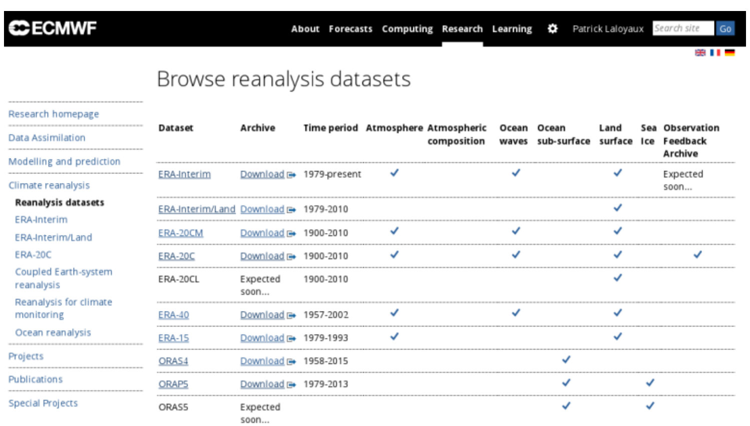


http://reanalyses.org/

Dataset overviews

https://climatedataguide.ucar.edu/

Dataset overviews



Dataset	Archive	Time period	Atmosphere	Atmospheric composition	Ocean waves	Ocean sub-surface	Land surface	Sea Ice	Observation Feedback Archive
ERA-Interim	Download	1979-present	✓		✓		✓		Expected soon...
ERA-Interim/Land	Download	1979-2010					✓		
ERA-20CM	Download	1900-2010	✓		✓		✓		
ERA-20C	Download	1900-2010	✓		✓		✓		✓
ERA-20CL	Expected soon...	1900-2010					✓		
ERA-40	Download	1957-2002	✓		✓		✓		
ERA-15	Download	1979-1993	✓				✓		
ORAS4	Download	1958-2015				✓			
ORAP5	Download	1979-2013				✓		✓	
ORAS5	Expected soon...					✓		✓	

http://www.ecmwf.int/en/research/climate-reanalysis

ECMWF reanalysis datasets

Further readings

Reanalysis:

Toward a consistent reanalysis of the climate system.
D.Dee et al., BAMS, 95, 1235–1248, 2014.

The need for a dynamical climate reanalysis.
Bengtsson *et al.*, BAMS, 88, 495-501, 2007.

The ERA-Interim reanalysis: configuration and performance of the data assimilation system.
Dee *et al.,* QJRMS, 137, 553-597, 2011.

Estimating low-frequency variability and trends in atmospheric temperature using ERA-Interim.
Simmons *et al., QJRMS, 140, 329-353, 2014.*

ERA-20CM: A twentieth-century atmospheric model ensemble.
H. Hersbach et al., QJRMS, 141, 2350-2375, 2015.

ERA-20C: An atmospheric reanalysis of the 20th century.
P. Poli et al., Journal of Climate, In press.

Observation recovery:

ERA-CLIM: Historical Surface and Upper-Air Data for Future Reanalyses
A. Stickler et al., Bull. Amer. Meteor. Soc., 95, 1419-1430, 2014.

Coupled assimilation:

A coupled data assimilation system for climate reanalysis.
P. Laloyaux, M. Balmaseda, D. Dee, K. Mogensen and P. Janssen. QJRMS. In press

Origin and impact of initialisation shocks in coupled atmosphere-ocean forecasts.
D. Mulholland, P. Laloyaux, K. Haines, M. Balmaseda. MWR. In Press.

Impact of scatterometer wind data in the ECMWF coupled assimilation system.
P. Laloyaux, J.-N. Thépaut and D. Dee. MWR. In Press.