Data Assimilation
(for atmospheric and climate monitoring)

Jean-Noël Thépaut

Acknowlegments: Clément Albergel, Erik Andersson, Tom Auligné, Magdalena Balmaseda, Peter Bauer, Gianpaolo Balsamo, Niels Bormann, Carla Cardinali, Dick Dee, John Derber, Patricia De Rosnay, Stephen English, John Eyre, Mike Fisher, Sean Healy, Andras Horanyi, Lars Isaksen, Marta Janisková, Erland Källén, Jérôme Lafeuille, Patrick Laloyaux, Philippe Lopez, Cristina Lupu, Pierre-Philippe Matthieu, Tony McNally, Paul Poli, Samuel Rémy, Roger Saunders, Yannick Trémolet, and others…

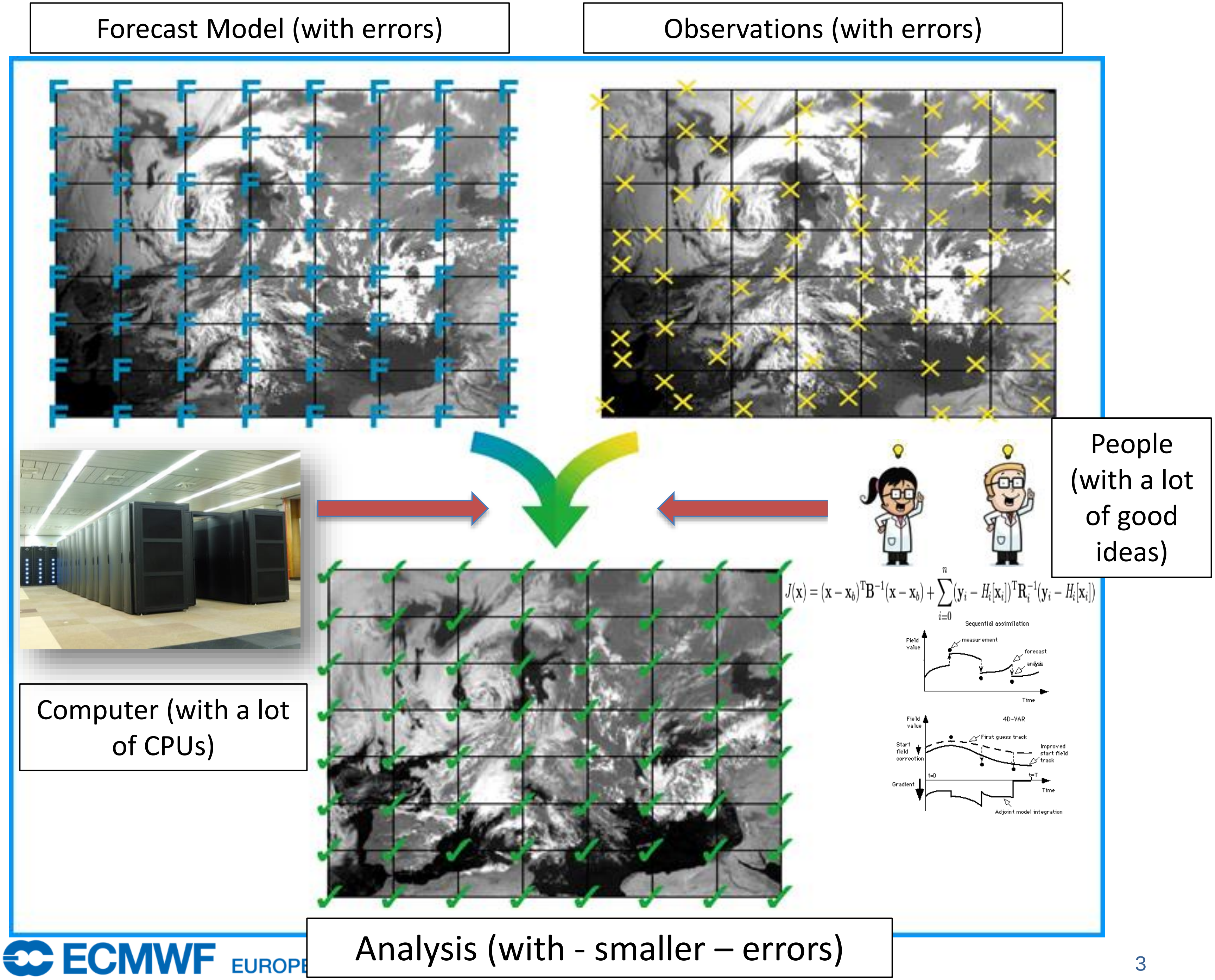
Data Assimilation

Wikipedia definition: Process by which observations are incorporated into a computer model of a real system

NWP definition: Process by which “optimal” initial conditions for numerical forecasts are defined.

- The best analysis (initial conditions) is the analysis that leads to the best forecast
- Makes “quickly” the best out of all information available

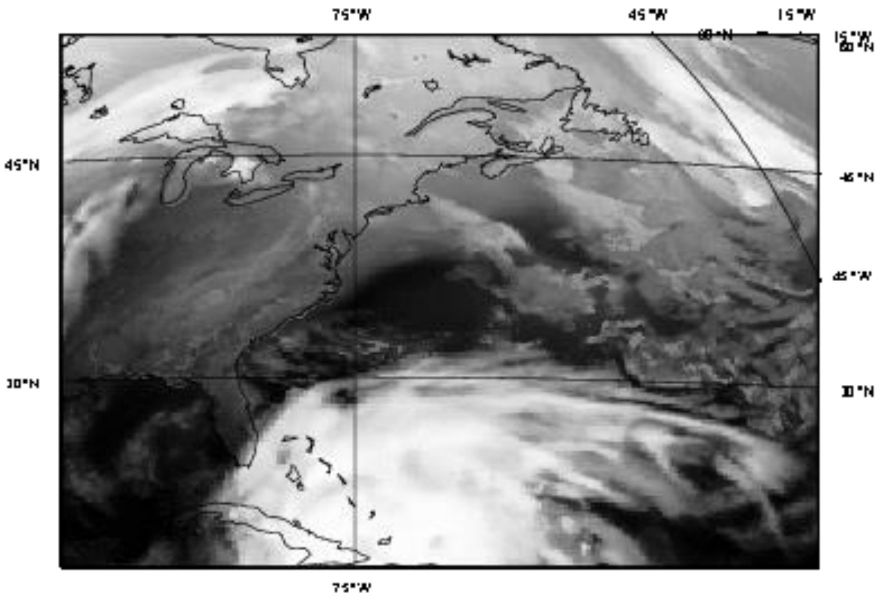
Climate definition: Process that provides a complete and physically consistent four-dimensional picture of the earth system out of a rich variety of heterogeneous and asynchronous sources of information

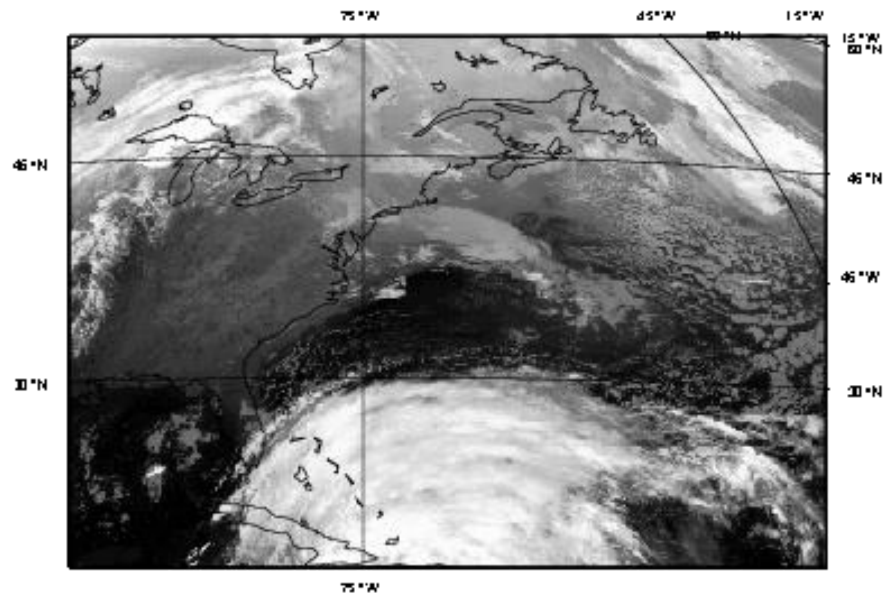

Forecast Model (with errors)
Observations (with errors)
People (with a lot of good ideas)
Computer (with a lot of CPUs)
$J(x) = (x - x_b)^T B^{-1} (x - x_b) + \sum_{i=0}^{n} (y_i - H_i[x_i])^T R_i^{-1} (y_i - H_i[x_i])$
Sequential assimilation
Field value
measurement
forecast
analysis
Time
4D-VAR
Field value
First guess track
Start field correction
Improved start field track
t=0
t=T
Gradient
Time
Adjoint model integration
Analysis (with - smaller – errors)

Models and observation operators have become much more realistic and accurate

ECMWF Fc 20121025 00UTC
Model simulated satellite images

GOES-13 IR10.8 20121025-20121030





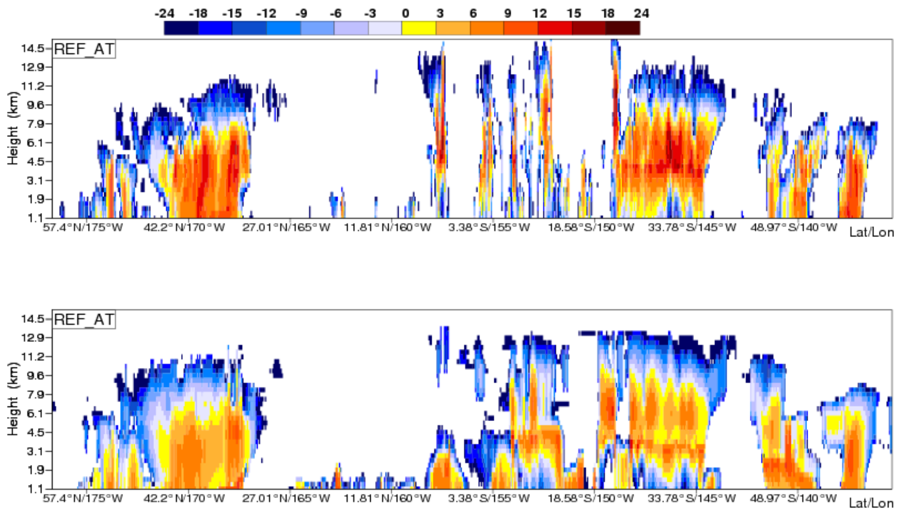
model

observation

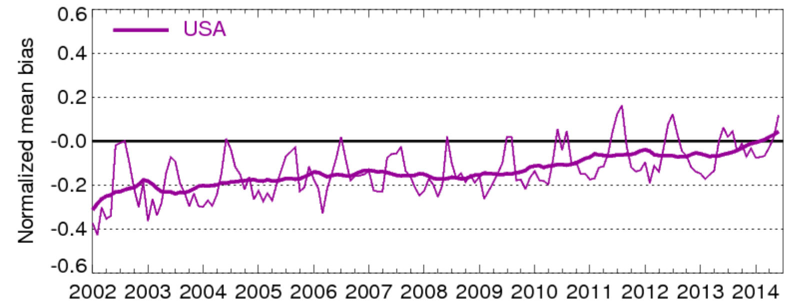
This widens opportunities to assimilate new data

Models and observation operators have become much more realistic and accurate

C
L
O
U
D
S
A
T



Cloud Radar Reflectivity



Continual improvement of ECMWF short-range precipitation forecasts with respect to ground-based radar data.

These improvements and associated opportunities are particularly relevant for exploring initialisation of new model variables

The Global Observing System

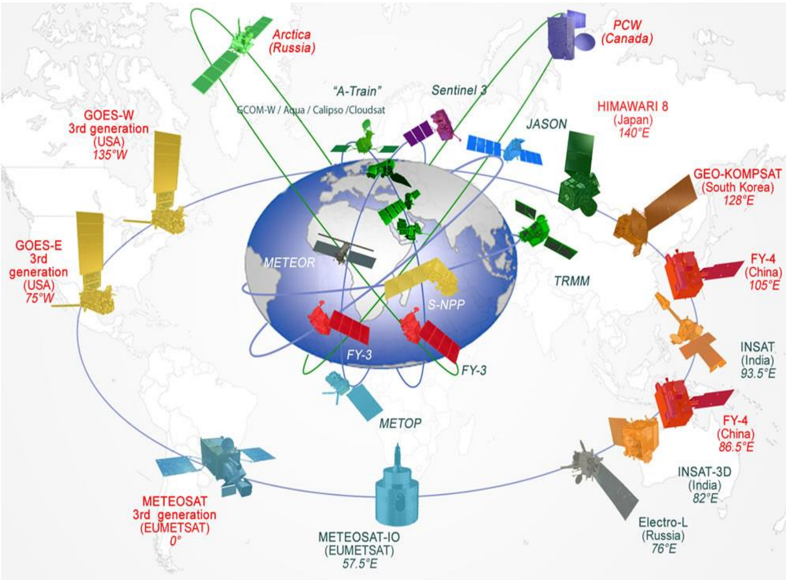
Maximizing the impact of new observations:

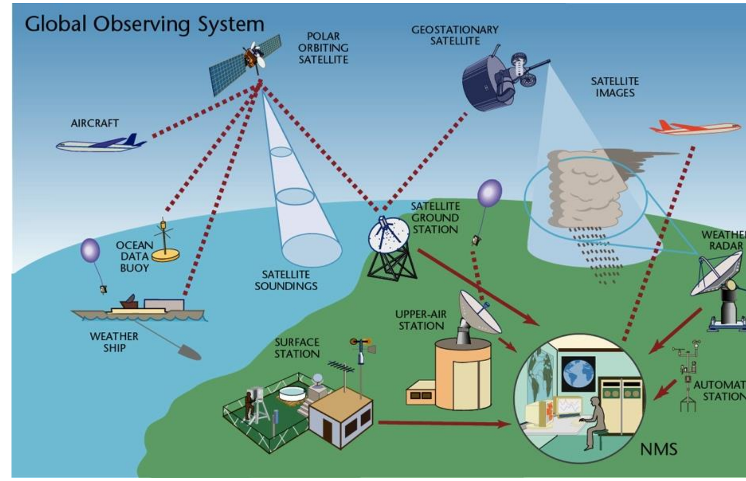
- Need to plan
- Need to assess
- Need to proof-of-concept
- Need to consolidate:
  - transfer from Research to Operations

Underpinning requirement: Full exploitation of new observations require sustained investments in model and data assimilation developments

Not to forget: Observations are also essential for verification (not only assimilation)!

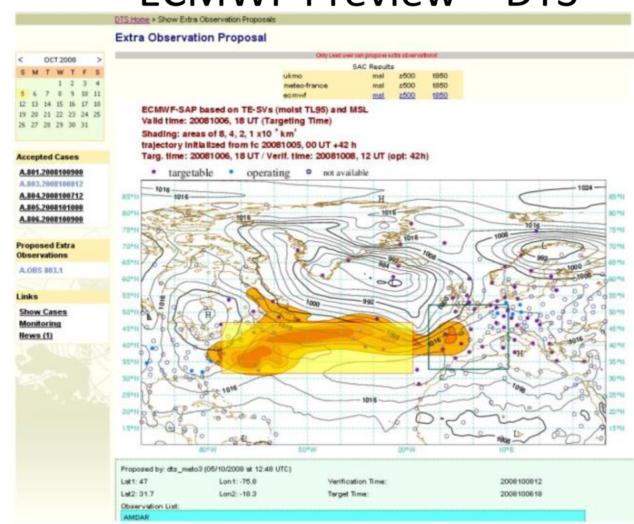
WMO Integrated Global Observing System



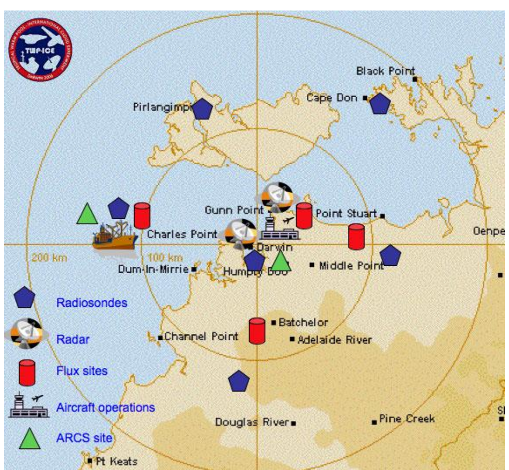


Courtesy: WMO

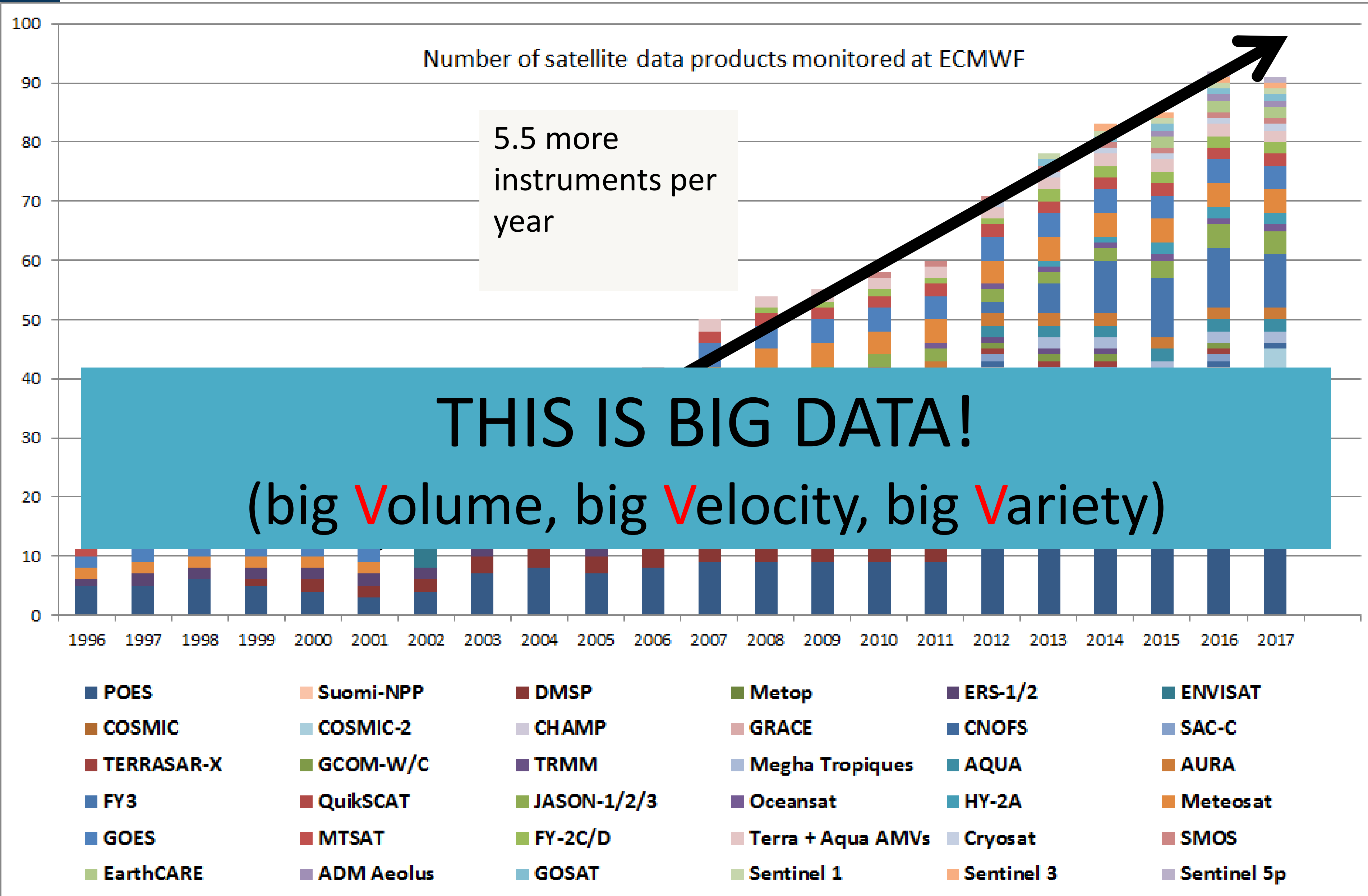
ECMWF Preview – DTS



T
W
P
-
I
C
E

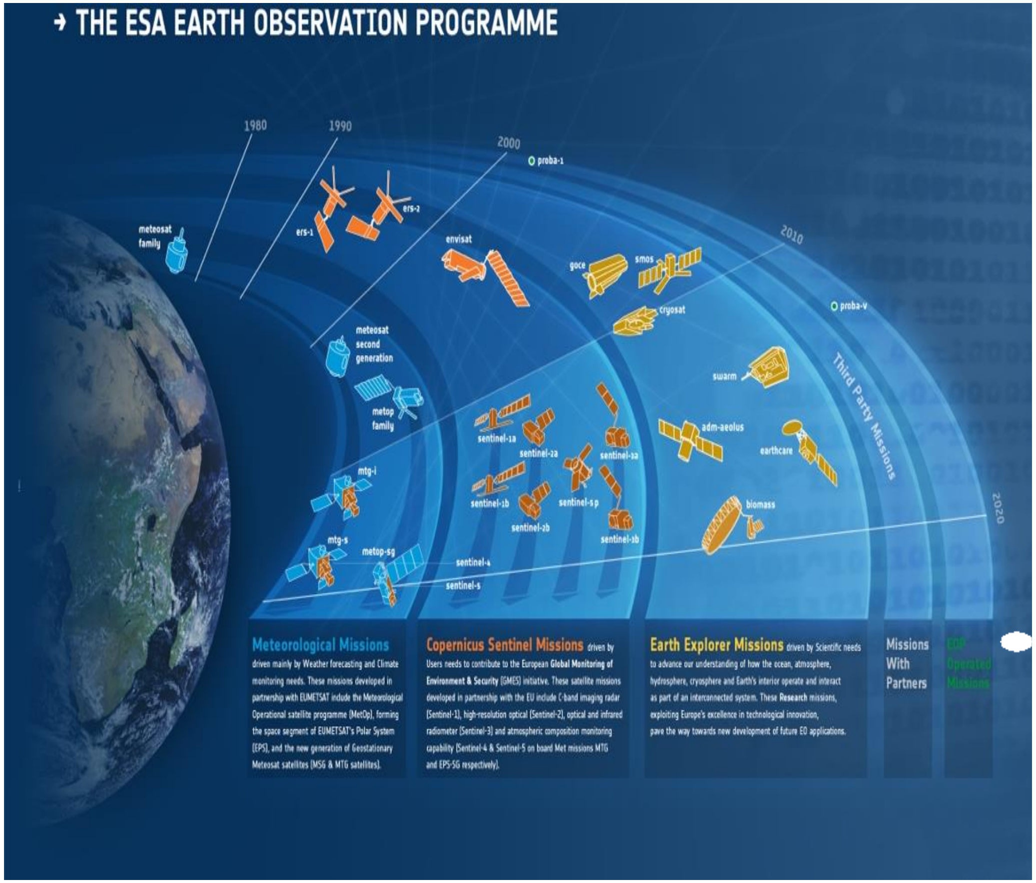


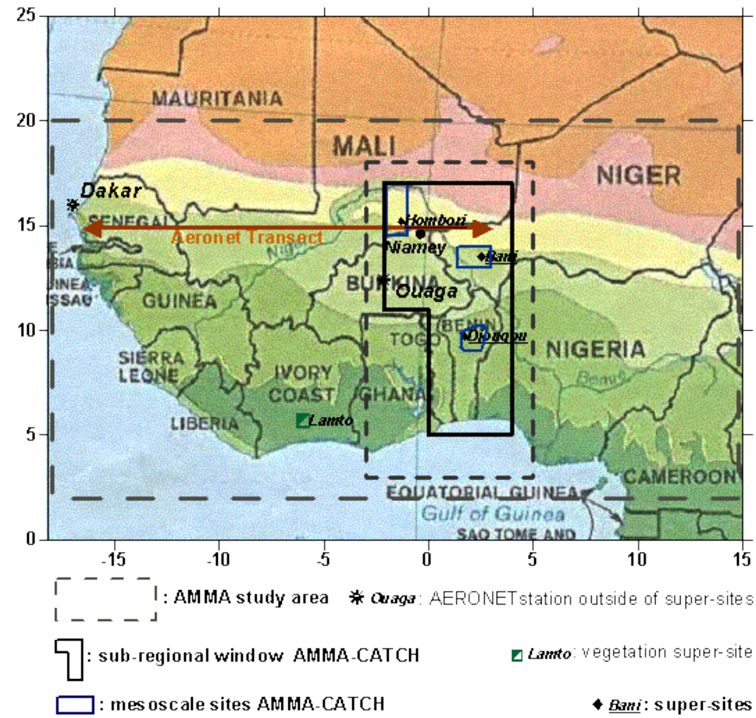
Supported by field campaign experiments, Data targeting studies, etc.



Number of satellite data products monitored at ECMWF
5.5 more instruments per year
THIS IS BIG DATA!
(big Volume, big Velocity, big Variety)
0 10 20 30 40 50 60 70 80 90 100
1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016 2017
POES
Suomi-NPP
DMSP
Metop
ERS-1/2
ENVISAT
COSMIC
COSMIC-2
CHAMP
GRACE
CNOFS
SAC-C
TERRASAR-X
GCOM-W/C
TRMM
Megha Tropiques
AQUA
AURA
FY3
QuikSCAT
JASON-1/2/3
Oceansat
HY-2A
Meteosat
GOES
MTSAT
FY-2C/D
Terra + Aqua AMVs
Cryosat
SMOS
EarthCARE
ADM Aeolus
GOSAT
Sentinel 1
Sentinel 3
Sentinel 5p

Research programmes pioneering new technologies and observing strategies



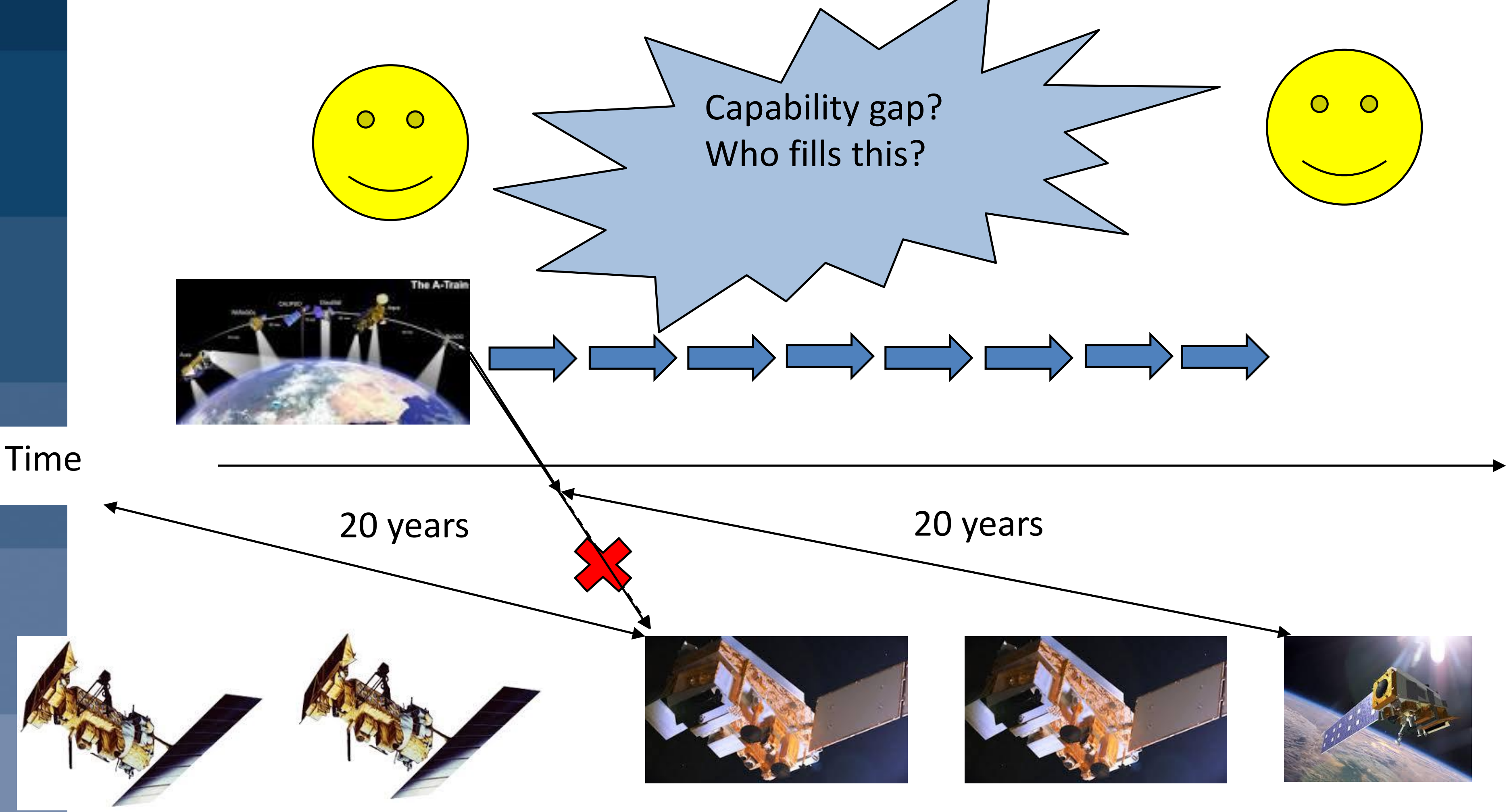


AMMA campaign

Courtesy ESA

How to transfer Research to Operations?

The Research to Operations gap



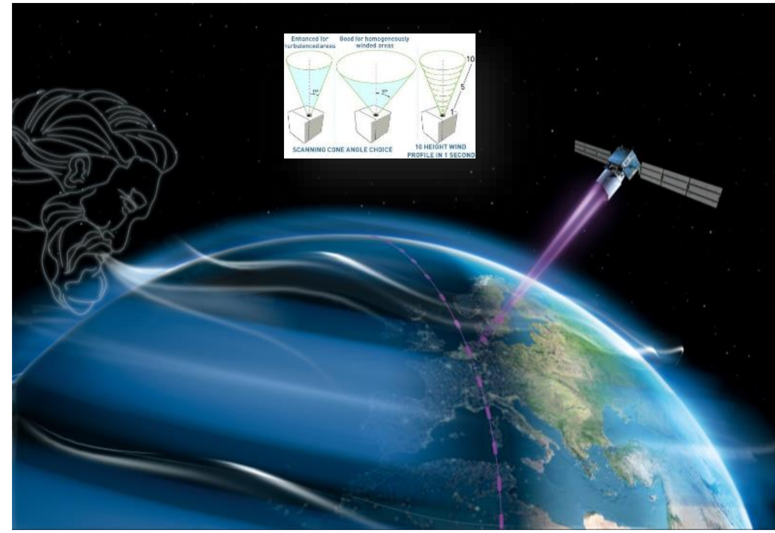
e.g. L-band. No operational plans yet, what happens post SMOS+SMAP?

ADM-AEOLUS: A new perspective for wind distribution understanding and data assimilation

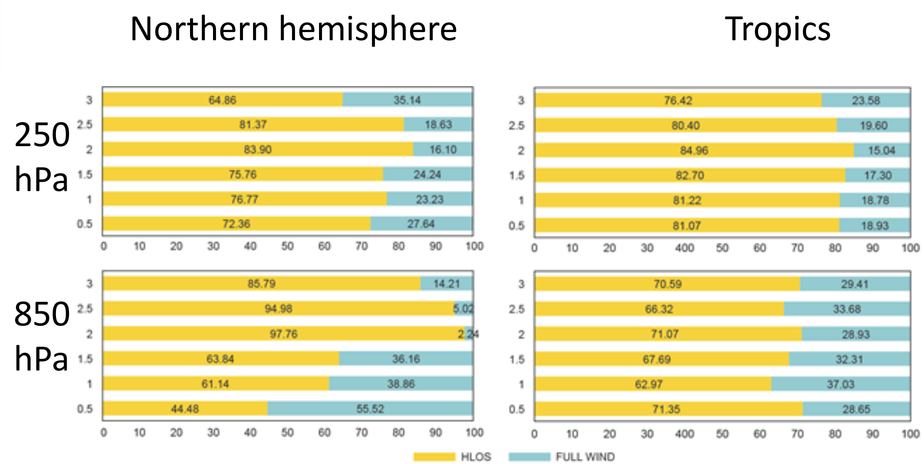
What if wind lidar from space delivers its promises (or more) during the lifetime of ADM-AEOLUS?

Transfer from research to operations should be by design, not by opportunity

Success stories: scatterometry, EU sentinel programmes, EPS-SG, etc.

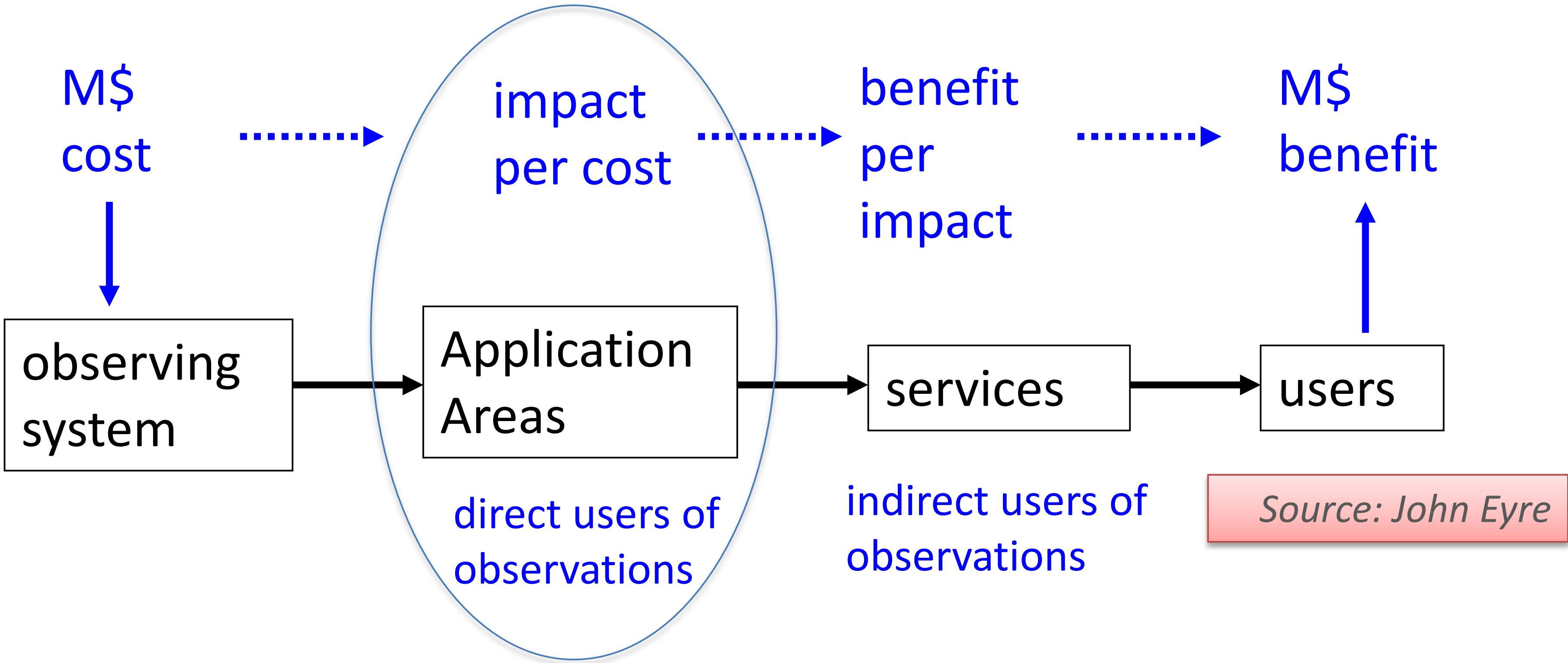


Courtesy: ESA



Using existing vector wind observations it has been shown that line of sight winds will be useful.

Observing System Cost-Benefit Chain

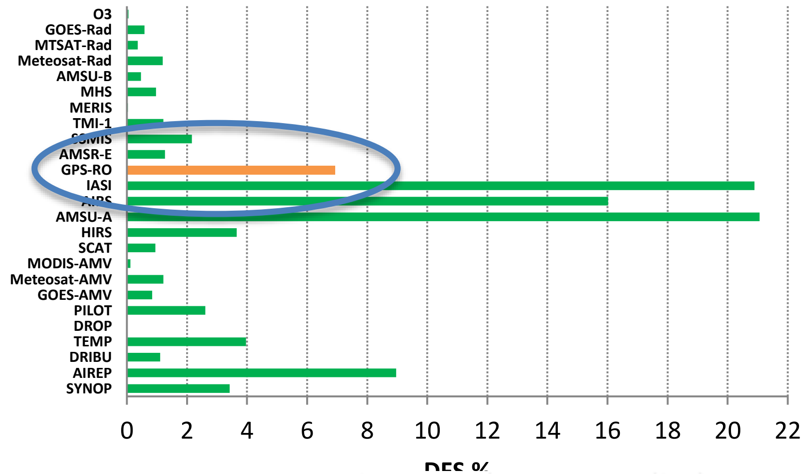


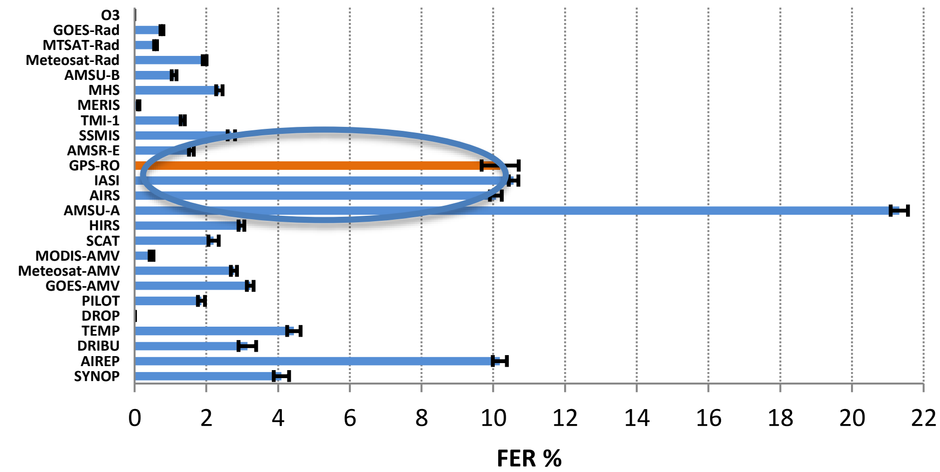
A number of tools exists to contribute to this evaluation:

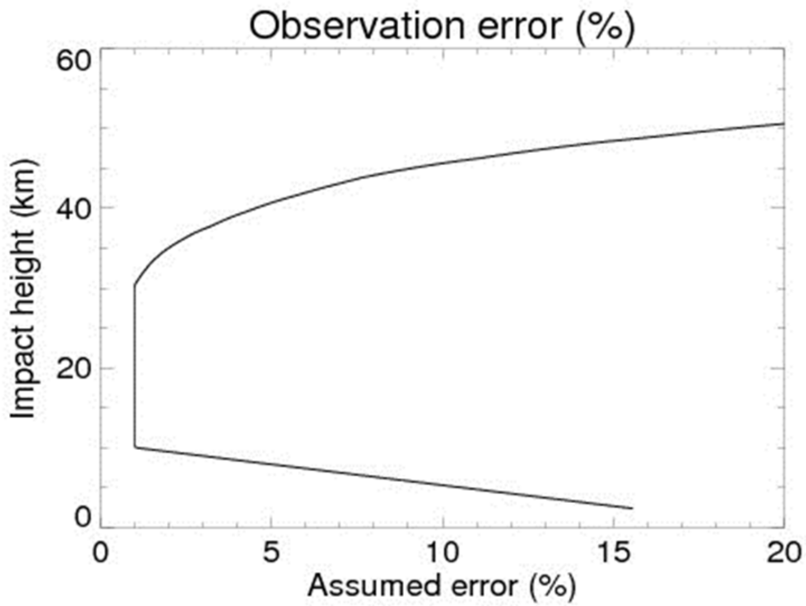
- Observing System Experiments (OSEs)
- Observing System Simulated Observations (OSSEs)
- Degree of Freedom for Signal (DFS)
- Forecast Sensitivity to Observations (FSO): adjoint or ensemble based

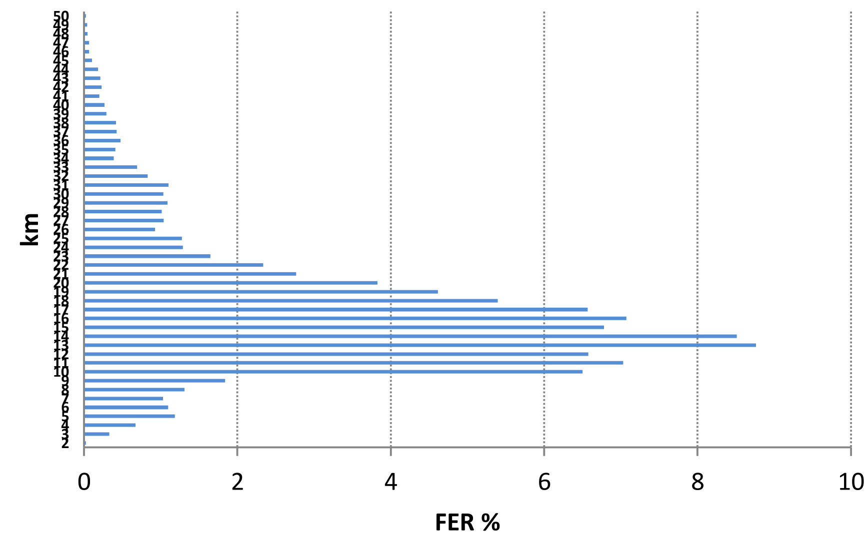
Added value of observations: ! Verification !

Observation Impact at ECMWF









Possible zoom on observation type, level, geographical area, etc.

Methodologies

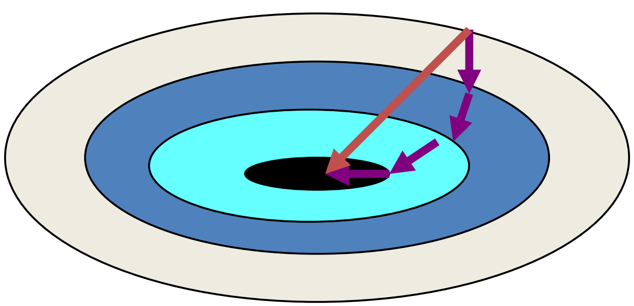
in a nutshell

Optimal Interpolation

- linear combination of observations

Variational methods

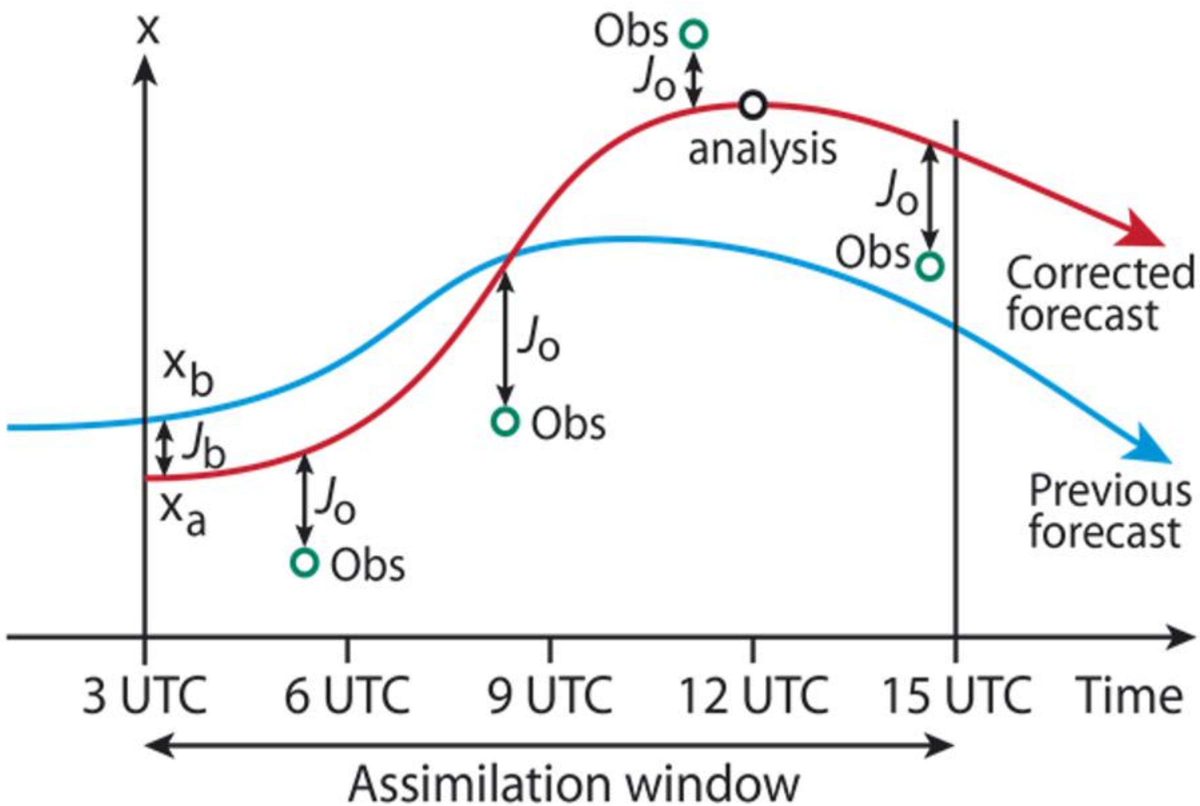
- Minimise distance to observations



Methodologies

4D-Var

Find the model trajectory that best fits the available observations



Methodologies

Over the past decades, operational DA techniques have evolved from:

- Cressman type methods (1960/1970s)

to:

- Hybrid methods exploiting the best of both variational and ensemble worlds

Choice remains application dependent

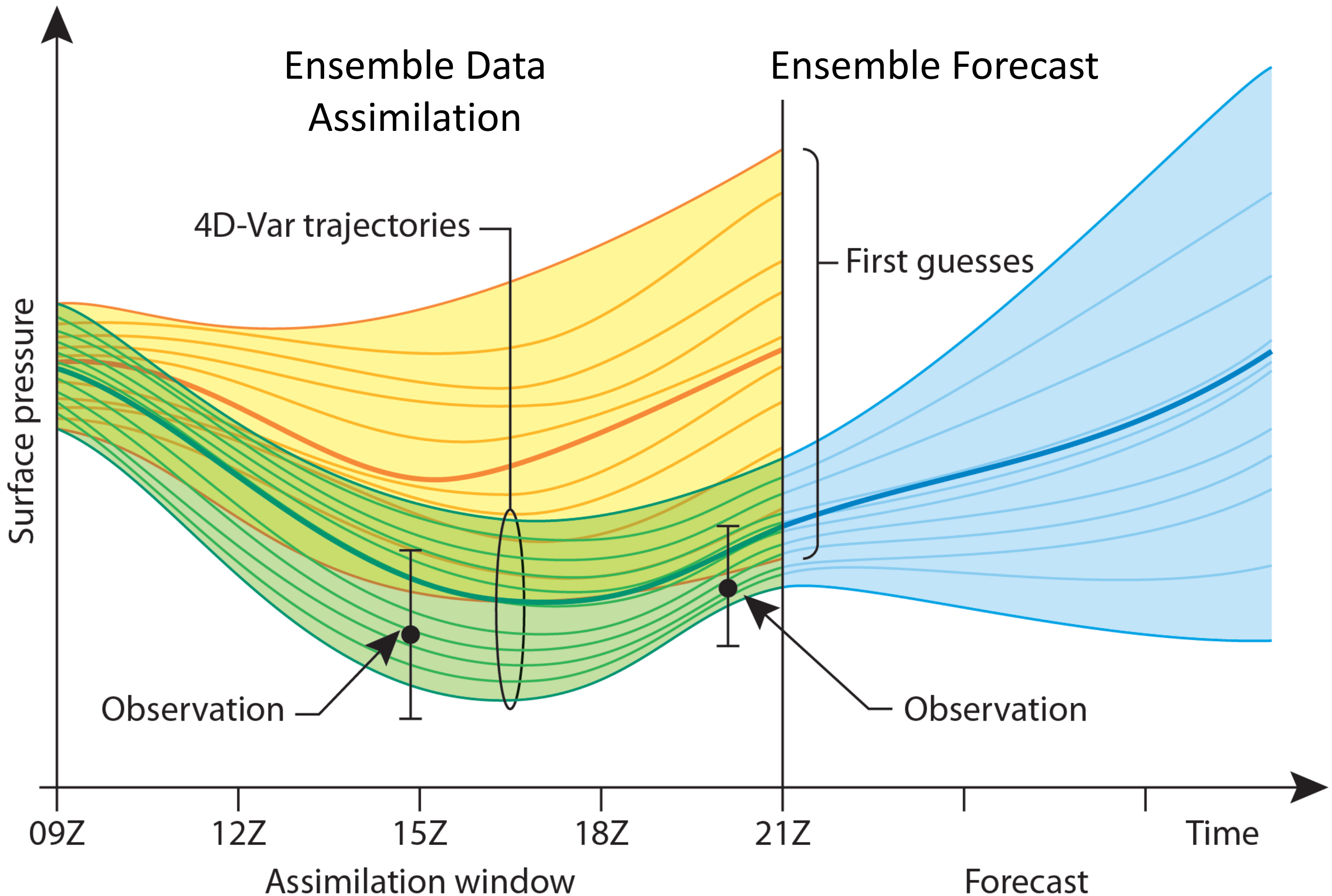
- Atmosphere, waves, sea-ice, ocean, composition, land…

Methodologies

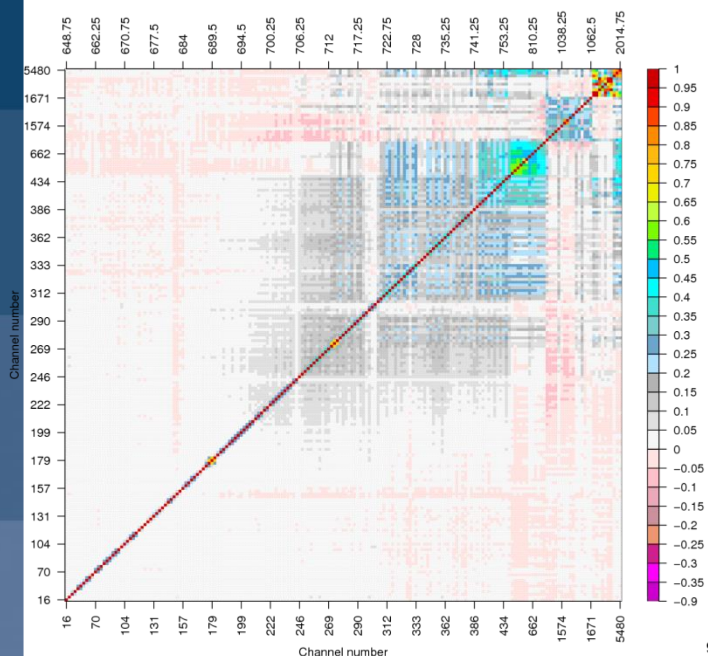
Overarching considerations include:

- Seamless quantification of uncertainty estimation (present to future)
- Improved specification of a priori errors
  - Model, background, observations - systematic and random
  - Errors of the day
- Covariance modeling
  - More variables (aerosols, trace gases, clouds)
  - Non gaussianity
  - Higher resolution
- Data Assimilation for a coupled earth system

Seamless EDA/ENS



Observation error specification:
Impact on FG-departures for other observations

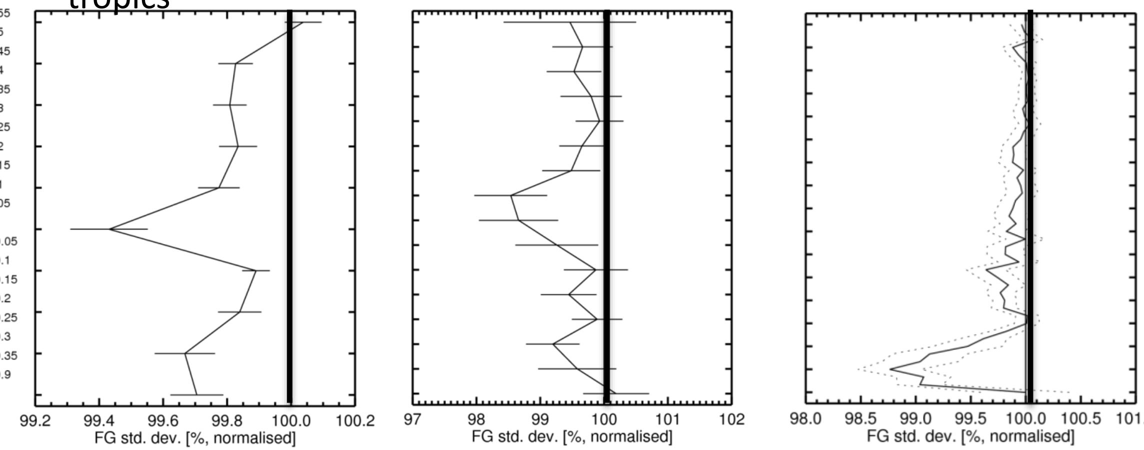


IASI inter-channel error correlation matrix

AMSU-A, tropics

Radiosondes - T, tropics

GPSRO, global



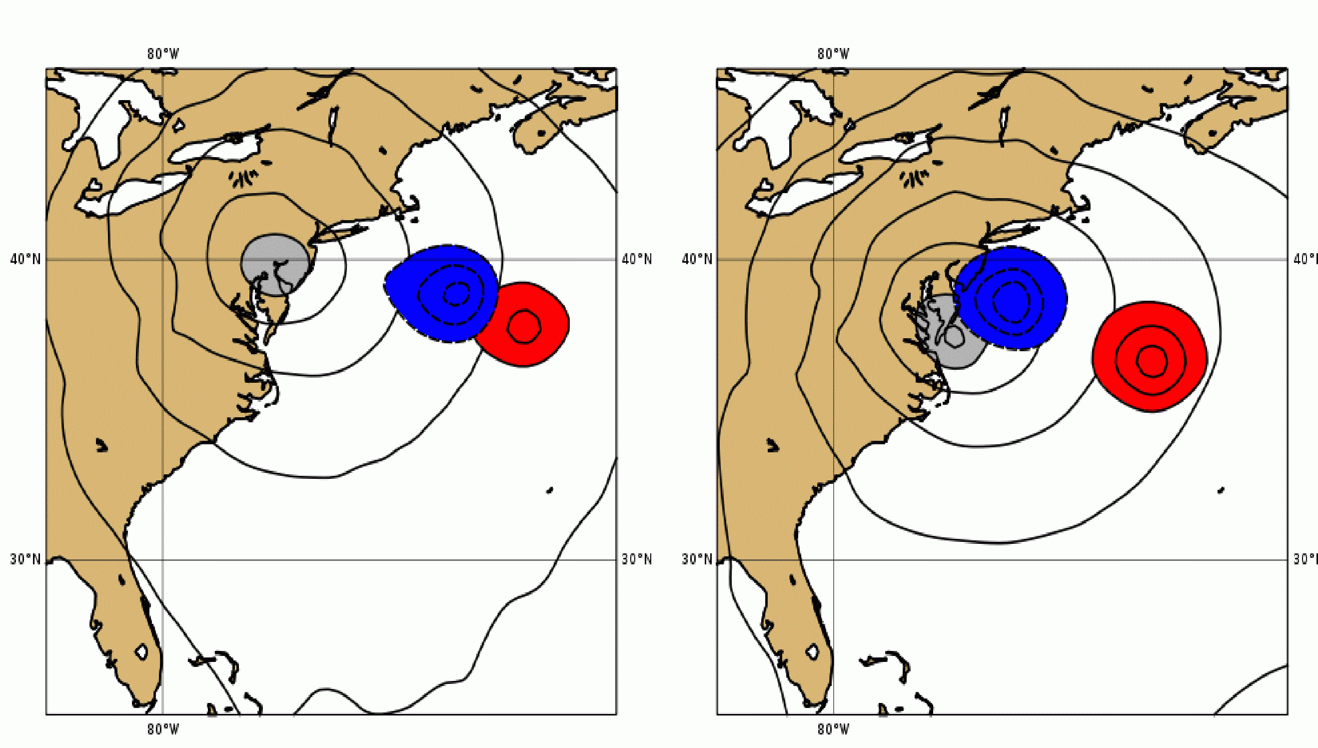
< 100% :
improved model first-guess fit to observations

Control

No Polar: Brute force

No Polar: EDA-based background error covariances

Sandy: impact of background error specifications

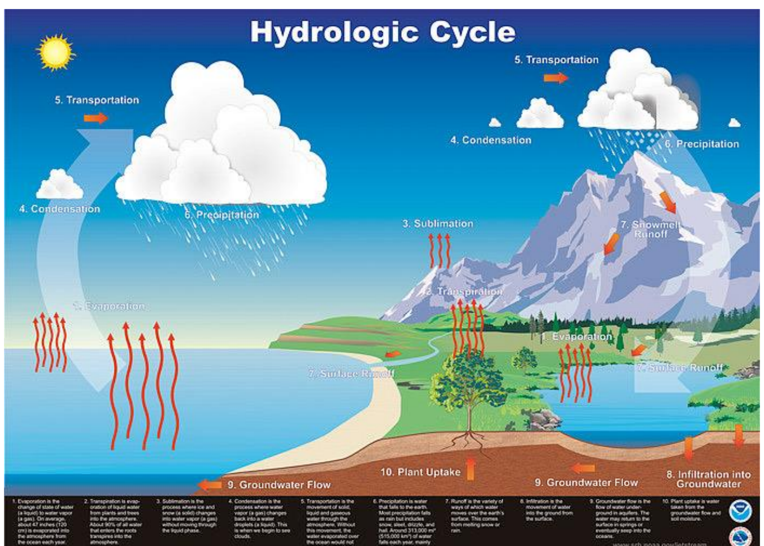


Four day forecasts of surface pressure launched from 26th October (left) and five day forecasts from the 25th October (right) for the control (grey), NOPOLAR (red) and NOPOLAR-EDA (blue). Contours at 10hPa intervals with shading below 970hPa).

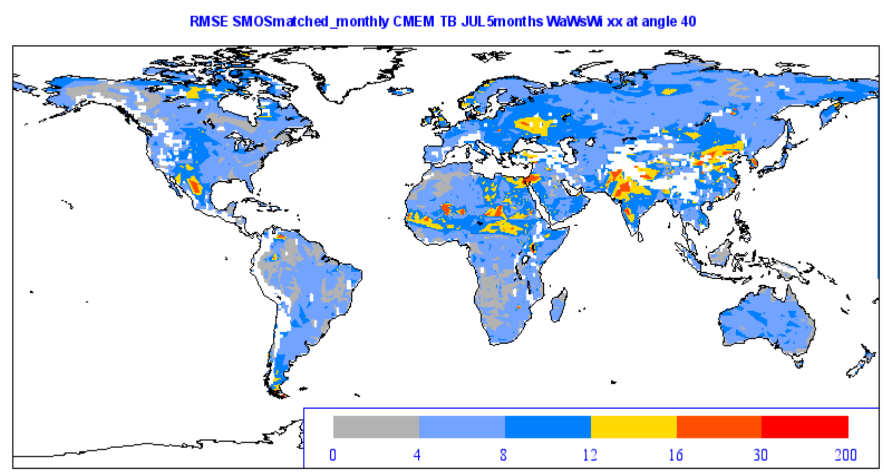
Land Data Assimilation

- Land surfaces: heterogeneities, range of spatial and time scales controlling the processes, reservoirs and fluxes.

- The Land Data Assimilation Systems (LDAS) make use of:
  - Processes and feedbacks represented with coupled land-atmosphere models (extension to carbon cycle)
  - Data assimilation schemes, such as nudging, OI, EKF, EnKF, that update models states variables and/or surface parameters for NWP and climate applications
  - Routine NRT observations with high information content about land surface variables (insitu, SMOS, ASCAT, SMAP, etc.)

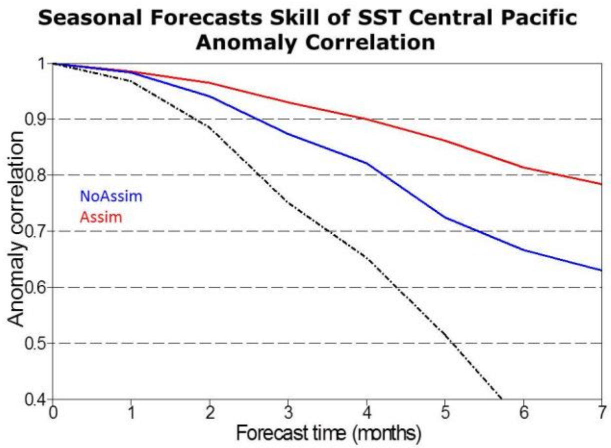


SMOS TB First Guess Departure (K) July 2012, RMSD=6.7K



Ocean Data Assimilation for

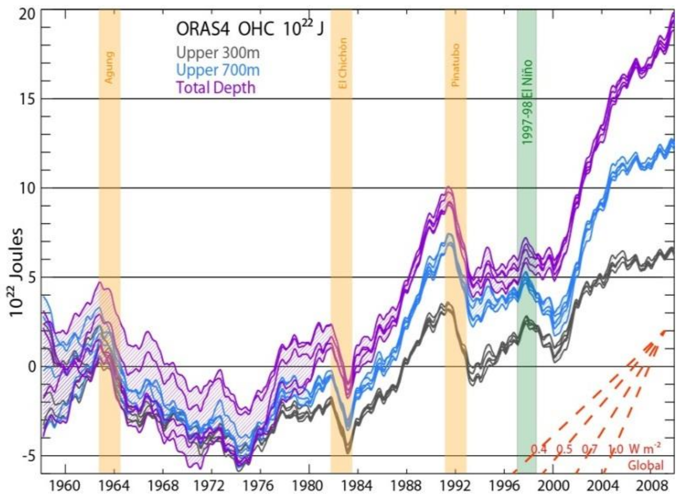
Seasonal Forecasts



NoAssim is ocean model simulation with SST constrain

Balmaseda et al, 2013, QJ

Climate Monitoring



Ocean Heat Content from ORAS4

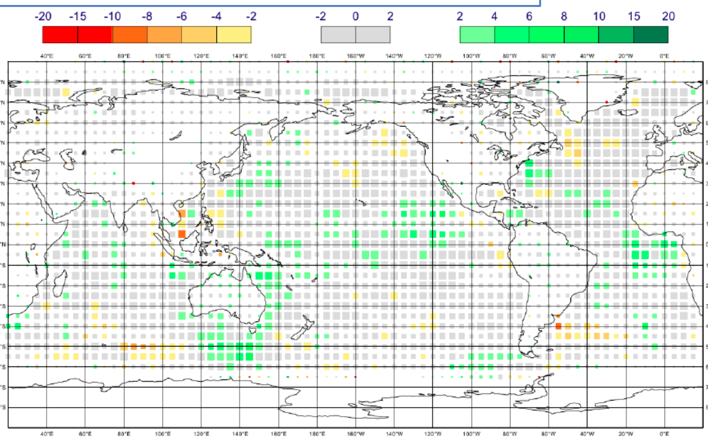
Balmaseda, Trenberth and Kallen, GRL 2013

Decadal Forecasts
Medium-range forecasts

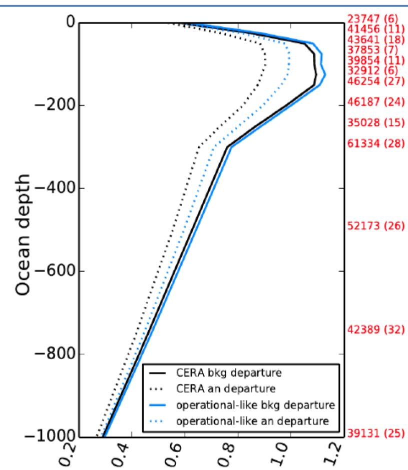
- Relative data paucity
- Slow and fast dynamics to handle

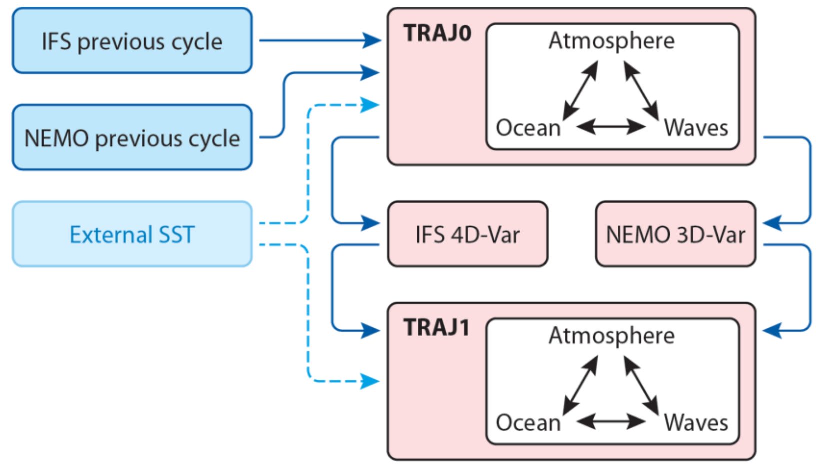
Ex:1 - Benefit of a coupled ocean-atmosphere assimilation

Improved fit to AMSU-A ch-5



Improved fit to ocean T obs. profiles



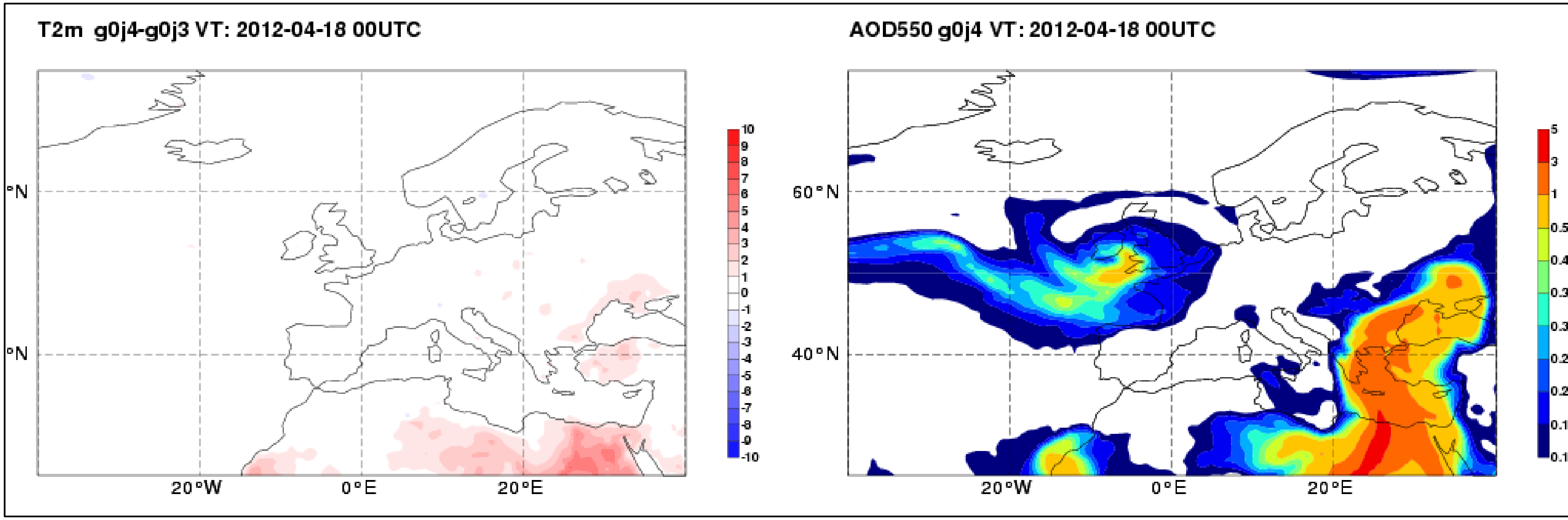


- IFS coupled with NEMO ocean model in 4D-Var outer loop
- External SST/SIC product to correct model bias
- Embed NEMOVAR in inner loop

Coupled assimilation:

- Better balance at the interfaces
- Consistent surface fluxes, mass and energy budgets.
- Observations influence both components of the earth system

Ex:2 - Impact of coupled aerosol assimilation on classical NWP parameters



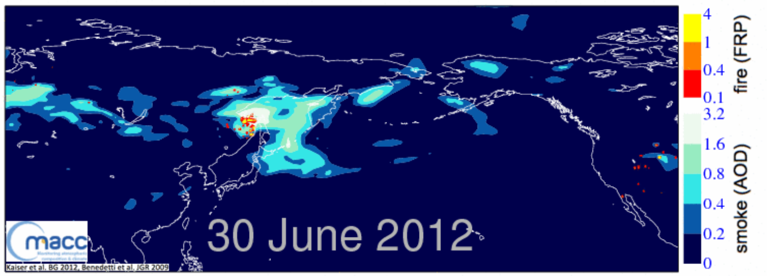
A WGNE initiative

- Taking into account the direct effect of aerosol brings warmer night-time temperatures over land, by up to 4 degrees
- Near-perfect collocation with AOD patterns
- For most stations in desertic area, it reduces night time cold bias

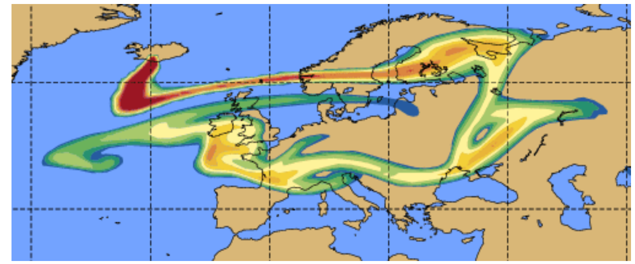
As Earth System modeling and DA become more integrated,
Earth Observations serve a wide range of communities and applications
DIRECTLY or INDIRECTLY

Tomorrow : “super-seamless” predictions

Weather +...

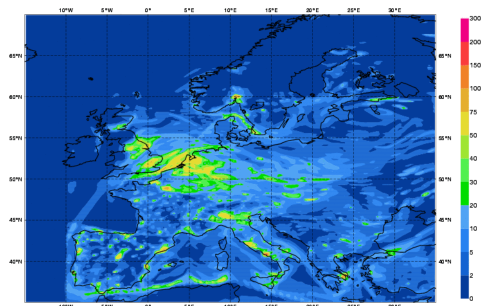


Fires and smoke

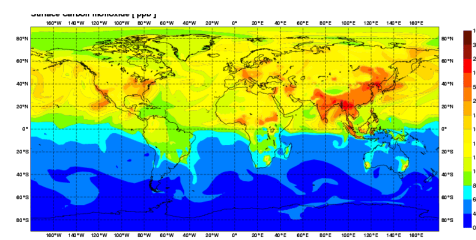


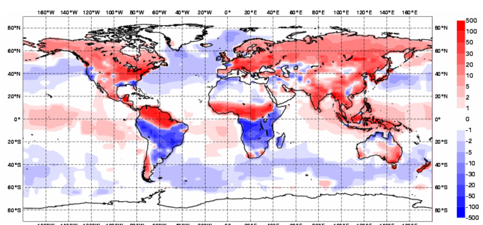
Volcanic eruptions

Air quality

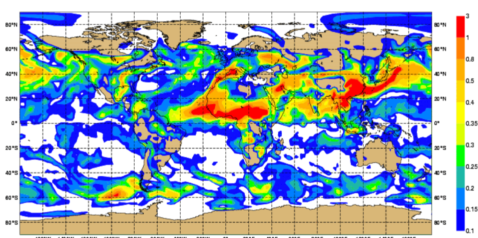


Global pollution



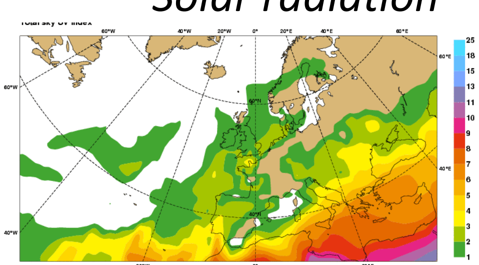


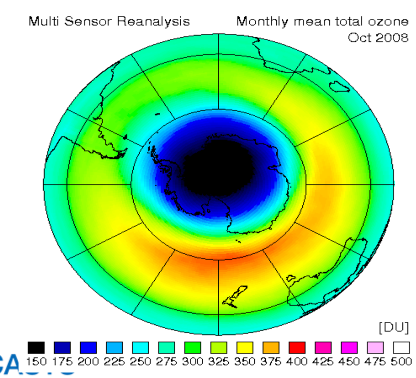
Surface fluxes



Aerosol

Solar radiation

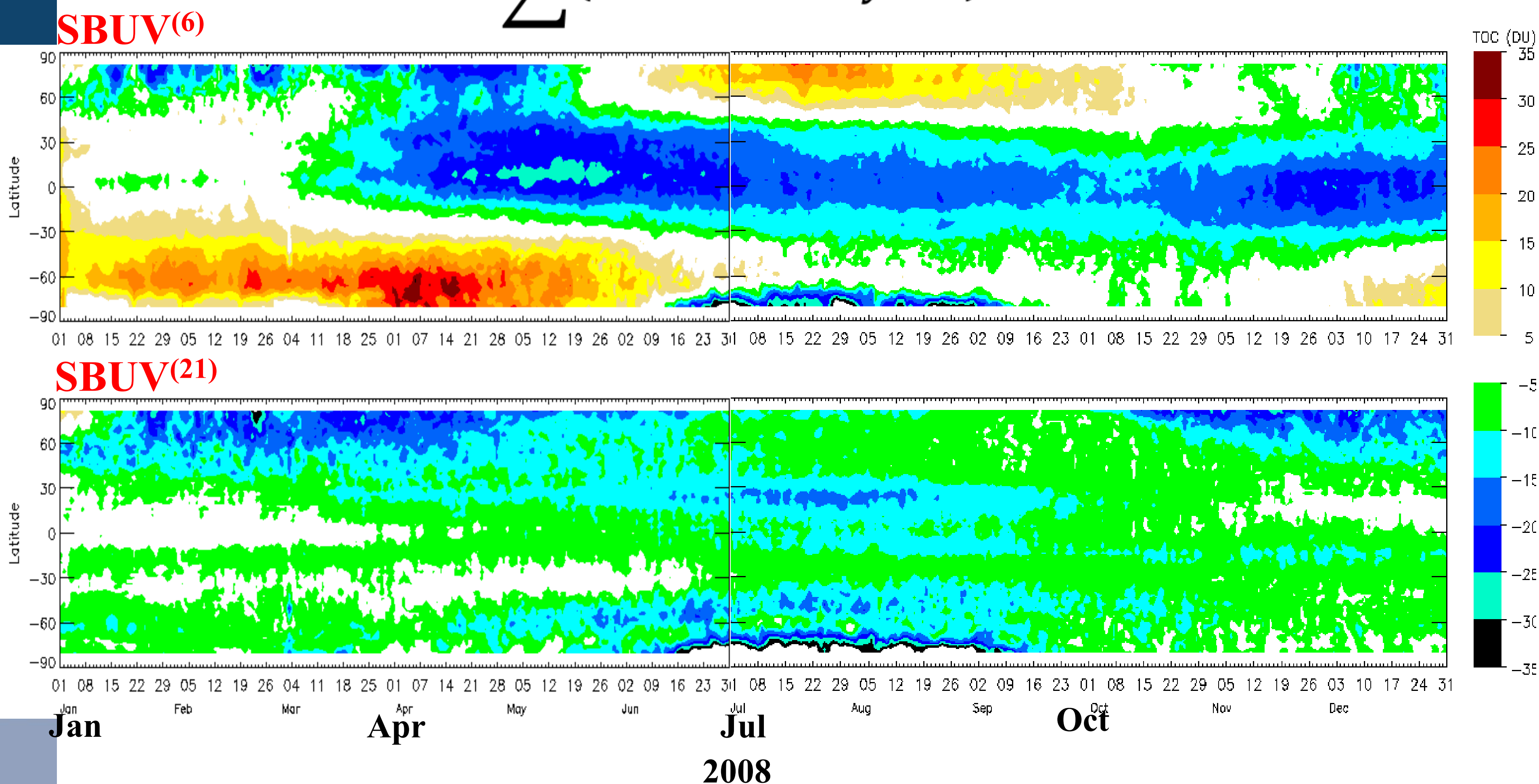




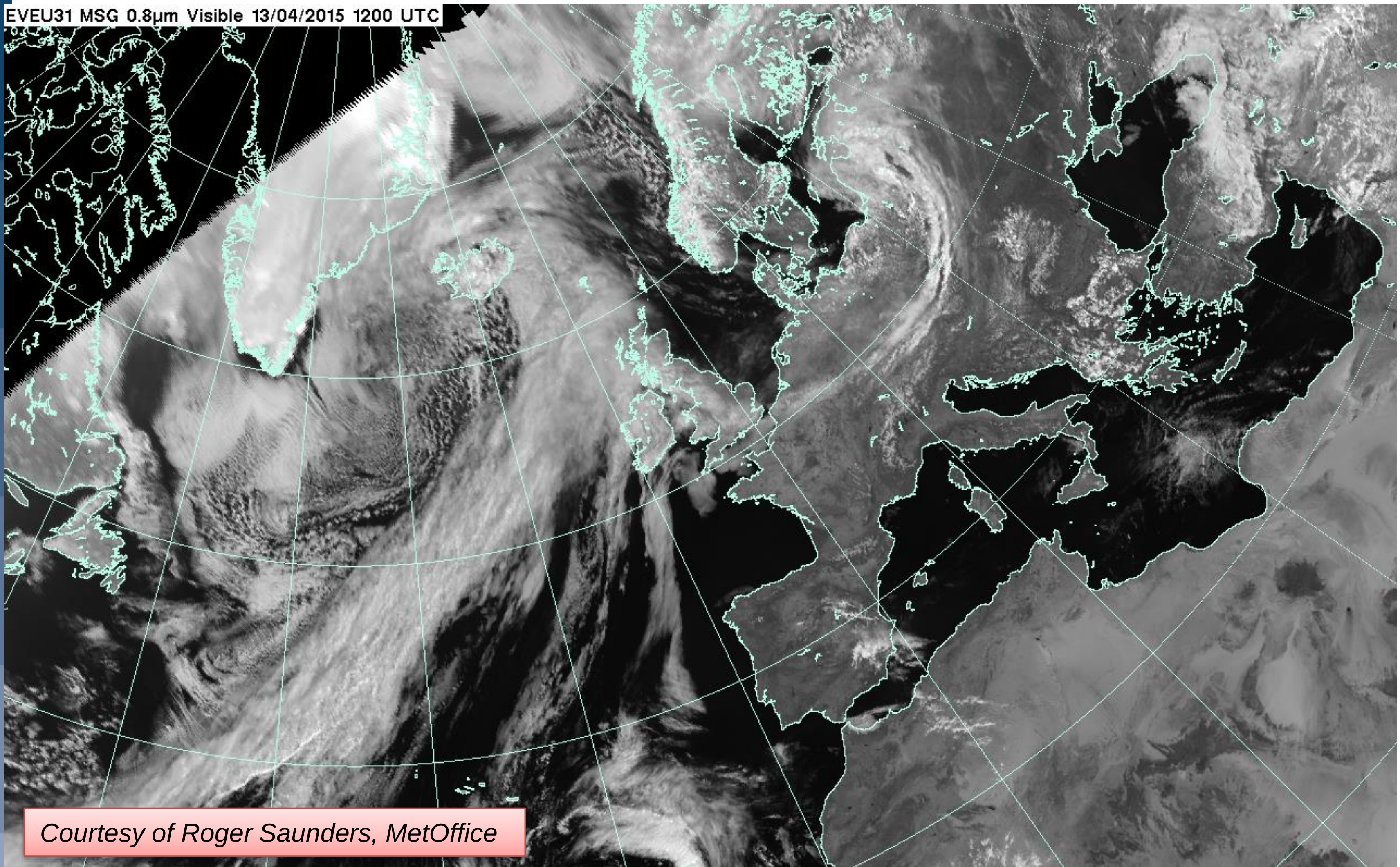
Ozone layer

Impact on the ozone column (NOAA-16, -17, -18)

$$\overline{\sum(MLS - Analyses)}$$

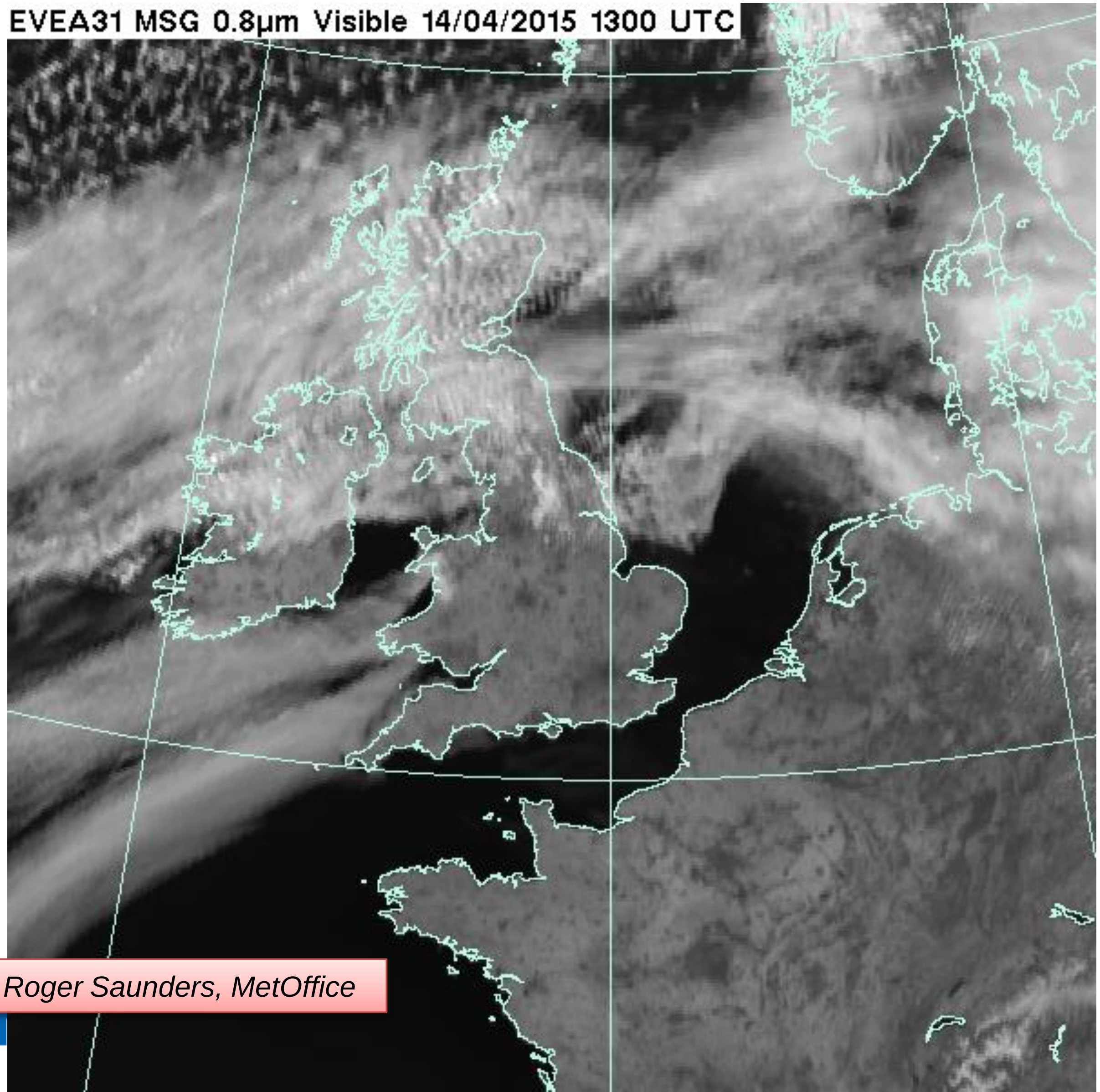


What next: latest improvements in radiative transfer for VIS/NIR simulations



Approx time taken to run forward model for single channel was 4 mins

What next: latest improvements in radiative transfer for VIS/NIR simulations



Courtesy of Roger Saunders, MetOffice

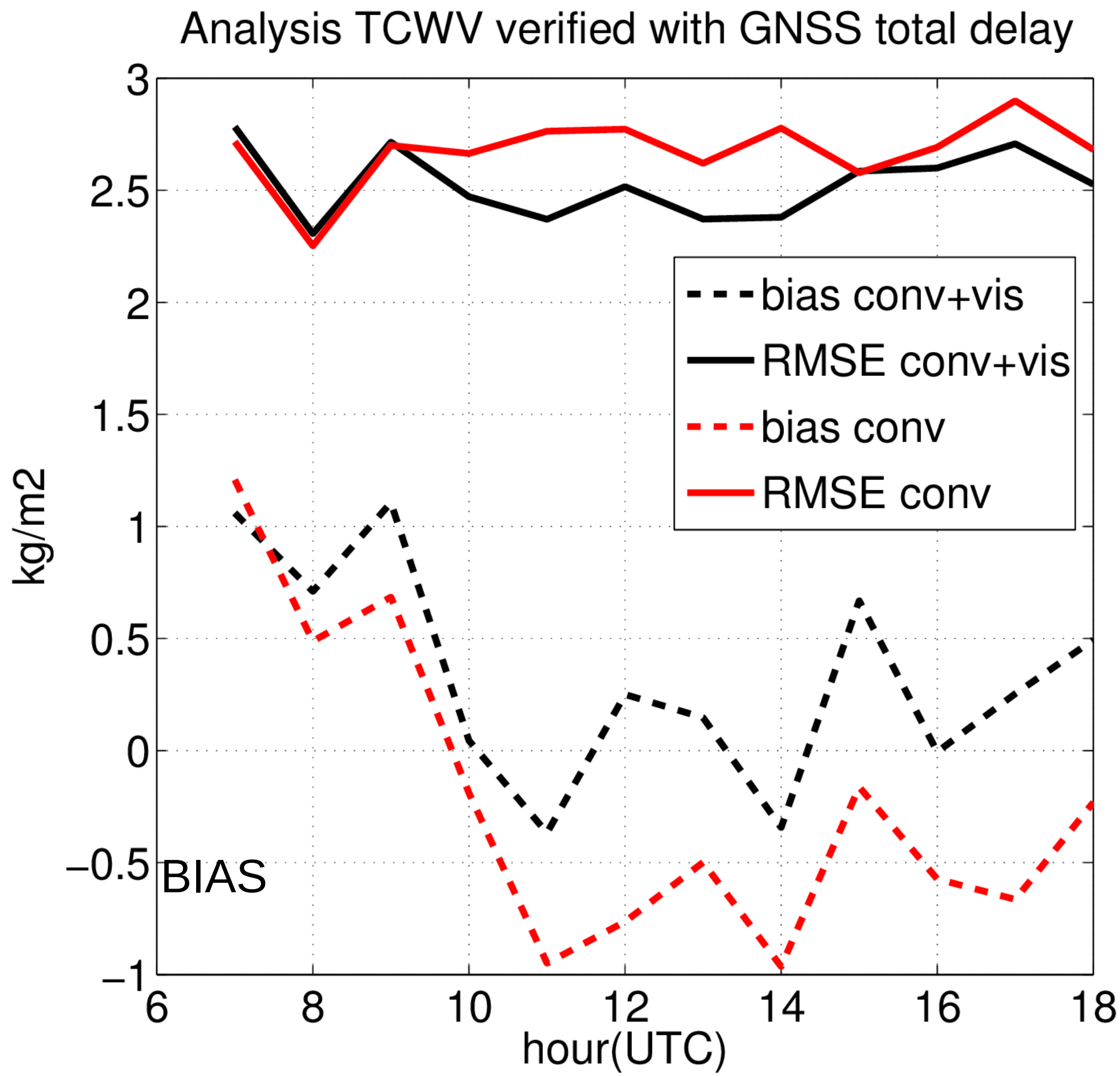
First assimilation results

Assimilation of conventional and/or SEVIRI obs. in COSMO/KENDA

Setup:
40 member LETKF
1h assimilation interval
600nm observations
Observation error 0.2
Superobbing (radius 3 pixels)
Horiz. localization 100km
No vertical localization

Assimilation of SEVIRI observations:
lower reflectance RMSE and bias

Independent GPS humidity observations: reduced error



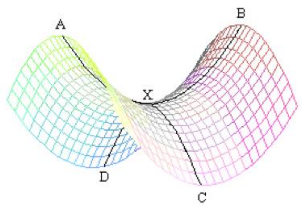
Courtesy of Leonhard Scheck, LMU

Methodologies

More specific challenges and opportunities

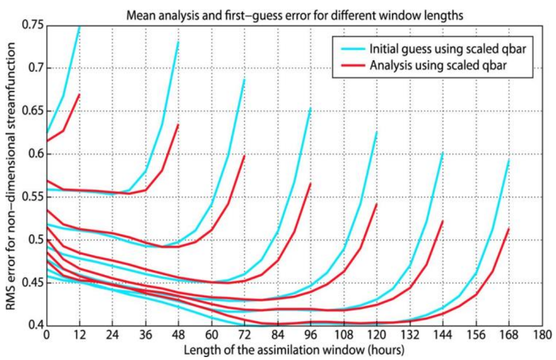
- Scalability
- Meso-scale Data Assimilation
- Climate Monitoring Applications

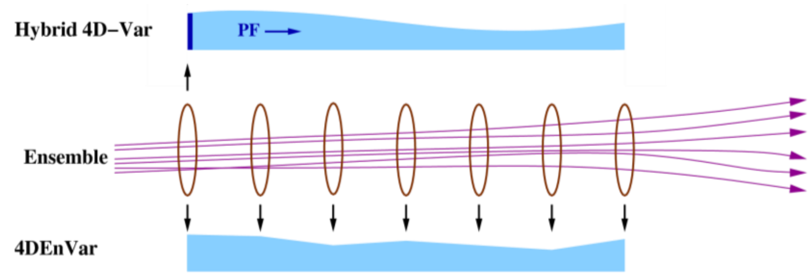
DA has to remain efficient on massively parallel computers



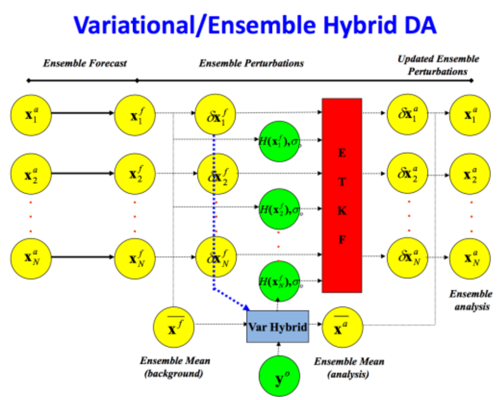
Lagrangian: $\mathcal{L}(\delta\mathbf{x}, \delta\mathbf{p}, \delta\mathbf{w}, \lambda, \mu)$

Long window weak-constraint 4D-Var (saddle point algorithm)





4D-en-Var (no TL/AD needed, ensembles run in parallel, I/O costs to be managed)



ENKF ("embarrassingly parallel") and various hybrid EDA/VAR methods

Pre- and Post-processing of big data are part of the story!

→ Seamless EDA/ENS forecasting

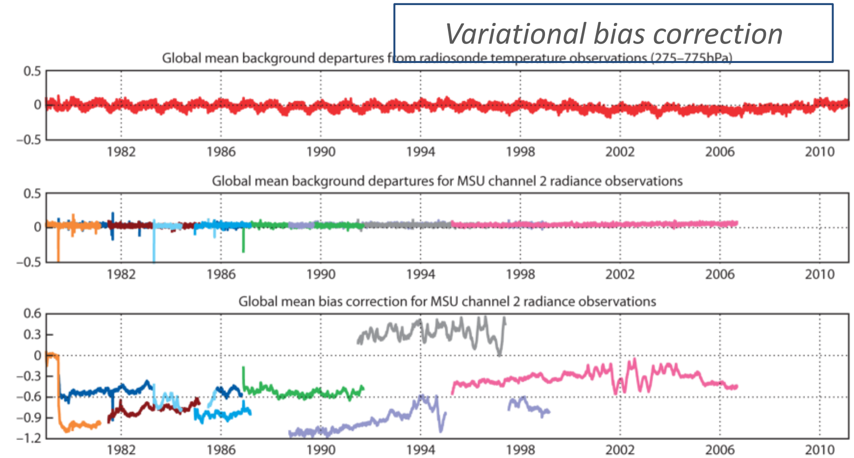
Data Assimilation for climate monitoring

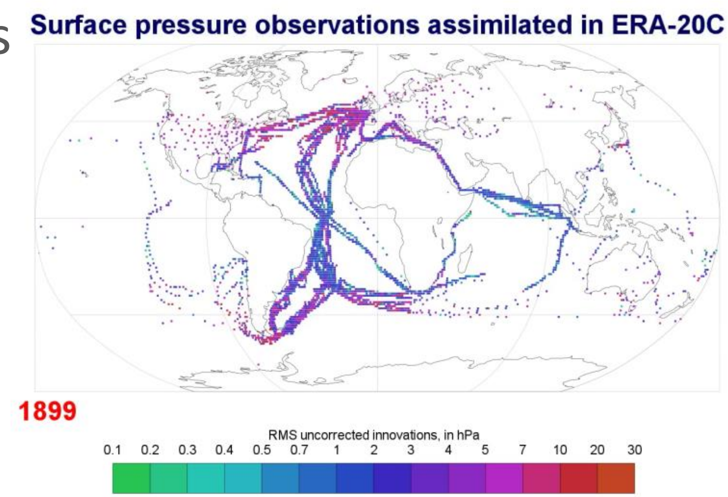
Scientific challenges for reanalysis

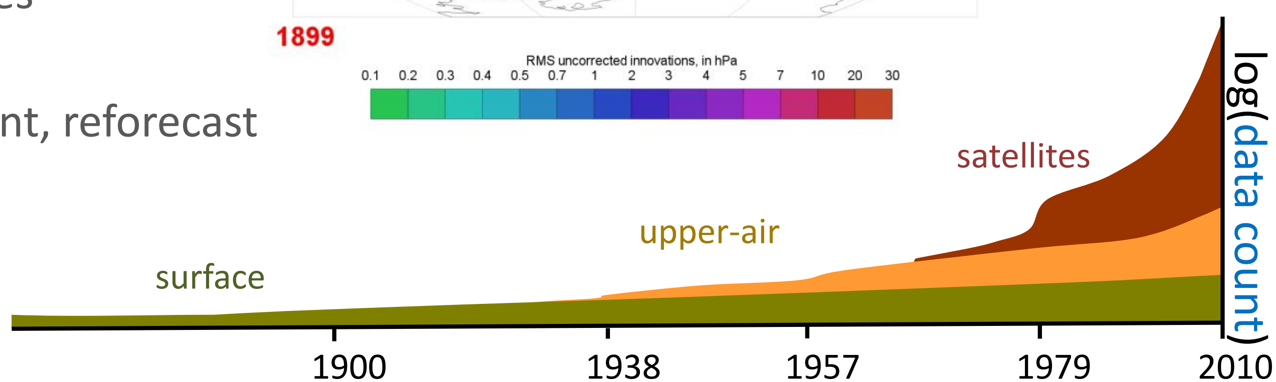
- How to make best use of existing observations (great increase in number and variety of observations over the course of the 20^{th} century)
- How to detect automatically data issues (breaks in station time-series, stuck sensors…)
- How to handle greater uncertainties in the Earth System components as we go back in time

Data Assimilation for climate monitoring

- A science in itself with dedicated methodologies
  - Exploration of long assimilation windows
  - bias correction paramount
- Long term stability requirements for observations
  - Importance of reprocessing
- Handling of changing Observing System
- Estimating uncertainties
  - Need for ensembles
- Fundamental for:
  - Model development, reforecast initialization
  - Climate Services

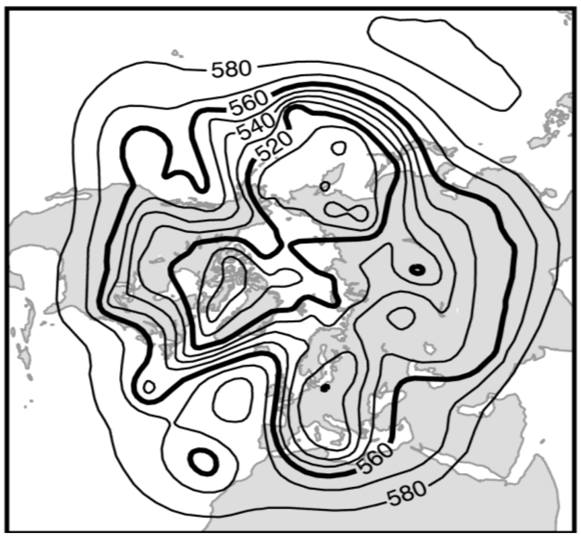




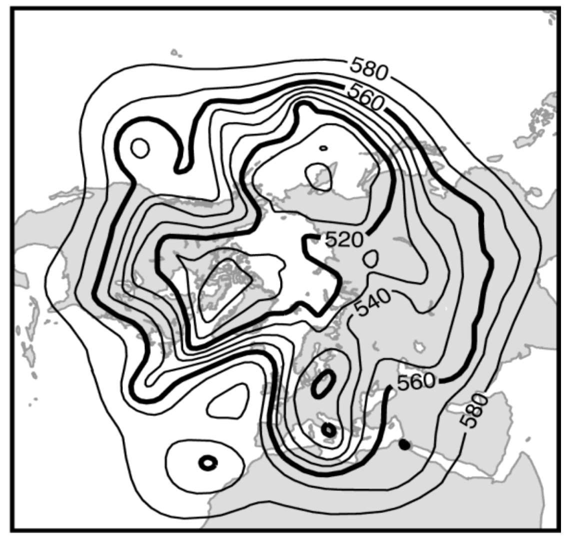


The power of data assimilation

Two modern analyses of geopotential height at 500hPa

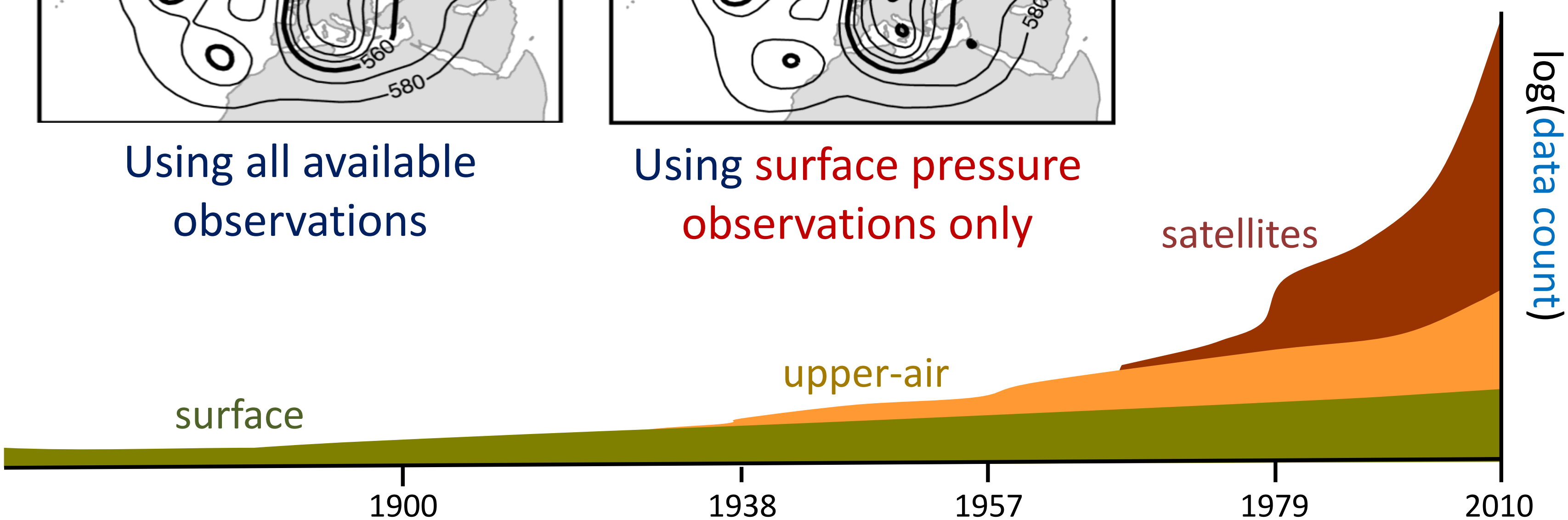


Using all available observations



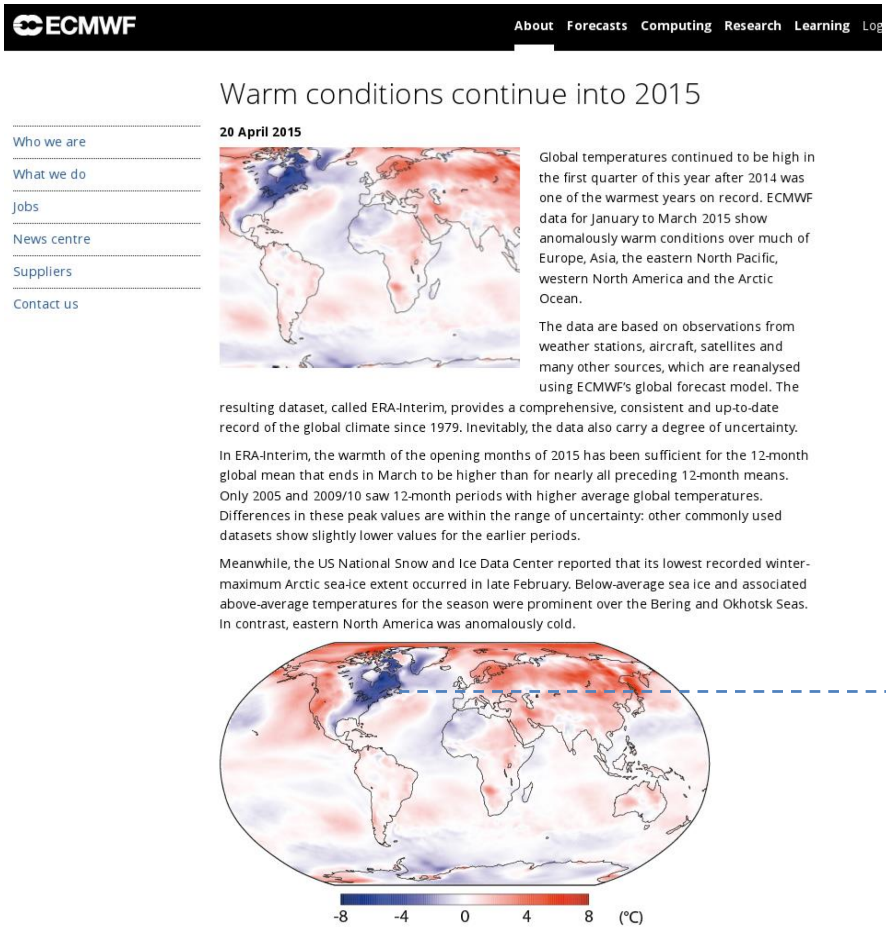
Using surface pressure observations only

Whitaker, Compo, and Thépaut 2009



Reanalysis for Climate Monitoring:

How do we **optimally combine** all the **information** at our disposal to **quickly** place current **weather** events in the **climate** context?



ECMWF | About | Forecasts | Computing | Research | Learning | Log

- Who we are
- What we do
- Jobs
- News centre
- Suppliers
- Contact us

Warm conditions continue into 2015

20 April 2015

Global temperatures continued to be high in the first quarter of this year after 2014 was one of the warmest years on record. ECMWF data for January to March 2015 show anomalously warm conditions over much of Europe, Asia, the eastern North Pacific, western North America and the Arctic Ocean.

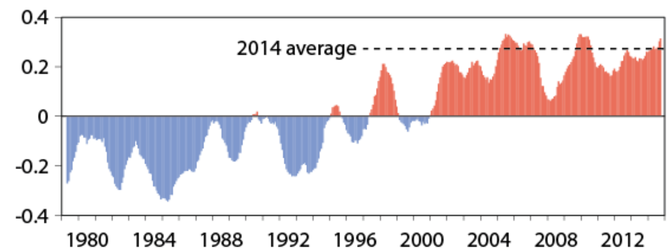
The data are based on observations from weather stations, aircraft, satellites and many other sources, which are reanalysed using ECMWF's global forecast model. The resulting dataset, called ERA-Interim, provides a comprehensive, consistent and up-to-date record of the global climate since 1979. Inevitably, the data also carry a degree of uncertainty.

In ERA-Interim, the warmth of the opening months of 2015 has been sufficient for the 12-month global mean that ends in March to be higher than for nearly all preceding 12-month means. Only 2005 and 2009/10 saw 12-month periods with higher average global temperatures. Differences in these peak values are within the range of uncertainty: other commonly used datasets show slightly lower values for the earlier periods.

Meanwhile, the US National Snow and Ice Data Center reported that its lowest recorded winter-maximum Arctic sea-ice extent occurred in late February. Below-average sea ice and associated above-average temperatures for the season were prominent over the Bering and Okhotsk Seas. In contrast, eastern North America was anomalously cold.

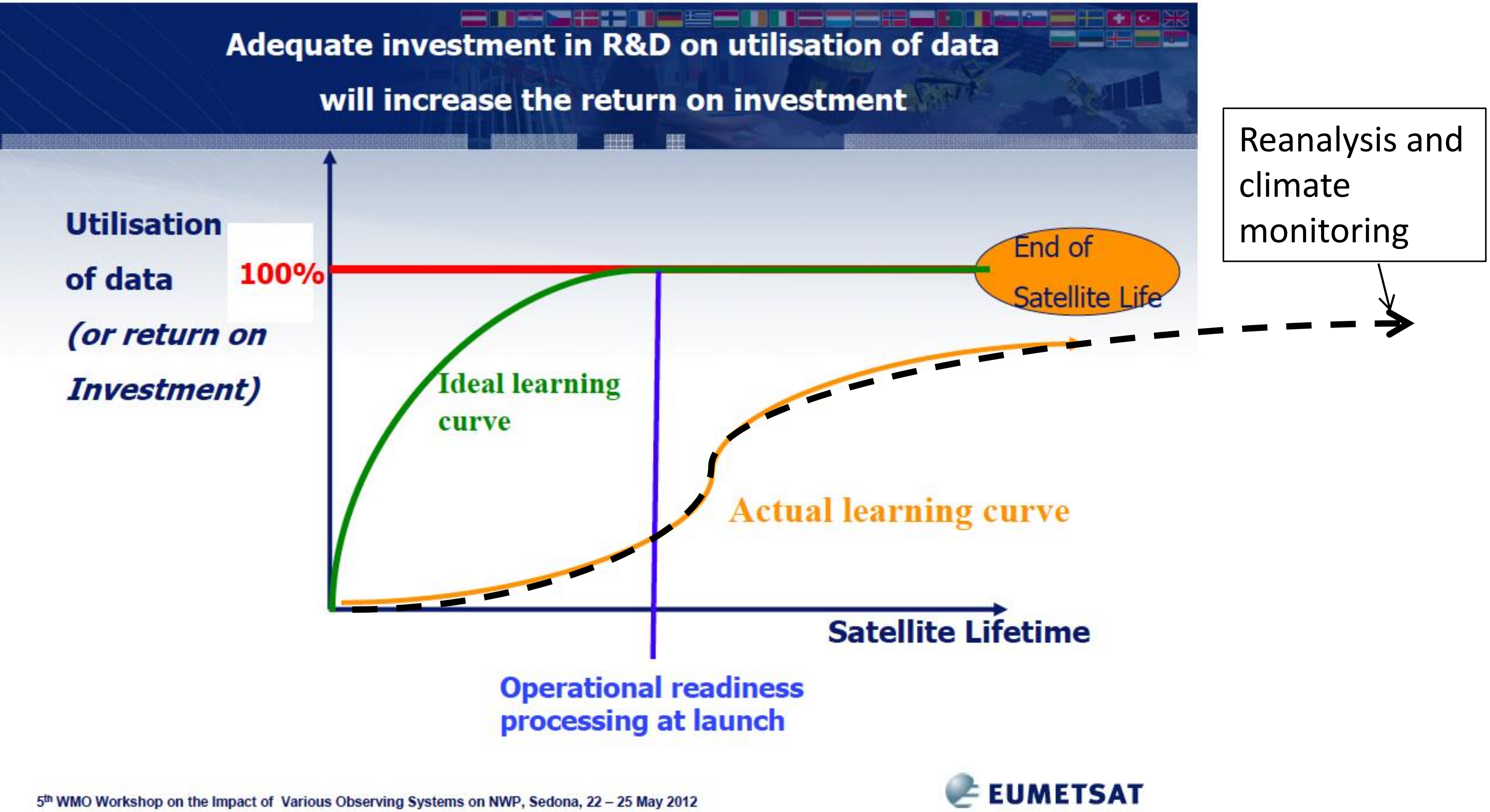
Differences between two-metre temperature for January–March 2015 and the 1981–2010 average for

Reanalysis continued into the present



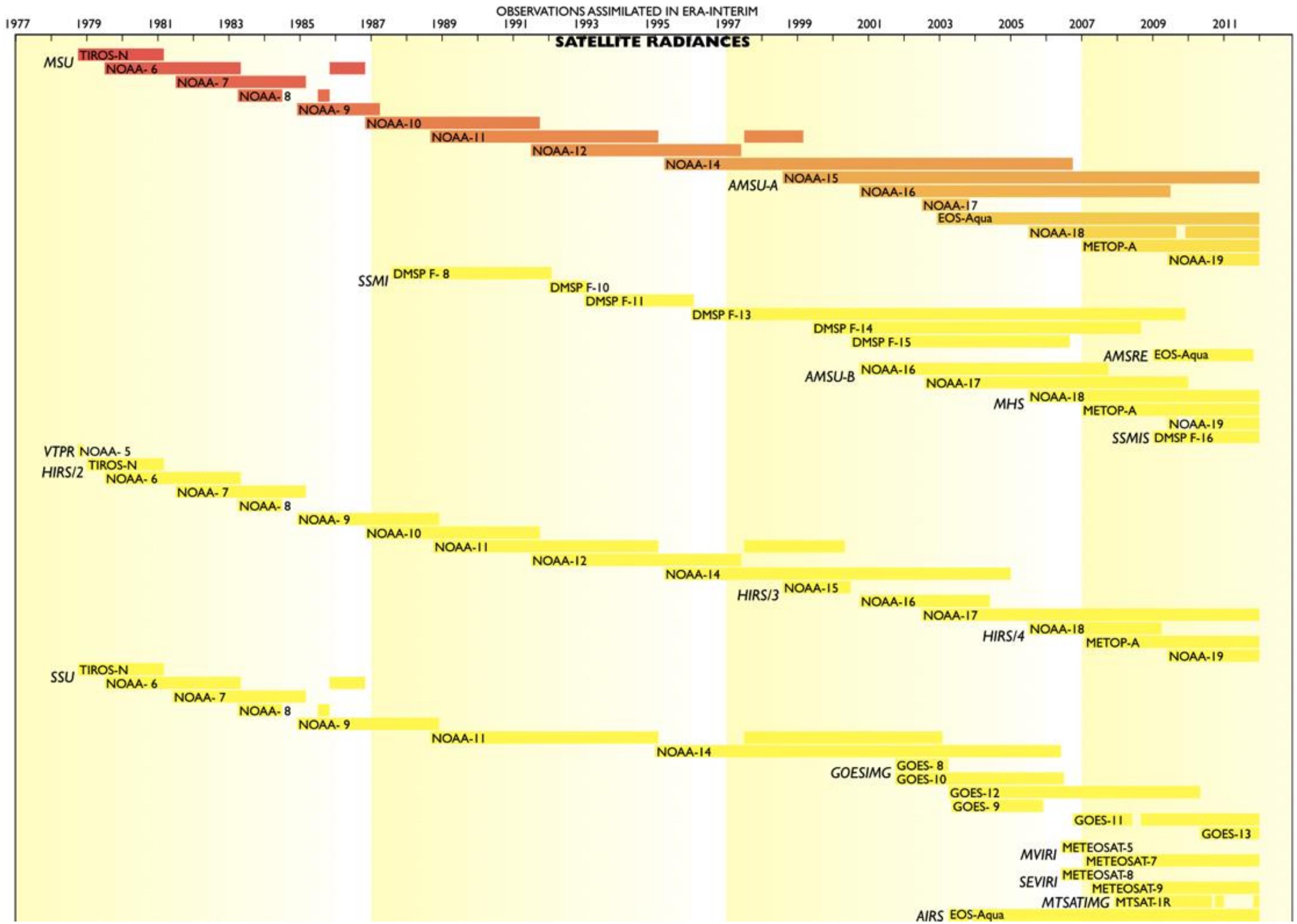
Twelve-month running-mean anomalies relative to 1981–2010 in global-mean surface temperature from the ERA-Interim reanalysis

http://www.ecmwf.int/en/about/media-centre/news/2015/warm-conditions-continue-2015

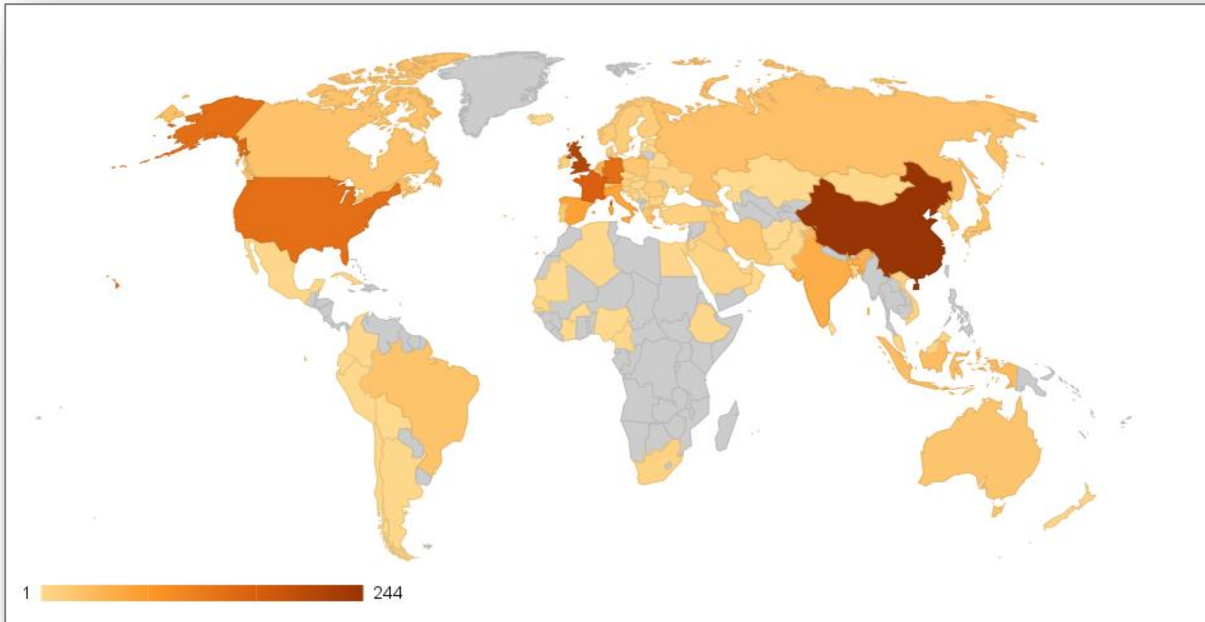


Source: Johannes Schmetz

Satellite radiance data used in ERA-Interim



MACC ATMOSPHERIC COMPOSITION REANALYSES

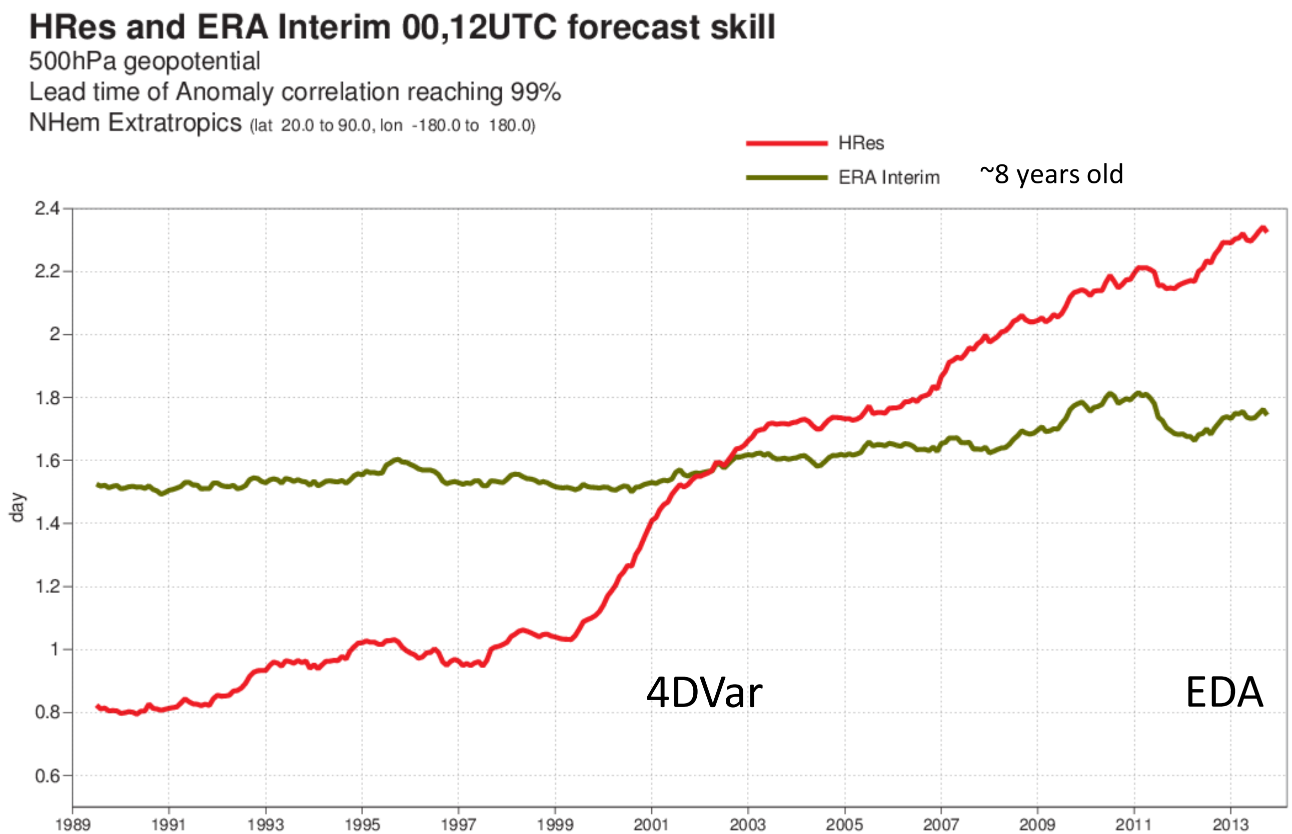


Users of the global composition re-analysis product

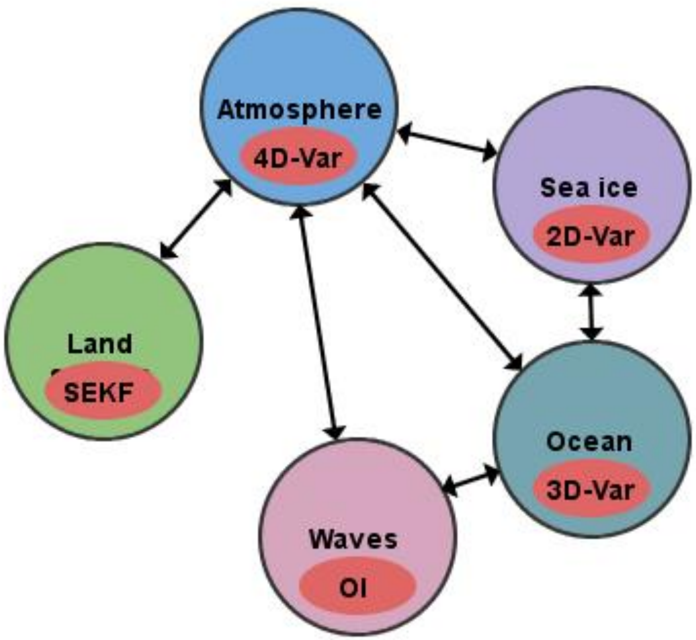
Similar initiatives around the world (e.g. MERRA(AERO)(2)

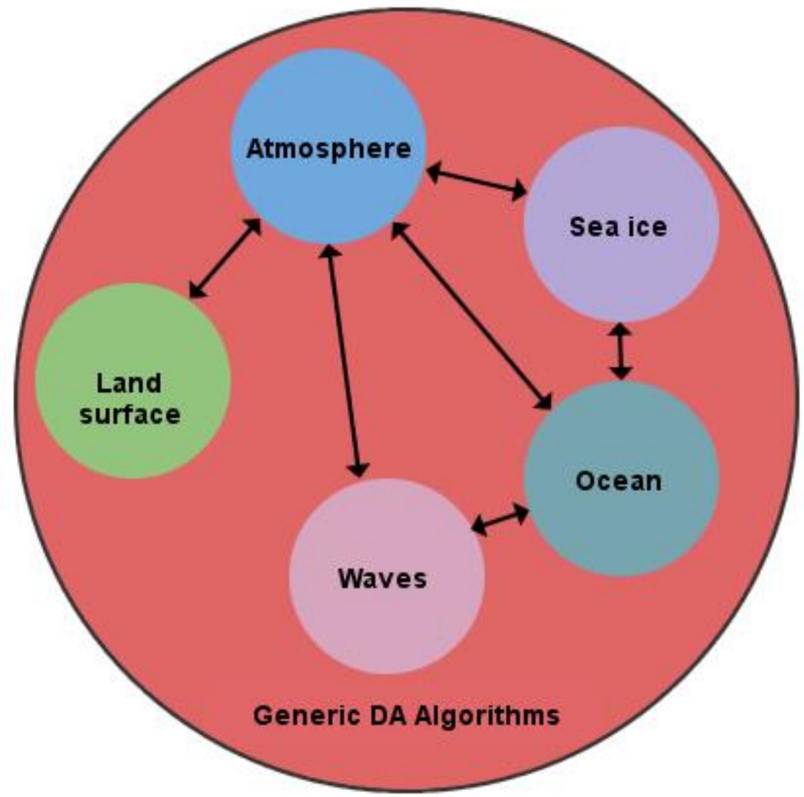
Where are we today?

Short range forecast skill: We are still progressing



Positioning Data Assimilation at the heart of earth-system modeling





Conclusions

- Prospects of reducing further initial condition errors are great
- DA is the natural vehicle to confront models and observations
- DA contribute to a seamless quantification of uncertainty estimation
- Full exploitation of the GOS needs:
  - Careful planning and coordination with data providers
  - Sustained investment in model and DA developments
  - Factoring cost-benefit in an integrated earth-system context
- Efficiency on future HPCs will be a fundamental driver
- Future prospects for multi-coupled assimilation
- DA provides the framework and the tools for atmosphere and climate monitoring
- Further progress will require overarching integration of data assimilation developments at the heart of Numerical modelling

Thank You