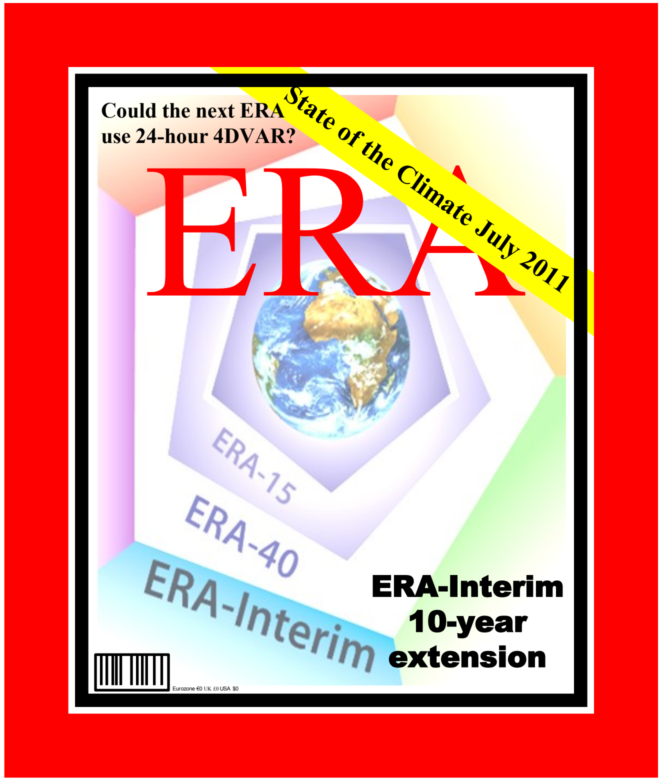
Data Assimilation for Atmospheric Reanalysis



P. Poli

ECMWF ReAnalysis (ERA) Team:
D. Dee, P. Berrisford, R. Brugge, H. Hersbach, C. Peubey, P. Poli, H. Sato, D. Tan

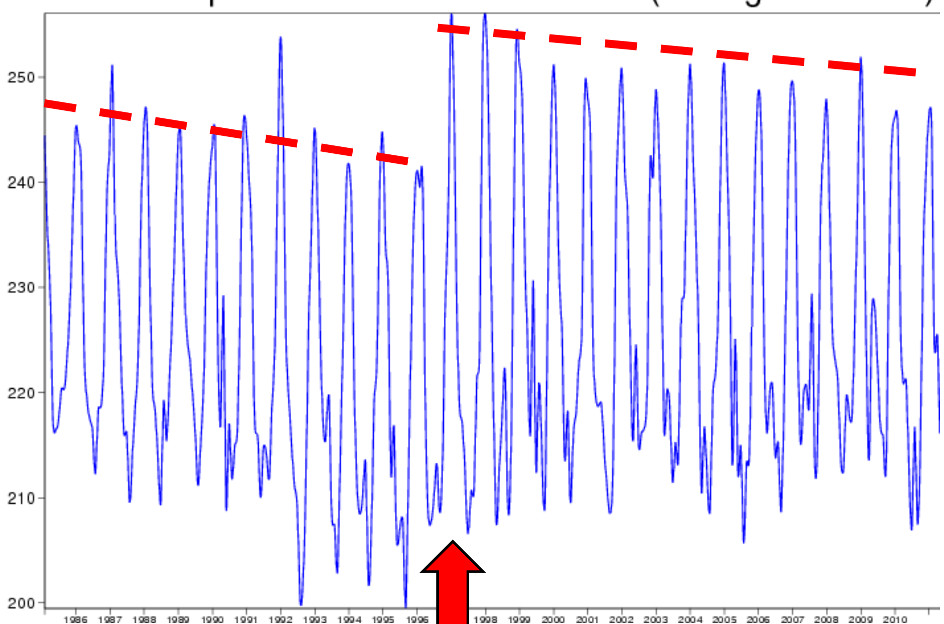
ECMWF

Outline

- Introduction:
  TIME, our great concern
- Reanalysis products
- Reanalysis process
- Recent developments
- Conclusion

ECMWF

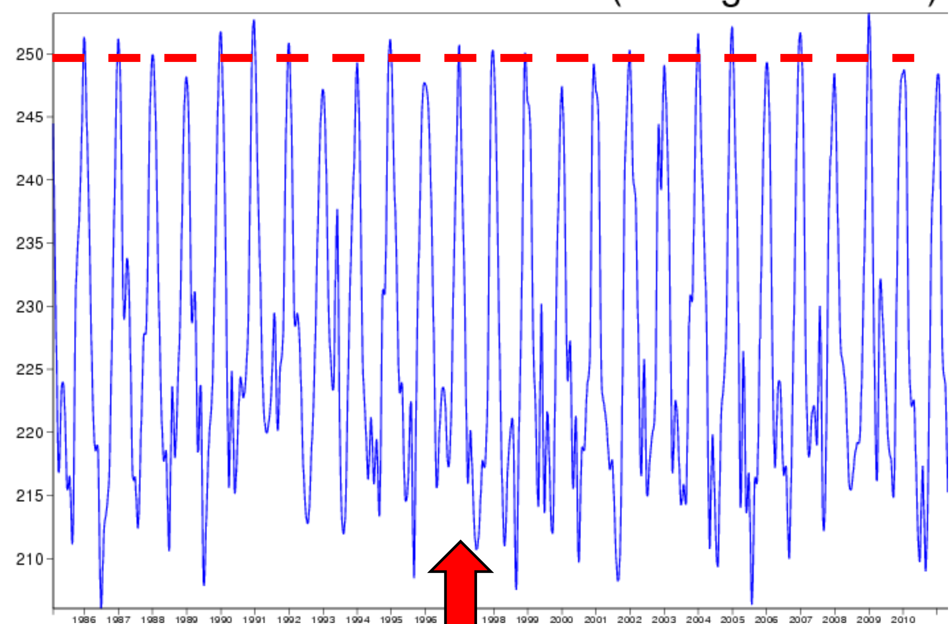
NWP already produced series of analyses – Why do we need REanalysis?



1 Feb 1985

1 May 2011

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



1 Feb 1985

1 May 2011

… probably not

Two time dimensions of reanalysis

Time period to re-analyze

Time it takes to re-analyze

ECMWF

The word “REANALYSIS” is used to refer to…

- Reanalysis product(s)
  - Gridded fields of NWP model
    - Control variables: vorticity, divergence, humidity, ozone…
    - Derived variables: precipitation, radiation…
  - Fit to observations
    - Before, and after, assimilation
    - Before, and after, bias correction
- Reanalysis process
  - 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

Outline

- Introduction:
  TIME, our great concern
- Reanalysis products
- Reanalysis process
- Recent developments
- Conclusion

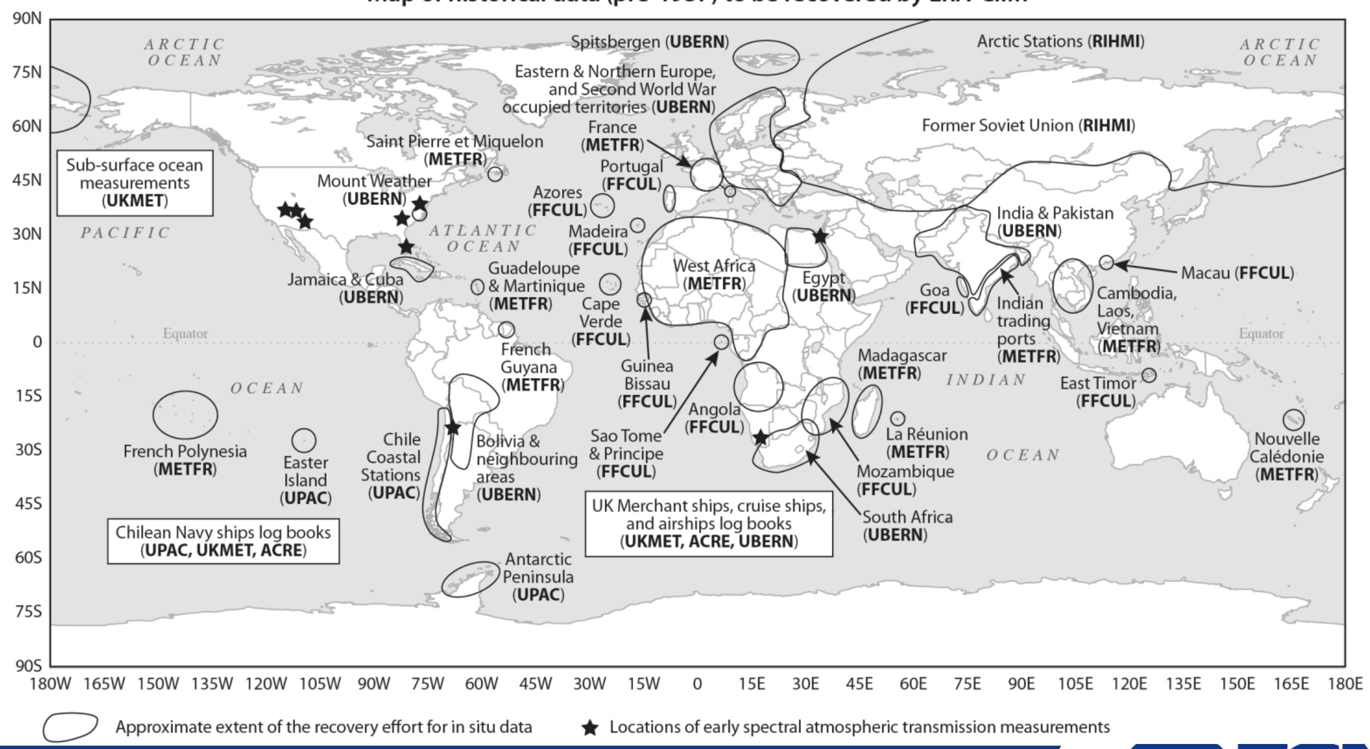
A short history of global atmospheric reanalyses

- **1979**: Observation datasets collected for the First GARP Global Atmospheric Research Program Experiment (FGGE): used *a posteriori* for several years, to initialize models, compare performance and track progress in NWP.
- **1983**: Reanalysis concept proposed by Daley for monitoring the impact of forecasting system changes on the accuracy of forecasts
- **1988**: Concept proposed again, but for climate-change studies, in two separate papers: by Bengtsson and Shukla, and by Trenberth and Olson
- **1990s**: First-generation comprehensive global reanalysis products *(~OI-based)*
  - NASA/DAO (1980 - 1993) from USA
  - NCEP/NCAR (1948 - present) from USA
  - ERA-15 (1979 - 1993) from ECMWF – with significant funding from USA
- **Mid 2000s**: Second-generation products *(~3DVAR)*
  - JRA-25 (1979 - 2004) from Japan
  - NCEP/DOE (1979 - present) from USA
  - ERA-40 (1958 - 2001) from ECMWF – with significant funding from EU FP5
- **Today**: third generation of comprehensive global reanalyses (*~better than 3DVAR*)
  - NASA/GMAO-MERRA (1979 - present) from USA (IAU)
  - NCEP-CFSRR (1979 - 2008) from USA (land/ocean/ice coupling)
  - ERA-Interim (1979 - present) from ECMWF (4DVAR)
  - JRA-55 (1958 - 2012) from Japan (4DVAR)
  - 20-CR from USA (Ensemble Kalman Filter, surface pressure obs. only)

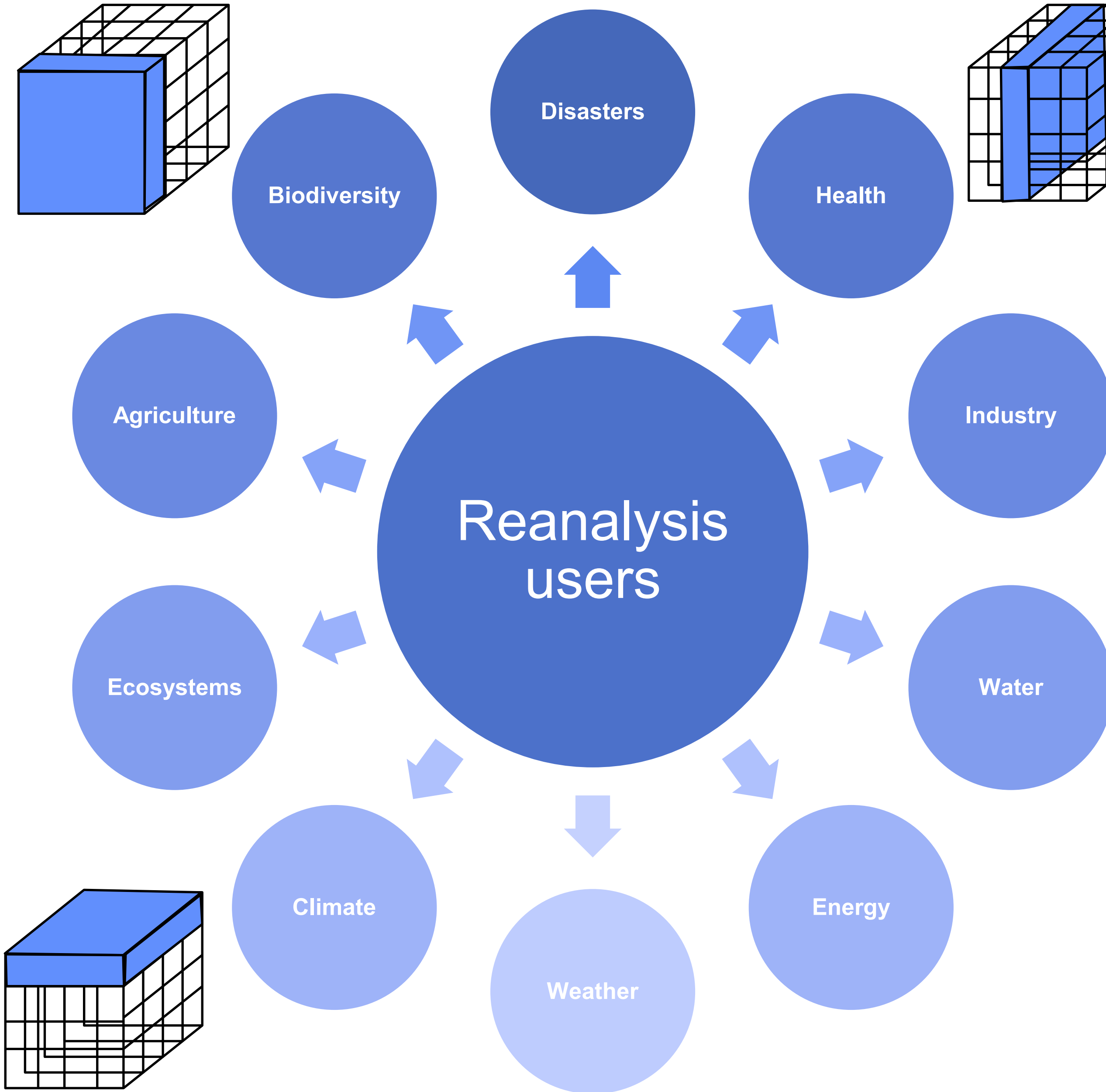
Ongoing reanalyses at ECMWF

- **ERA-Interim**
  - Initial intent: 1989-2008
  - Since 2009: updated monthly (into the present)
  - In 2011: extended back to 1979
- **In preparation**
  - With EU FP7 ERA-CLIM partners
  - ERA-20C: reanalysis of 20th Century
  - ERA-SAT: 1979-present
  - Ambitious data rescue effort

Map of historical data (pre-1957) to be recovered by ERA-Clim



ECMWF



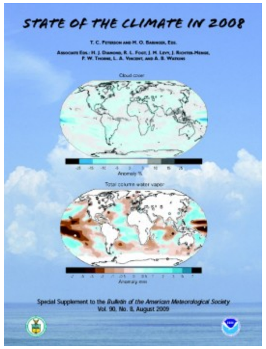
- **Thousands of users**
  - 12000 registered users of ERA public data server
  - ≥5M fields retrieved daily by ECMWF and Member-State users
  - National mirror sites for ERA in several countries
- **Many citations**
  - Paper on NCEP/NCAR and ERA-40 reanalyses highly cited paper in geosciences
  - Increasing use in climate studies

ECMWF

Can you spot a trend?

Note: past performance is no guarantee of future results

BAMS State of the Climate in 2008



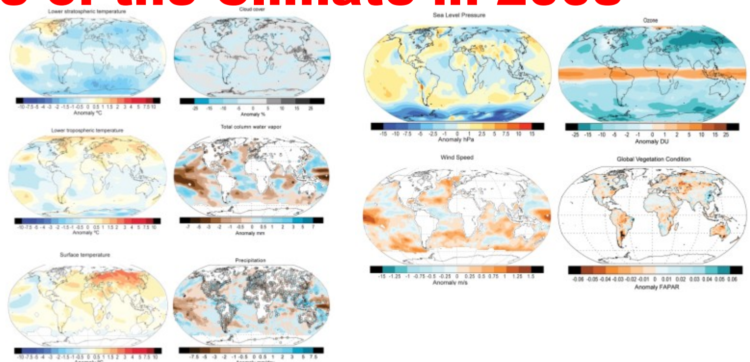
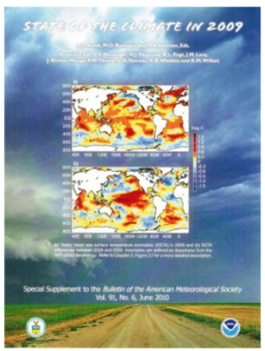


Plate 2.1. Global annual anomaly maps for those variables for which it was possible to create a meaningful anomaly estimate. Climatologies differ among variables, but spatial patterns should largely dominate over choices of climatology period. Dataset sources and climatologies are given in the form (dataset name/data source, start year–end year) for each variable. See relevant section text and figures for more details. Lower stratospheric temperature (RSS MSU 1981–90); lower tropospheric temperature (UAH MSU 1981–90); surface temperature (NCDC 1961–90); cloud cover (PATMOS-x 1982–2008); total column water vapor (SSM/I/GPS 1997–2008); precipitation (RSS/GHCN 1989–2008); mean sea level pressure (HadSLP2r 1961–90); wind speed (SSM/I1988–2007); total column ozone (annual mean global total ozone anomaly for 2008 from SCIAMACHY. The annual mean anomalies were calculated from 1 1.25 gridded monthly data after removing the seasonal mean calculated from GOME (1996–2003) and SCIAMACHY (2003–07)]; vegetation condition [annual FAPAR anomalies relative to Jan 1998 to Dec 2008 from monthly FAPAR products at 0.5 0.5 [derived from SeaWiFS (NASA) and MERIS (ESA) data].

BAMS State of the Climate in 2009



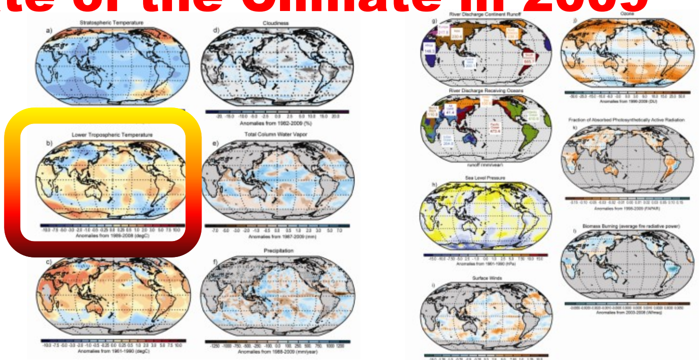
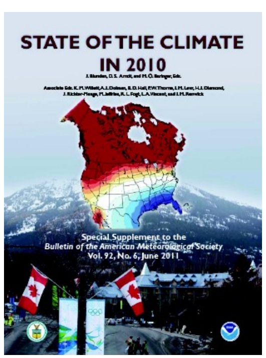


Plate 2.1. Global annual anomaly maps for those variables for which it is possible to create a meaningful 2009 anomaly estimate. Climatologies differ among variables, but spatial patterns should largely dominate over choices of climatology period. Dataset sources/names are as follows: lower stratospheric temperature (RSS MSU); lower tropospheric temperature (ERA-interim); surface temperature (NOAA NCDC); cloudiness (PATMOS-x); total column water vapor (SSM/I over ocean, ground based GPS over land); precipitation (RSS over ocean, GHCN (gridded) over land); river discharge (authors); mean sea level pressure (HadSLP2r); wind speed (AMSR-E); ozone (GOME2); FAPAR (SeaWIFS); Biomass Burning (GEMS/MACC). See relevant section text and figures for more details.

BAMS State of the Climate in 2010



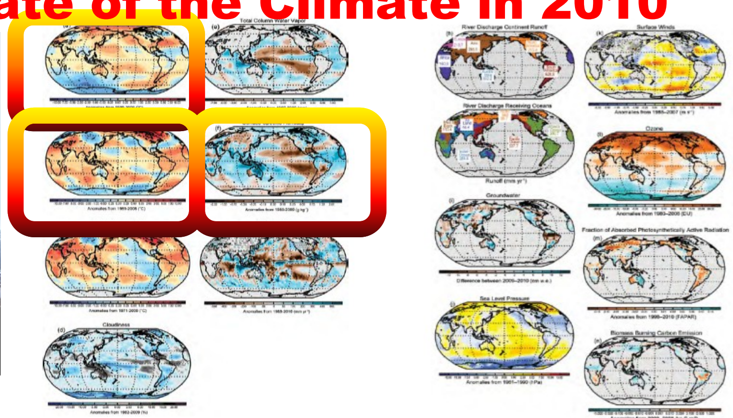
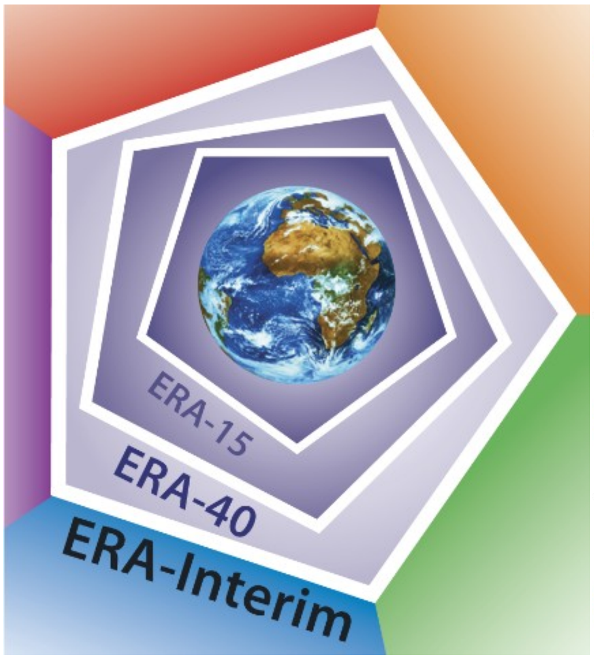


Plate 2.1. Global annual anomaly maps for those variables for which it is possible to create a meaningful 2010 anomaly estimate. Reference base periods differ among variables, but spatial patterns should largely dominate over choices of base period. Dataset sources/names are as follows: lower stratospheric temperature (ERA-Interim); lower tropospheric temperature (ERA-Interim); surface temperature (NOAA/NCDC); cloudiness (PATMOS-x); total column water vapor (AMSR-E over ocean, ground-based GPS over land); surface specific humidity (ERA-Interim); precipitation (RSS over ocean, GHCN (gridded) over land); groundwater 2010–2009 differences (the sum of groundwater, soil water, surface water, snow, and ice, as an equivalent height of water in cm) (GRACE); river discharge absolute values (authors); mean sea level pressure (HadSLP2r); surface wind speed (AMSR-E over ocean, authors in situ over land); ozone (SBUVs/OMI/TOMS/GOME1/SCIAMACHY/GOME2, base period data from the multi-sensor reanalysis, MSR); FAPAR [SeaWiFS (NASA) and MERIS (ESA) sensors]; biomass burning (GFAS). See relevant section text and figures for more details.

"ERA-Interim State of the Climate" July 2011



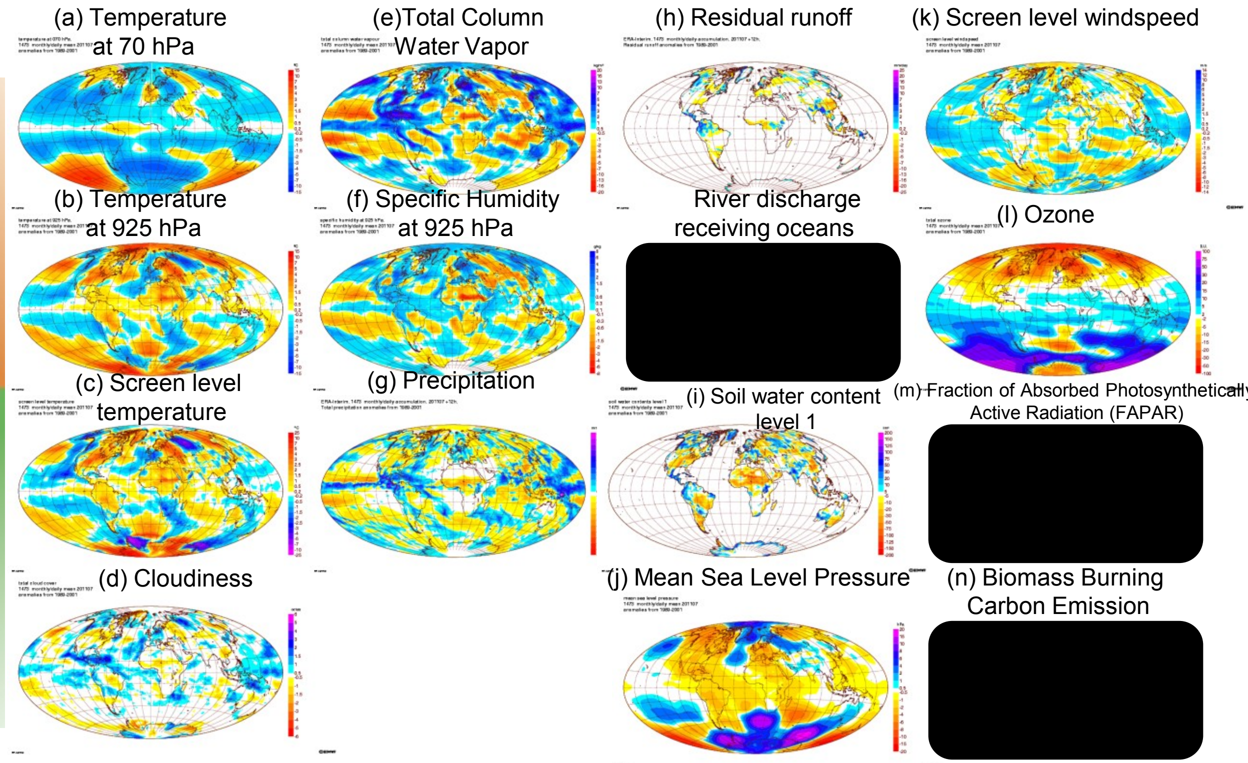


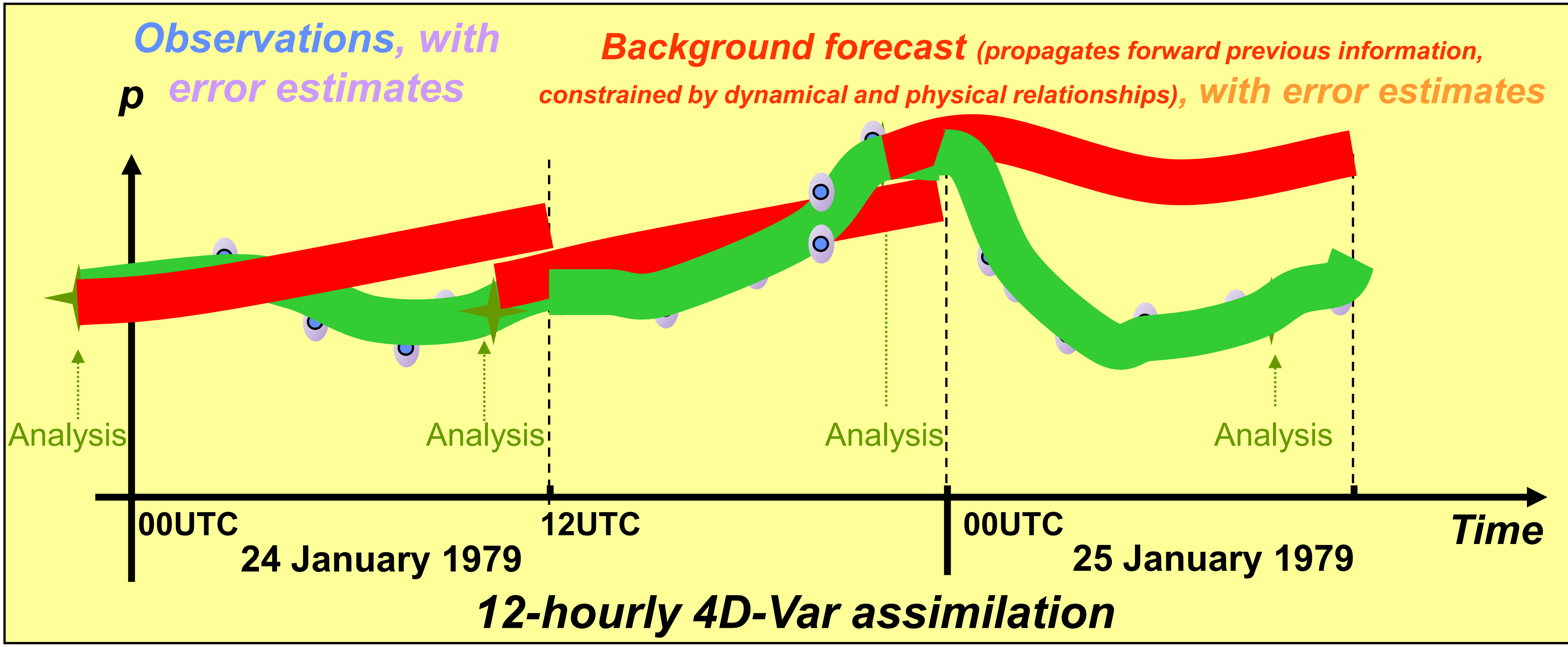
Plate 1. Global annual anomaly maps for July 2011.
Reference base period is 1989-2001. Datasets are as follows: lower stratospheric temperature at 70 hPa (ERA-Interim); lower tropospheric temperature at 925 hPa (ERA-Interim); screen level temperature (ERA-Interim); cloudiness (ERA-Interim); total column water vapor (ERA-Interim); specific humidity at 925 hPa (ERA-Interim); precipitation (ERA-Interim); soil water content level 1 (ERA-Interim); residual runoff (ERA-Interim); river discharge receiving oceans (??); mean sea level pressure (ERA-Interim); surface wind speed (ERA-Interim); ozone (ERA-Interim); FAPAR (??); biomass burning (??).

… Climate quality ?

Outline

- Introduction:
  TIME, our great concern
- Reanalysis products
- Reanalysis process
- Recent developments
- Conclusion

The ubiquitous data assimilation slide



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

background constraint — observation constraint

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

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

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

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

This produces the “most probable” atmospheric state *

* In a maximum-likelihood sense, which is equivalent to the minimum variance, provided that **background and observation errors are Gaussian**, **unbiased**, **uncorrelated with each other**; all error covariances are correctly specified; model errors are negligible within the 12-h analysis window

ECMWF

Special features of reanalysis: All have to do with TIME

Use a fixed DA scheme

- Think ahead and put in it all that is needed, observation-wise:
  - A **blacklist** that covers the entire reanalysis period
  - Observation operators, thinning, obs. errors etc…
- Should use a **B matrix that's fit for the job** (ideally, time-varying)
- Unlike NWP (future unknown), **should aim to benefit from future obs.**

Test the DA scheme with various amounts of observations

- With few observations (1), and with all observations (2)
- To get a feeling for how the products will be affected when going from (1) to (2)

Biases affect climate signal

- **Observation bias** correction algorithms should have some underlying physical foundations: radiation corrections for radiosondes, radiative transfer model biases, …
- **Model biases** should ideally be corrected, to prevent spurious changes in climate trends as the observation coverage changes

Keep that setup throughout the reanalysis

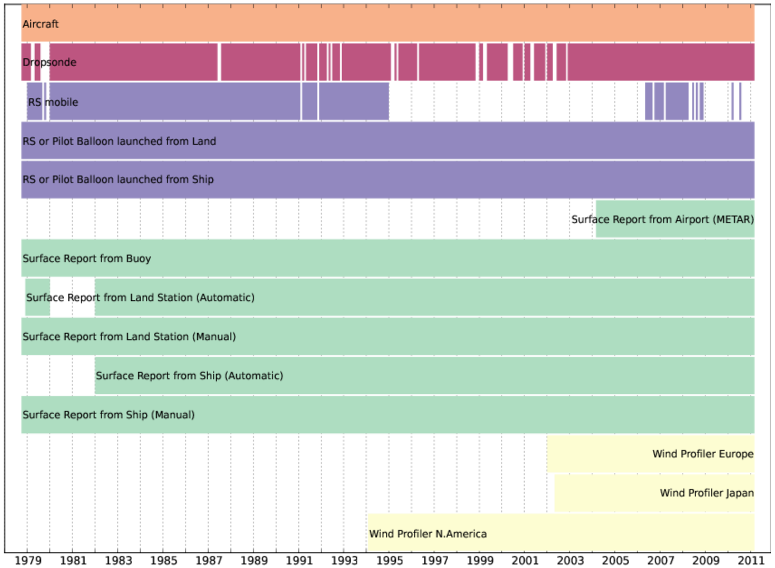
- Be extra careful during **run-time** etc… so as not to lose the **backward-compatibility** (important in order to be able to rerun!)
- Continuing a reanalysis into the present implies to **back-phase new developments** in order to be able to use new satellites

Don't assume someone else will do the monitoring

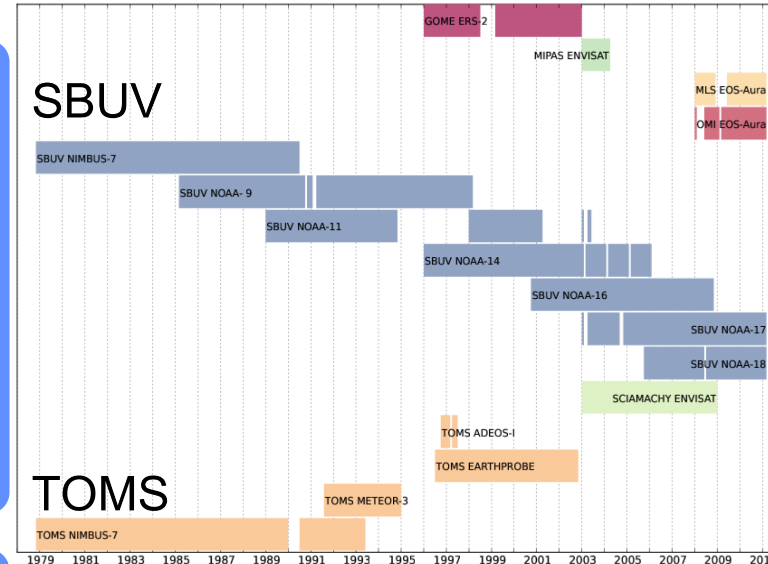
- Think ahead and develop tools to gather statistics as the run proceeds
- Develop solutions to display synthetic and detailed results as needed
- Hire as many eyes as you can
- Be ready to jump on the "stop" and "rerun" buttons

Time-lines of observing systems used in ERA-Interim

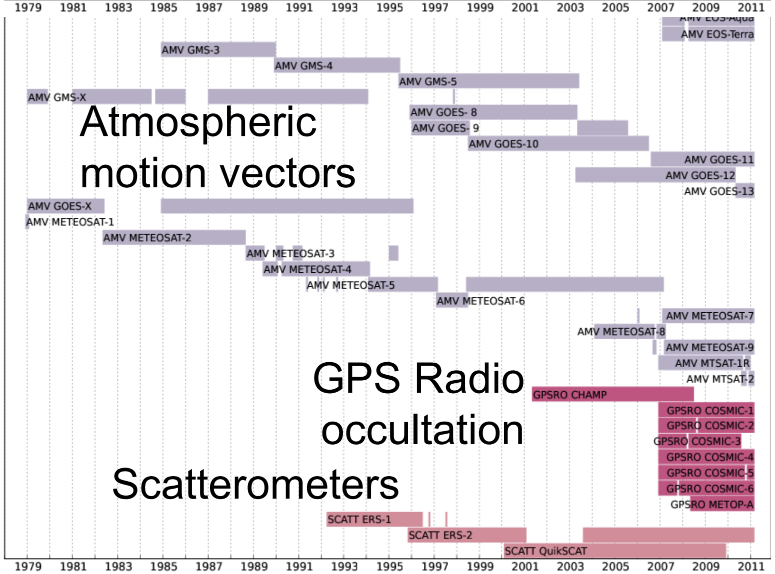
Conventional



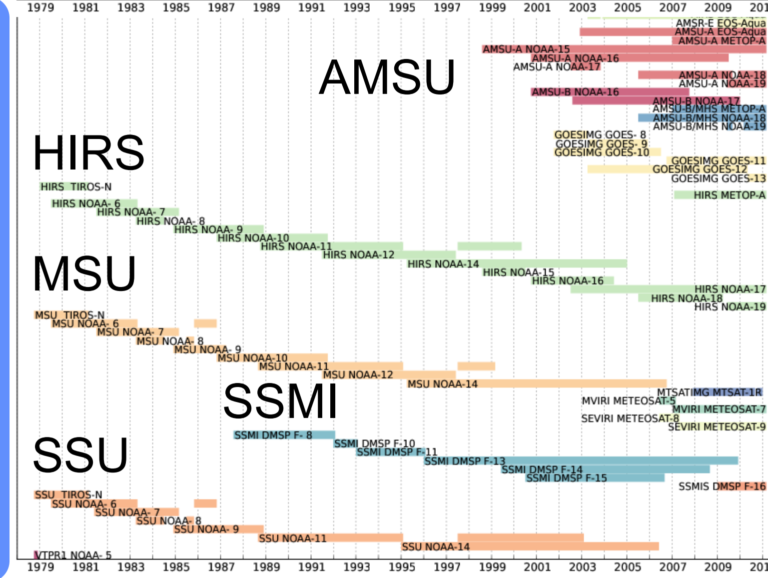
Satellite ozone



Other satellite data



Satellite radiances



Land Surface

- T2m, RH2m, snow depth

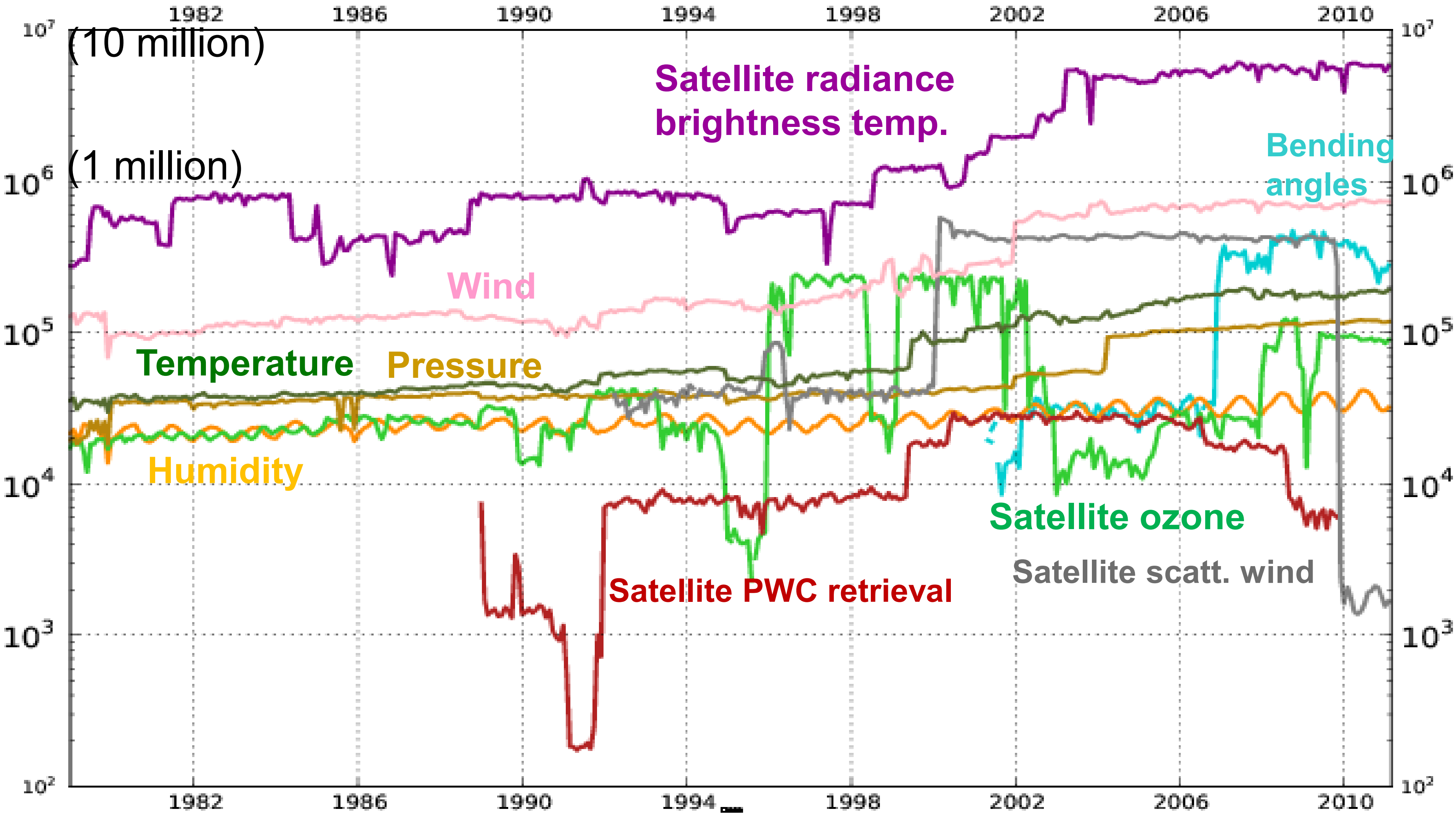
Altimeter Wave Height

- ERS-1, ERS-2, ENVISAT, JASON-1, JASON-2

ECMWF

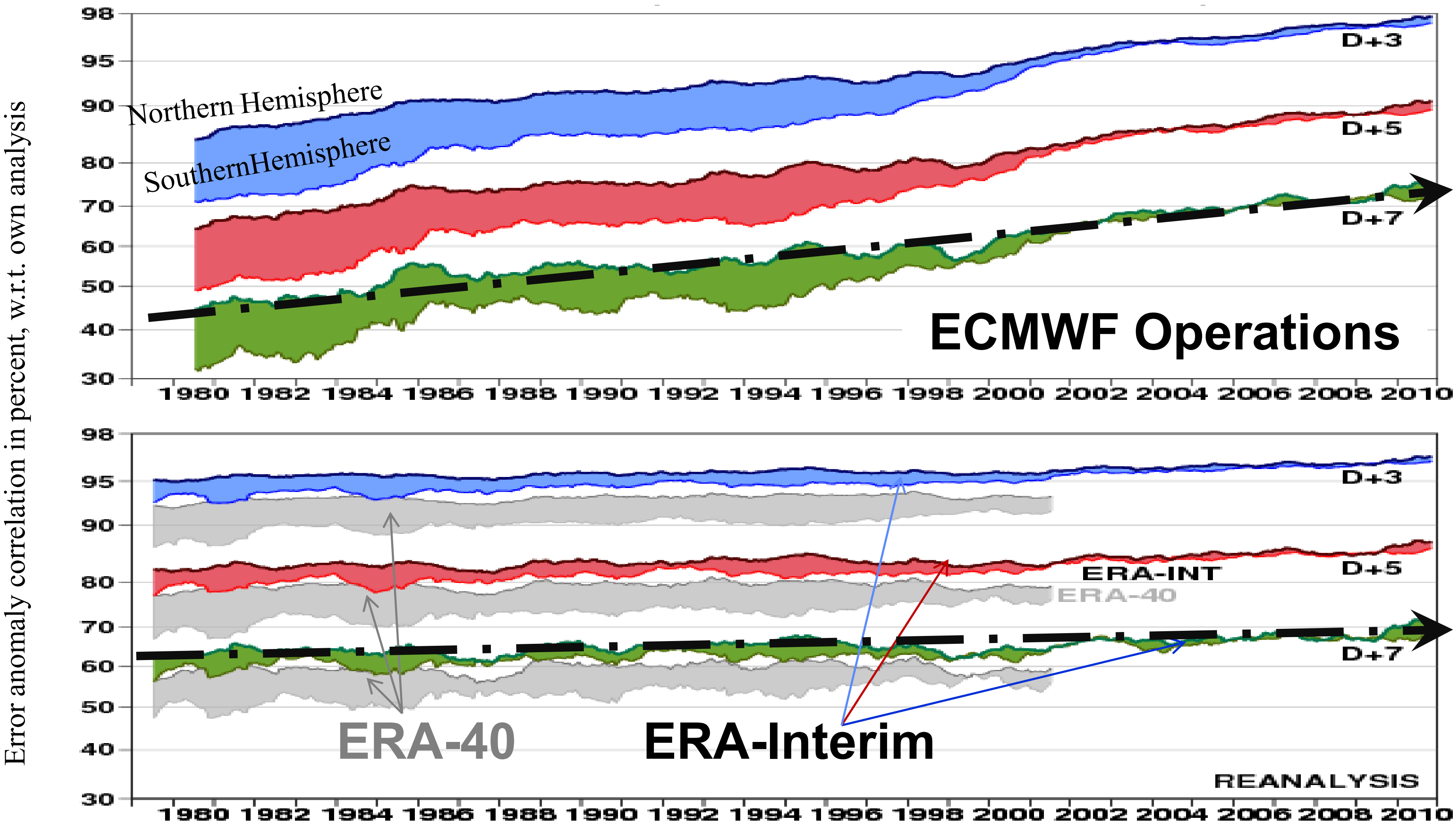
Observations diversity and evolution

Number of obs. assimilated per day in ERA-Interim 4D-Var



ECMWF

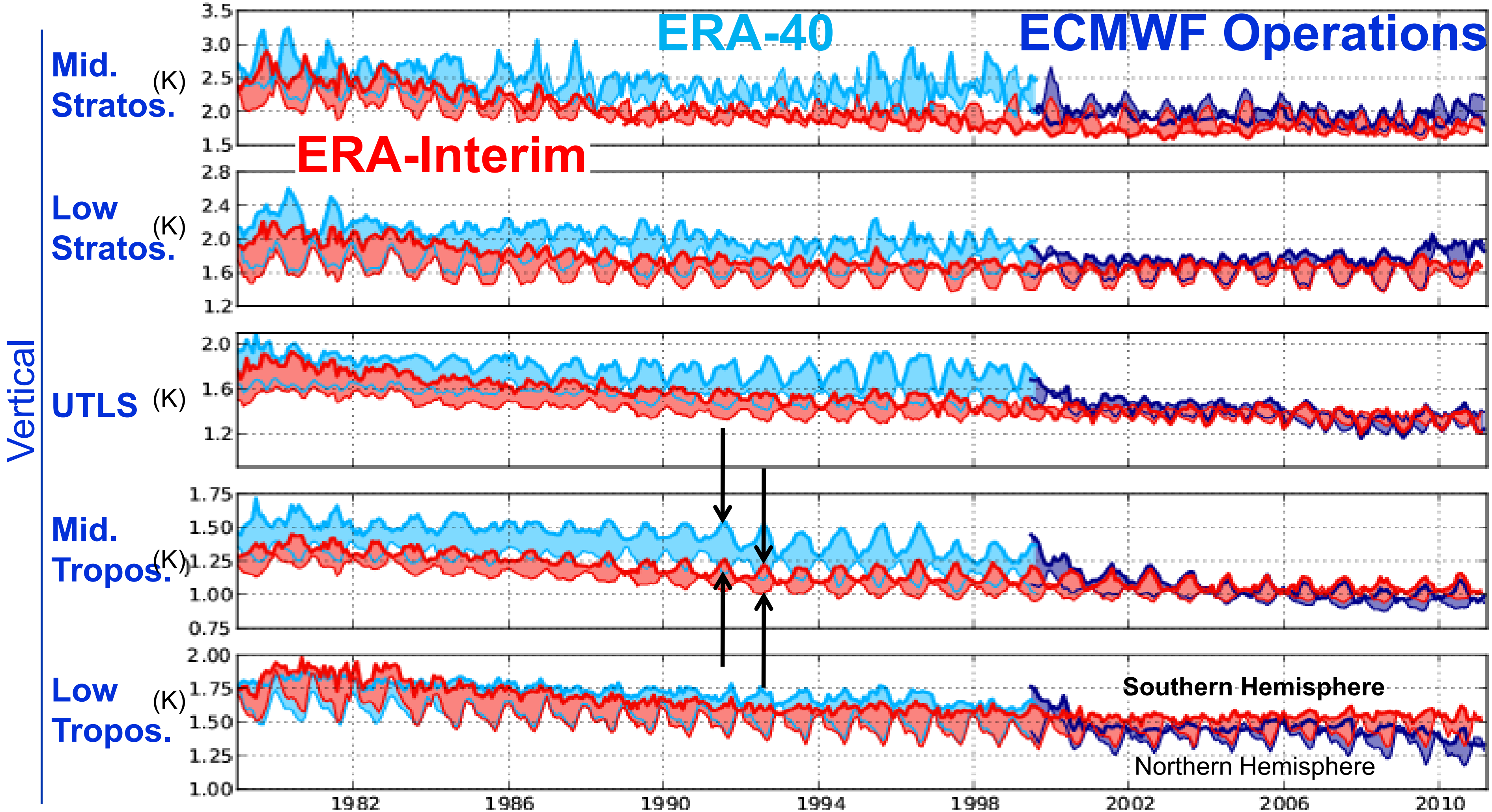
Re-forecasts
Anomaly correlation of Z 500



A. Simmons

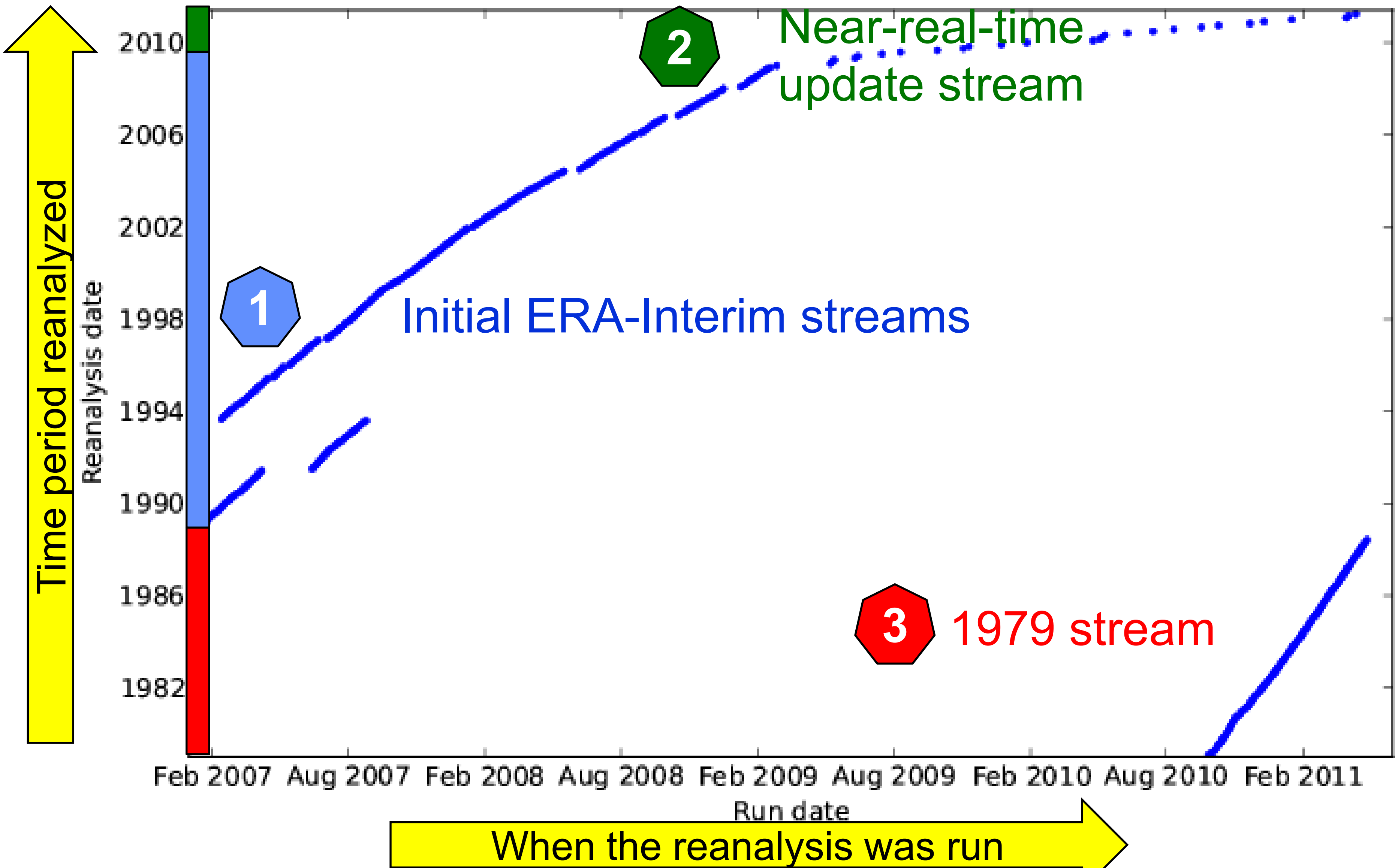
Spatio-temporal consistency of reanalysis quality

RMS of (radiosonde temperature obs.) minus (reanalyses short-term forecasts)



ECMWF

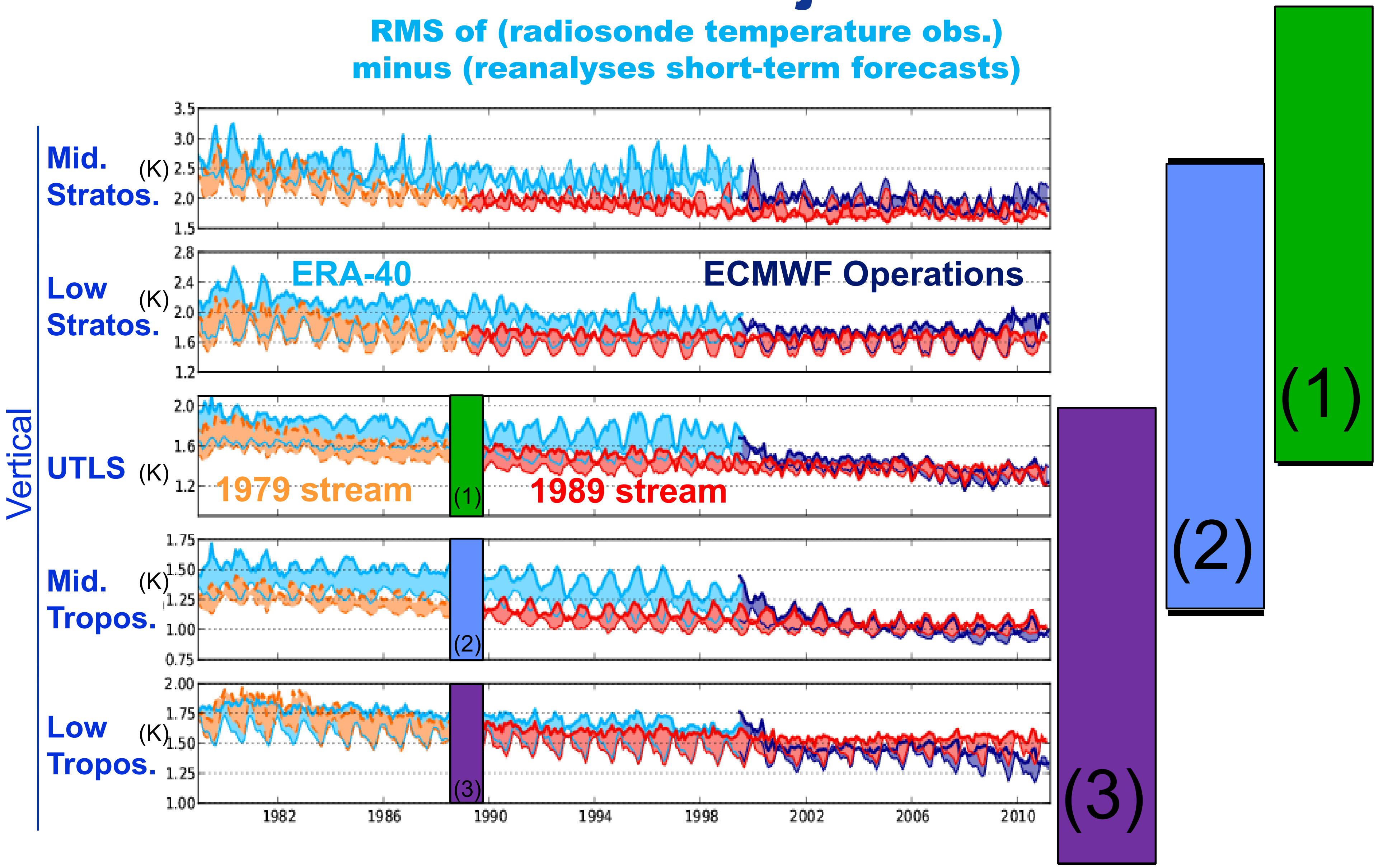
The two reanalysis time dimensions: streams



ECMWF

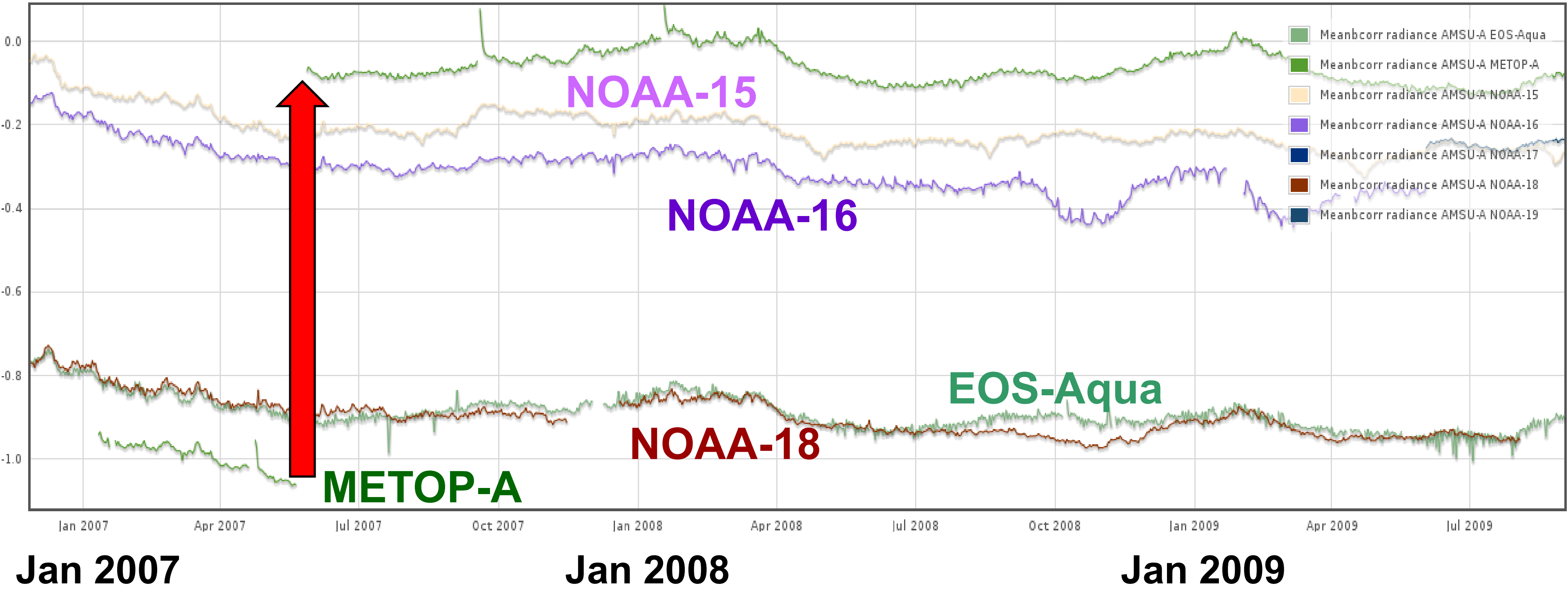
When two streams join in 1989

RMS of (radiosonde temperature obs.) minus (reanalyses short-term forecasts)



Adaptive estimation of satellite biases: Variational bias correction

Mean global bias correction of AMSU-A channel 9 (lower stratos.)

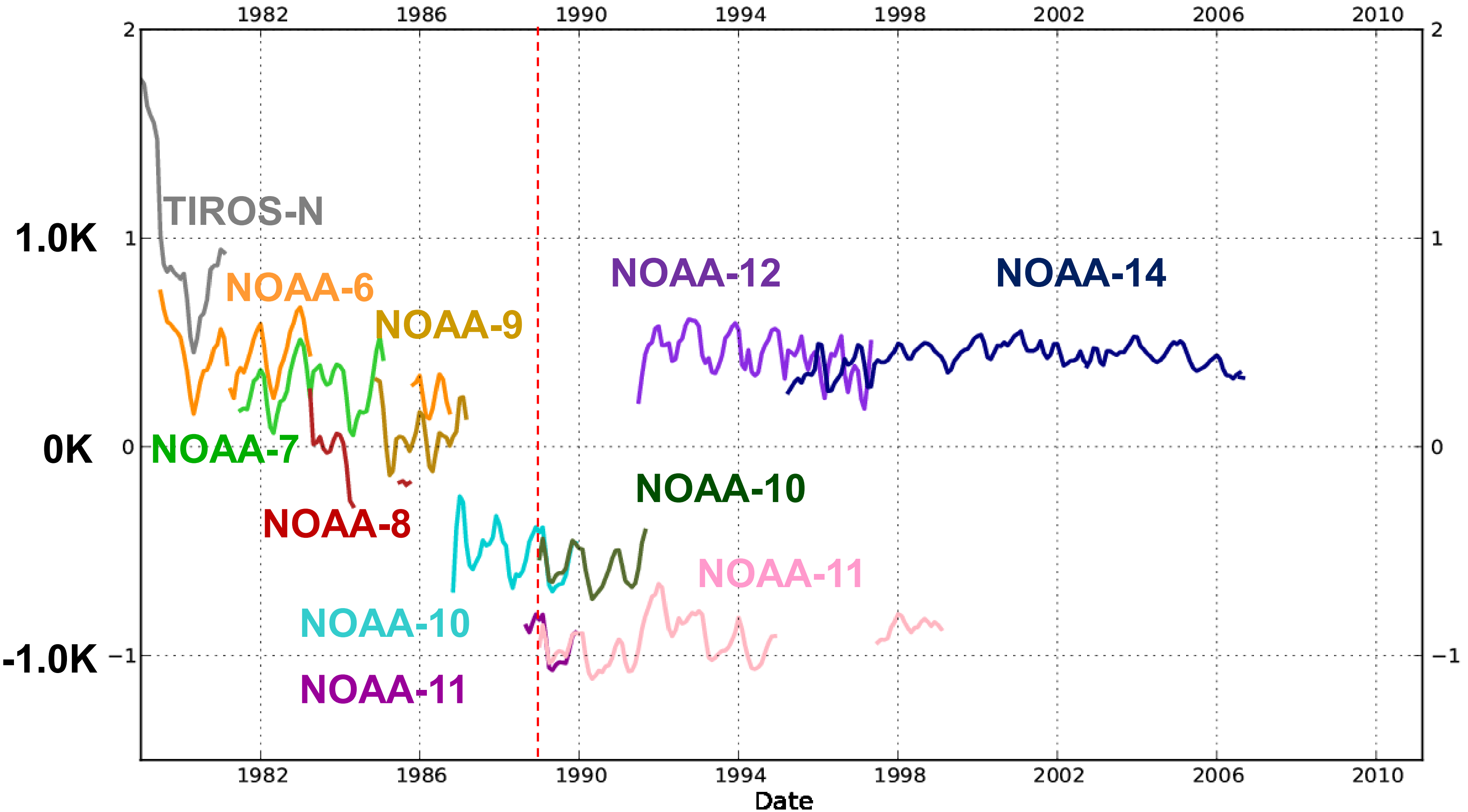


Context: METOP-A instruments were recalibrated a few months after launch

ECMWF

Adaptive estimation of satellite biases

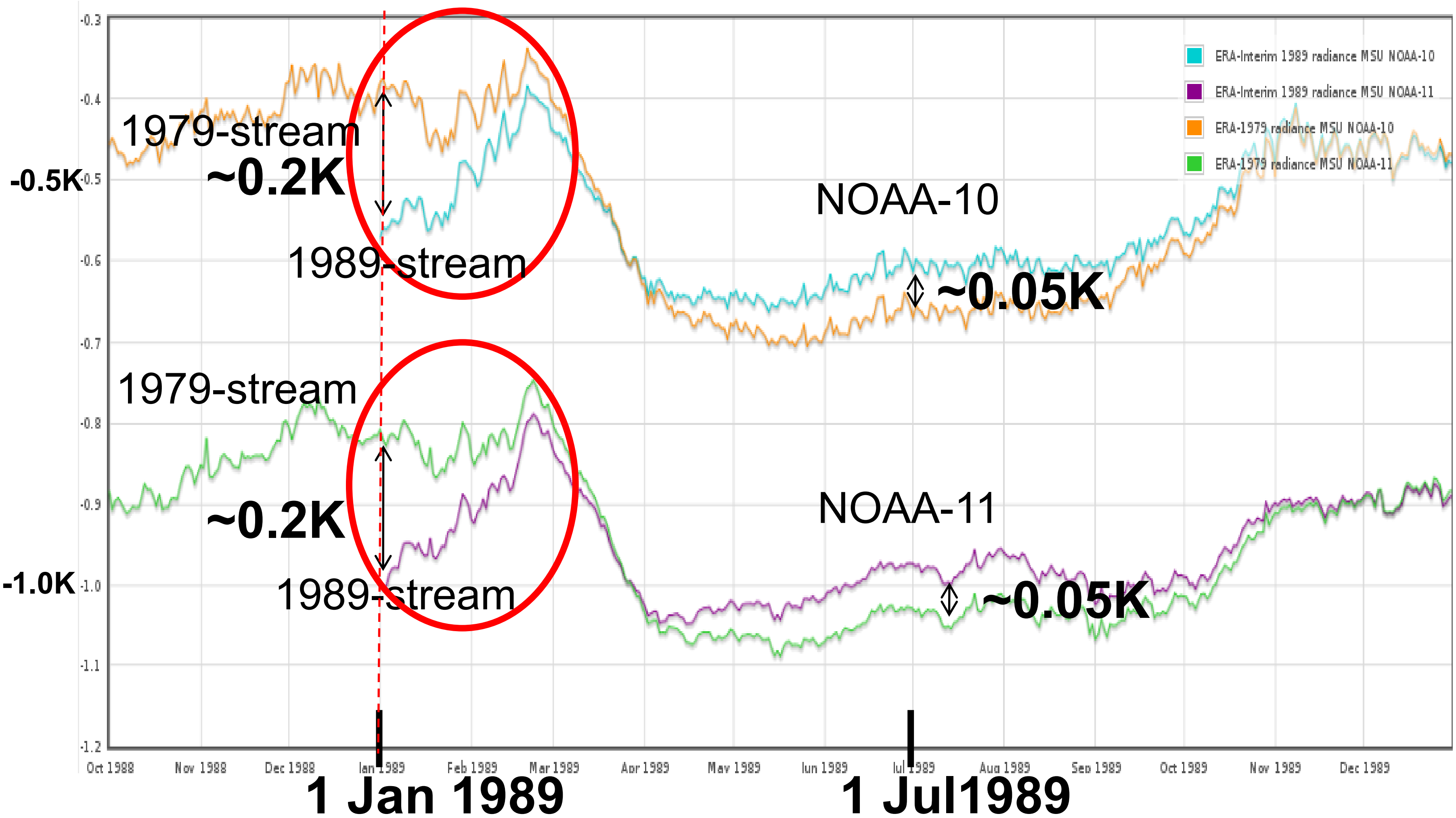
Mean global bias correction of MSU channel 4 (lower stratos.)



ECMWF

For improved time continuity in stratosphere between reanalysis streams

Mean global bias correction of MSU channel 4



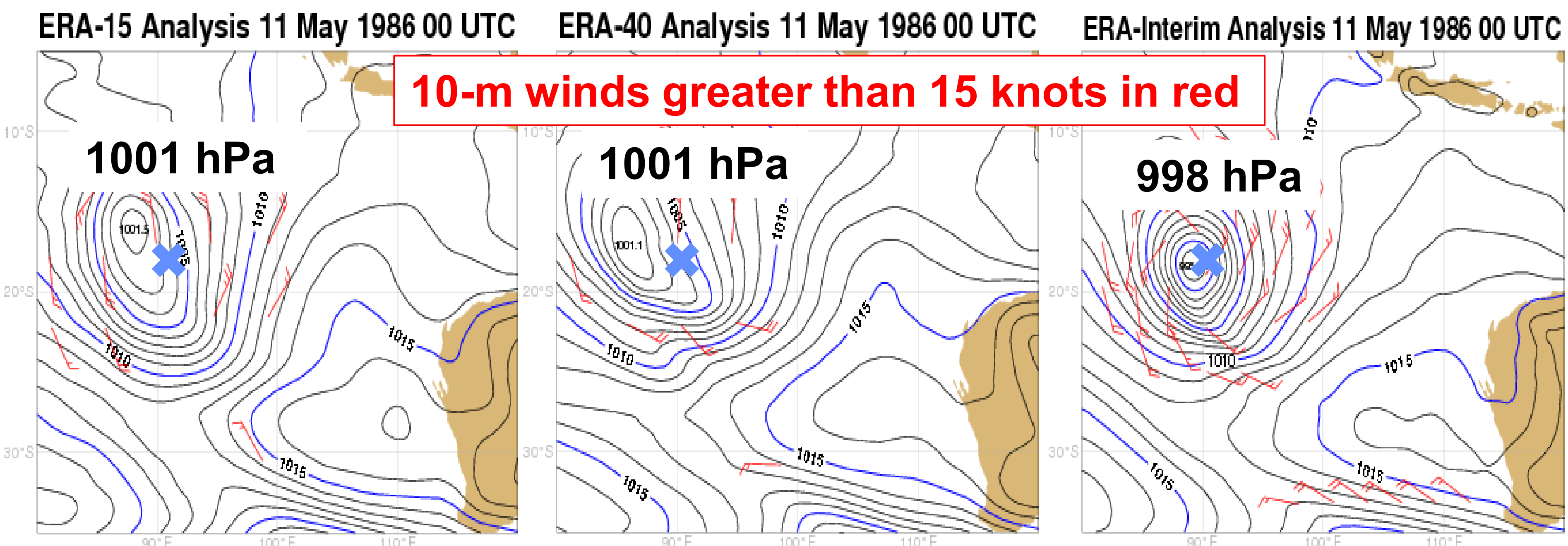
- **Solution:** longer overlap between streams, 3 months?

ECMWF

The other time dimension of reanalysis

Three successive ERAs reanalyzing tropical Cyclone Billy (11 May 1986)

ERA-15 (~1994)	ERA-40 (~2002)	ERA-Interim (~2006)
190 km hor. resol.	**125 km hor. resol.**	**80 km hor. resol.**
31 vertical levels	60 vertical levels	60 vertical levels
Assimilation every 6 hrs	Assimilation every 6 hrs	Assimilation every 12 hrs
Optimal interpolation	3D-Var	4D-Var 30 min. time-slots

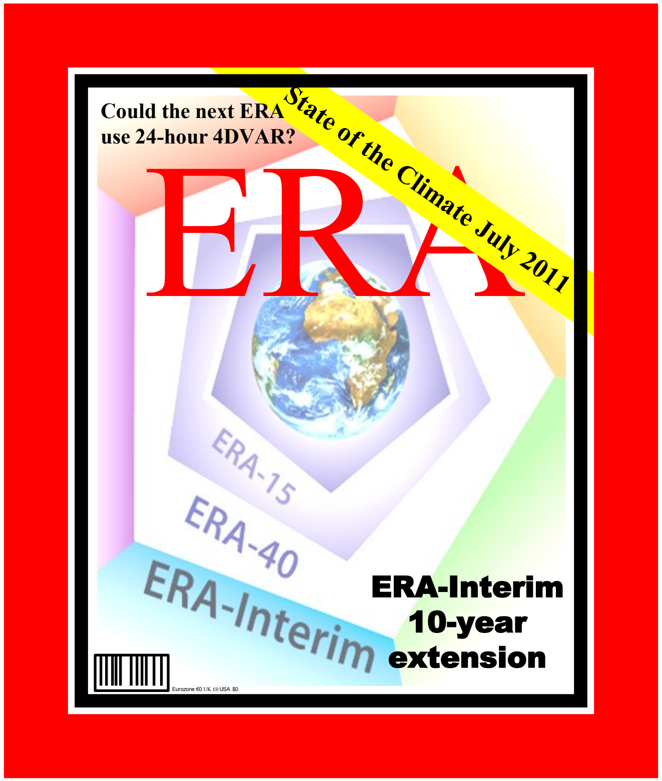


Estimated central pressure (Australian Bureau of Meteorology): 954 hPa at blue cross

C. Kingston (Bureau of Meteorology), 1986: The Australian tropical cyclone season 1985-86, *Austral. Met. Mag.* **34**, 103-115

ECMWF

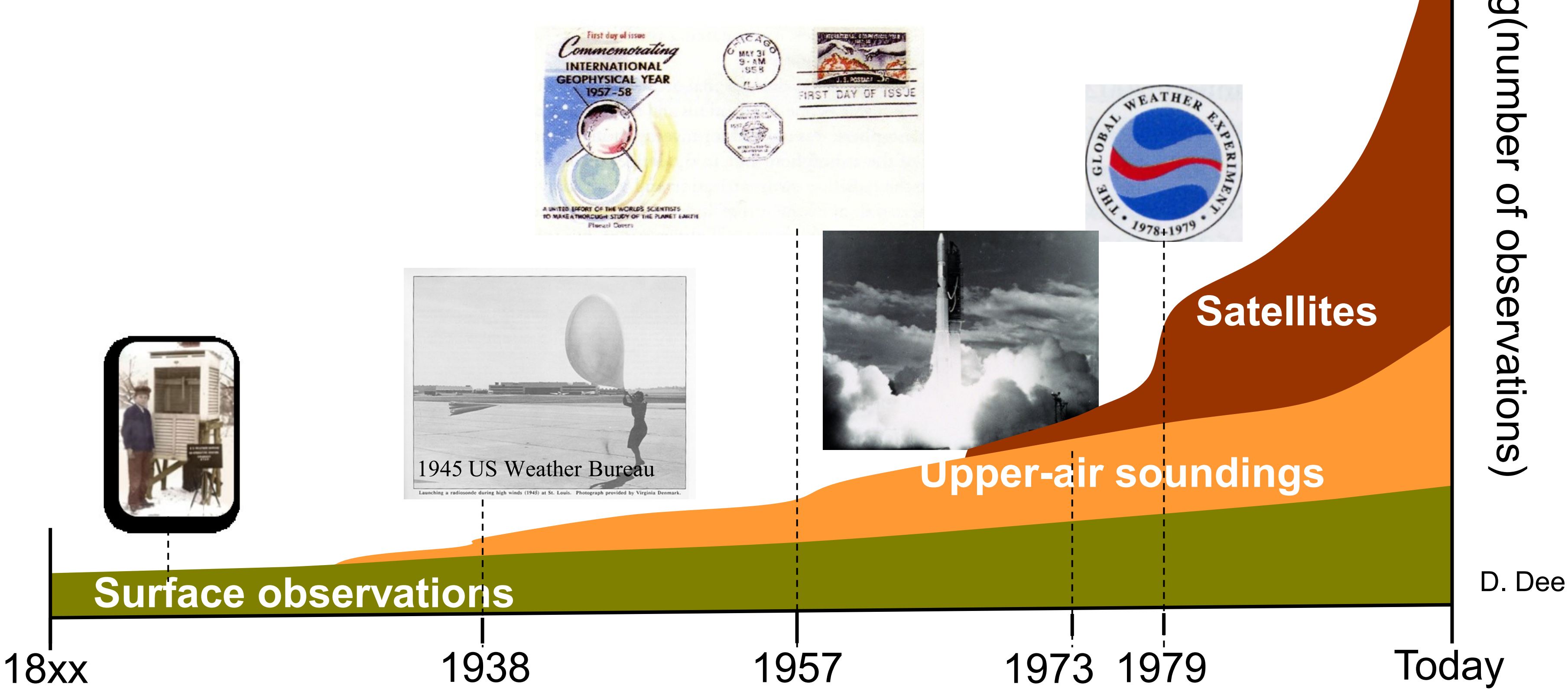
Outline



- Introduction:
  TIME, our great concern
- Reanalysis products
- Reanalysis process
- Recent developments
- Conclusion

ECMWF

Evolution of the observing system



Manual stations, limited data exchange

First radiosonde networks, systematic soundings

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

First operational satellite soundings (NOAA-2)

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

More satellites, aircraft, buoys, ocean drifters. Fewer radio-sondes (but reach higher). Better knowledge of instrument errors. More obs. per hour per sensor.

Developments in reanalysis

- **Target: 20th Century global reanalysis capability in Europe**
  - Building on continuous improvements in models
    - European modelling community, via ECMWF IFS
  - Assimilating reprocessed, recovered observations
    - Satellite agencies (EUMETSAT…), national met services…
  - To be continued in near-real-time to serve GMES Climate Services
  - To be repeated at regular intervals as new developments warrant much improved reanalyses

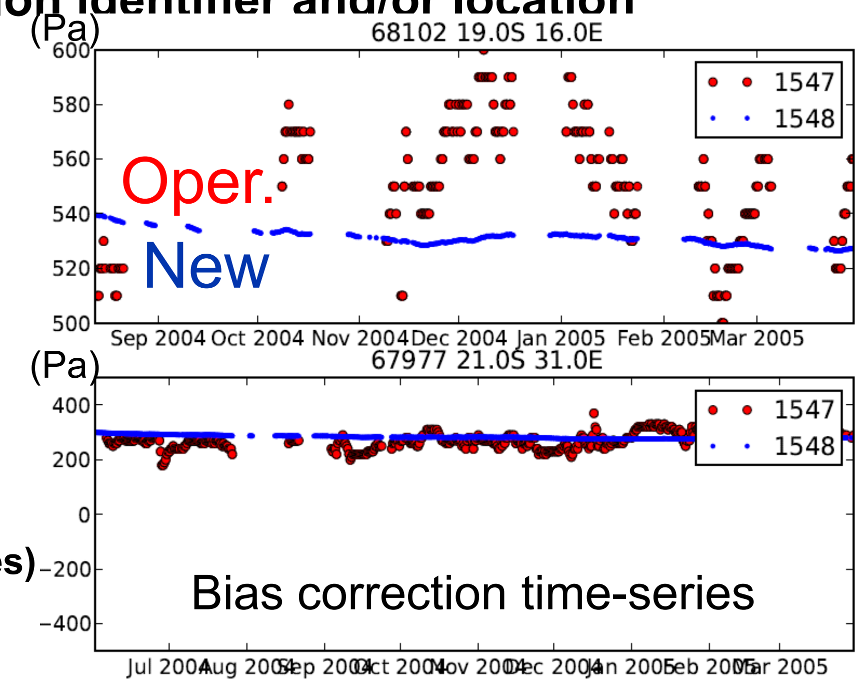
- **Capacity building: EU FP7 ERA-CLIM**
  - Surface obs. assimilation: topic expanded in following slides
  - Upper-air obs. assimilation
  - Satellite obs. Assimilation

ECMWF

Towards a surface-pressure-only reanalysis

- **Already done in the US: 20CR (Compo et al., QJRMS 2011)**
- **Our production window: 1 year to reanalyze 100 years**
  - **Production speed target is 100 days/day**
  - **This implies a low resolution (horizontal: T159 ~ 125 km)**
- ***New* bias correction scheme for surface station observations**
  - **Variational bias correction, by station identifier and/or location**
  - **Runs *within* the analysis; observations can be prepared ahead of time without waiting for previous analysis**
  - **Reproduces the behaviour of the old scheme for sfc. P., except**
    - **All stations get corrected**

      (previously only those with large biases)
    - **Greater stiffness**

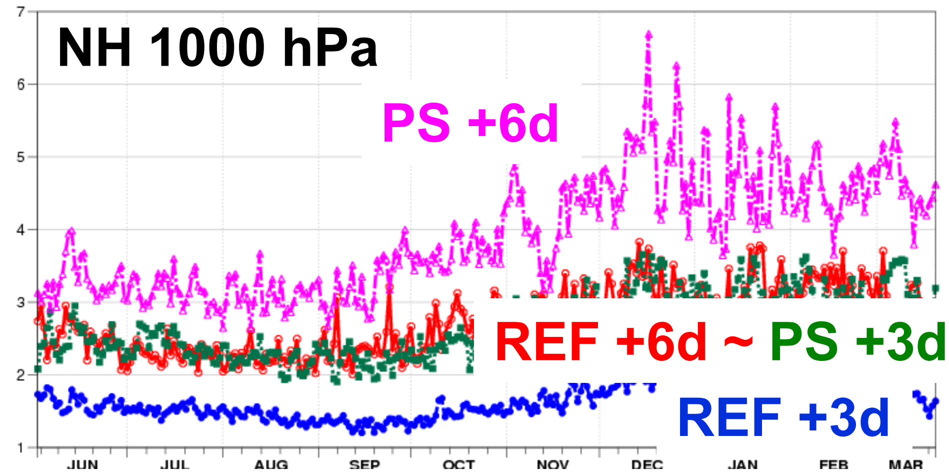


Surface-pressure-only T159 (an T159/T95)

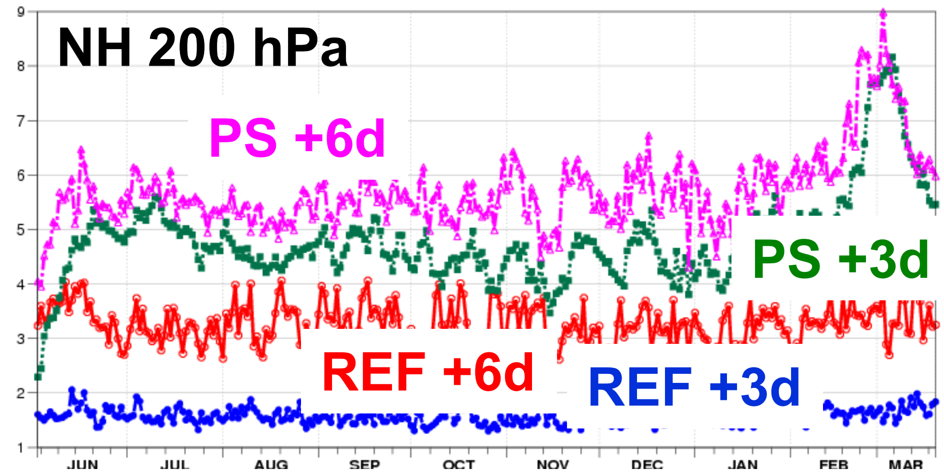
vs. All obs T255 (an T159/T95)

Forecast RMSE Temperature +3 days and +6 days

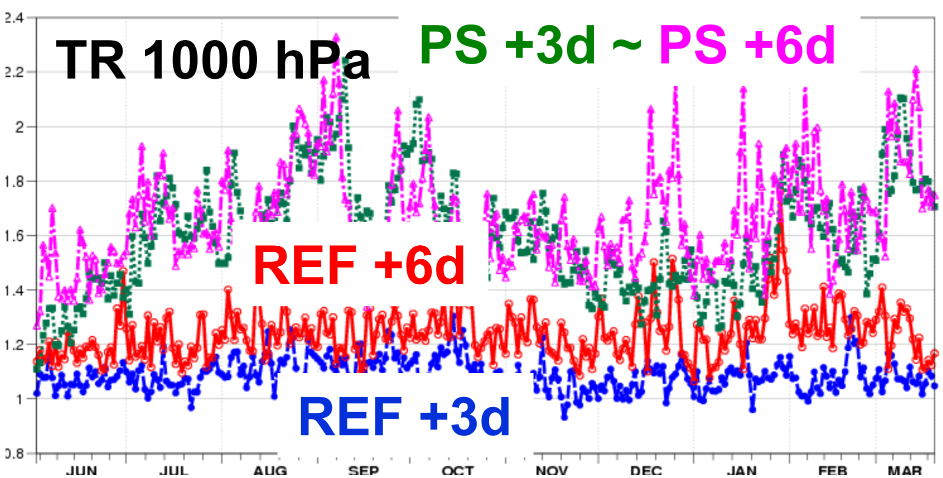
(K)



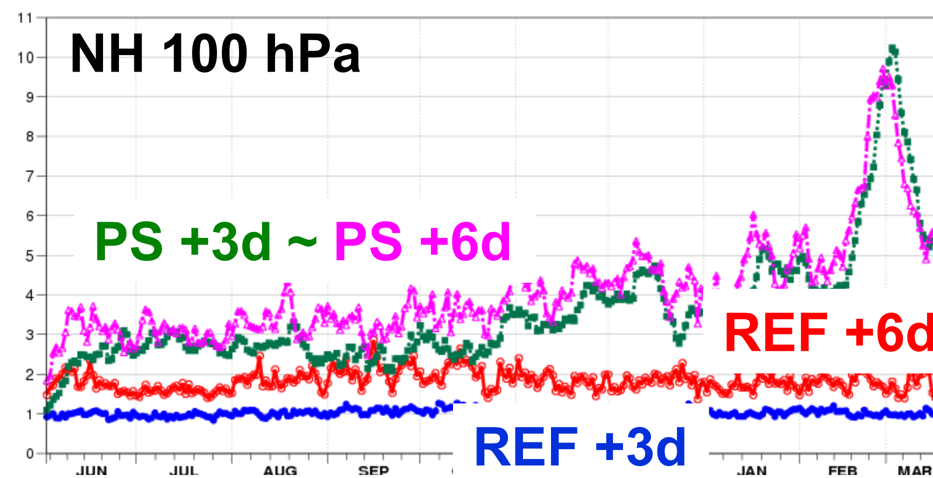
For NH (and SH), there is some forecast skill up for 1000 (and 500) hPa



For NH (and SH), model variability above 200 hPa



For TR, model variability (no forecast skill)

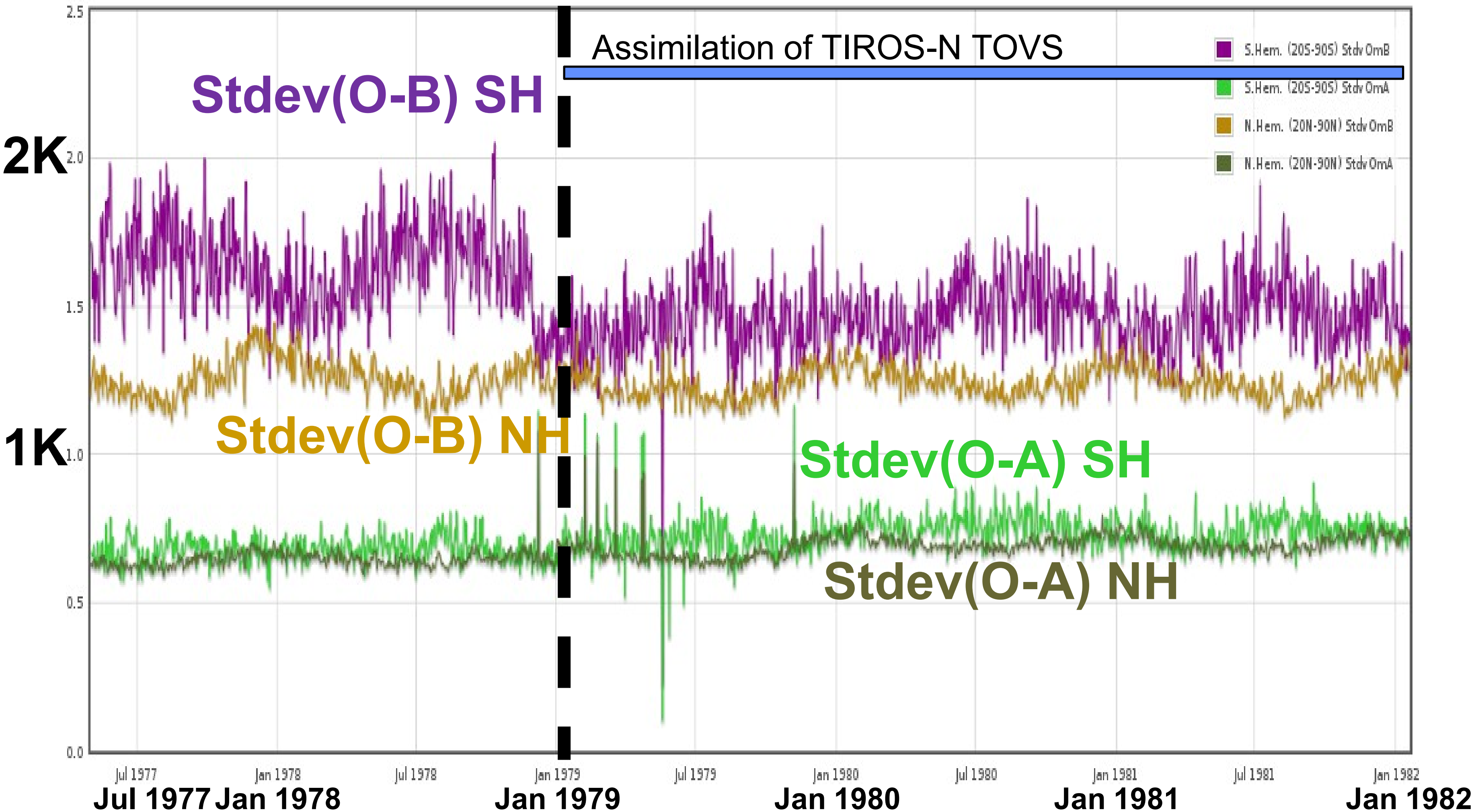


Problem with sudden strat. warming events

ECMWF

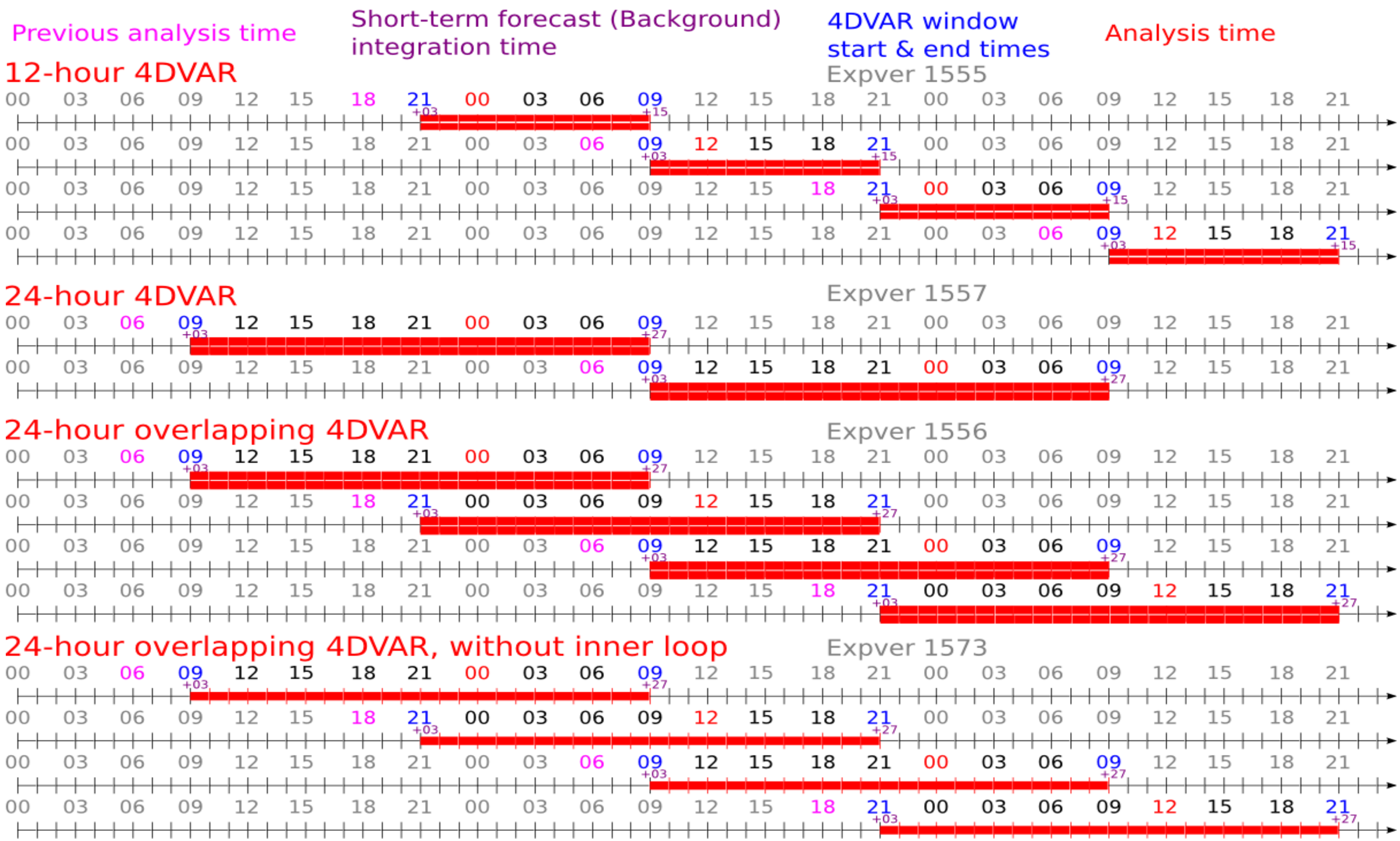
Ideally, we should use a time-varying background error covariance matrix

ERA-40 fit to assimilated radiosonde temperatures at 500hPa



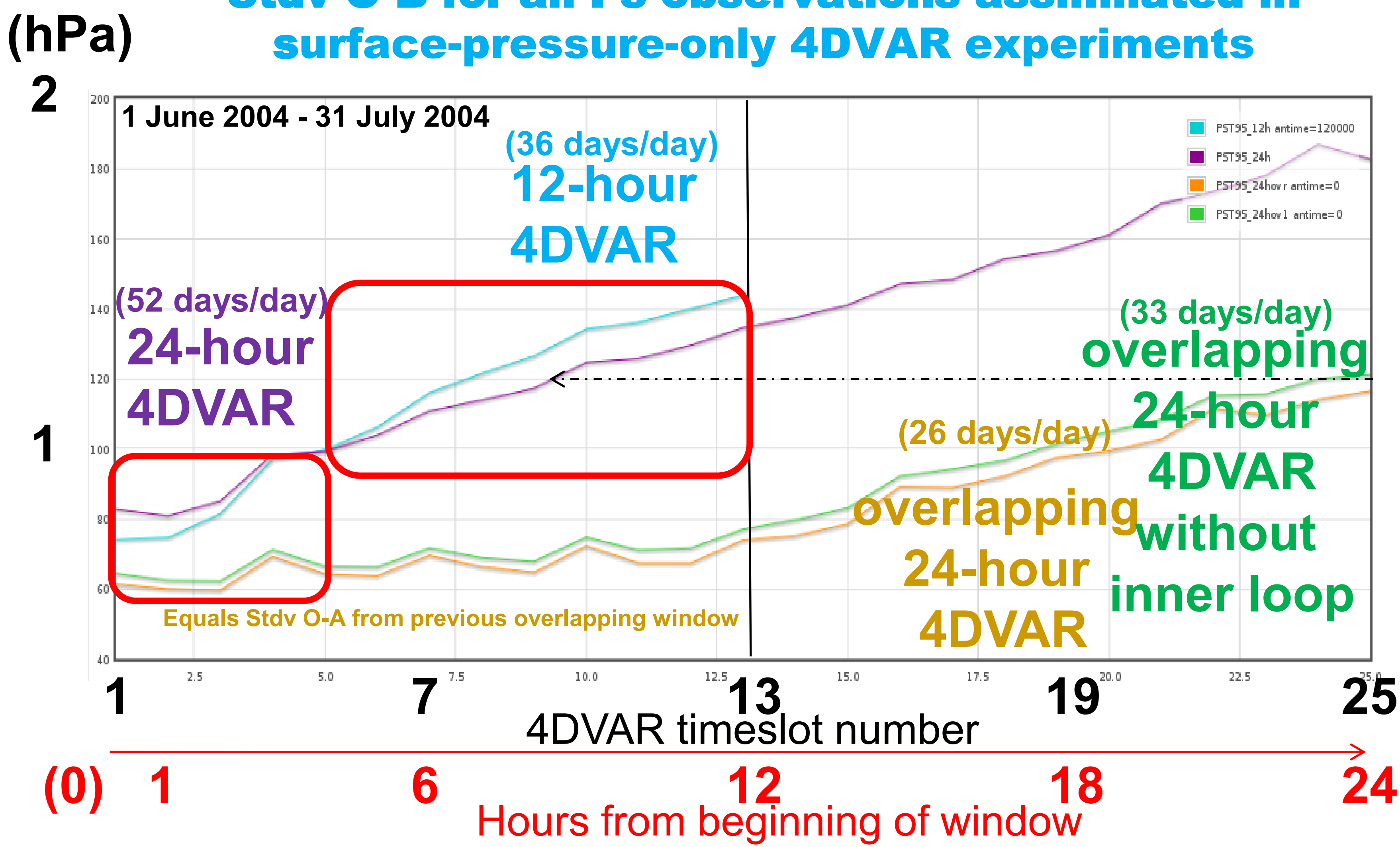
ECMWF

Towards 24-hour 4DVAR: Code developments from Yannick Trémolet

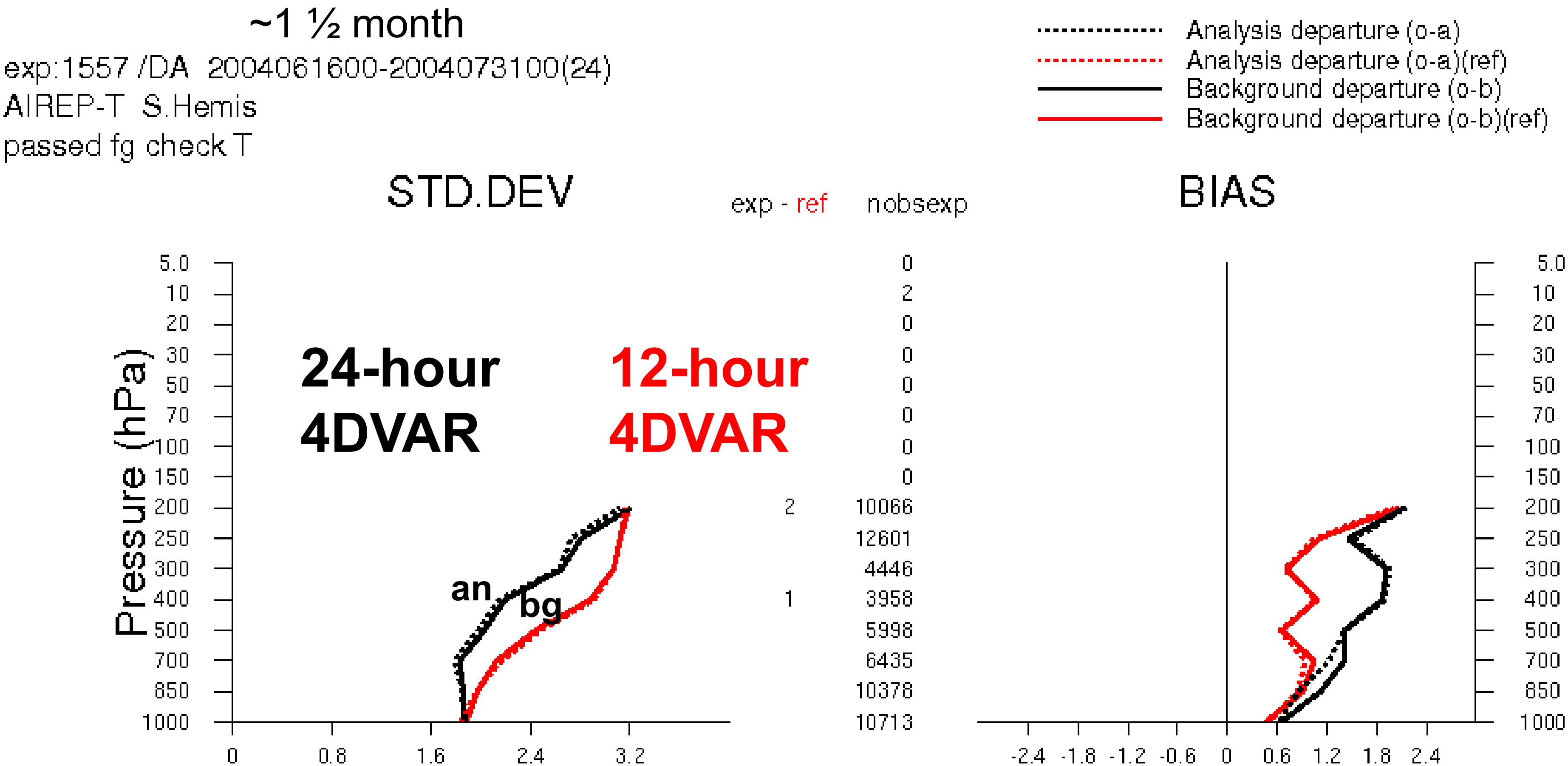


12-hour *vs* 24-hour *vs* overlapping 24-hour *vs* overlapping 24-hour without inner loop

Stdv O-B for all Ps observations assimilated in surface-pressure-only 4DVAR experiments



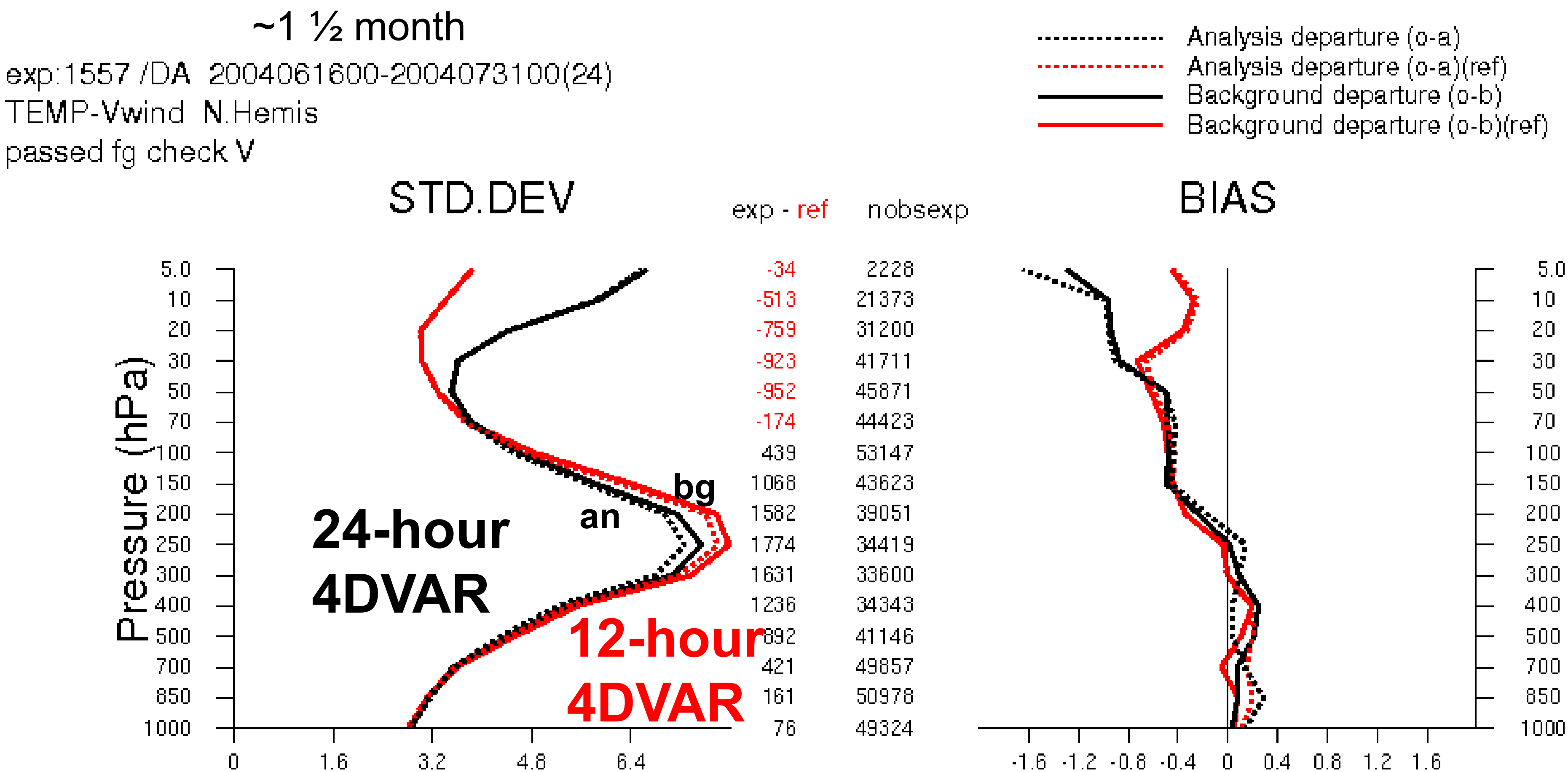
Comparison to non-assimilated observations
Aircraft temperatures in SH



1. Some marginal redundancy of information between Ps and T in troposphere
2. 24-hour 4DVAR seems to better fit variability of independent temperatures, but shows bigger mean errors (model bias?)

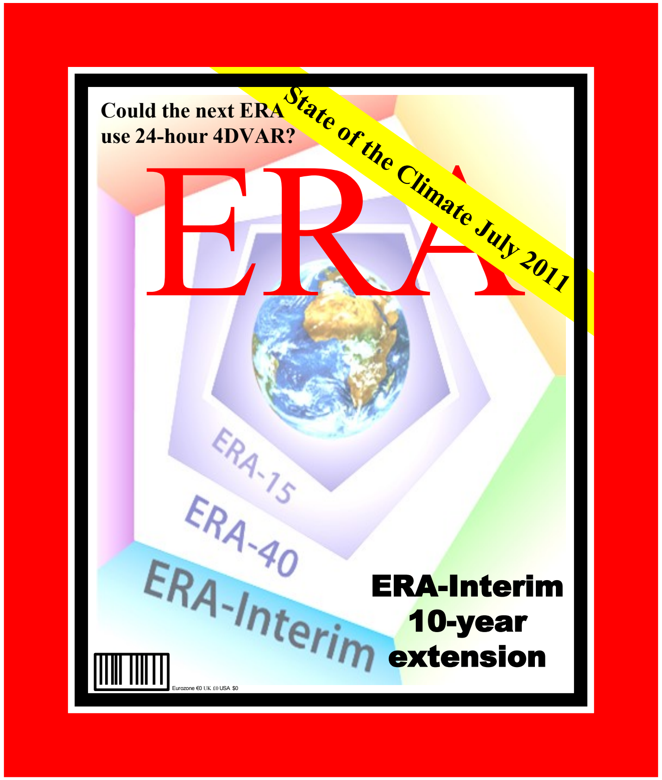
ECMWF

Comparison to non-assimilated observations
Radiosonde meridional wind (v) in NH



1. Some redundancy of information between Ps and wind up to tropopause
2. 24-hour 4DVAR seems to better fit independent winds in troposphere, but degradation in stratosphere (model bias? incorrect B matrix?)

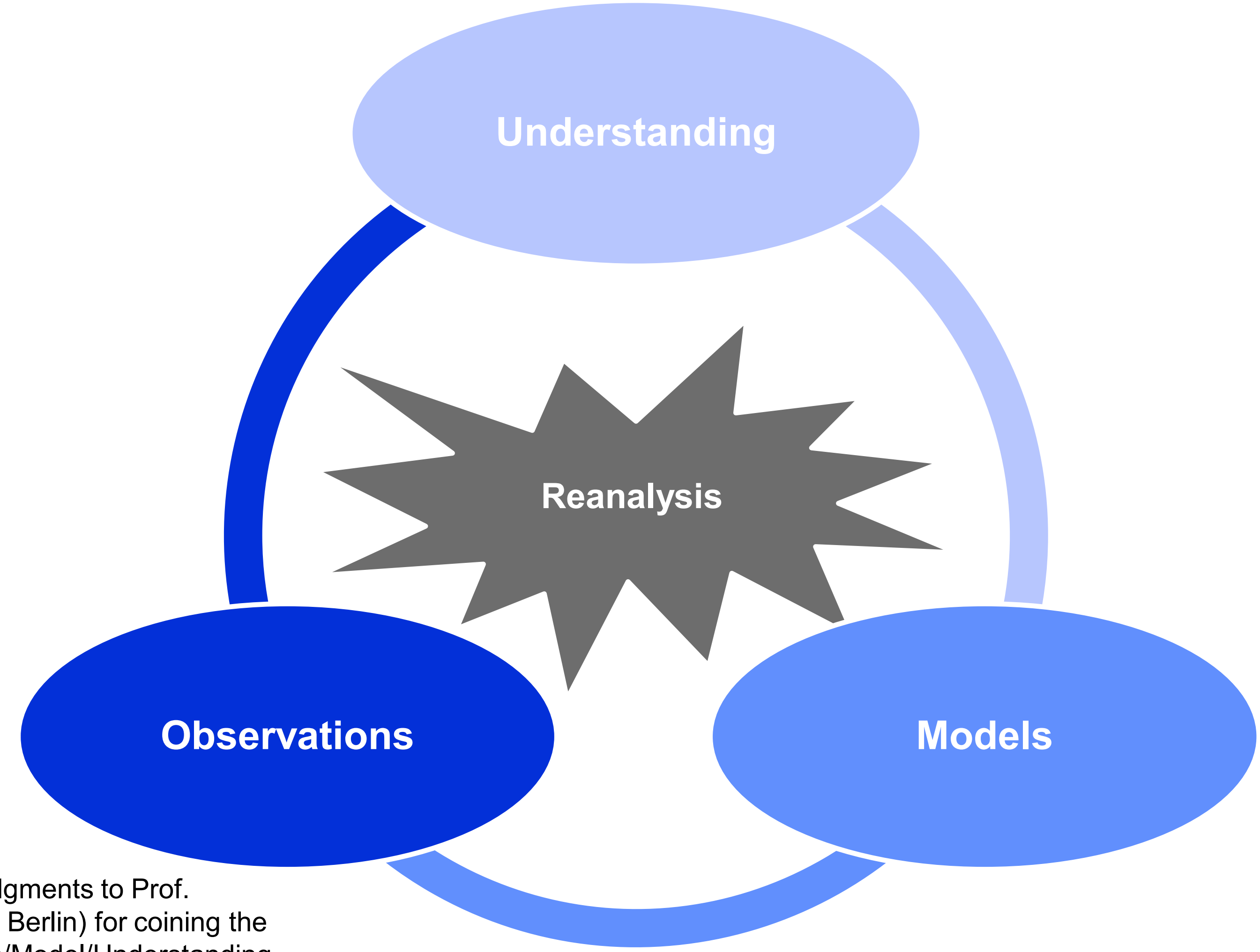
Outline



- Introduction:
  TIME, our great concern
- Reanalysis products
- Reanalysis process
- Recent developments
- Conclusion

ECMWF

Conclusion
Observations, Models, Understanding



Acknowledgments to Prof.
Ulbrich (U. Berlin) for coining the
triplet Data/Model/Understanding

ECMWF

Back to our original goal: TIME consistency
Key ingredients to bridge the gap between "NWP reanalysis" and "Climate Reanalysis"

- **Background and observation error covariances**
  - Need to be adaptive (space or type, and time-varying); solutions:
  - Pre-determined, or
  - Ensemble Data Assimilation, Desroziers' method…
- **Model bias**
  - Need some form of correction; solutions:
  - Pre-determined model bias correction, or
  - Weak constraint 4DVAR
- **Model improvement**
  - Include other elements of the Earth system: ocean coupling…
- **Observations**
  - Need reprocessing whenever possible
  - Need more of them!

ADDITIONAL MATERIAL