Operational Data Assimilation at ECMWF

Yannick Trémolet
(currently on leave at NASA/GMAO)

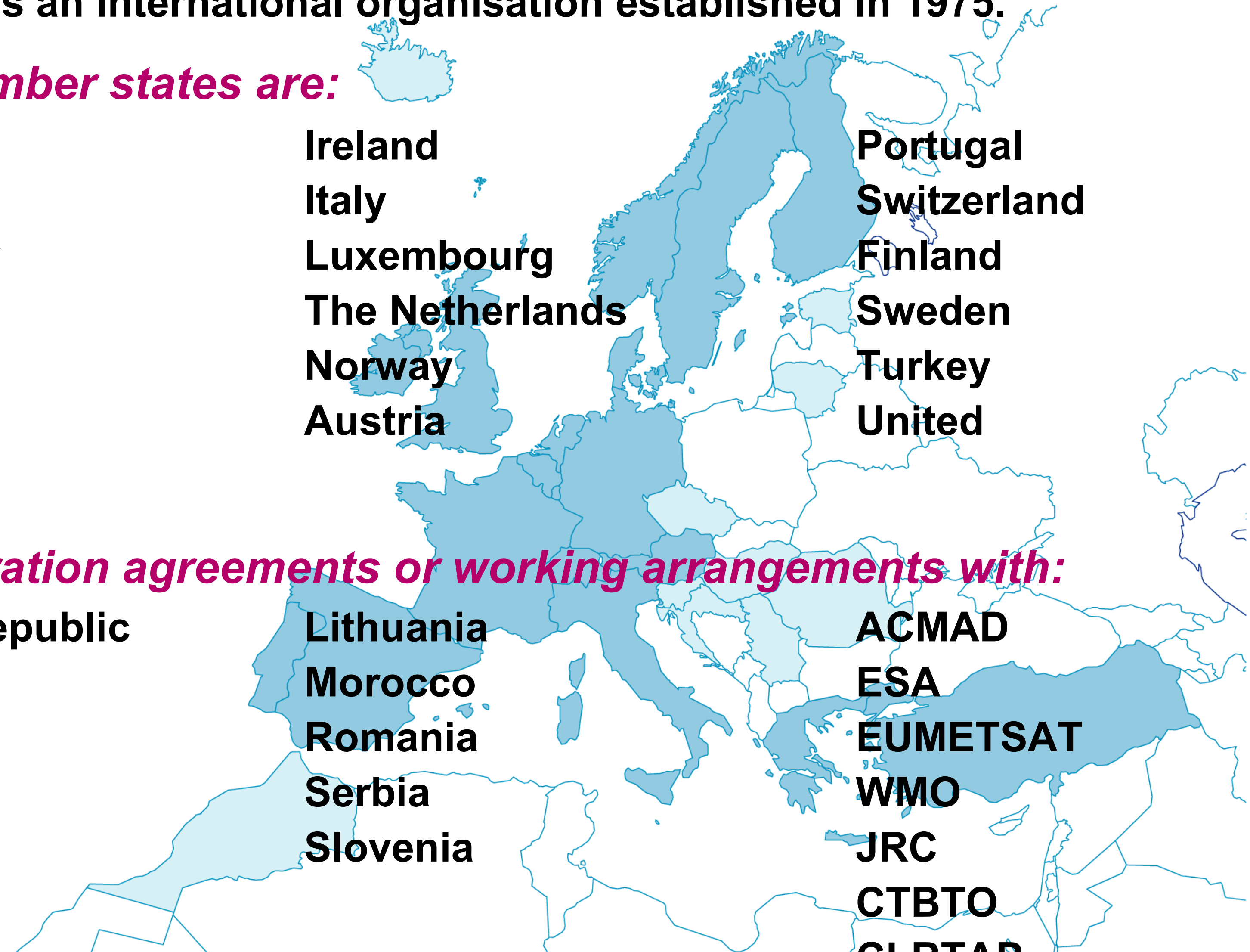
With help from many colleagues.

Outline

- Introduction: ECMWF
- Operational data assimilation
  - Model
  - Observations
  - Computational aspects
- The future of 4D-Var
  - Flow dependent background error statistics
  - Weak constraint 4D-Var
- Conclusions

Supporting States and Co-operation

ECMWF is an international organisation established in 1975.



The member states are:

Belgium
Ireland
Portugal
Denmark
Italy
Switzerland
Germany
Luxembourg
Finland
Spain
The Netherlands
Sweden
France
Norway
Turkey
Greece
Austria
United Kingdom

Co-operation agreements or working arrangements with:

Czech Republic
Croatia
Estonia
Hungary
Iceland
Lithuania
Morocco
Romania
Serbia
Slovenia
ACMAD
ESA
EUMETSAT
WMO
JRC
CTBTO
CLRTAP

The Centre's principal objectives

- The development of numerical methods for medium-range weather forecasting;
- The preparation, on a regular basis, of medium-range weather forecasts for distribution to the meteorological services of the Member States;
- Scientific and technical research directed at the improvement of these forecasts;
- Collection and storage of appropriate meteorological data.

ECMWF Forecast Products

Atmosphere global forecasts

- Forecast to ten days from 00 and 12 UTC at 25 km resolution and 91 levels
- 50 ensemble forecasts to fifteen days from 00 and 12 UTC at 50 km resolution

Ocean wave forecasts

- Global forecast to ten days from 00 and 12 UTC at 50 km resolution
- European waters forecast to five days from 00 and 12 UTC at 25 km resolution

Monthly forecasts: Atmosphere-ocean coupled model

- Global forecasts to one month:

  atmosphere: 1.125° resolution, 62 levels

  ocean: horizontally-varying resolution (⅓° to 1°), 9 levels

Seasonal forecasts: Atmosphere-ocean coupled model

- Global forecasts to six months:

  atmosphere: 1.8° resolution, 40 levels

  ocean: horizontally-varying resolution (⅓° to 1°), 9 levels

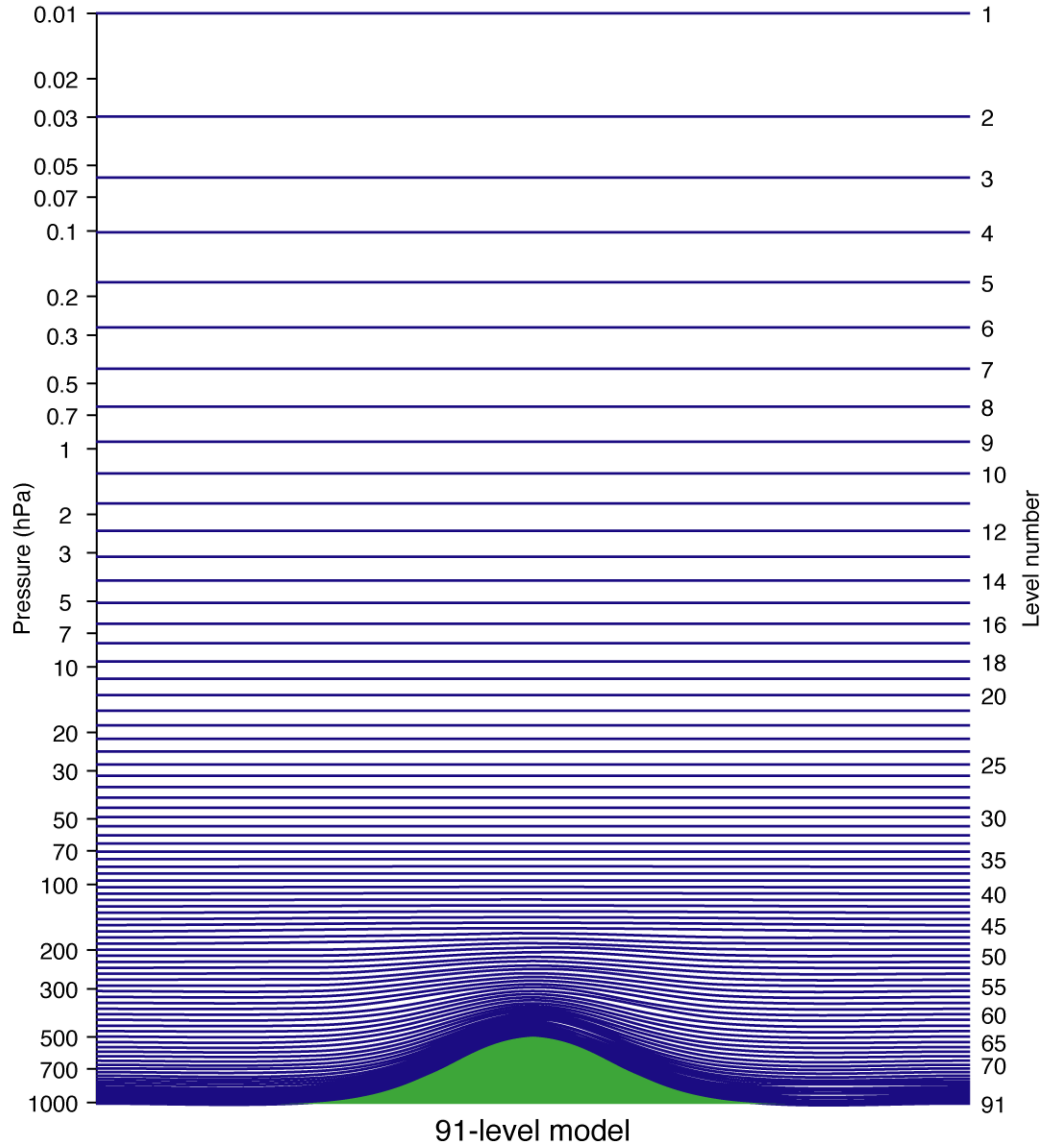
Outline

- Introduction: ECMWF
- Operational data assimilation
  - ➔ Model
  - ➔ Observations
  - ➔ Computational aspects
- The future of 4D-Var
  - ➔ Flow dependent background error statistics
  - ➔ Weak constraint 4D-Var
- Conclusions

The medium-range model – numerical scheme

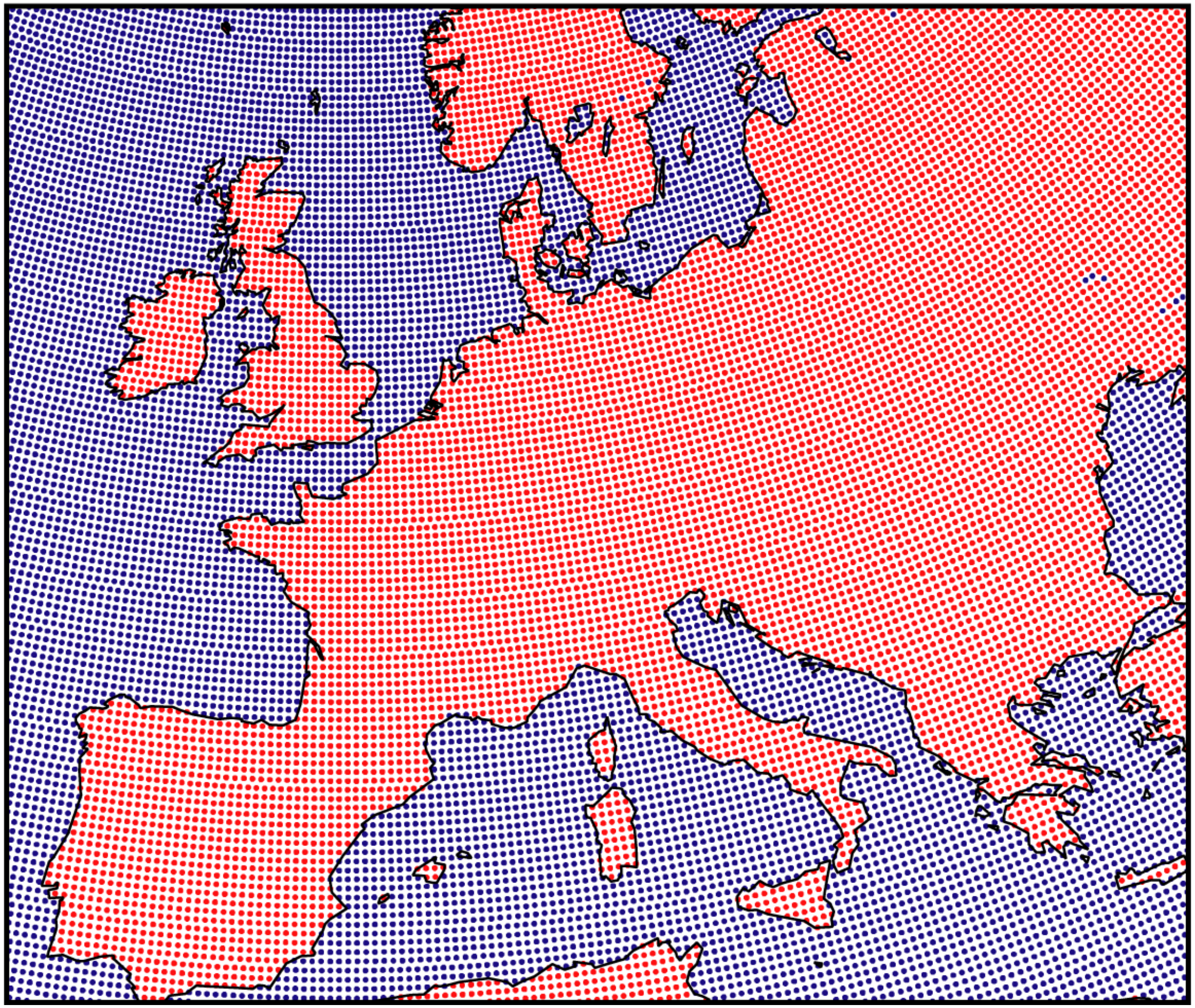
- A spectral T_L799L91 (triangular truncation, 799 waves around a great circle on the globe; 91 levels between the earth's surface and 80 km); semi-Lagrangian formulation. The grid points in this model are separated by about 25 km in the horizontal around the globe. Models of lower resolution are used for EPS and seasonal forecasts.
- Variables at each grid point (re-calculated at each time-step): wind (including vertical velocity), temperature, humidity, cloud water and ice and cloud fraction, ozone (also pressure at surface grid-points).

Operational model levels
(91-level model)



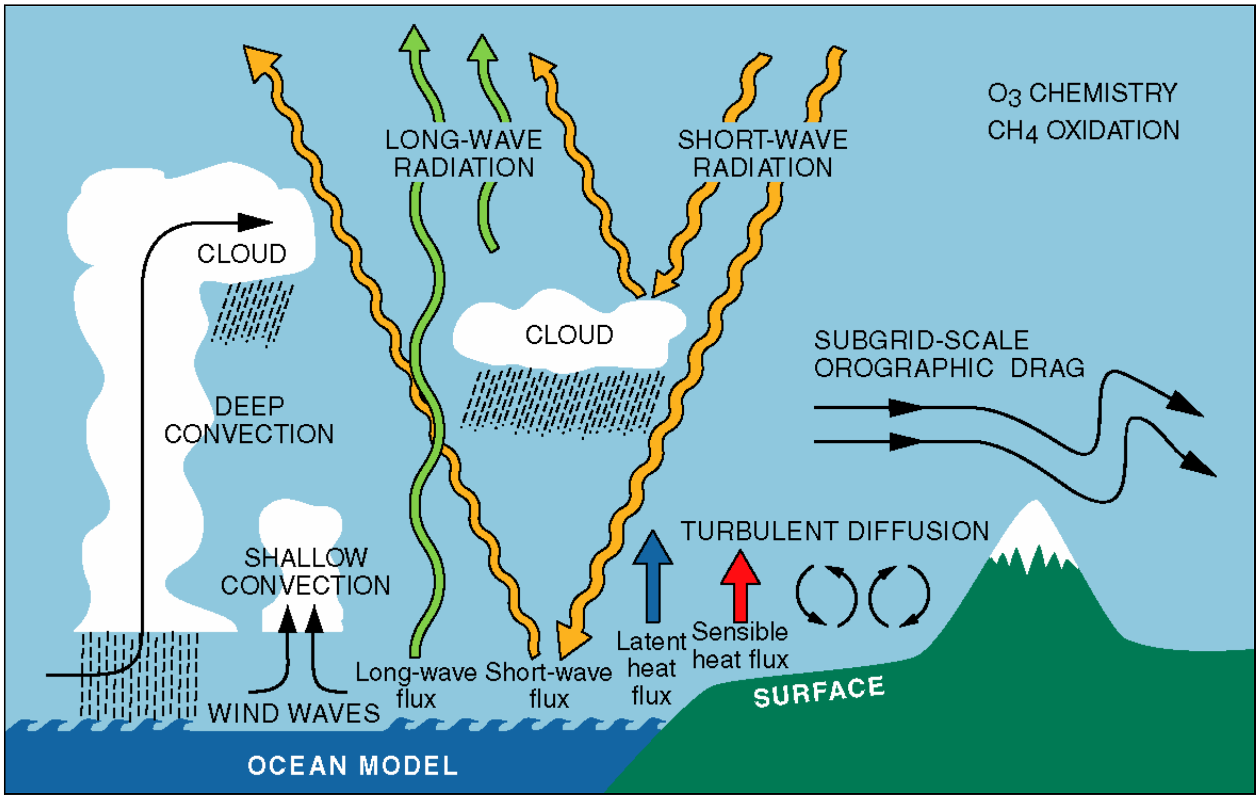
ECMWF

Operational model grid - T799 (25km)



The ECMWF numerical weather prediction model

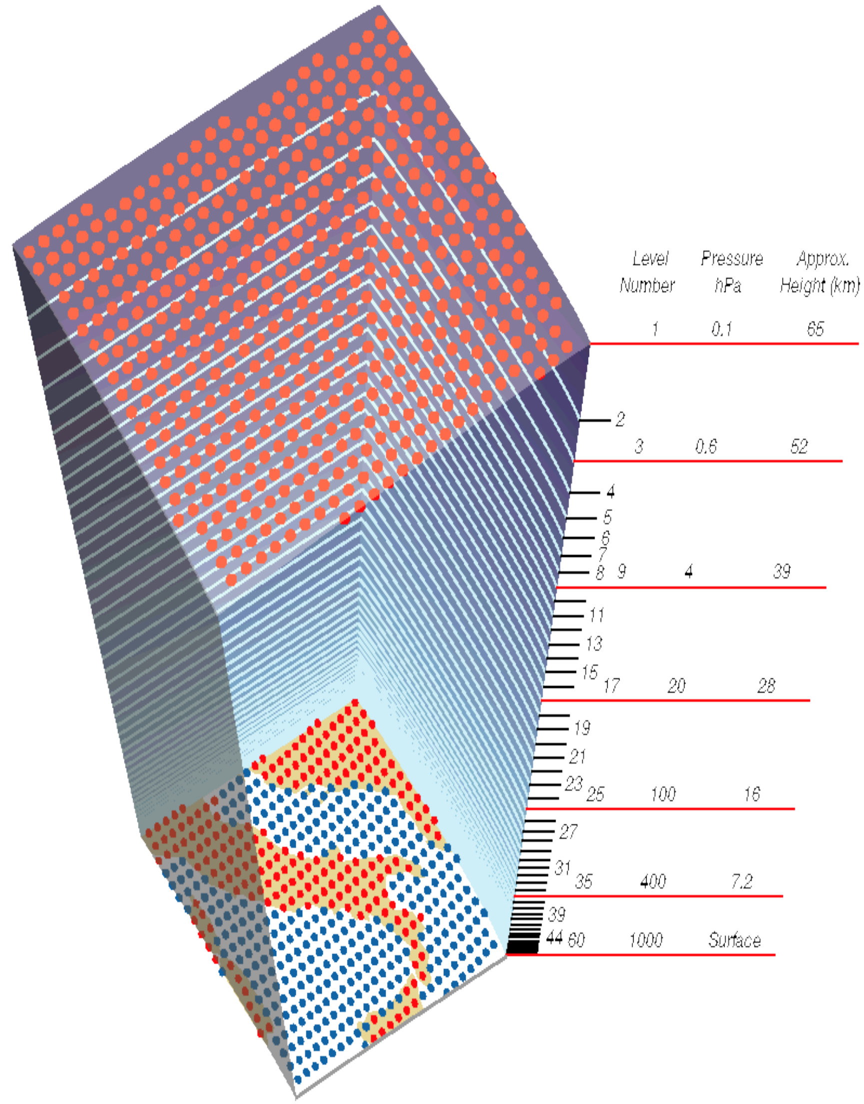
The atmosphere does not evolve in isolation, interactions between the atmosphere and the underlying land and ocean are also important in determining the weather. Ocean ice processes, ocean surface waves, land surface, soil, hydrological and snow processes are all represented at ECMWF in the operational Earth-system model.



These physical processes have smaller scales than the model grid (40 km) and are therefore represented by so-called "Parametrization Schemes" which represent the effect of the small-scale processes on the large-scale flow.

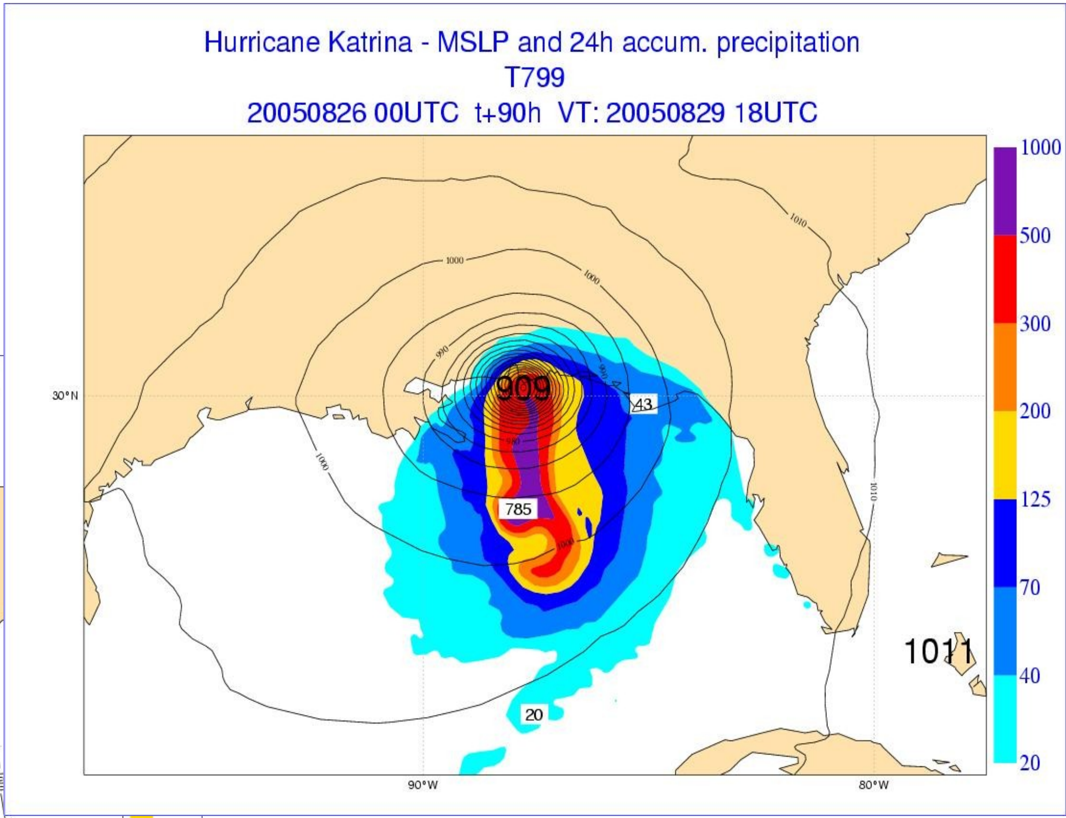
The ECMWF forecast model

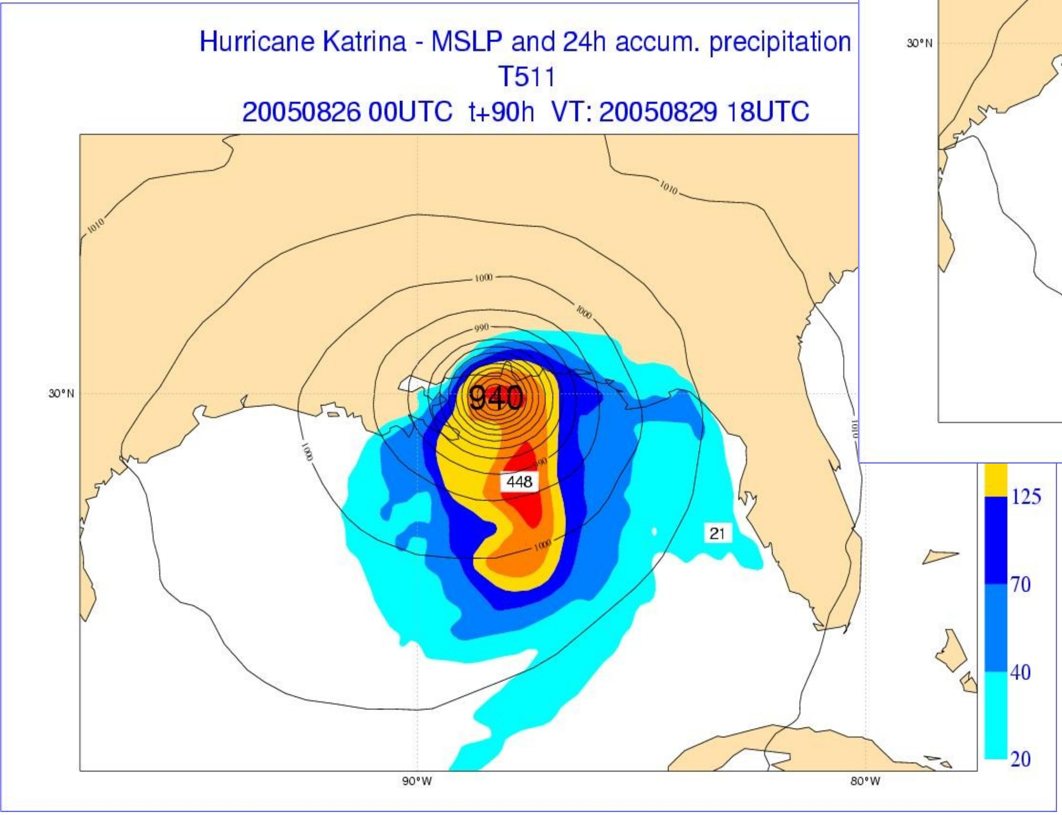
- The model has 76,757,590 grid points
- Spacing of grid points: 25 km
- 91 levels from surface to 85 km (1Pa)
- Temperature, wind, humidity and ozone are specified at each point.
- This is done in 12 minute time-steps.
- Number of computations required to make a ten-day forecast: 1,630,000,000,000,000



Katrina 90h forecasts at T511 and T799

T799 (25km):
- **Central pressure 909 hPa**
- **785mm rain in 24h**





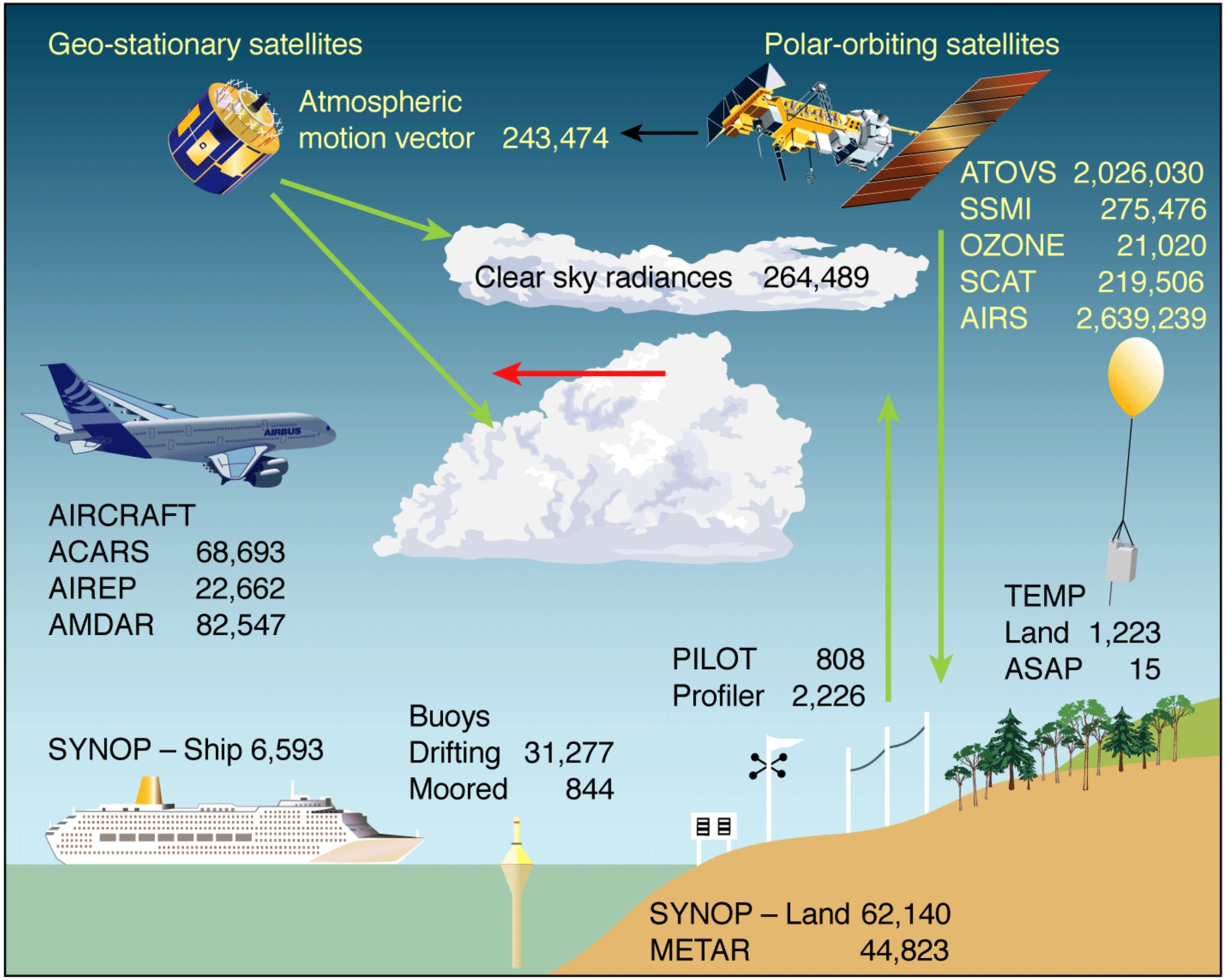
T511 (40km):
- **Central pressure 940 hPa**
- **448 mm rain in 24h**

Outline

- Introduction: ECMWF
- Operational data assimilation
  - → Model
  - → Observations
  - → Computational aspects
- The future of 4D-Var
  - → Flow dependent background error statistics
  - → Weak constraint 4D-Var
- Conclusions

The medium-range model
Data assimilation: atmospheric fields

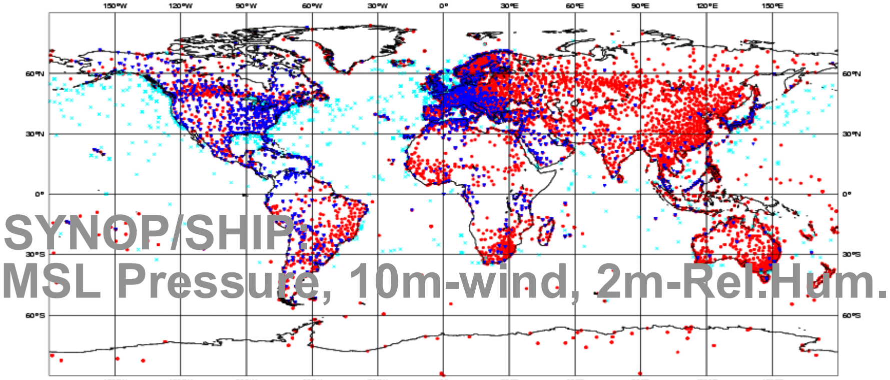
- Analyses of the wind, temperature, humidity, ozone and surface pressure of the atmosphere are produced by a four-dimensional variational assimilation system.
- The wind data in the free atmosphere are provided by balloon sondes, aircraft, profilers, dropsondes, and feature-tracking by geostationary satellites. Low-level wind data are used over sea only, and rely on ship reports, buoys and satellite information from scatterometer radar and microwave sensors.
- Temperature data come from balloon sondes and aircraft measurements, and by satellite remote-sensing of microwave and infrared radiances.
- Humidity data are provided by land station reports, by balloon sondes, and by satellite microwave and infrared sensors. Ozone data are inferred from satellite sensors.
- Surface pressure data are provided by land and ship stations and buoys.

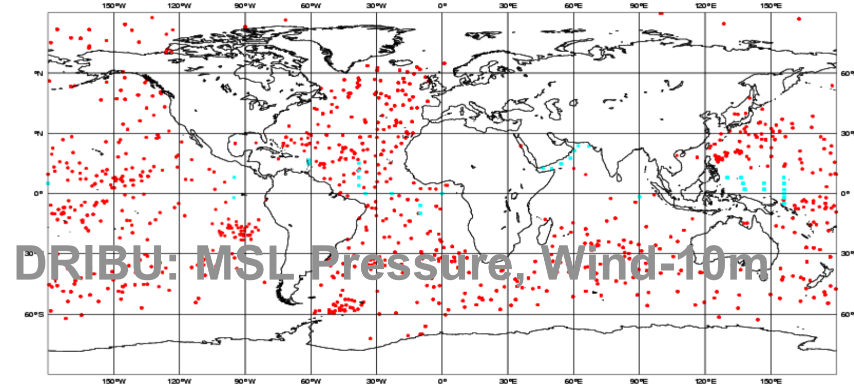


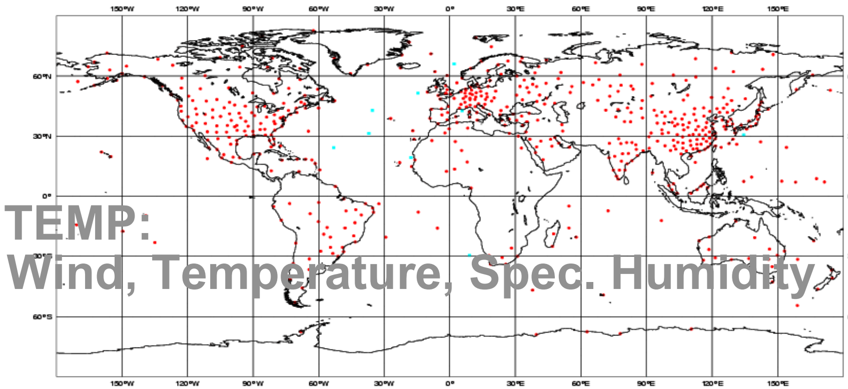
Data sources for the ECMWF Meteorological Operational System (EMOS)

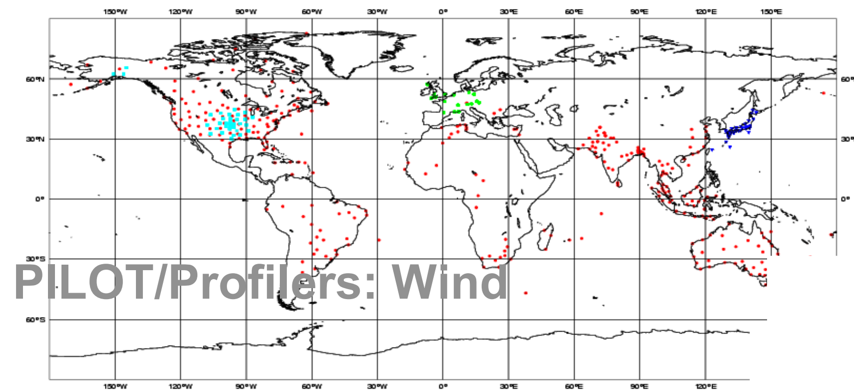
Number of observational items assimilated over 24 hours on 13th February 2006

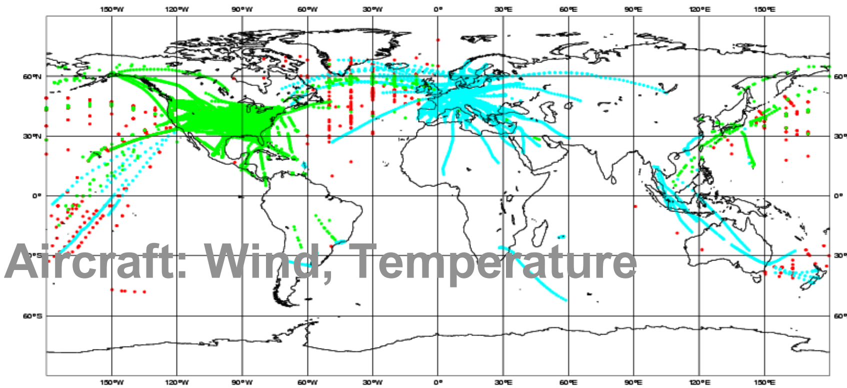
Conventional observations used

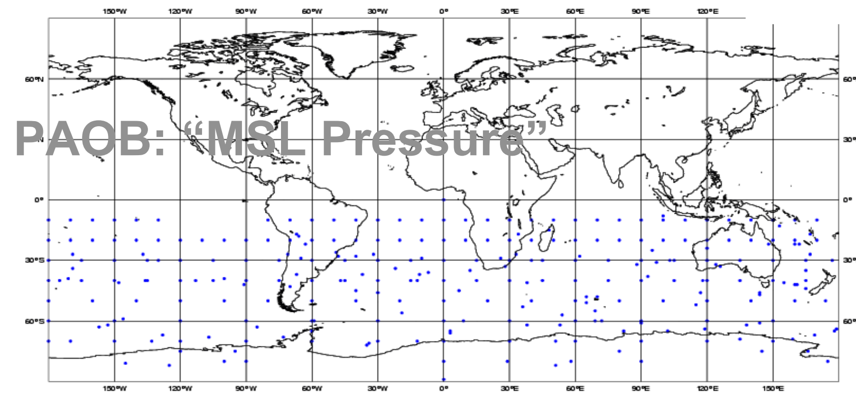






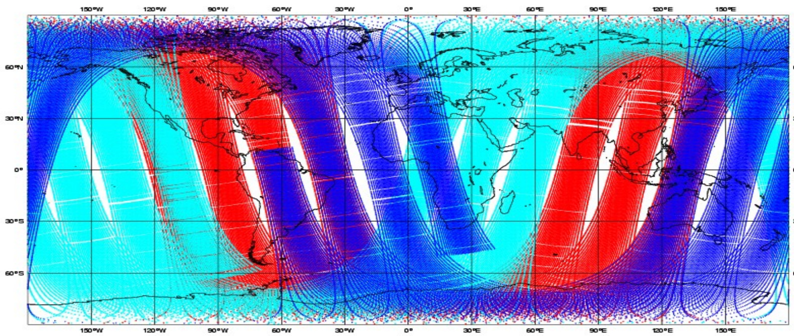




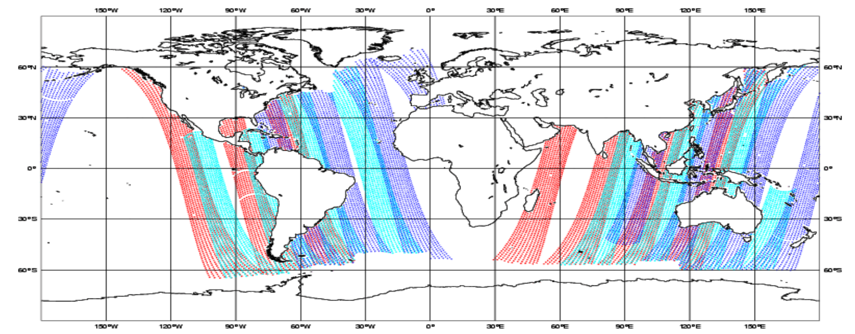


27 satellite data sources used in 4D-Var

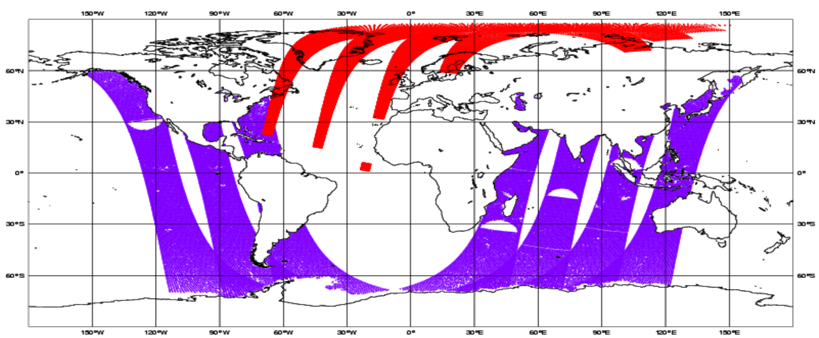
NOAA AMSUA/B HIRS, AQUA AIRS



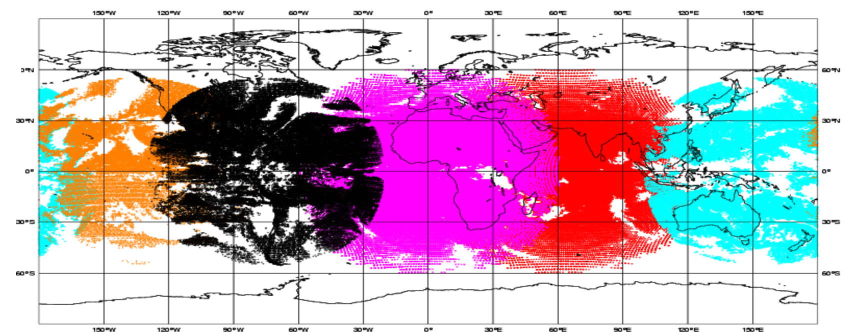
DMSP SSM/I



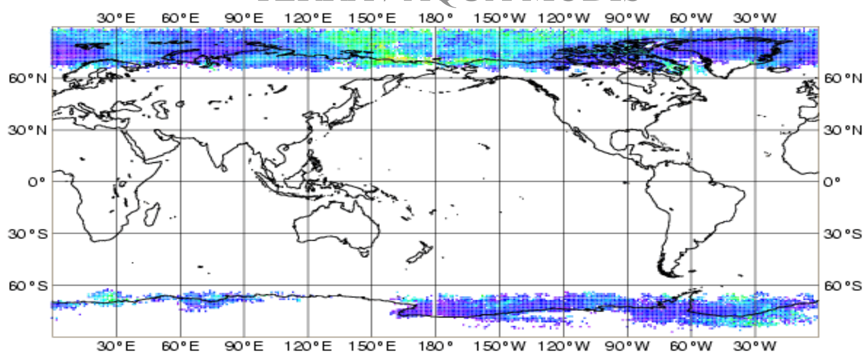
SCATTEROMETERS



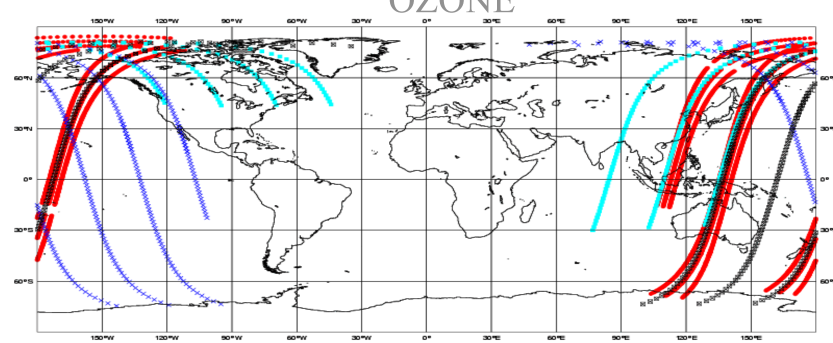
GEOS



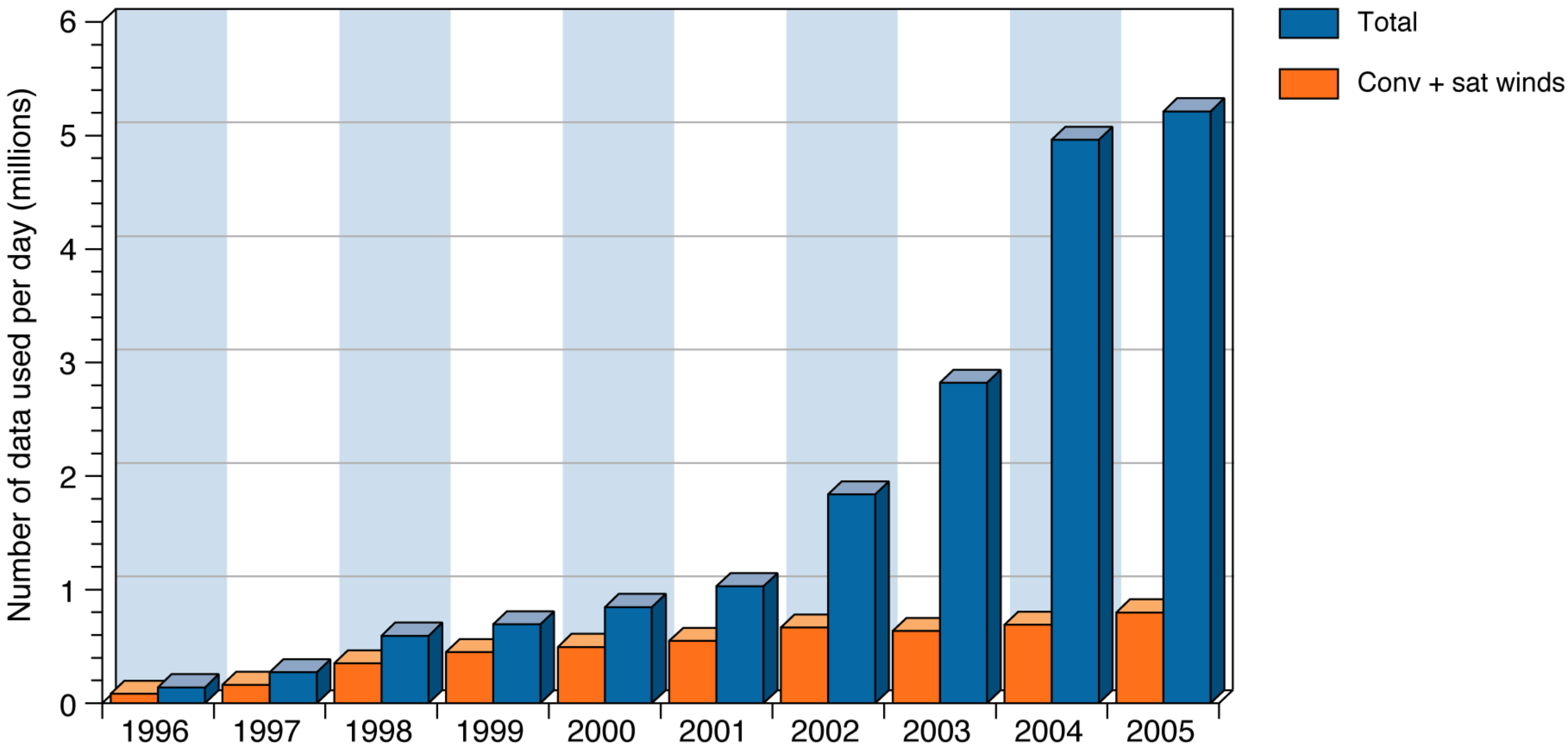
TERRA / AQUA MODIS



OZONE



Number of Used Data per Day



Recent revisions to observation usage

- Use MODIS winds from AQUA satellite (04/2005)
- METAR surface pressure data active (04/2005)
- Surface pressure adaptative bias correction (04/2005)
- Assimilation of “Cloudy and rainy radiances” from SSM/I (06/2005)
- Atmosp. feature track. winds from Meteosat-8 (MSG) (06/2005), MTSAT (12/2006)
- Thinning of low level AMDAR data (07/2006)
- Use GPS radio occultation data (COSMIC, CHAMP, GRACE) (12/2006)
- Assimilation of IASI data (06/2007)

Recent revisions to the assimilation system

- Wavelet Jb formulation (04/2005)
- Adaptive bias correction scheme for surface pressure data (04/2005)
- Jb statistics from latest ensemble data assimilation (06/2005)
- Increased resolution from T511-T95/T159 L60 to T799-T95/T255 L91 (02/2006)
- Use grid-point humidity and ozone in 4D-Var analysis (02/2006)
- Variational bias correction of satellite radiances (07/2006)
- Three outer loops in 4D-Var (T799-T95/T159/T255) (06/2007)
- Convection in moist tangent linear physics (06/2007)

Observation data count (27/07/07-00UTC)

Screened

Synop	407,812	0.26%
Aircraft	487,435	0.31%
Dribu	19,494	0.01%
Temp	164,880	0.11%
Pilot	107,004	0.07%
AMVs	2,201,118	1.40%
Radiances	152,125,646	97.06%
Scat.	820,830	0.52%
GPS occult.	209,501	0.13%
Total	156,734,720	100.00%

99% of screened data is from satellites

Assimilated

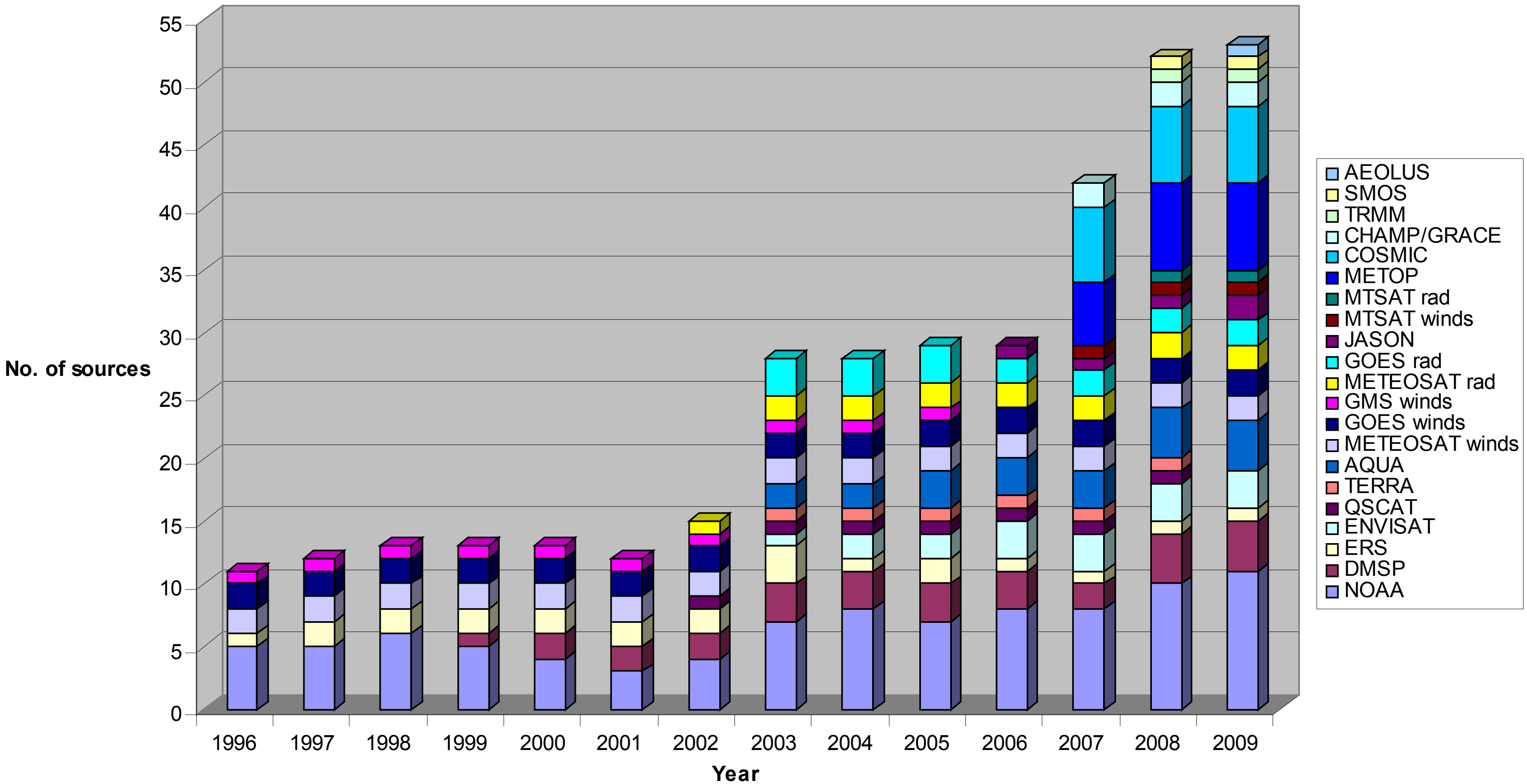
Synop	60,683	0.68%
Aircraft	235,741	2.65%
Dribu	5,901	0.07%
Temp	82,569	0.93%
Pilot	48,870	0.55%
AMVs	95,466	1.07%
Radiances	8,137,481	91.37%
Scat.	149,000	1.67%
GPS occult.	90,716	1.02%
Total	8,906,427	100.00%

95% of assimilated data is from satellites

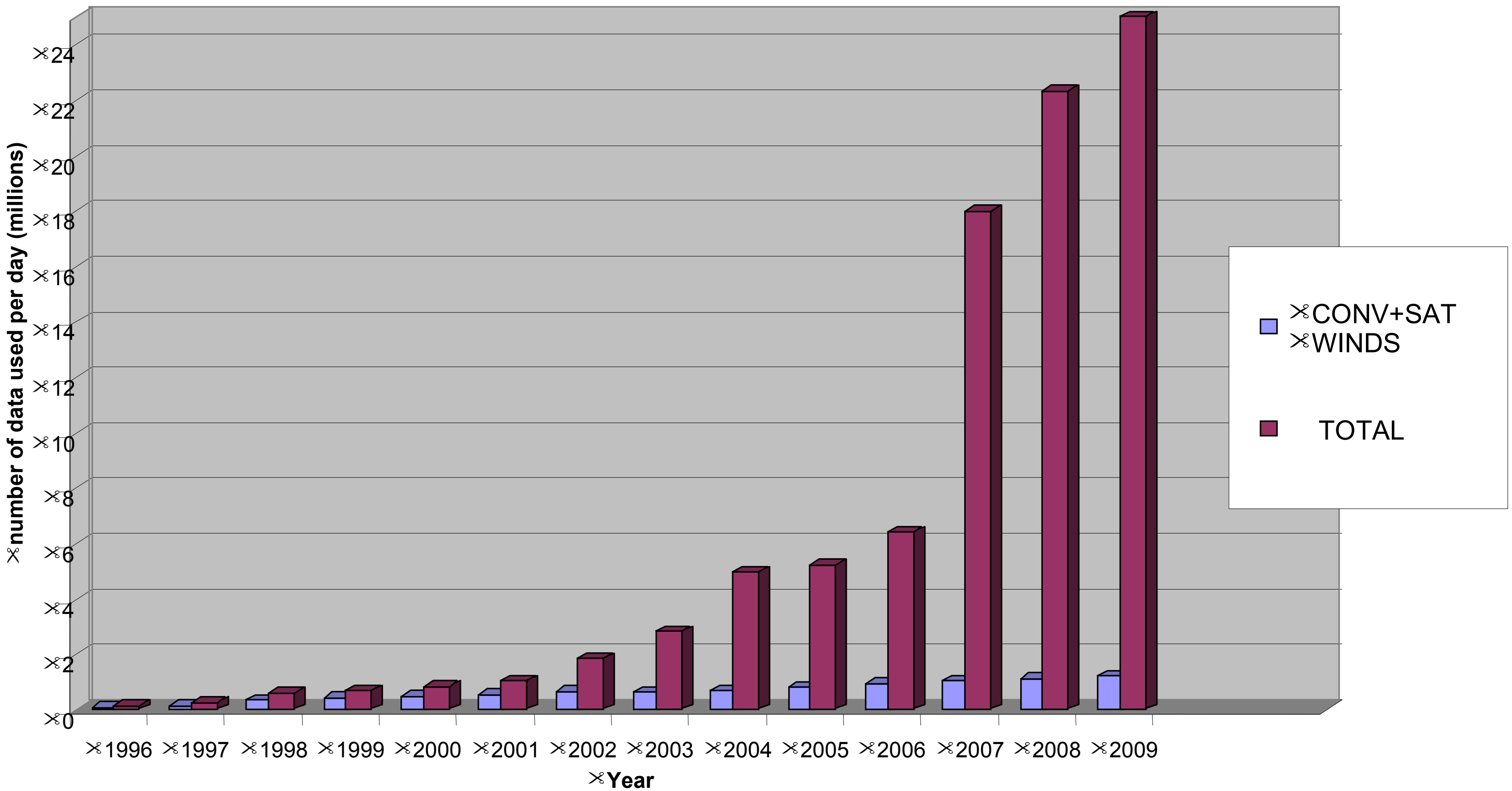
Only 5.7% of screened data is assimilated.

ECMWF

Mid 2007 we use 41 different satellite data sources, and by 2009 we should use more than 50



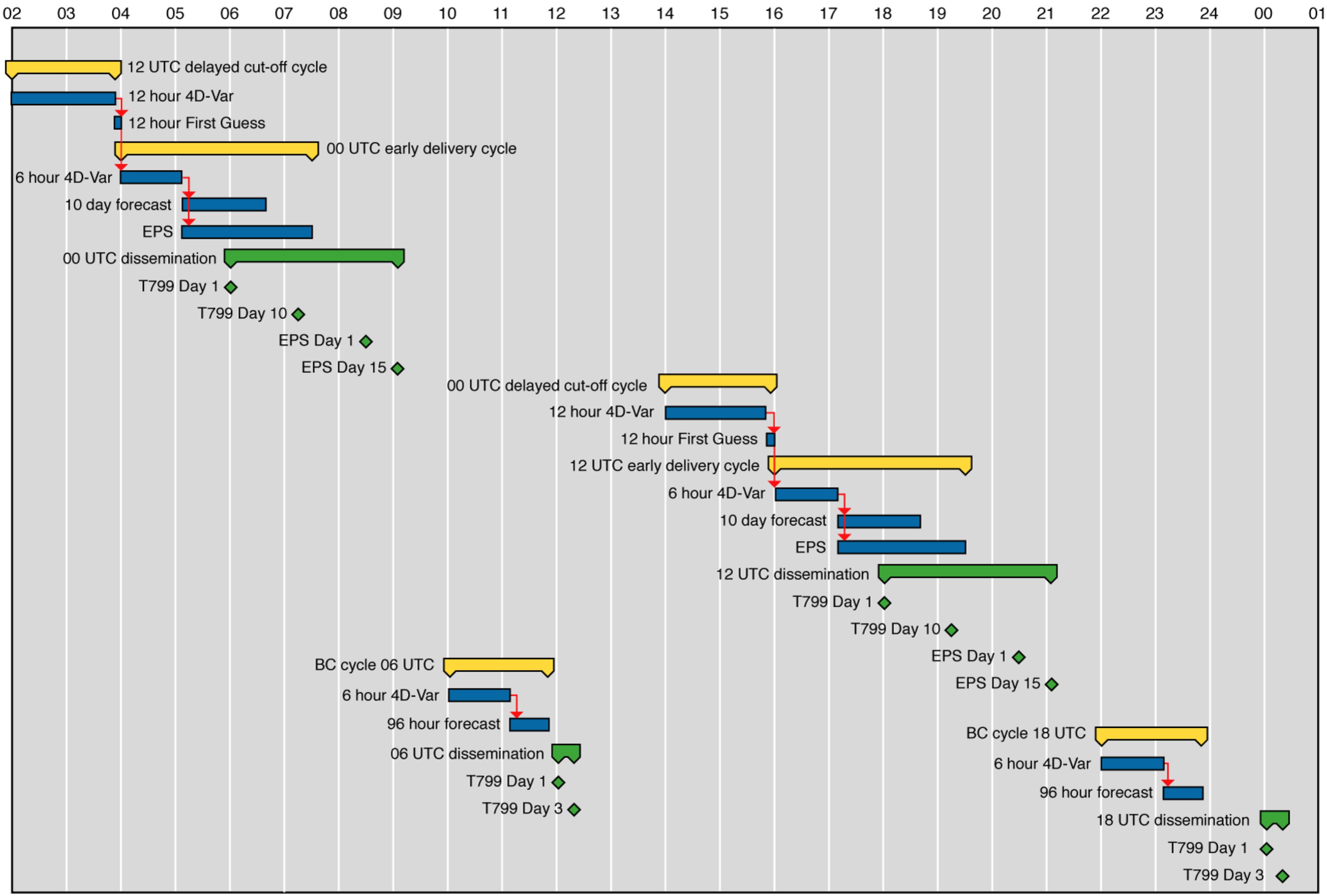
Mid 2007: Satellite data volumes used: around 18 millions per day today, and probably around 25 millions by 2009



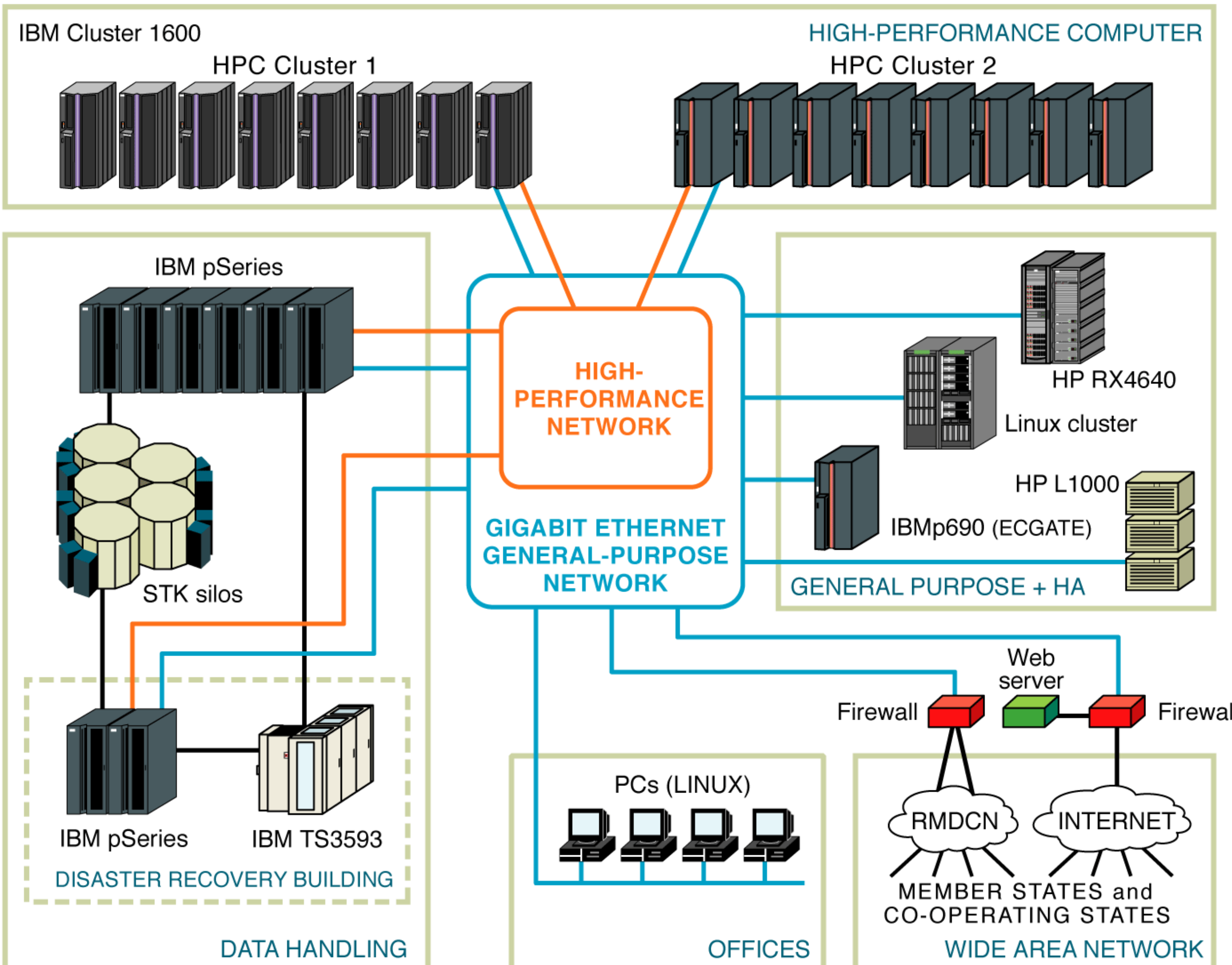
Outline

- Introduction: ECMWF
- Operational data assimilation
  - Model
  - Observations
  - Computational aspects
- The future of 4D-Var
  - Flow dependent background error statistics
  - Weak constraint 4D-Var
- Conclusions

The main operational suites on ECMWF's HPCF



Current computer configuration



ECMWF

Supercomputer Configuration

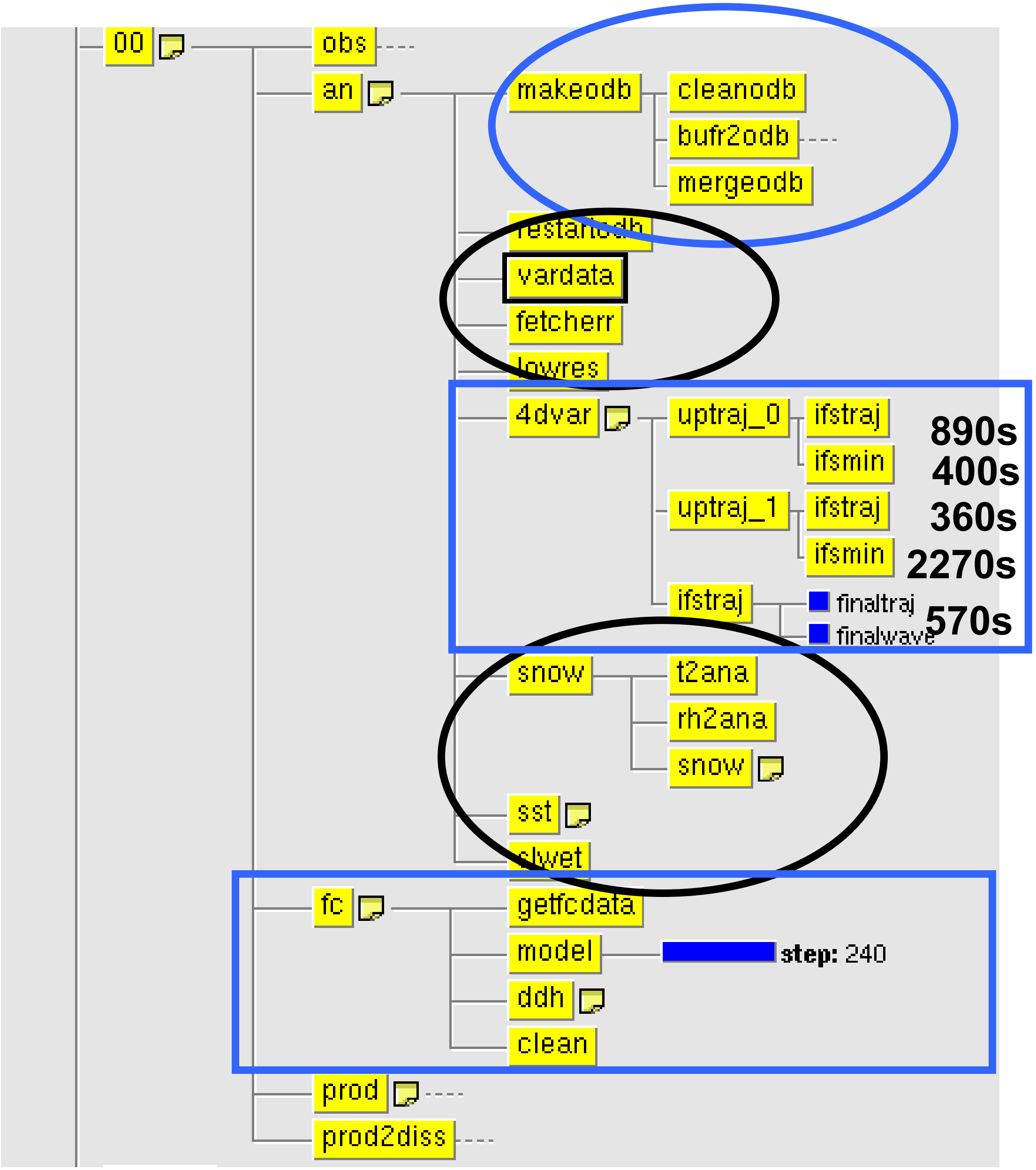
IBM p5-575+ (x2)	2006-2007
Number of processors	2480
Type of processor	1.9Ghz Power5+
Performance per CPU	7.6 Gflops
Number of Nodes	155
CPUs per node	16
Memory per Node	32 Gb
Total Memory	4.5 Tb
Switch Bandwidth	2 Gb/s
Total Performance (peak)	19 Tflops
Total Performance (sustained)	4 Tflops

Performance of Operational Runs

	Tasks	Threads	CPUs	Time
4D-Var	96	8	768	1h25min
10 days T799 Forecast	96	8	768	1h15min
T399 EPS Forecast	24	4	96	35min

The total cost of the 51-members EPS is roughly twice the cost of the 10 days T799 4D-Var+forecast.

Analysis and forecast for one cycle



BUFR to ODB.

200sec 4x(8-16PEs)

Fetch T799 fields and interpolate to T95 and T255

360sec (1-128PEs)

Analysis: trajectory, minimization and update.

4490sec (768PEs)

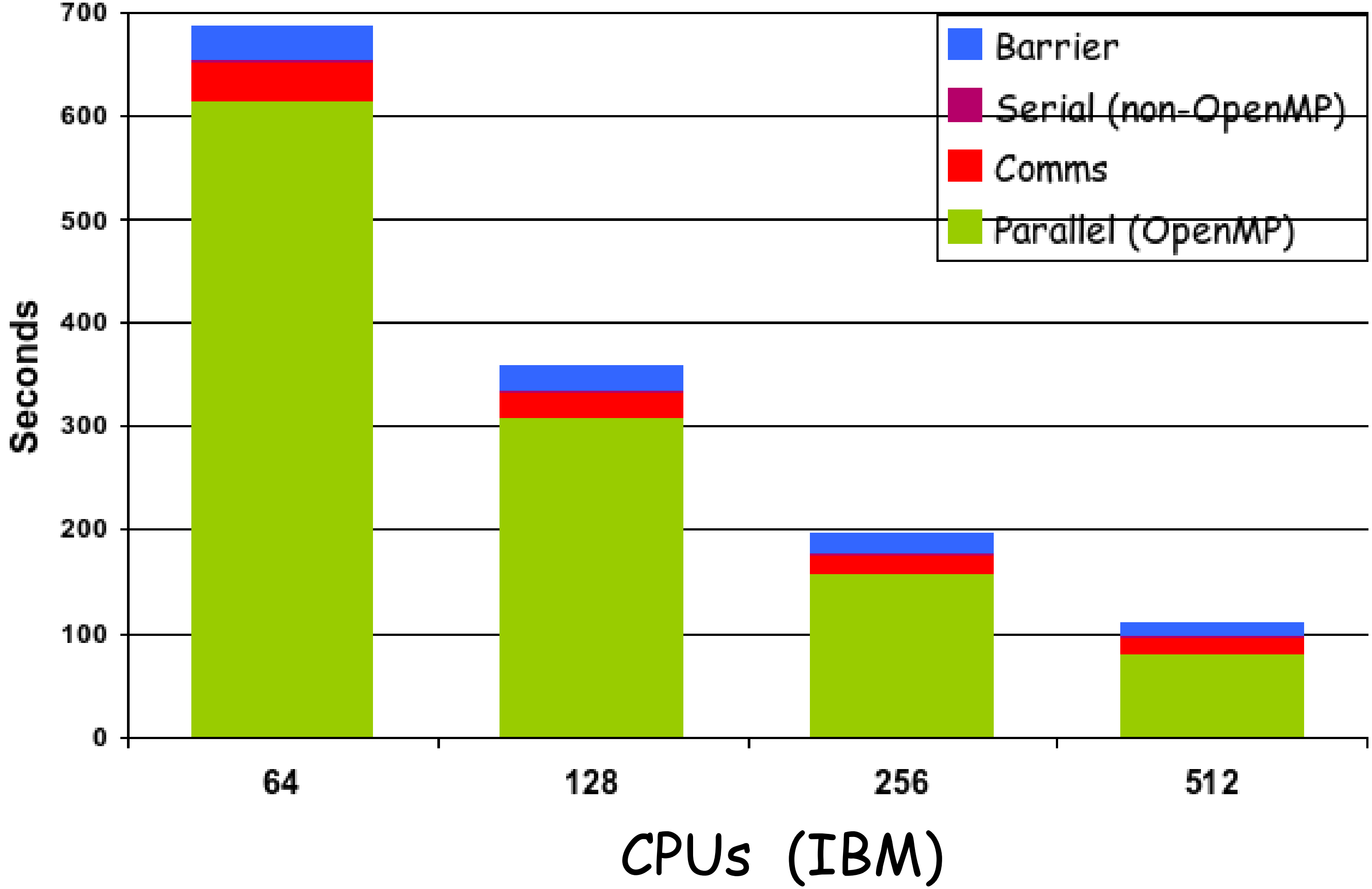
Surface analysis.

940sec 4x(1PEs)

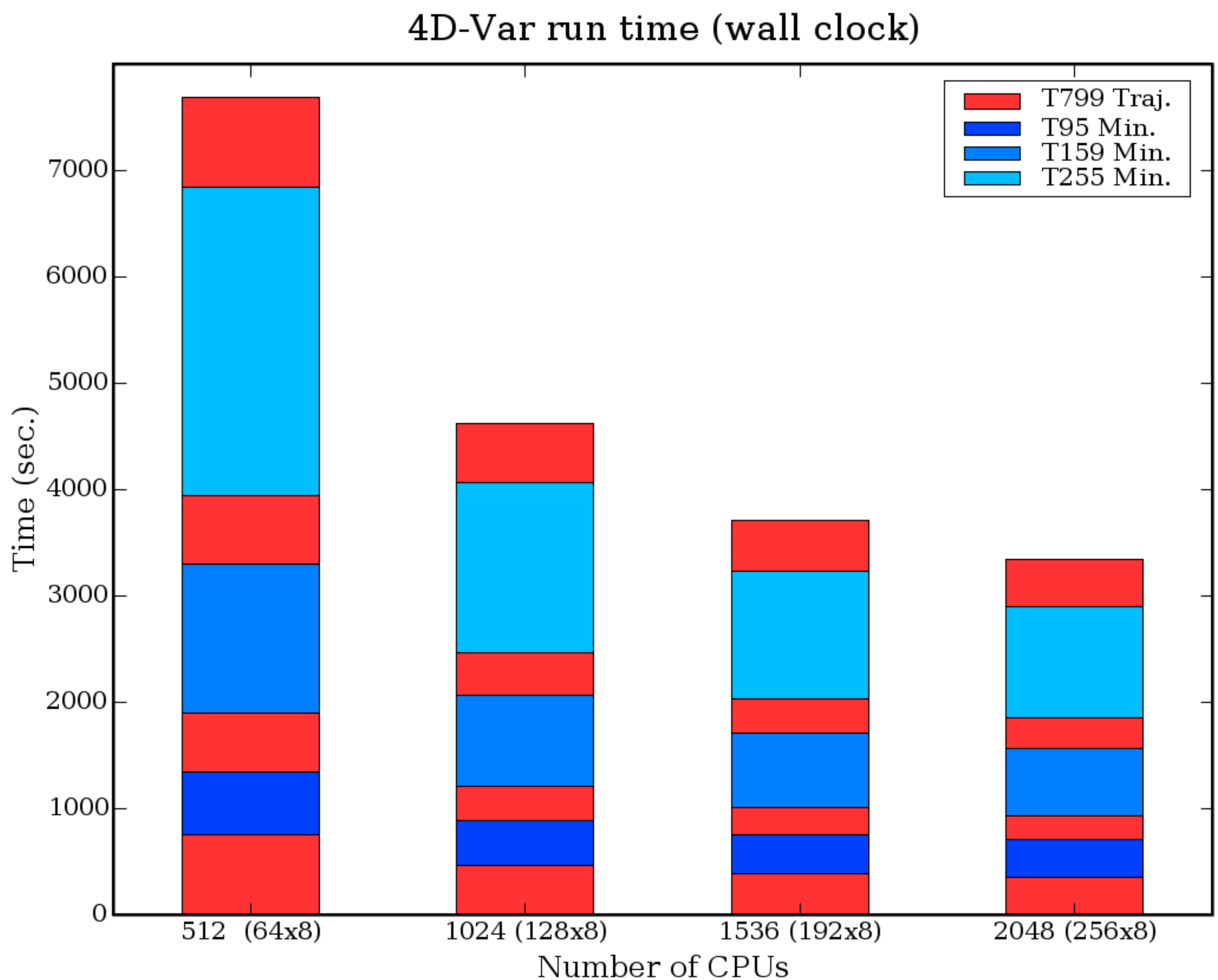
10 day forecast.

4980sec (768PEs)

Scaling of Forecast (T511 per day)



Parallel computing and 4D-Var



Outline

- Introduction: ECMWF
- Operational data assimilation
  - ➔ Model
  - ➔ Observations
  - ➔ Computational aspects
- The future of 4D-Var
  - ➔ Flow dependent background error statistics
  - ➔ Weak constraint 4D-Var
- Conclusions

The Future of 4D-Var at ECMWF

- Provide better atmospheric estimates for operational forecasting and reanalysis.
- New types and higher resolution of observations.
- Observation equivalents have to be computed accurately in the minimisation (model and observation operators) in terms of:
  - ➔ Resolution,
  - ➔ Physical processes.
- Errors that were ignored in the past become important !
- Observation error correlations should be taken into account (use randomisation method).
- Account for biases: Observations and Model.

The Future of 4D-Var at ECMWF

- Flow dependant background error statistics:
  - ➔Real time En4DV at reduced resolution.
- Weak constraint 4D-Var:
  - ➔Account for model error (including bias),
  - ➔Long window.
- All this will require even more computer power !

Outline

- Introduction: ECMWF
- Operational data assimilation
  - Model
  - Observations
  - Computational aspects
- The future of 4D-Var
  - Flow dependent background error statistics
  - Weak constraint 4D-Var
- Conclusions

J.-N. Thépaut

Need for tuning the 3DVAR and 4DVAR systems

- **Background error covariances have to reflect the poorer accuracy of analysis and short range forecast due to the sole use of surface pressure observations.**
- **This has been done "objectively" from statistics obtained from the blunt "surface pressure only " assimilation experiments, by computing the effective σ_b**

$$\sum_i (obs - guess)^2 = \sigma_o^2 + \sigma_{beff}^2$$

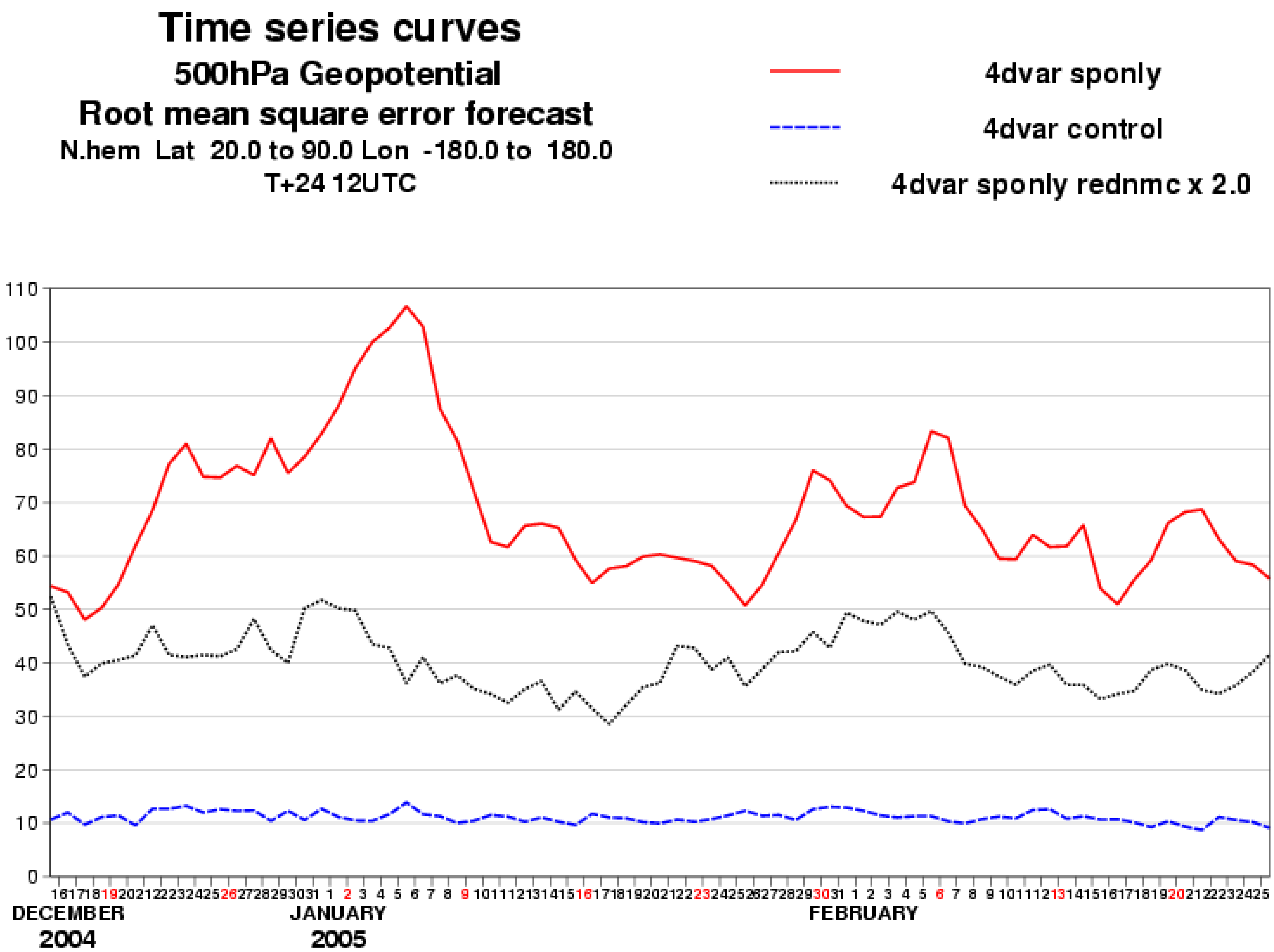
$$\sigma_{beff} = \alpha \times \sigma_{borig}$$

"REDNMC" factor

Surface pressure only experiment

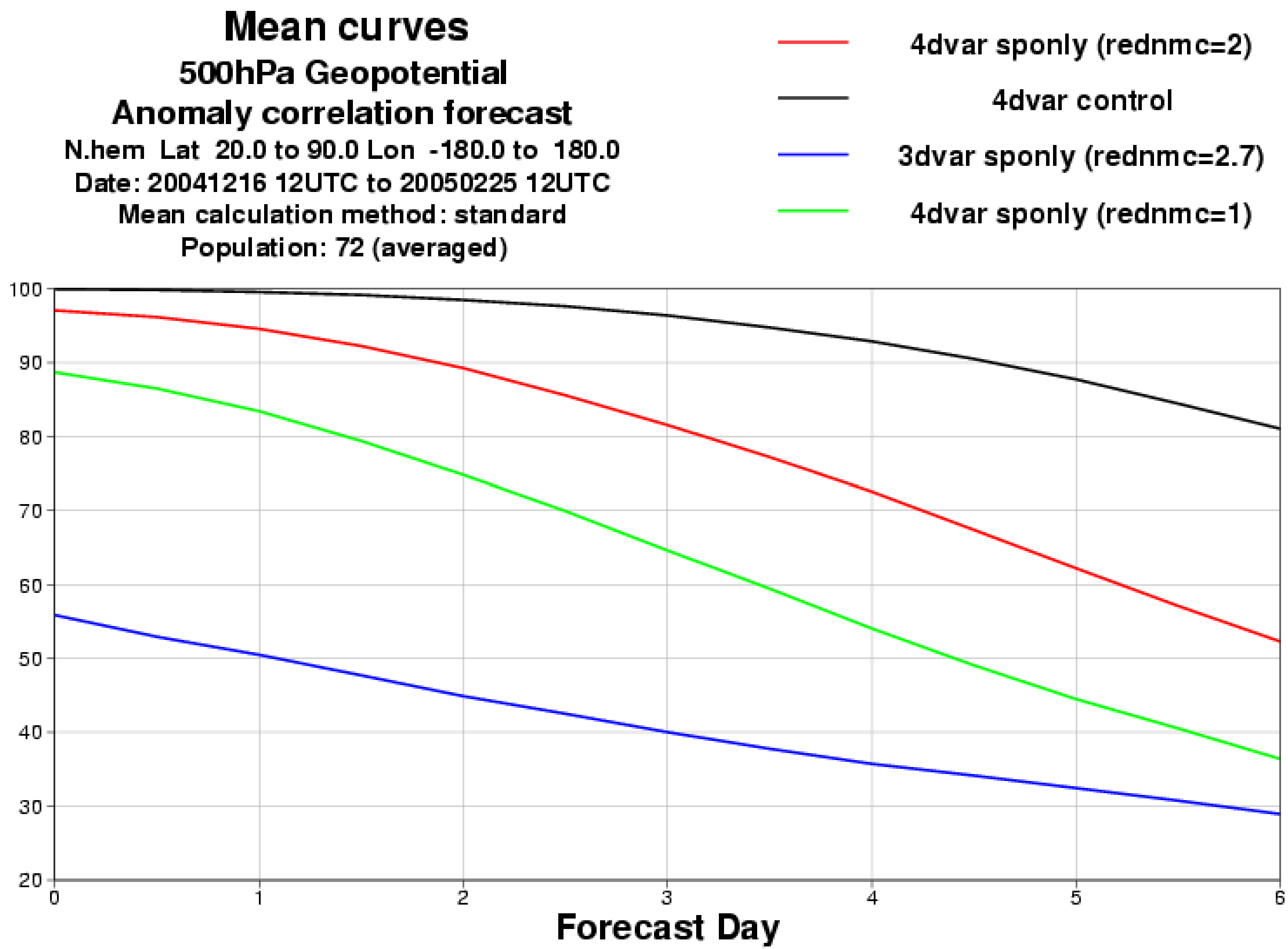
J.-N. Thépaut

Impact of tuning (4DVAR Surf. Press. only)



J.-N. Thépaut

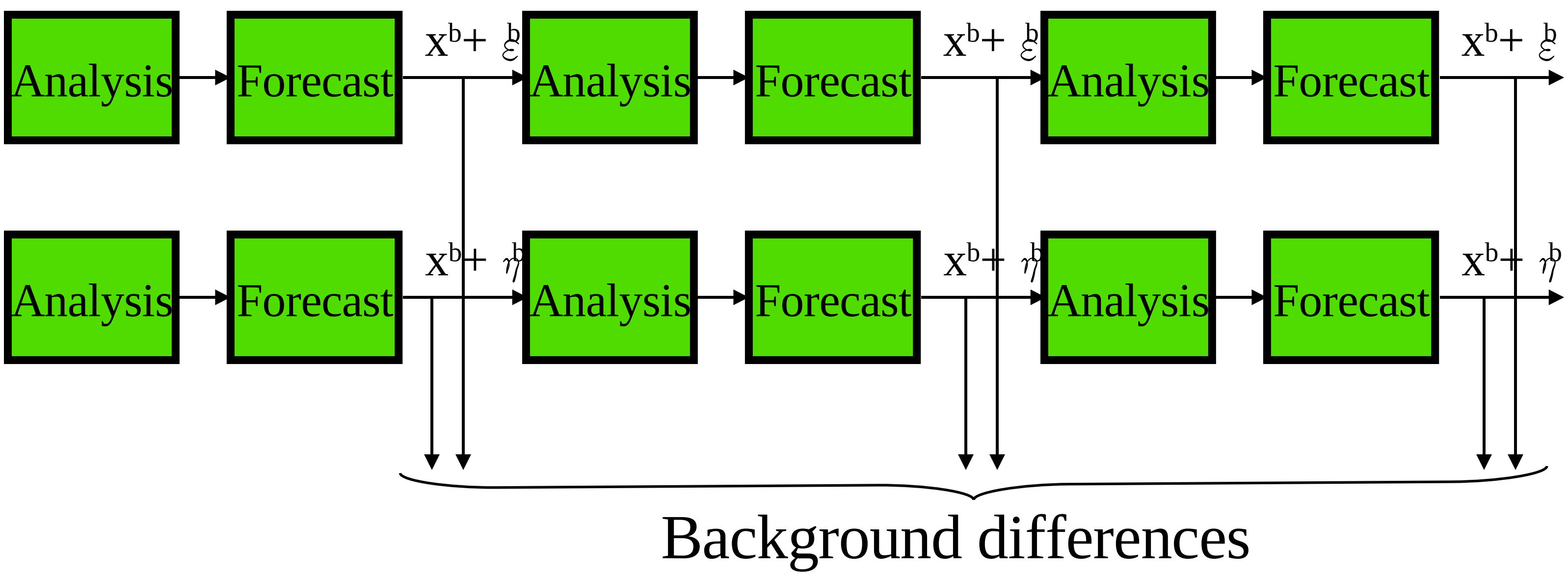
Impact of tuning (4DVAR Surf. Press. only)



ECMWF

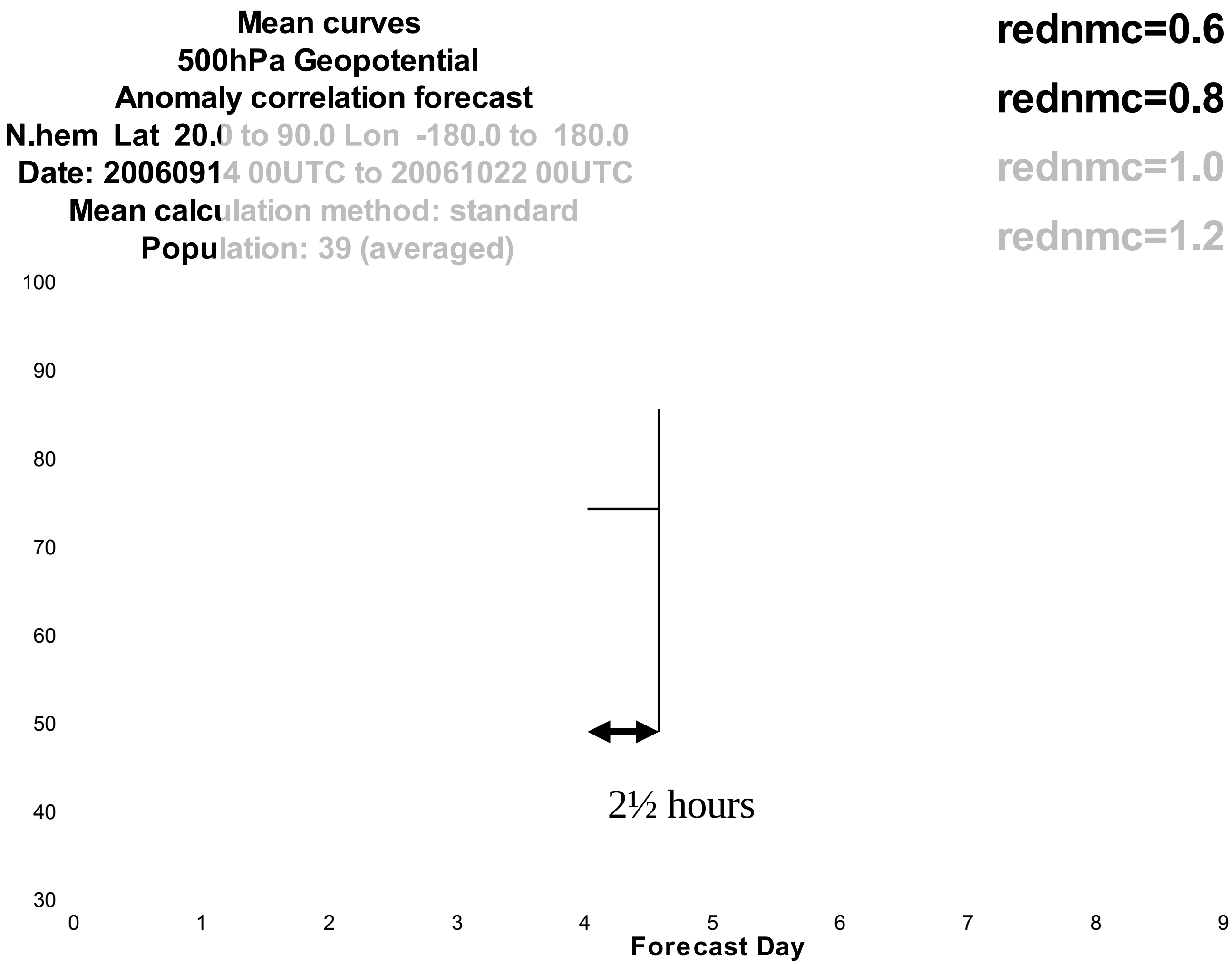
Estimating Flow dependent Background Error from an ensemble DA method

- Run an ensemble of analyses with random observation and SST perturbations, and form differences between pairs of background fields.
- These differences will have the statistical characteristics of background error (but twice the variance).



M. Fisher

Impact of a Global Scaling of σ_b



EnDA Conclusions

- We have developed a streamlined ensemble DA system
- We have included an improved representation of model error to get more realistic spread from the ensemble DA
- The use of flow dependent background error variances based on the ensemble DA spread does not improve the general scores; but impact near tropical cyclones, troughs and extra-tropical cyclones looks promising
- The general scores are at ECMWF determined by the broad temperature/wind structures that are well described by the high volume satellite data. Flow dependence does not matter for these structures. The forecast model can generate and evolve cyclones on its own from these accurate but broad initial conditions.

EnDA Future work

- We will try to take account of correlated radiance errors and improve representation of model error in the ensemble DA
- It would be beneficial to run a research mode ensemble DA system in real time mode to learn from daily monitoring and to calculate seasonal variance estimates
  - ➔ Resolution? TBD (likely T399 outer loop/T159 inner loop)
  - ➔ Number of members? TBD (likely 10)
- The wavelet J_b formulation used at ECMWF will ease introduction of flow dependent variances and structures
- It may be beneficial to use ensemble data assimilation based estimates of short range forecast errors in the EPS system

Outline

- Introduction to ECMWF
- Operational data assimilation
  - Model
  - Observations
  - Computational aspects
- The future of 4D-Var
  - Flow dependent background error statistics
  - Weak constraint 4D-Var
- Conclusions

Weak constraint 4D-Var

- Model error does affect 4D-Var.
- Several approaches are possible to account for model error:
  - → Model error forcing term,
  - → Model bias term,
  - → 4D control variable.
- Benefits in the stratosphere.
- Interactions with observation bias correction?
- Potential for use of long windows in 4D-Var.
- Research in progress: see talk from last week.

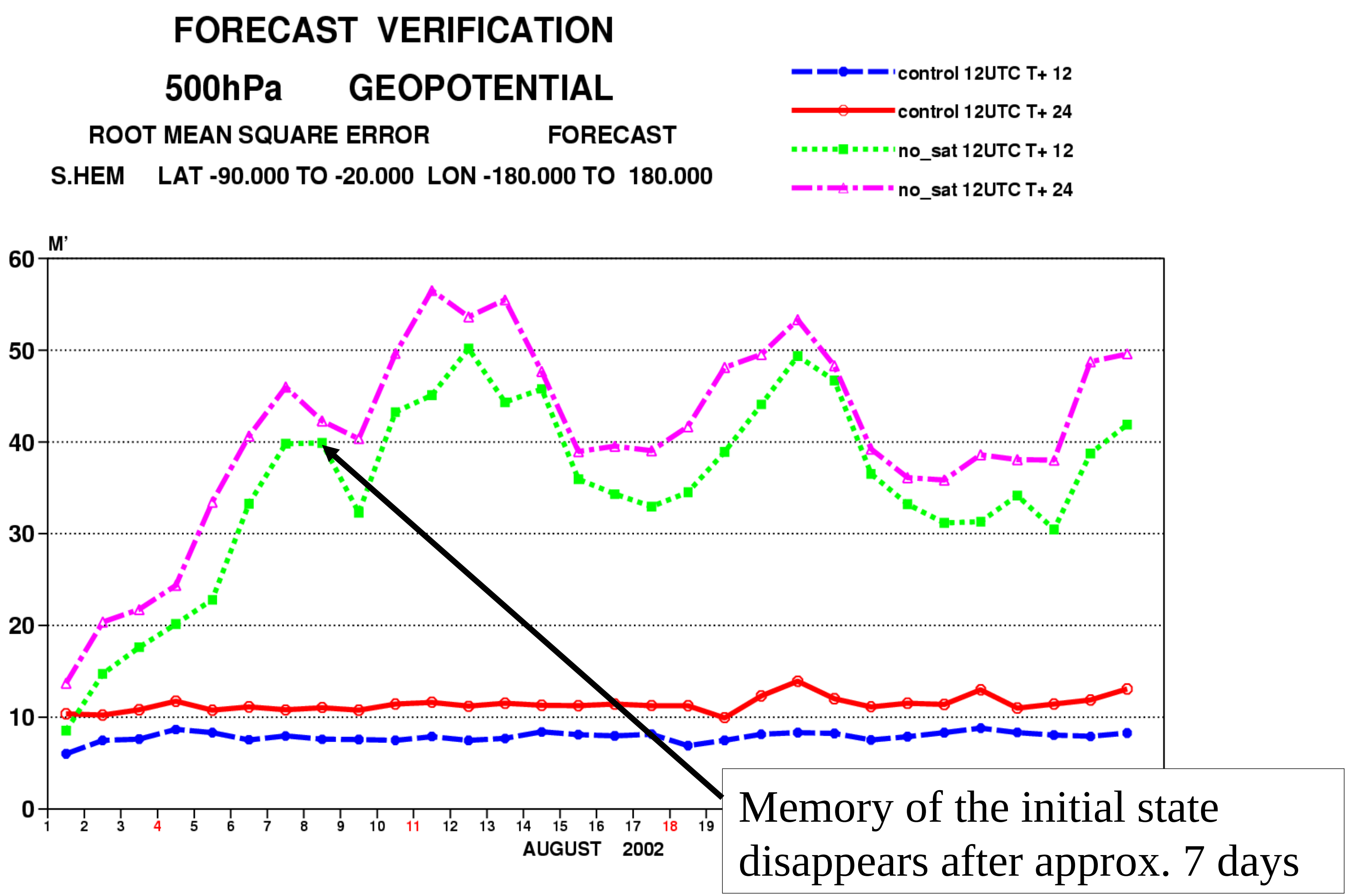
4D-Var = Kalman Smoother

- The equivalence "4dVar = Kalman smoother" is well known (and easy to prove) for a perfect, linear model.
- The equivalence is less well known for the case of an imperfect, linear model. A proof is given be Ménard and Daley (1996, Tellus):
    - ➔ Weak-constraint 4dVar with an imperfect, linear model and background covariance matrix B is equivalent to (i.e. gives the same state estimates as) a fixed-interval Kalman smoother that uses the same model, observations, observation operators, and initial covariance matrix B.
- In fact, 4dVar can handle more general pdf's (time-correlated model errors, non-Gaussian observation errors, nonlinear model, etc.) than the Kalman smoother.
- In this sense, 4dVar is a more fundamental method than the Kalman smoother!

Long window 4D-Var summary

- Long-window, weak-constraint 4dVar is an efficient algorithm for solving the Kalman smoothing problem for large-dimensional systems.
- No rank-reduction required.
- If the window is long enough, the analysis (and its covariance matrix) is independent of the background (and its covariance matrix).
- “Long enough” is probably somewhere between 3 and 10 days.
- Weak-constraint is a less stiff problem than strong-constraint: Minimization should be better conditioned (i.e. faster).
- Long-window, weak-constraint 4dVar isn’t cheap. But, you get a <u>full-rank</u> Kalman filter that should at least be a useful tool to evaluate other, sub-optimal methods.

Limited Memory

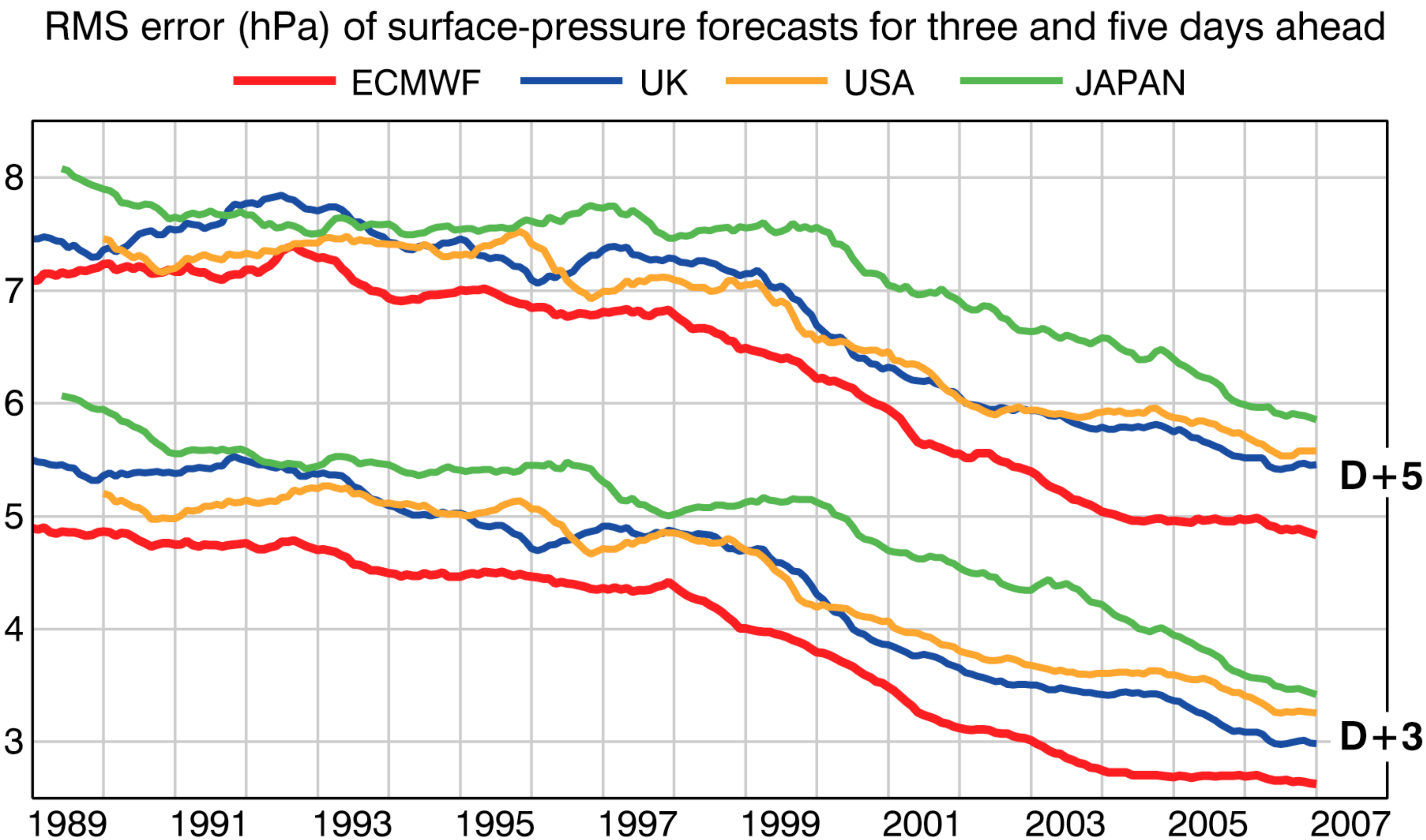


Analysis experiments started with/without satellite data on 1st August 2002

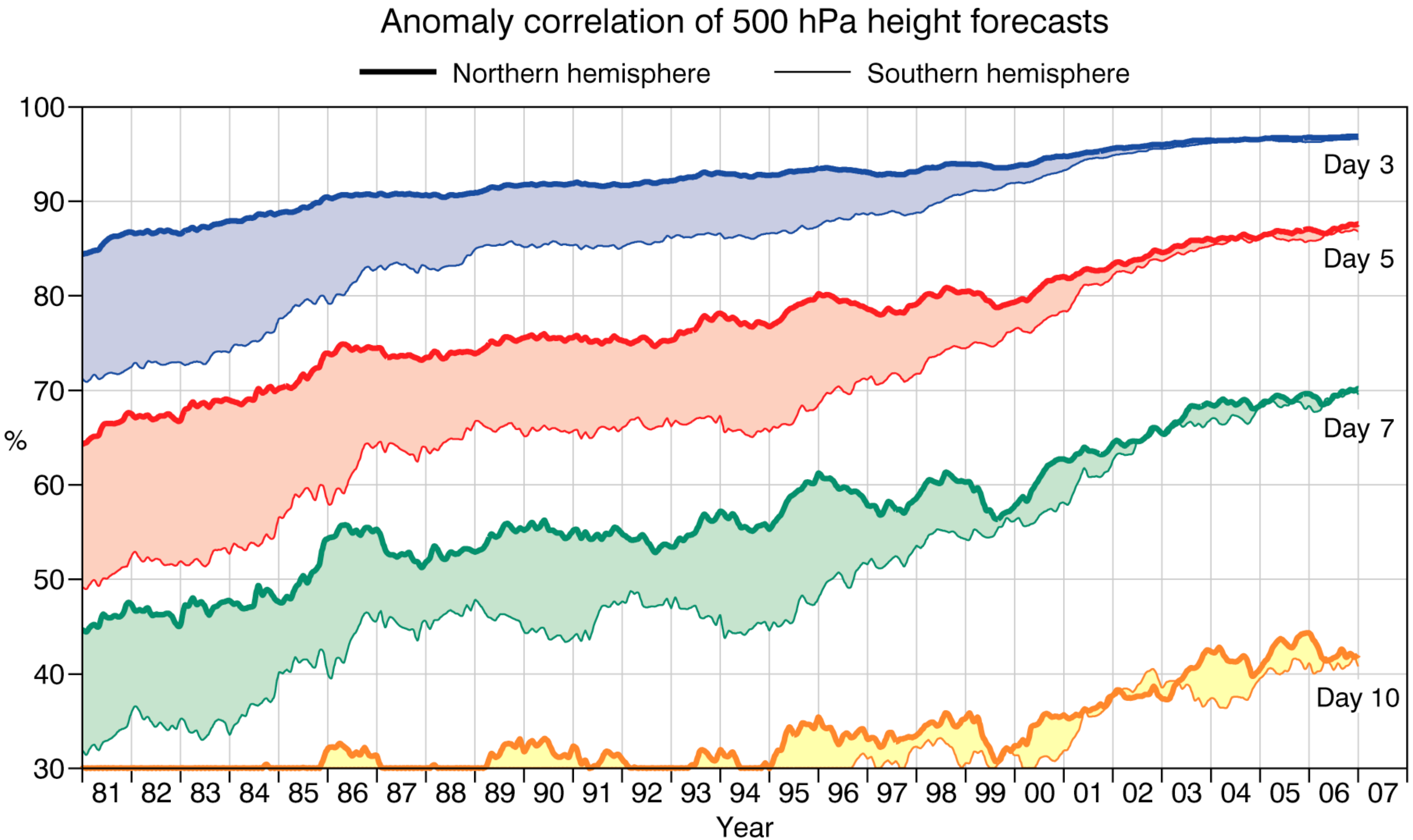
Outline

- Introduction to ECMWF
- Operational data assimilation
  - ➔ Model
  - ➔ Observations
  - ➔ Computational aspects
- The future of 4D-Var
  - ➔ Flow dependent background error statistics
  - ➔ Weak constraint 4D-Var
- Conclusions

ECMWF scores compared to other major global centres



Evolution of ECMWF scores comparison northern and southern hemispheres



Conclusions

- 4D-Var has performed well since its operational implementation in 1997.
- It should be improved further by new developments:
  - ➔Use higher resolution observations and new types of observations (clouds, rain…),
  - ➔Correction of biases (observations and model),
  - ➔Better modelling of errors (background and observations),
  - ➔Account for model error.