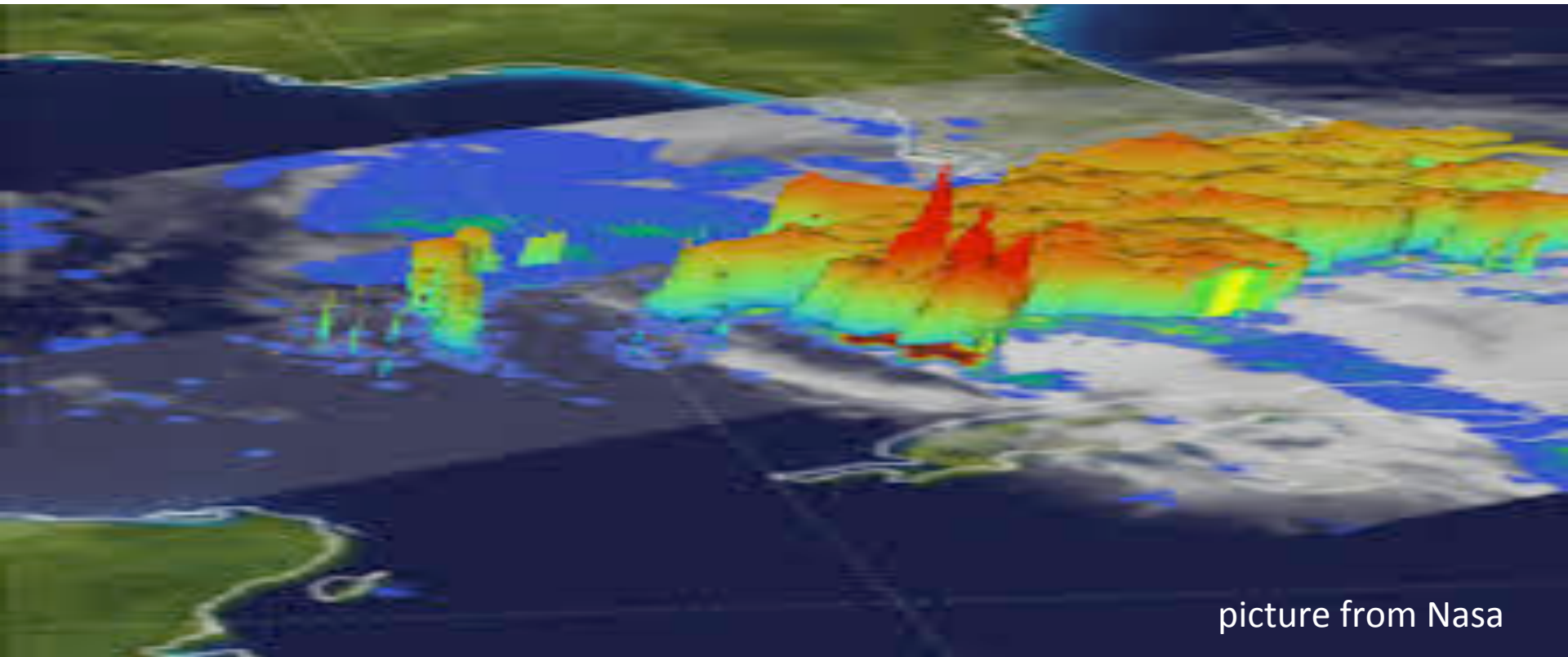


# 1. Analysis and Reanalysis Products

Adrian M Tompkins, ICTP

Tompkins@ictp.it



picture from Nasa

# Many hazards to health have climate drivers

- Precipitation:
  - Water-borne disease such as cholera,
  - Vector-borne diseases e.g. malaria, dengue, LF, schisto...
- Temperature:
  - heatstress, extremes, accidents
  - vector-borne diseases,
  - communicable diseases (behaviour)
- Humidity
  - Vector-borne disease
  - Virus transmission
- Wind
  - Meningitis
  - malaria (vector tracking)
  - transport of pathogens

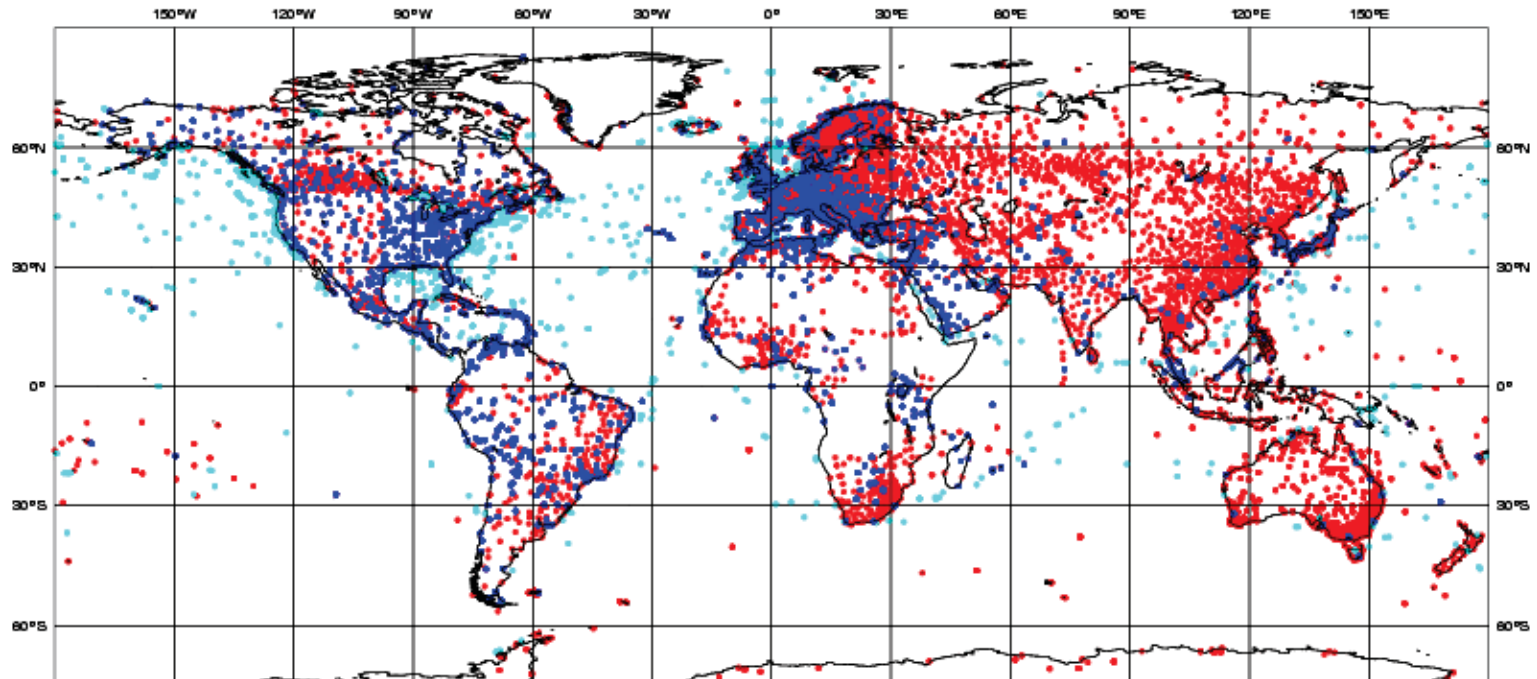
# Sources of data: stations

● 17092 SYNOP ● 2513 SHIP ● 12011 METAR

**ECMWF Data Coverage (All obs DA) - SYNOP/SHIP**

**20/JUL/2008; 12 UTC**

**Total number of obs = 31616**



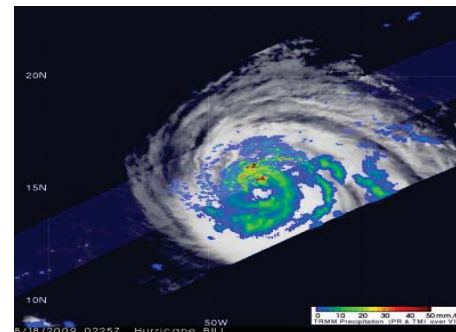
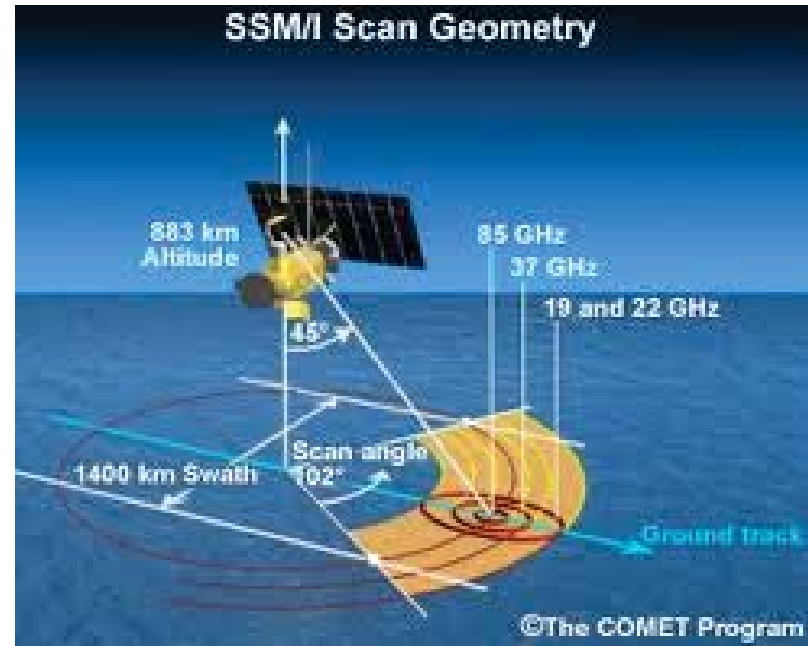
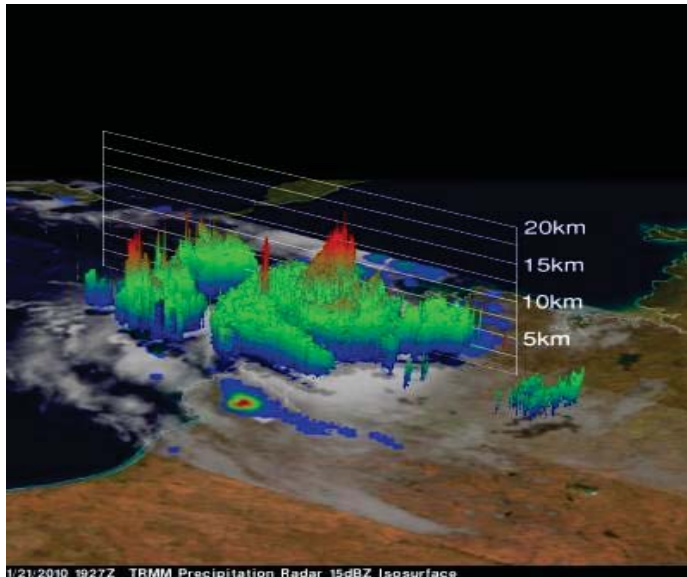
# Station Data: Advantages and Disadvantages

- ▶ Full array of variables
- ▶ Locally representative
- ▶ Not often available locally
- ▶ Potentially data gaps, handling of bad data
- ▶ Representativeness over complex terrain



# Satellite Data

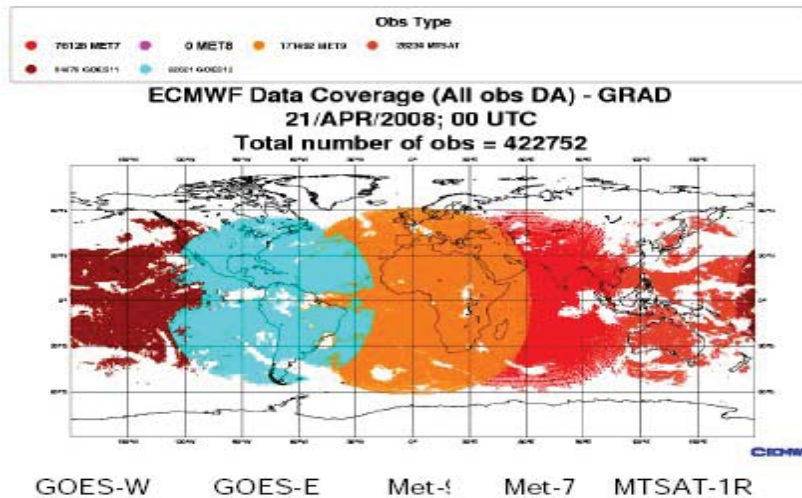
- ▶ Surface Temperature
- ▶ Precipitation
- ▶ Humidity
- ▶ Winds



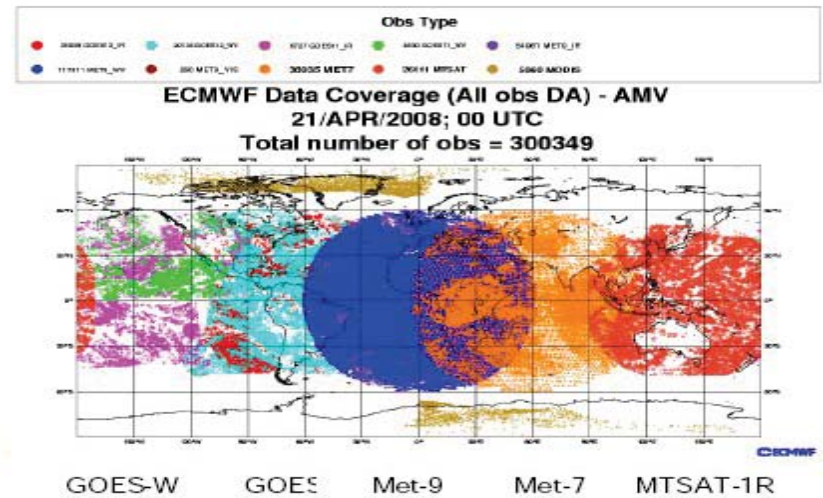


# For Satellite – coverage can be less of an issue (polar or geostationary – resolution, swathe, return times)

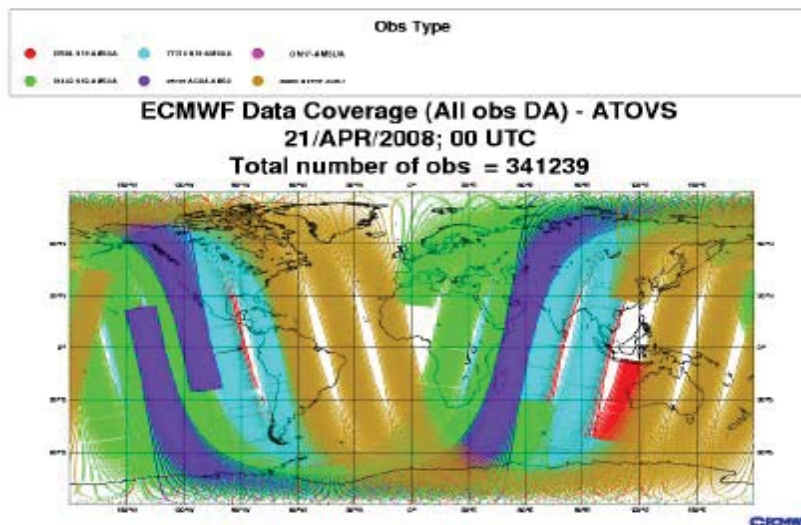
Clear-sky radiances



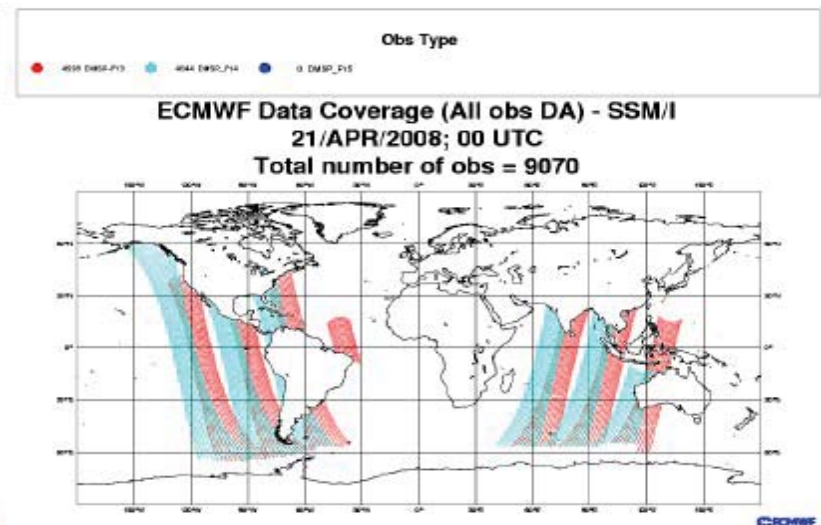
Atmospheric Motion Vectors



AMSU-A



SSM/I



# Satellite – advantages and disadvantages

- ▶ Good spatial and/or temporal coverage (depending on swathe, scan, orbit...)
- ▶ Only way to get regional information in conventional data-sparse regions
- ▶ Large uncertainties
- ▶ Temperature is skin temperature
- ▶ Problems for clouds, aerosols, insects etc
- ▶ Vertical resolution of atmospheric variables poor
- ▶ Problems in many retrievals mechanisms over land



# Wide Choice of retrievals: e.g. Precipitation

- ▶ GPCP – 1995 daily (1 deg), 1979 monthly (2.5 deg) – not real-time. Mix of IR and raingauge
- ▶ CMAP – Similar to GPCP – monthly only at 2.5deg
- ▶ CMORPH – 2003-present, realtime. 30 mins. based on microwave channels, using IR to provide temporal resolution. 25/8km.
- ▶ FEWS – daily, only over Africa, using gauge if nearby, otherwise combination of IR/microwave channels, 11km resolution. Realtime, 2000-present.
- ▶ TRMM – 25km resolution, 1998-present, 3 hourly.
  - ▶ 2A25 – precip radar product – not gridded
  - ▶ 3B42 – merged, radar, IR and microwave using gauge calibration – realtime

What to use?  
What is best?



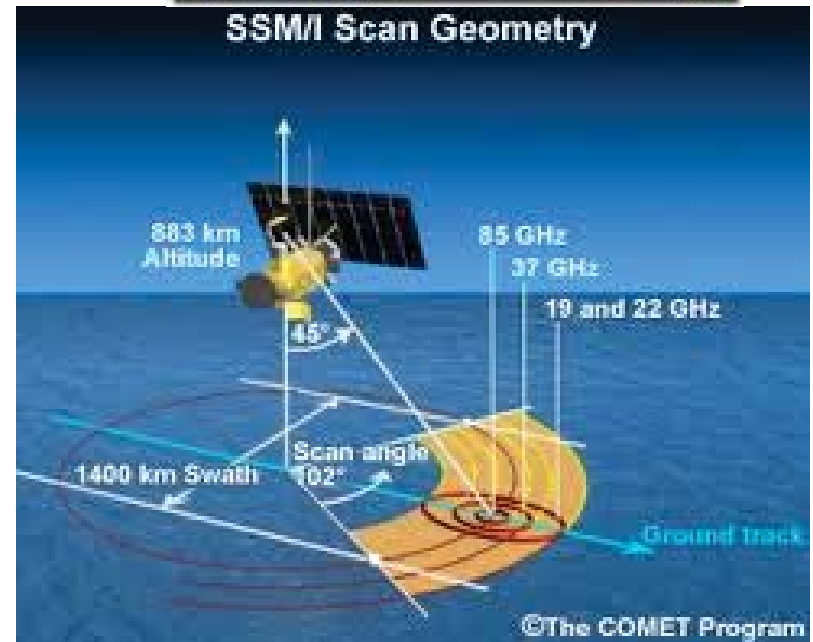


But some variables in contrast are difficult to get directly from Satellite

- **Surface temperature**: reliable over oceans using microwave. Some products over land, but uncertainty is large and not available daily
- **Winds**: reasonable over oceans using scatterometer data, surface winds over lands not possible. Upper level winds from feature tracking (cloud, humidity) but uncertainties high.
- **Humidity**: near surface only indirectly.
- Take home message: most (near) surface variables **over land** very difficult to infer from remote sensing

# Take home messages

- ▶ Station data are good where they exist, but they require careful treatment
- ▶ Satellite data useful for a regional view, but uncertainties are large, not all parameters are available, and the retrieval techniques are often obscure.



# A supplement source of climate information: **analysis** and **reanalysis**

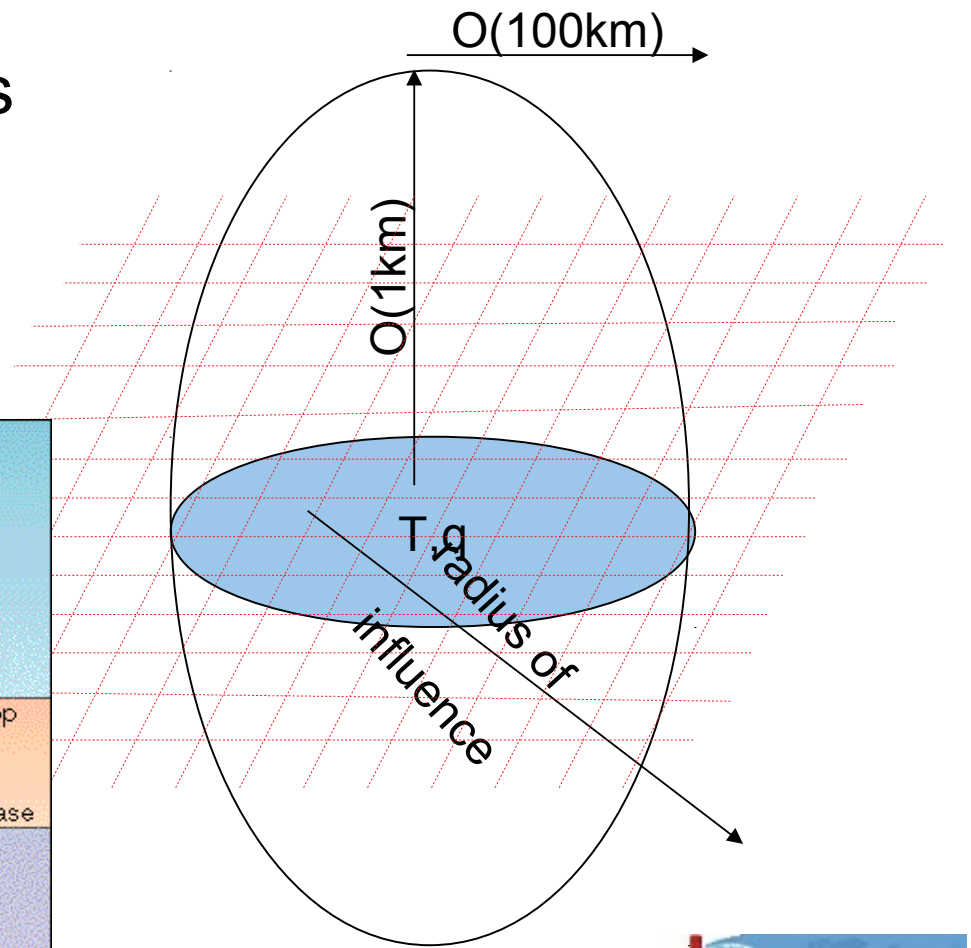
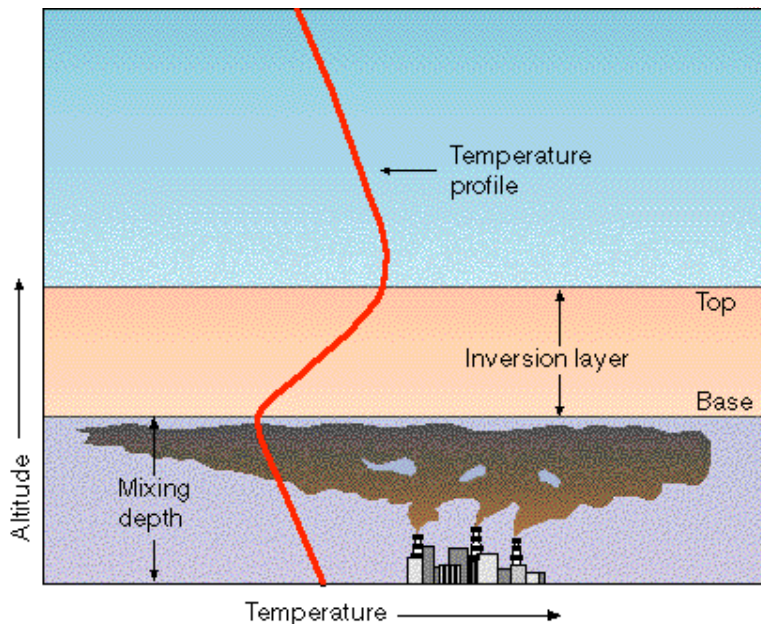
- To make forecasts of the future weather, knowledge of the present state is required
- This “picture” of the atmosphere needs to be “balanced” – Simple spatial and temporal interpolation of observations doesn’t work
- Hence the development of analysis systems

# Aim of “Data Assimilation” System

- ❑ To take a wide variety of variables (not necessarily model variables)...
- ❑ ...from a wide variety of instruments...
- ❑ ...with vastly different measurement densities...
- ❑ ...taking care to reject bad measurements...
- ❑ ...and combine them into an assessment of the atmospheric state, that is near balance with the forecast model “climate”
- ❑ Sounds Easy?

# Data assimilation

- ❑ Radius/distance of influence for each observation type needs to be defined
- ❑ Not obvious: e.g. Inversions, fronts etc.



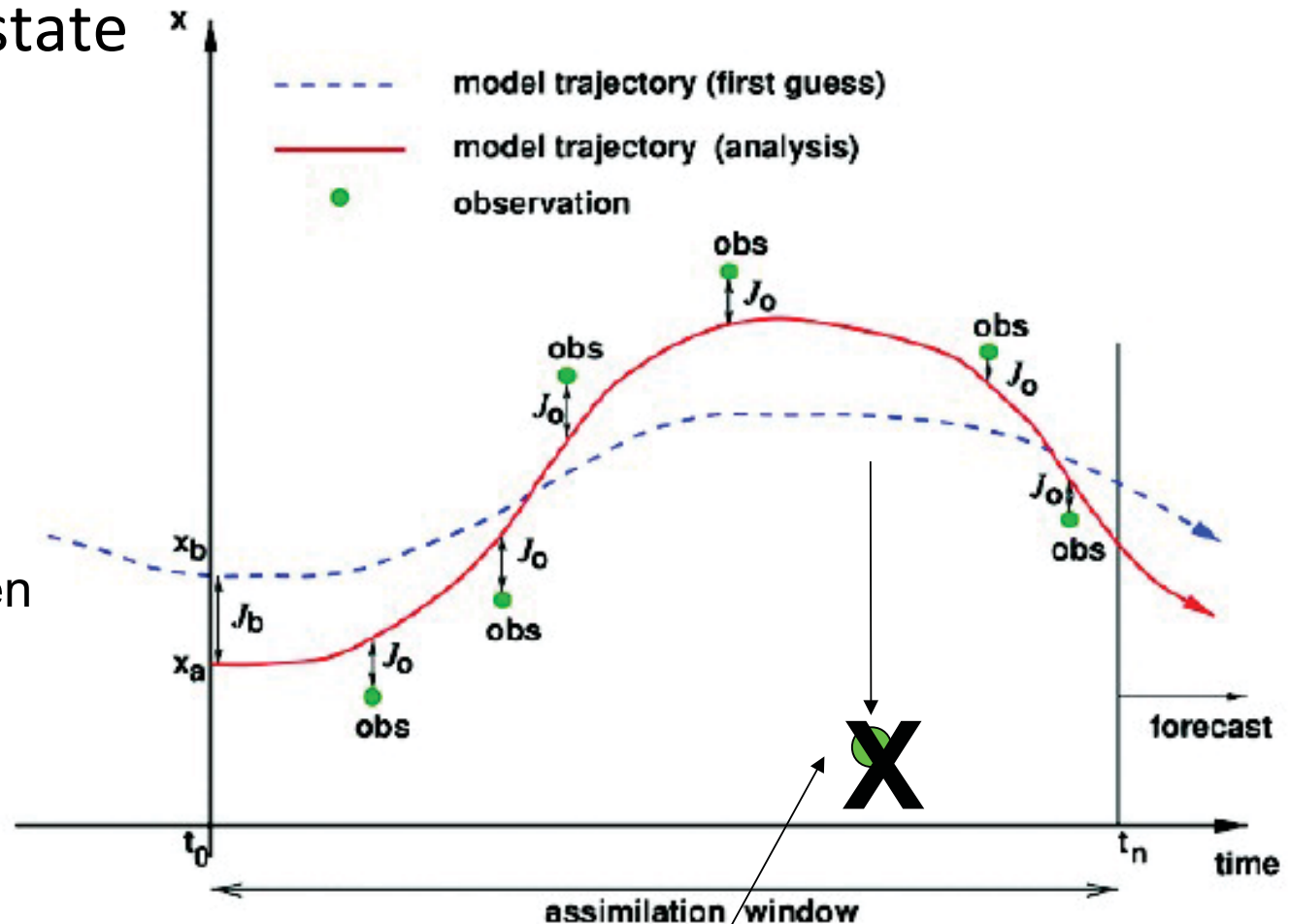


# Use of a forecast model is required to obtain balanced state

from Marta Janiskova

## 4DVAR assimilation

Goal: define atmospheric state  $x(t_0)$  such that the “distance” between the model trajectory and observations is minimum over a given time period

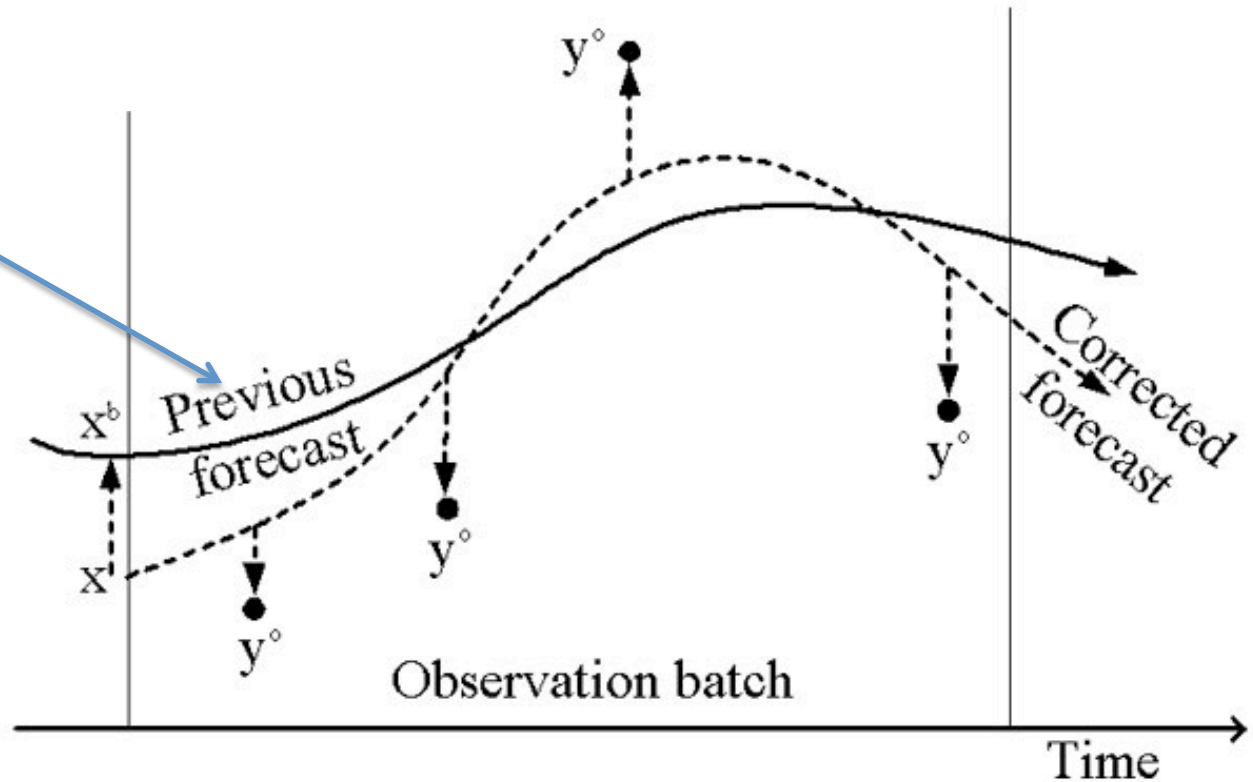


Note that the quality of the forecast model is important for a good analysis!

Observations “too far” from the background forecast are rejected as unreliable!!!

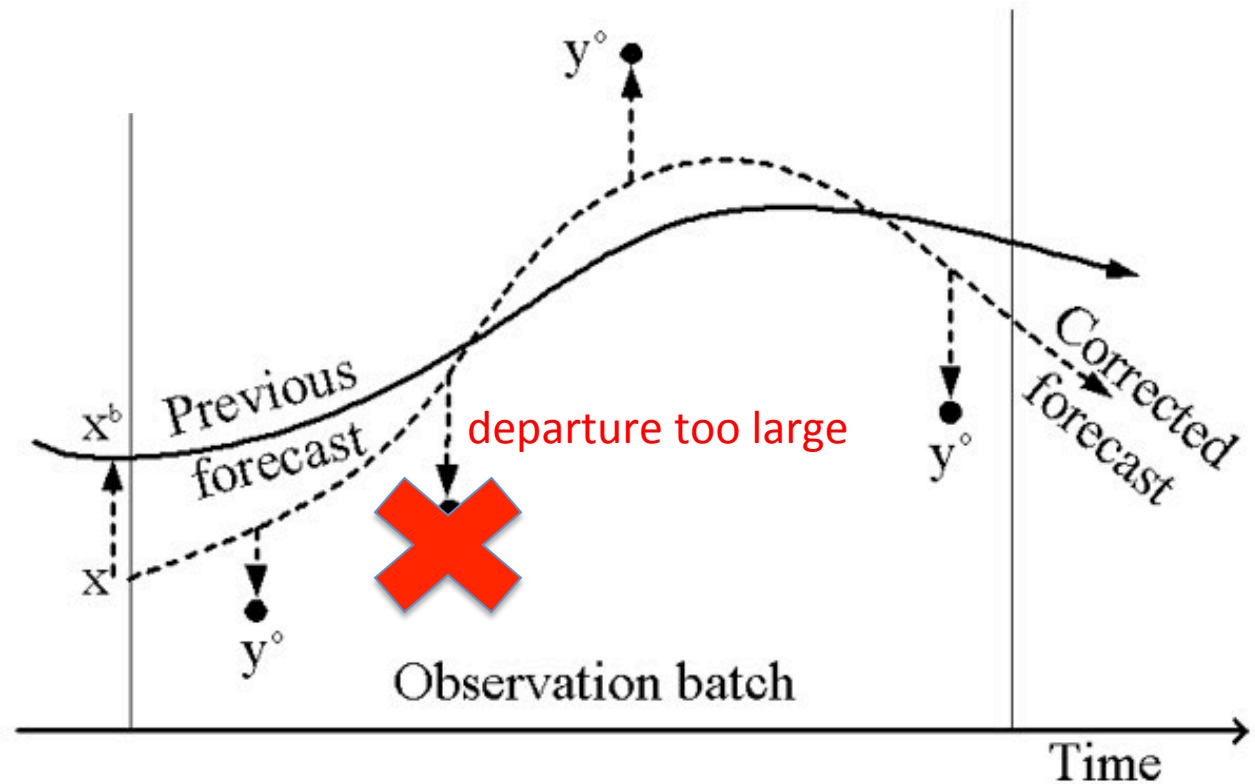
# recipe in a nutshell

1. Make a short forecast from previous “analysis”, call the “control”



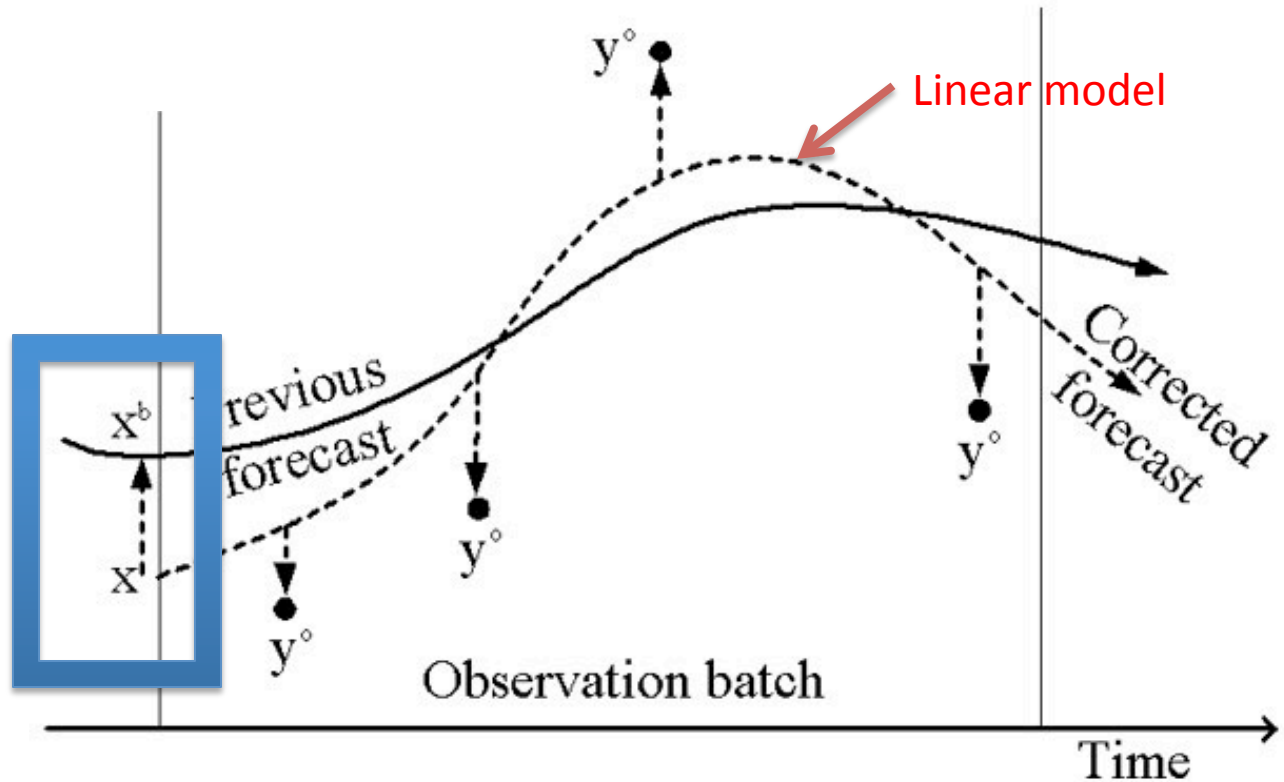
# recipe in a nutshell

2. Throw out  
“bad” data  
automatically



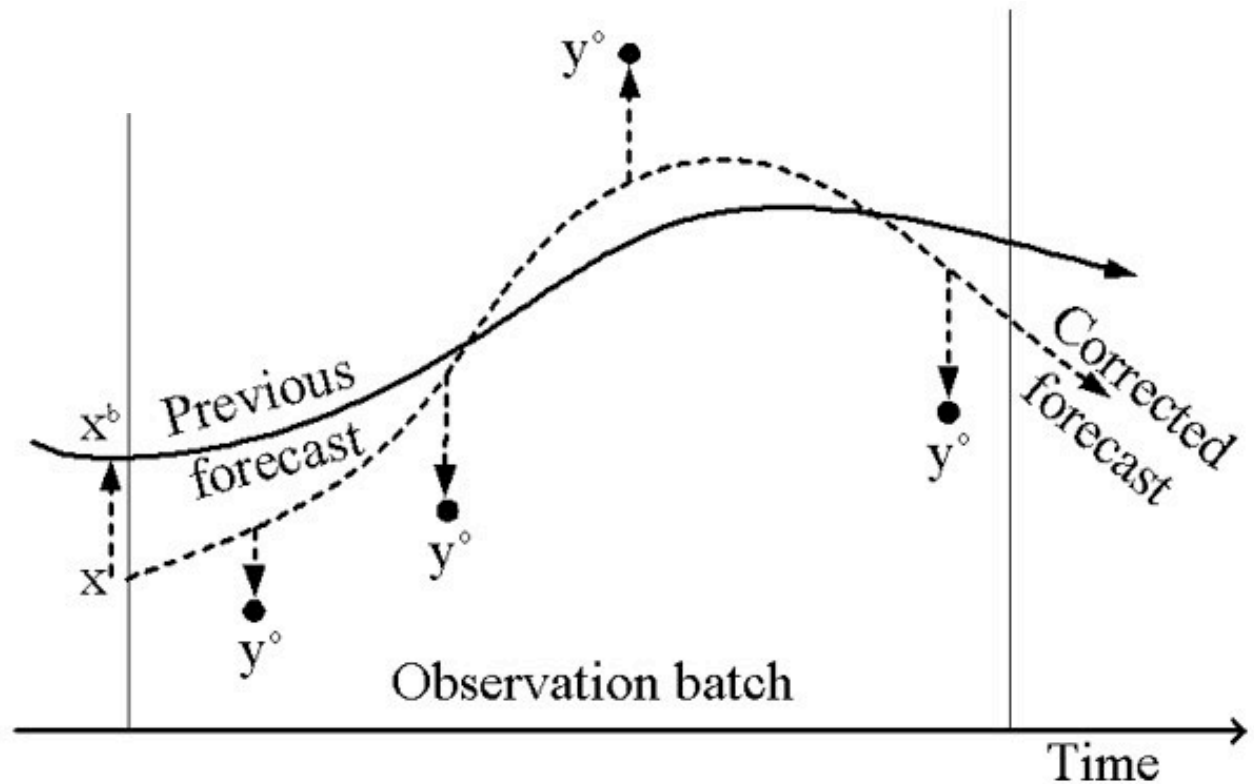
# recipe in a nutshell

3. Using a clever technique, find set of initial condition perturbations that minimize the departure of a revised **LINEAR** forecast from **both** the control and the set of “good” observations (translate model to observation space where necessary)



# recipe in a nutshell

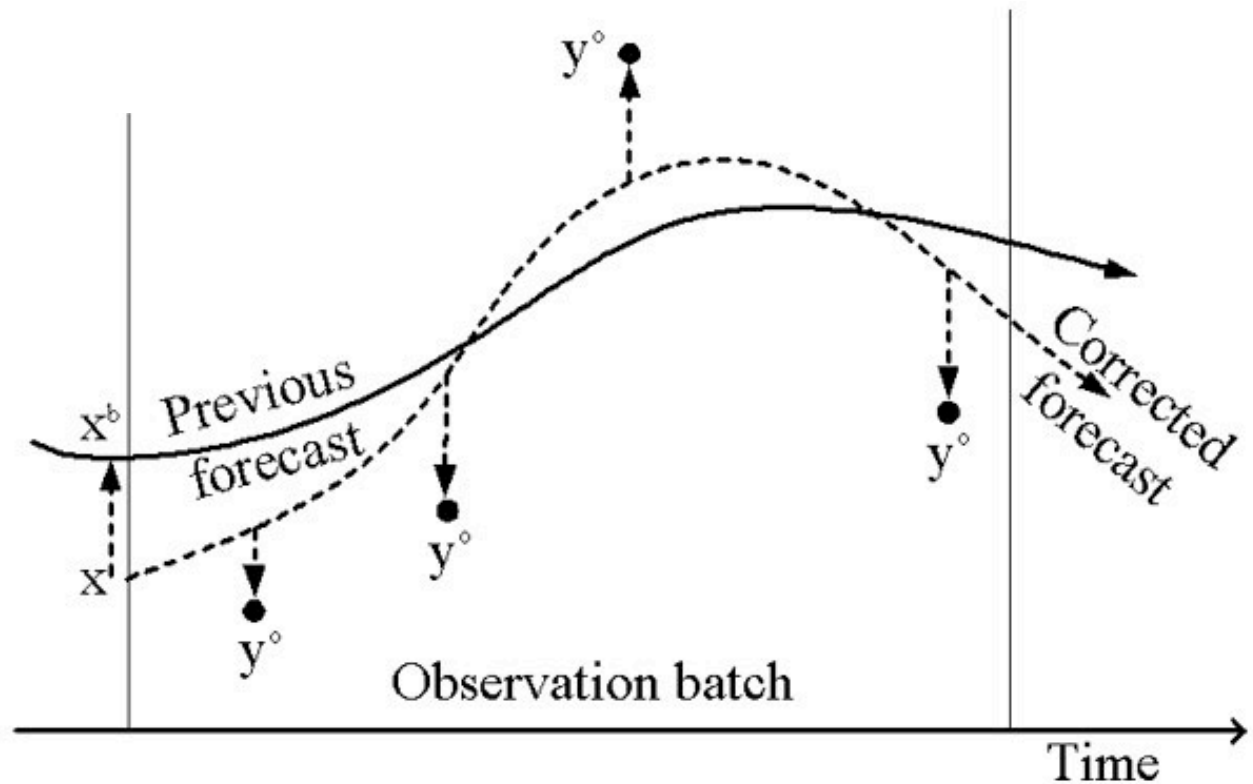
4. Perform revised “control” forecast starting from this new initial condition





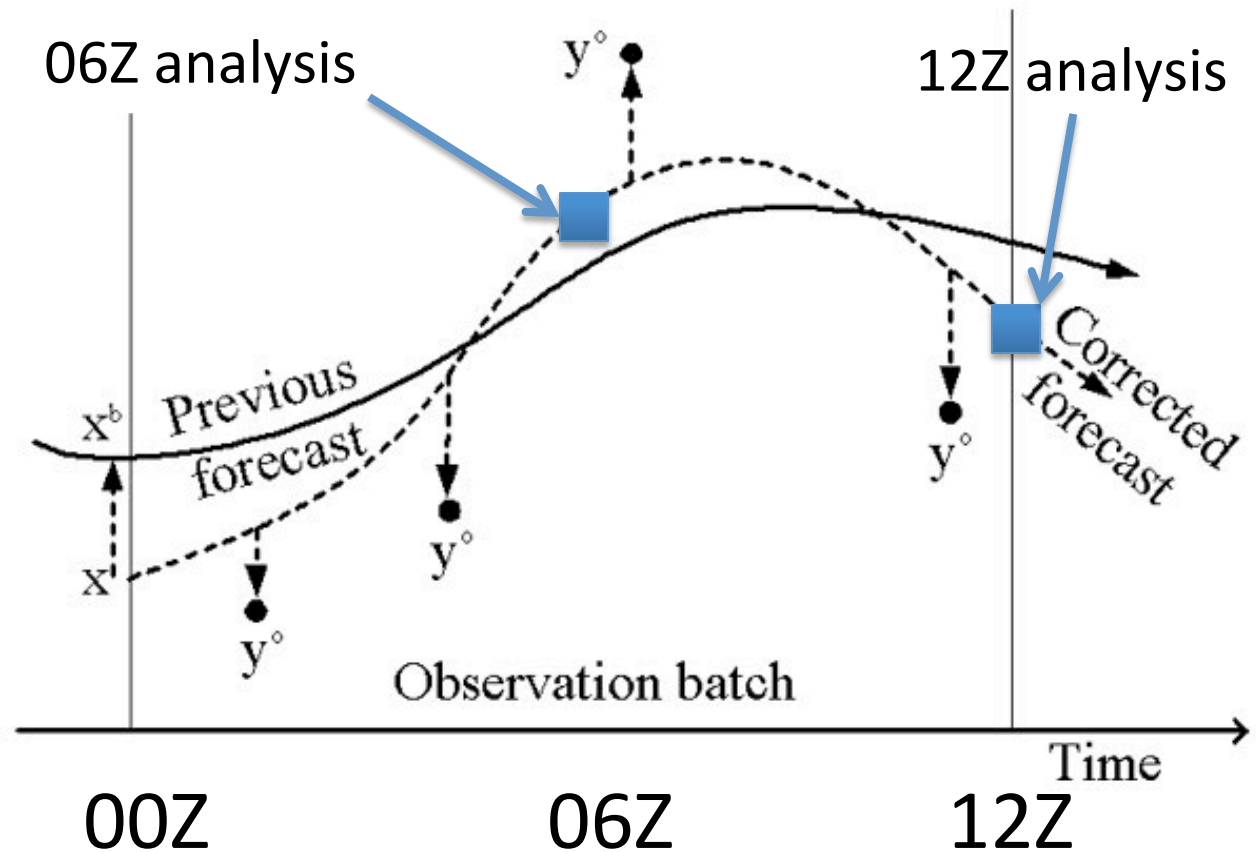
# recipe in a nutshell

5. Repeat  
step 1-4  
until (if!) the  
process  
converges  
(e.g. 3  
cycles)



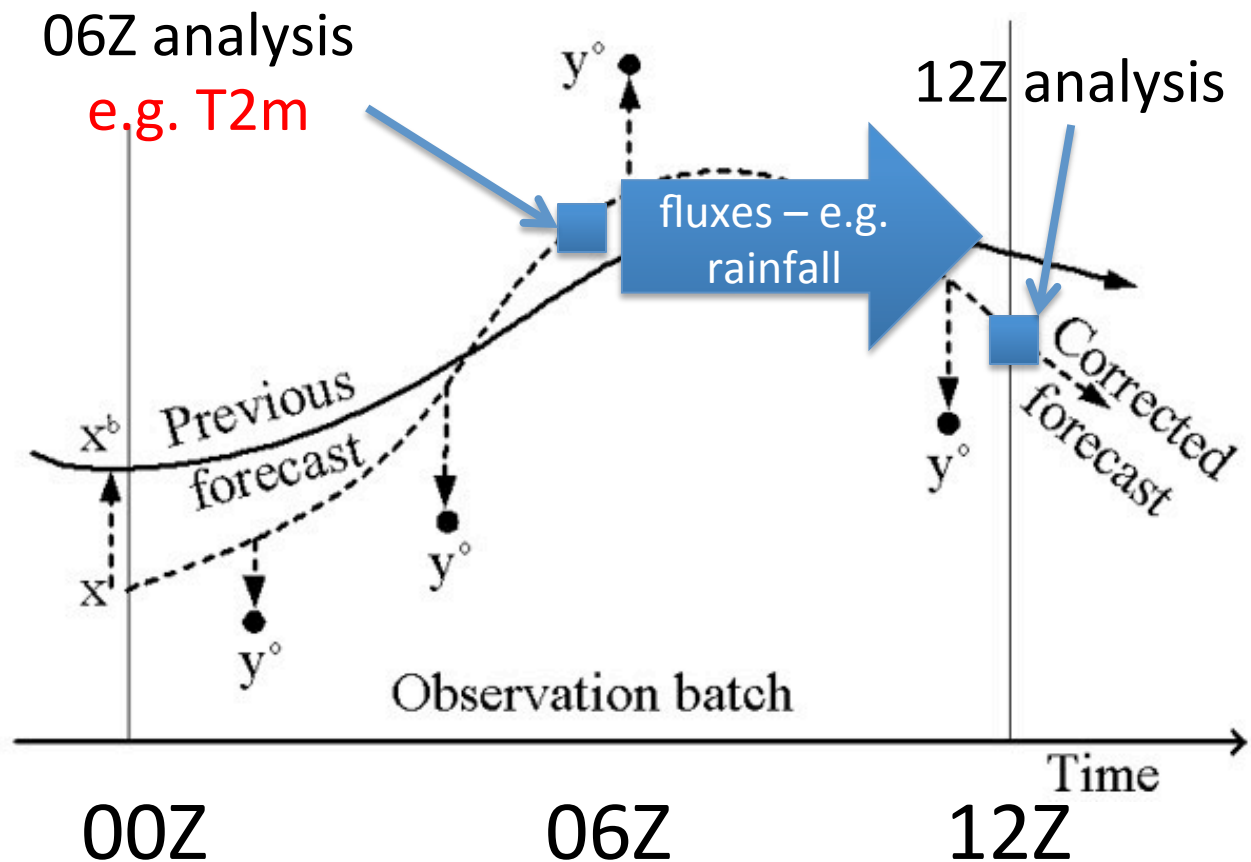
## recipe in a nutshell

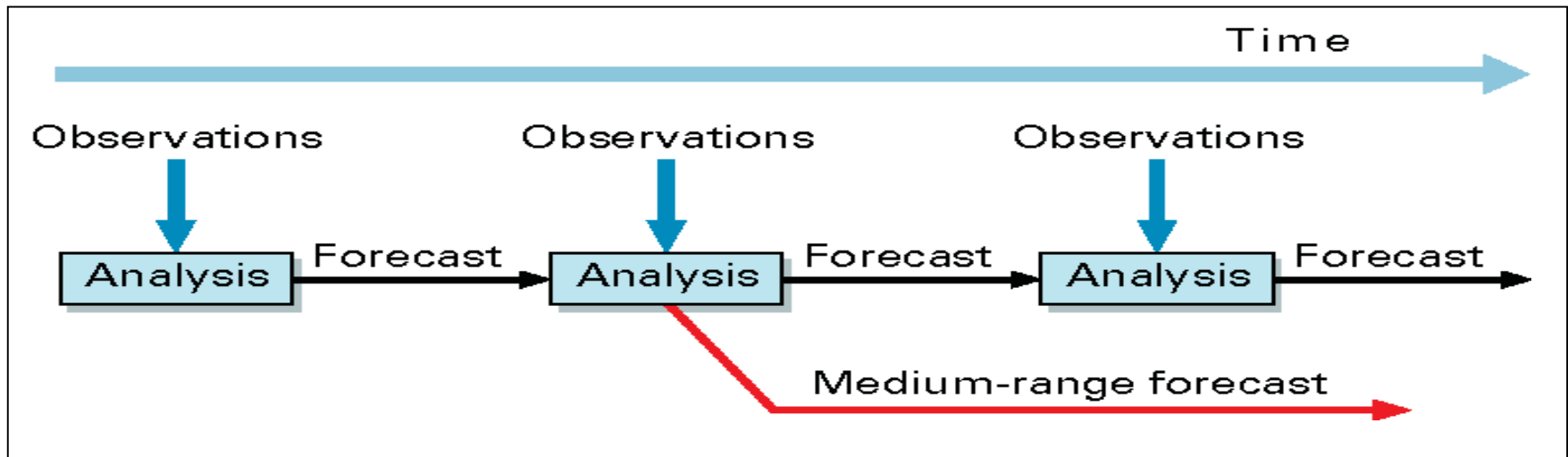
6. Now take any time point of the final “control” and use this as the “analysis”



# Fluxes and instantaneous fields

- Instantaneous fields such as temperature are taken from the “analysis”
- Pointless to do this for fluxes as can not calculate water and energy budgets – these are obtained from short-range forecasts (0-24hrs)





- Observations are used to “correct” errors in the short forecast from the previous analysis time.
- At ECMWF, every 12 hours 4 – 8,000,000 observations are assimilated

DATA USED: Pressure, humidity during day  
SYNOP T,q also used for soil moisture analysis

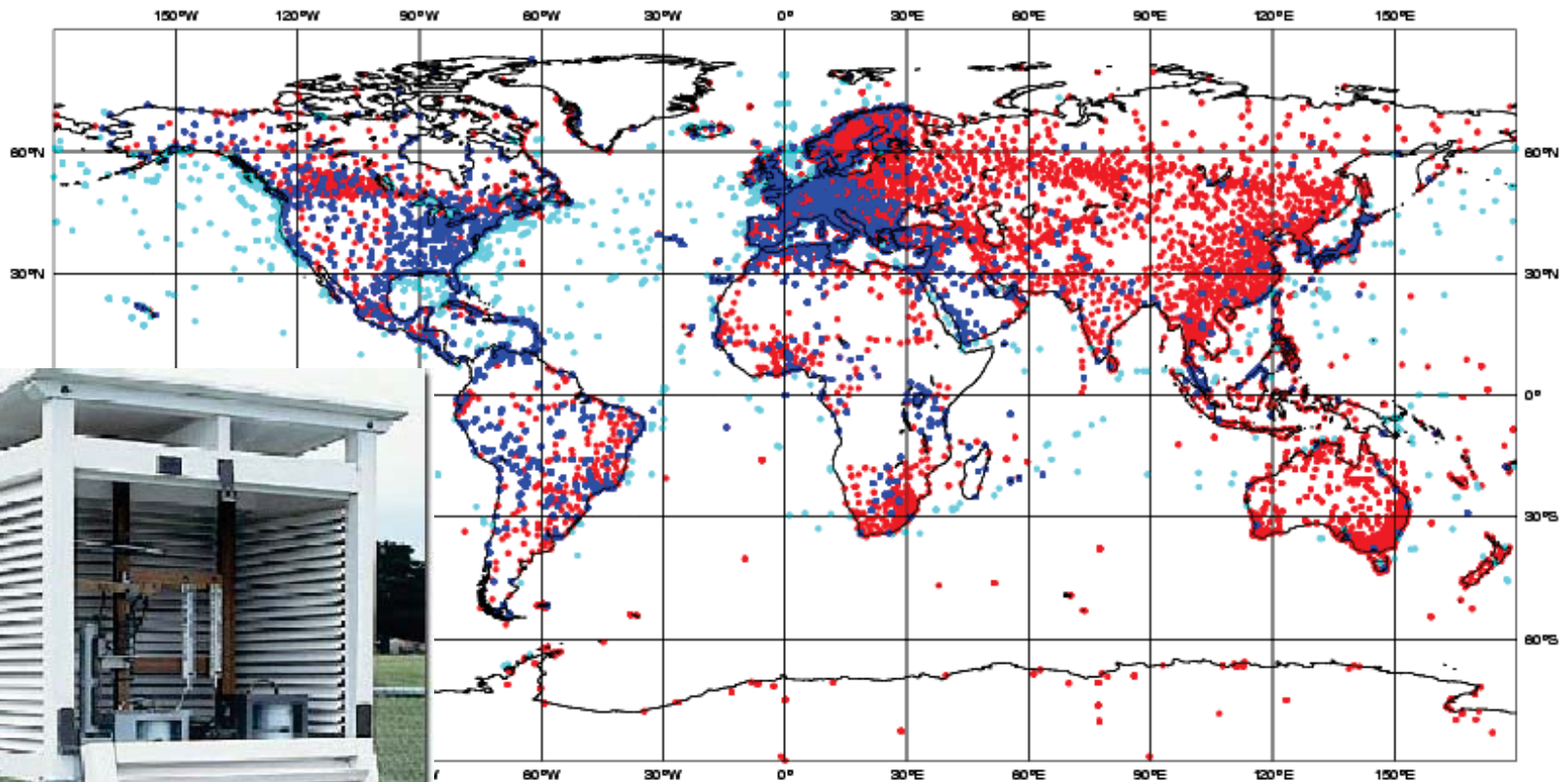
Obs Type

● 17092 SYNOP ● 2513 SHIP ● 12011 METAR

## ECMWF Data Coverage (All obs DA) - SYNOP/SHIP

20/JUL/2008; 12 UTC

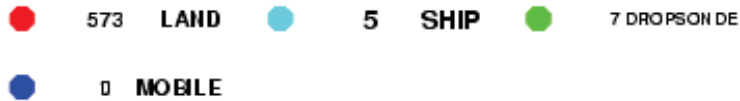
Total number of obs = 31616





DATA USED: T,q,u,v – humidity to 300 or 100hPa

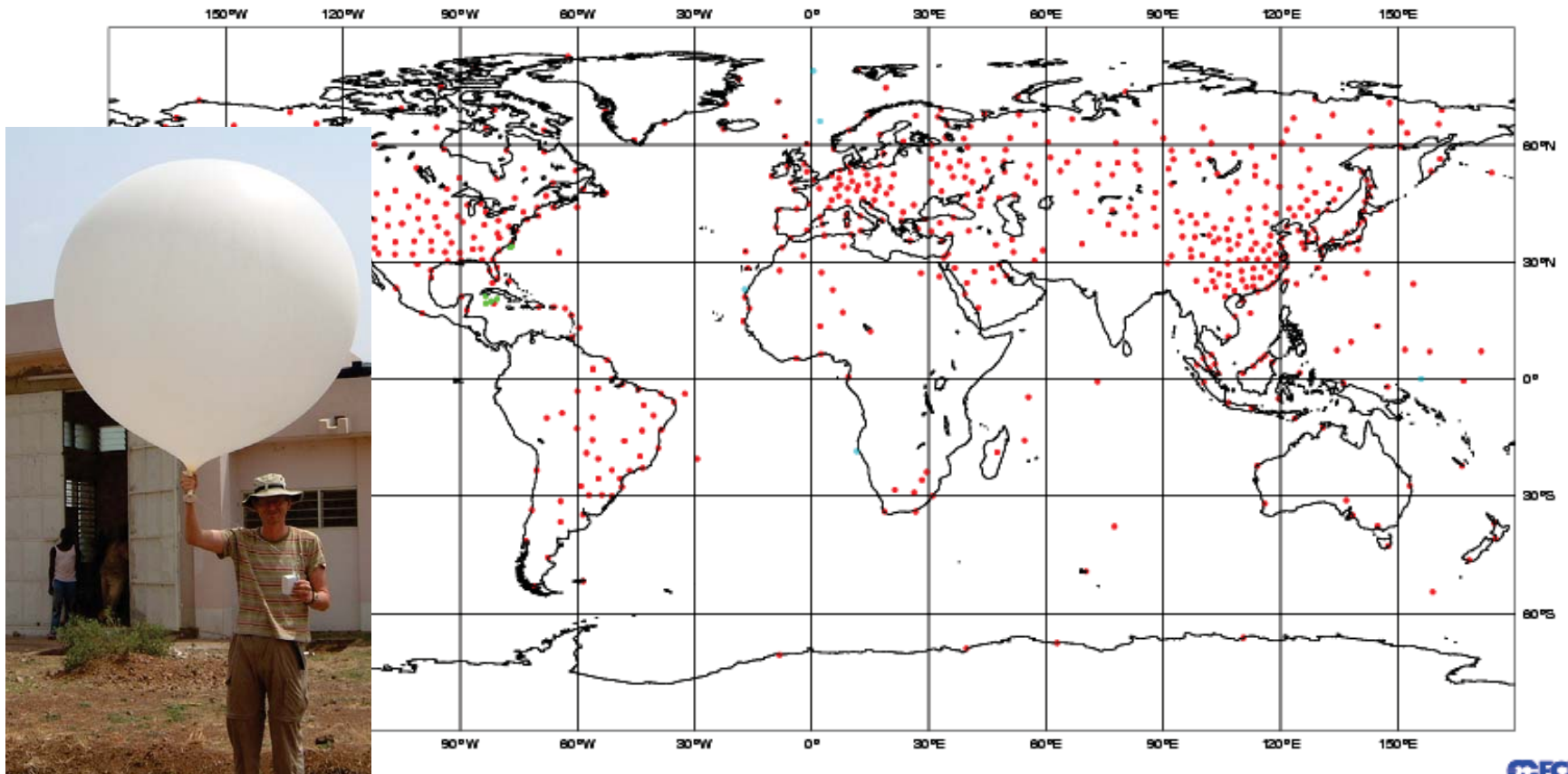
Obs Type



ECMWF Data Coverage (All obs DA) - TEMP

20/JUL/2008; 12 UTC

Total number of obs = 585





DATA USED: Temperature, winds  
(mozaic humidity research product)

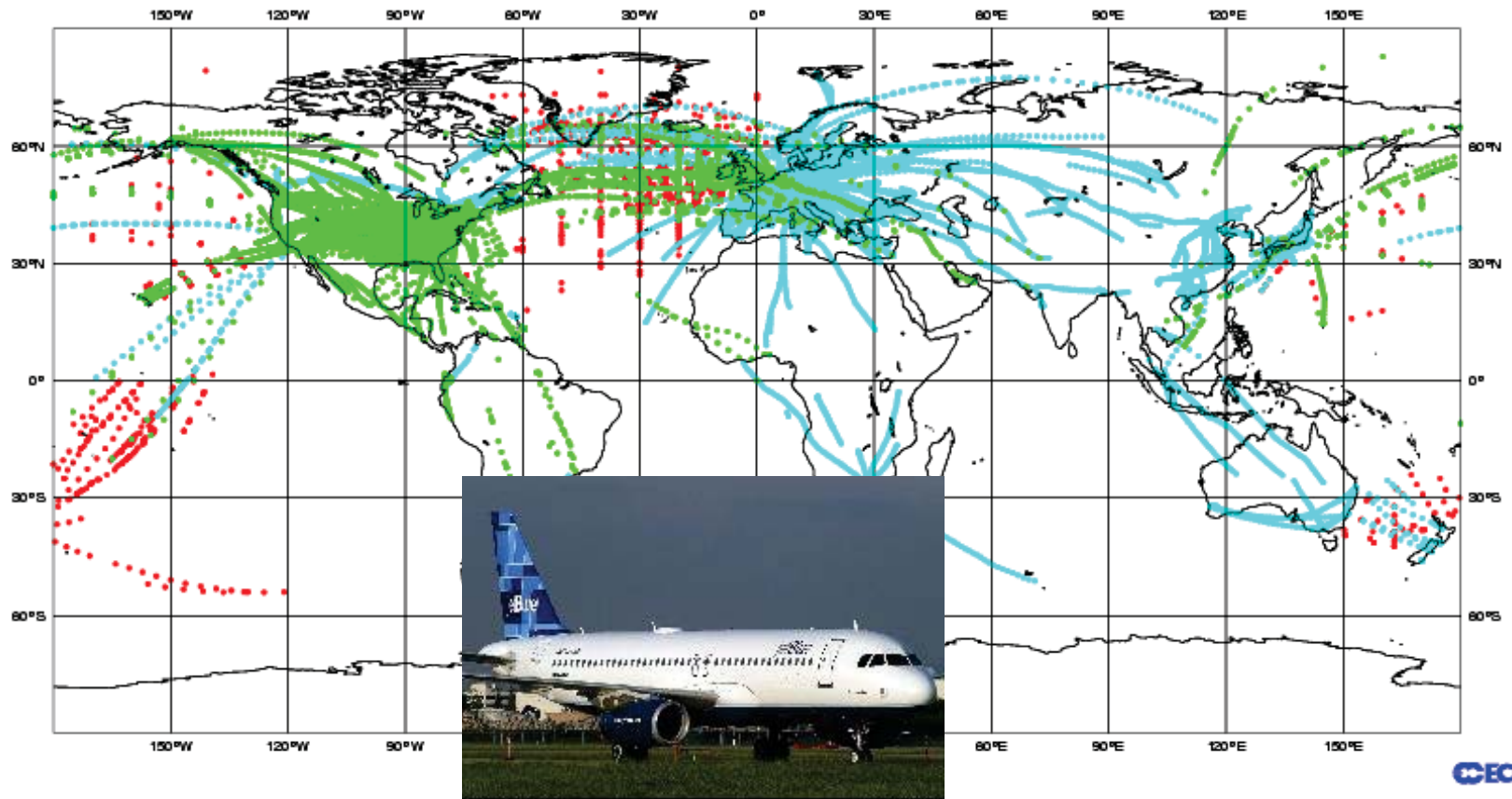
Obs Type

● 18035 ACARS

## ECMWF Data Coverage (All obs DA) - AIRCRAFT

20/JUL/2008; 12 UTC

Total number of obs = 50089

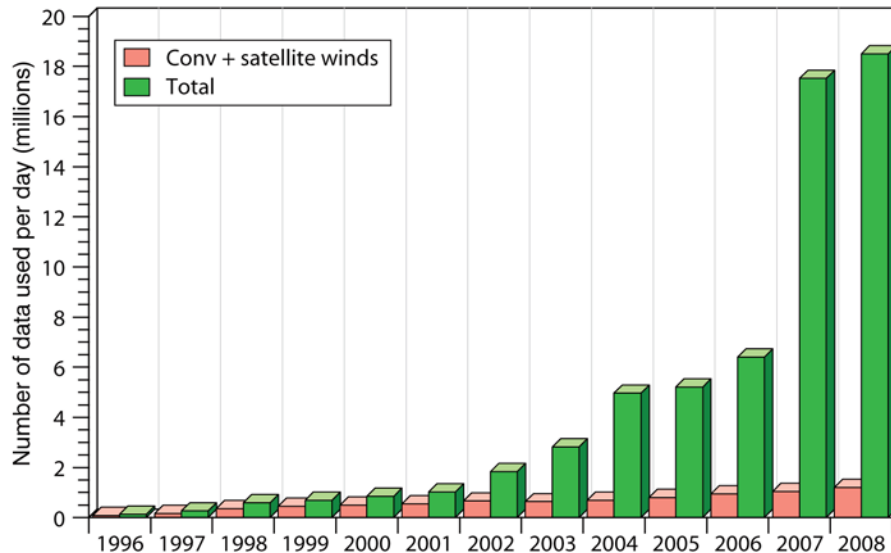
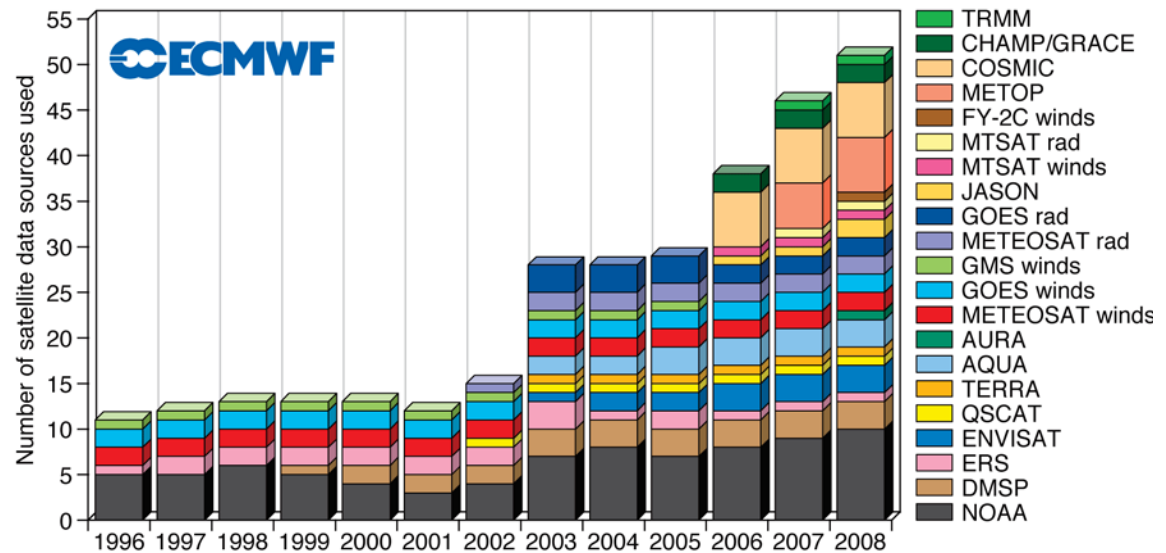


# Satellite data used at ECMWF

A key factor for the advance in NWP is increase availability of satellite data.

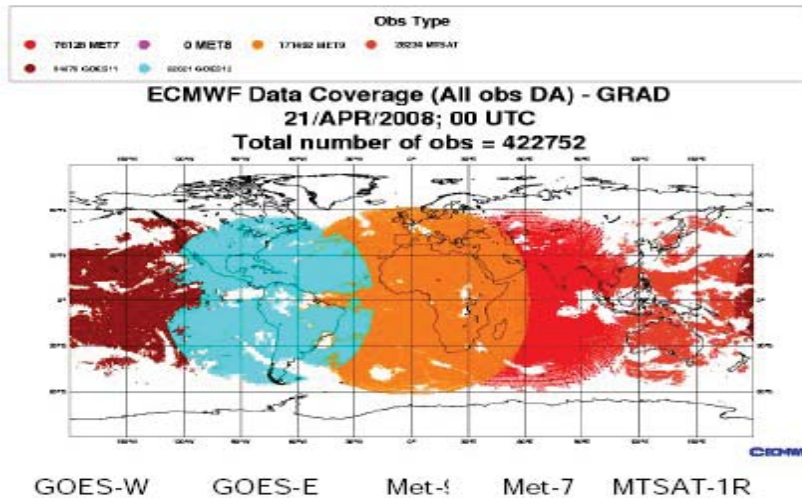
In 2008, ~ 300 million satellite observations from ~ 50 instruments have been received daily (top).

At ECMWF, ~ 6% of the available observations (~18 of the ~300 million) have been used daily (bottom).

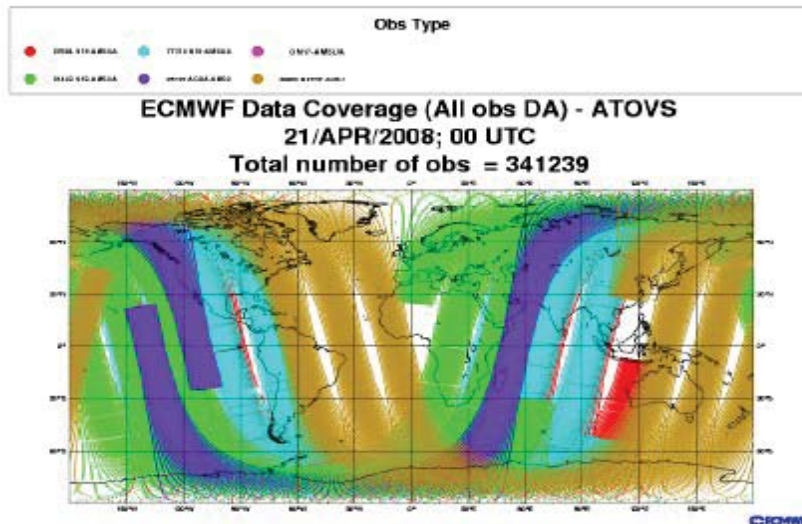




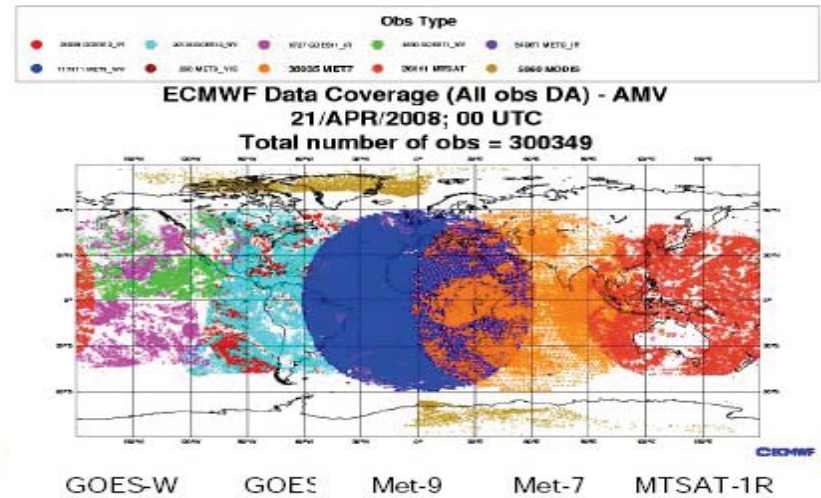
## Clear-sky radiances



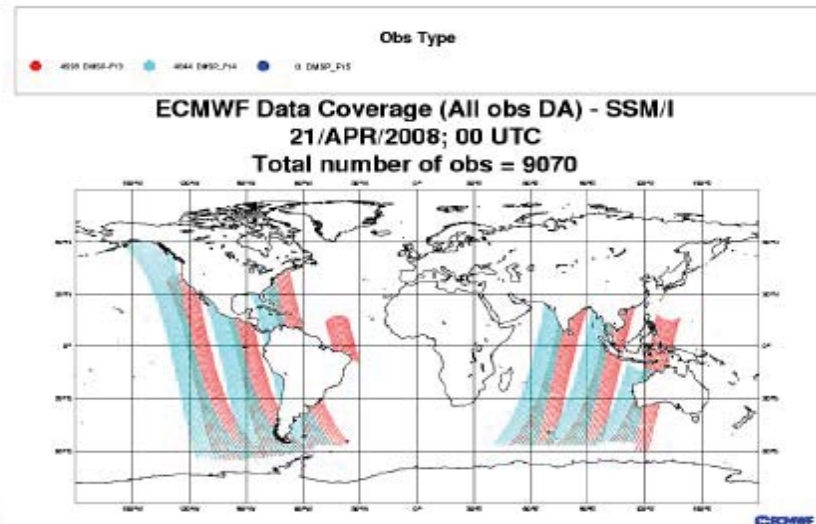
## AMSU-A



## Atmospheric Motion Vectors

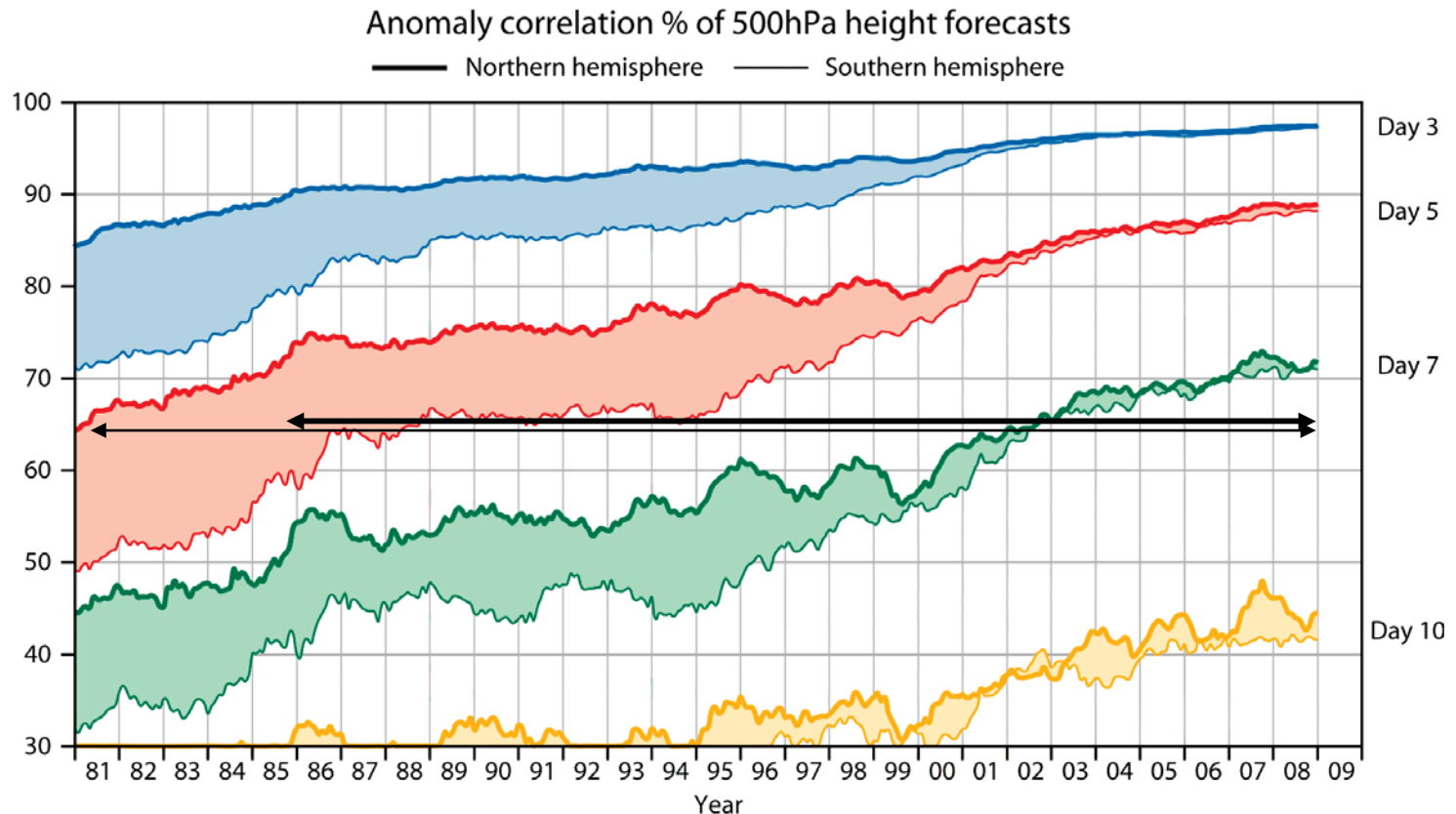


## SSM/I



# Evolution of ECMWF scores over NH and SH for Z500

- Over NH (SH) a day-7 single forecast of the upper-air atmospheric flow has the same accuracy as a day-5 in 1985 (day-3 in 1981).
- Note that Satellite data now implies equally good FC in NH and SH





## Some common misconceptions

- ❑ Very little information concerning clouds or precipitation is directly assimilated into the model
- ❑ Clouds in the analysis are a model product from the model physics, their location/properties determined by temperature, humidity and dynamics.
- ❑ Thus the parameters most important for impacts modelling (e.g. esp. temperature and precipitation) are all influenced by the model physics even in the analysis

# Advantages of analysis system

- all observations contribute to all variables
- Poor data can be automatically “sifted”

Example: Data denial experiments conducted over West Africa by Tompkins et al. 2003 QJRMS:

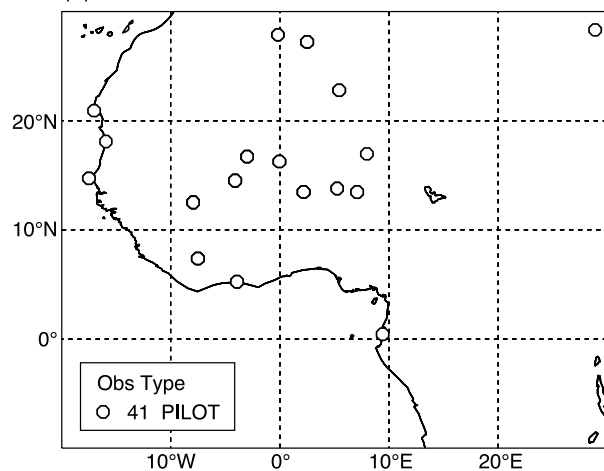
TABLE 2. DATA DENIAL EXPERIMENTS

Experiment	Data denied	Region
1	Radiosonde, pilot and aircraft	Local
2	Radiosonde, pilot and aircraft	Global
3	Satellite	Local
4	Surface SYNOP and drift sondes	Local
5	All wind information	Local

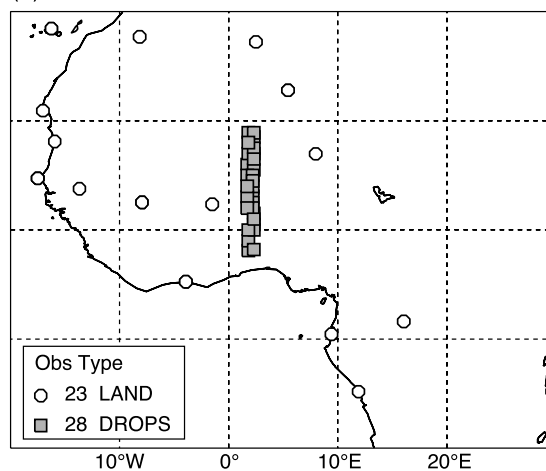
‘Local’ implies the region 0 to 30°N and from 30°W to 60°E.

# Observations assimilated in 2000

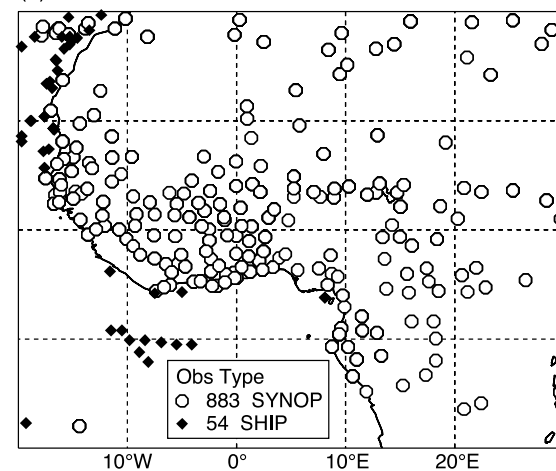
(a) PILOT/PROFILER



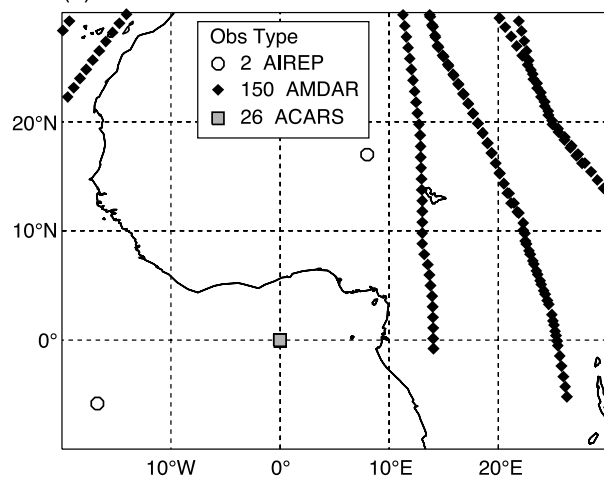
(b) TEMP



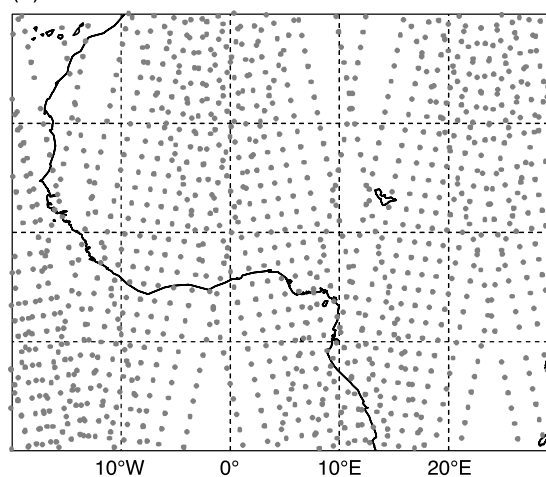
(c) SYNOP/SHIP



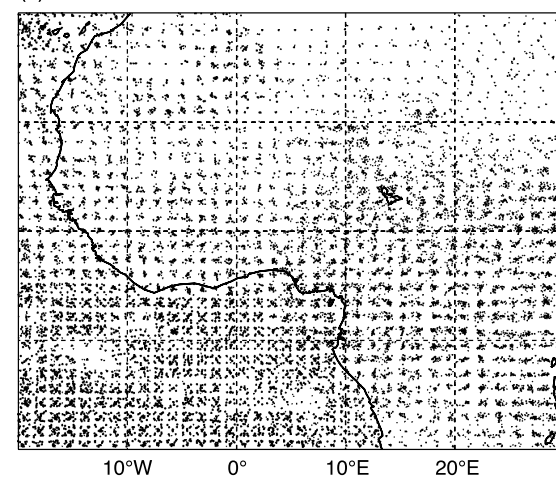
(d) AIRCRAFT



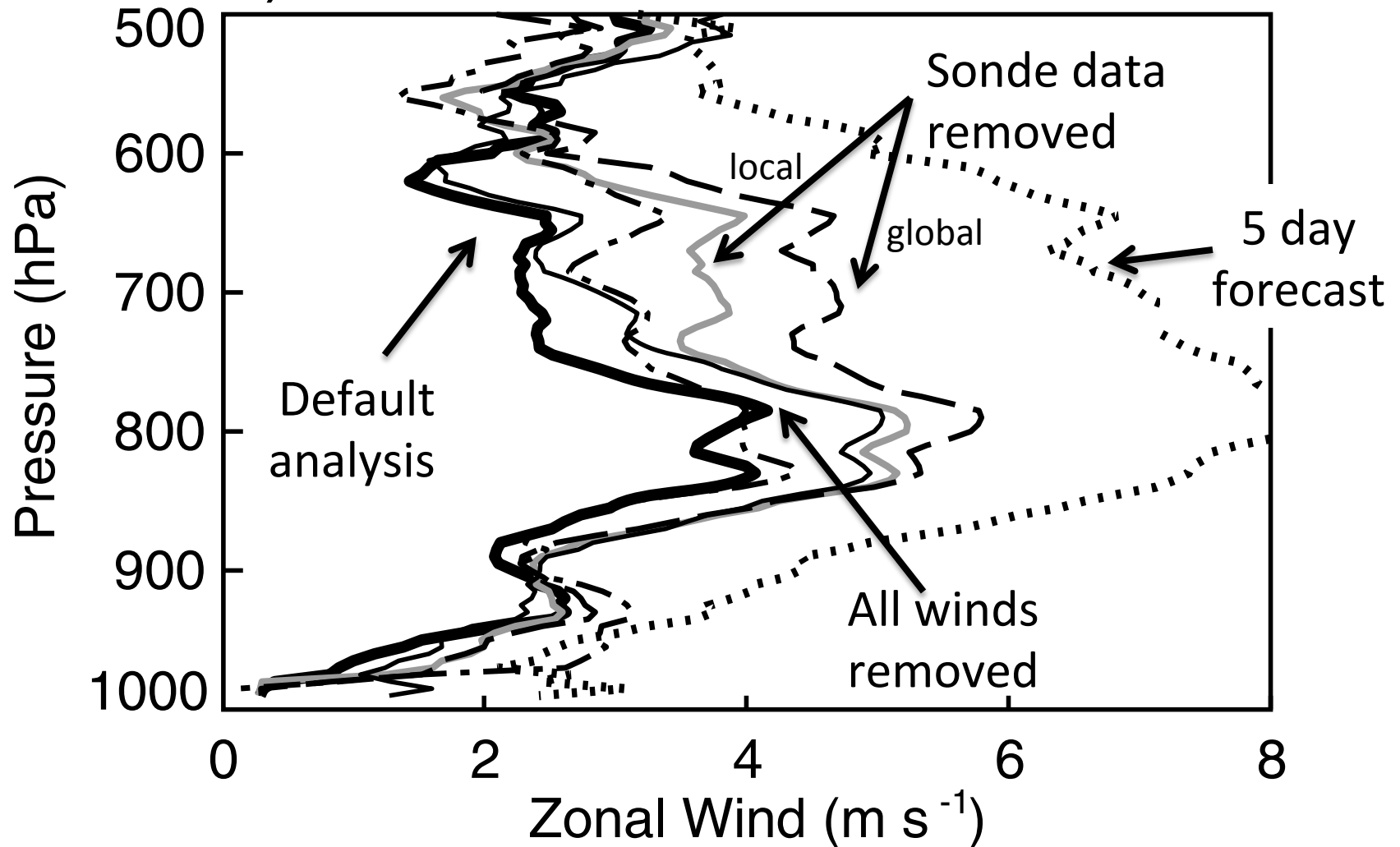
(e) ATOVS



(f) SATOB



# Root mean square wind errors – compared to independent data



Conclusion: Sonde **temperature** information more important for **wind** analysis than **winds**!

# To analysis or reanalysis – that is the question?

## **Analysis**

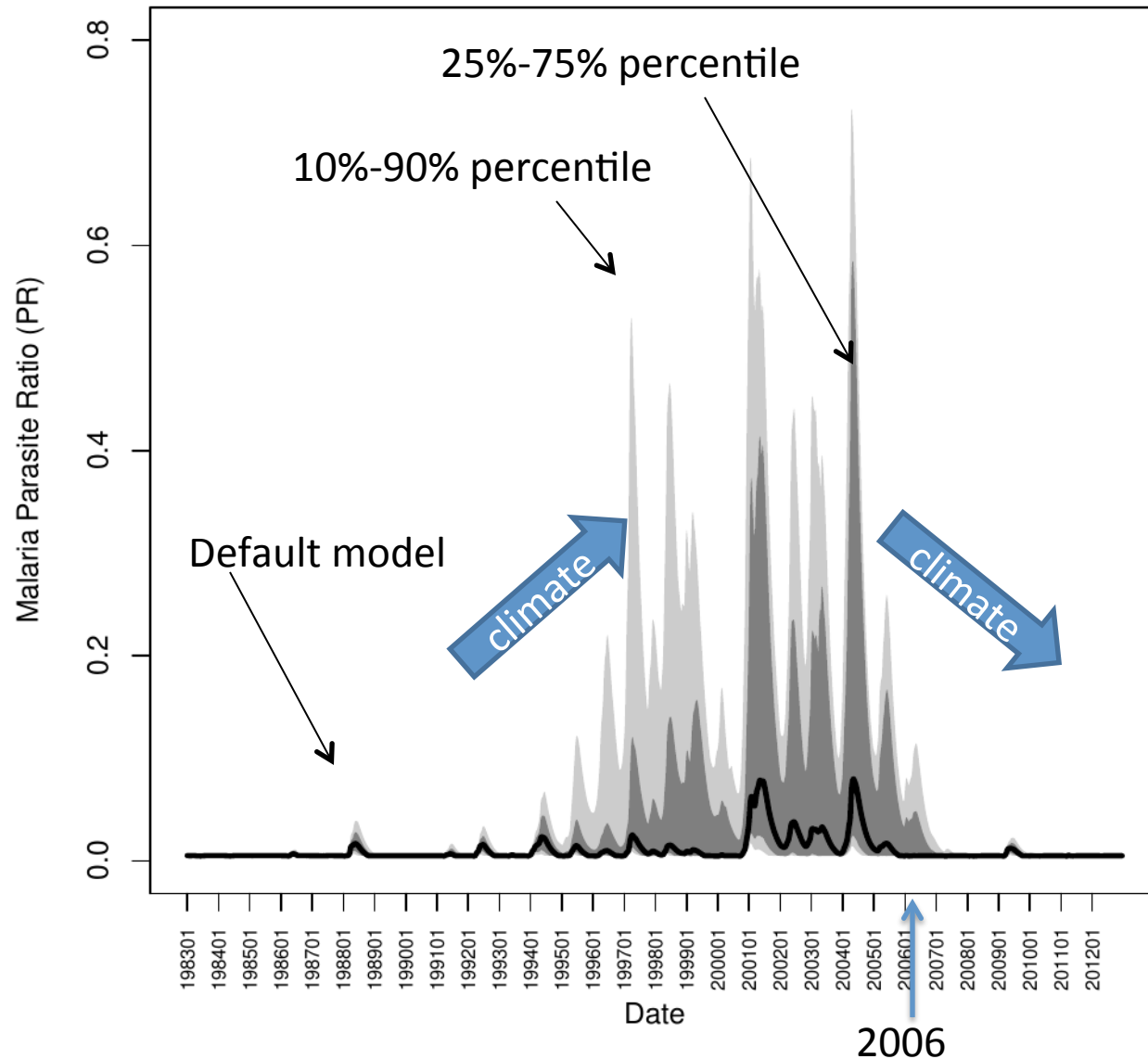
- ▶ Latest operational system
- ▶ High resolution
- ▶ Latest observation suite
- ▶ Model and observations change over time
- ▶ Not easily available
- ▶ Ideal for recent case study

## **Reanalysis (e.g. ERAI)**

- ▶ Using same model system, rerun for long period
- ▶ More continuity, although observations change over time
- ▶ Low resolution
- ▶ Obsolete model (ERAI from 10 years ago)
- ▶ Ideal for long term study

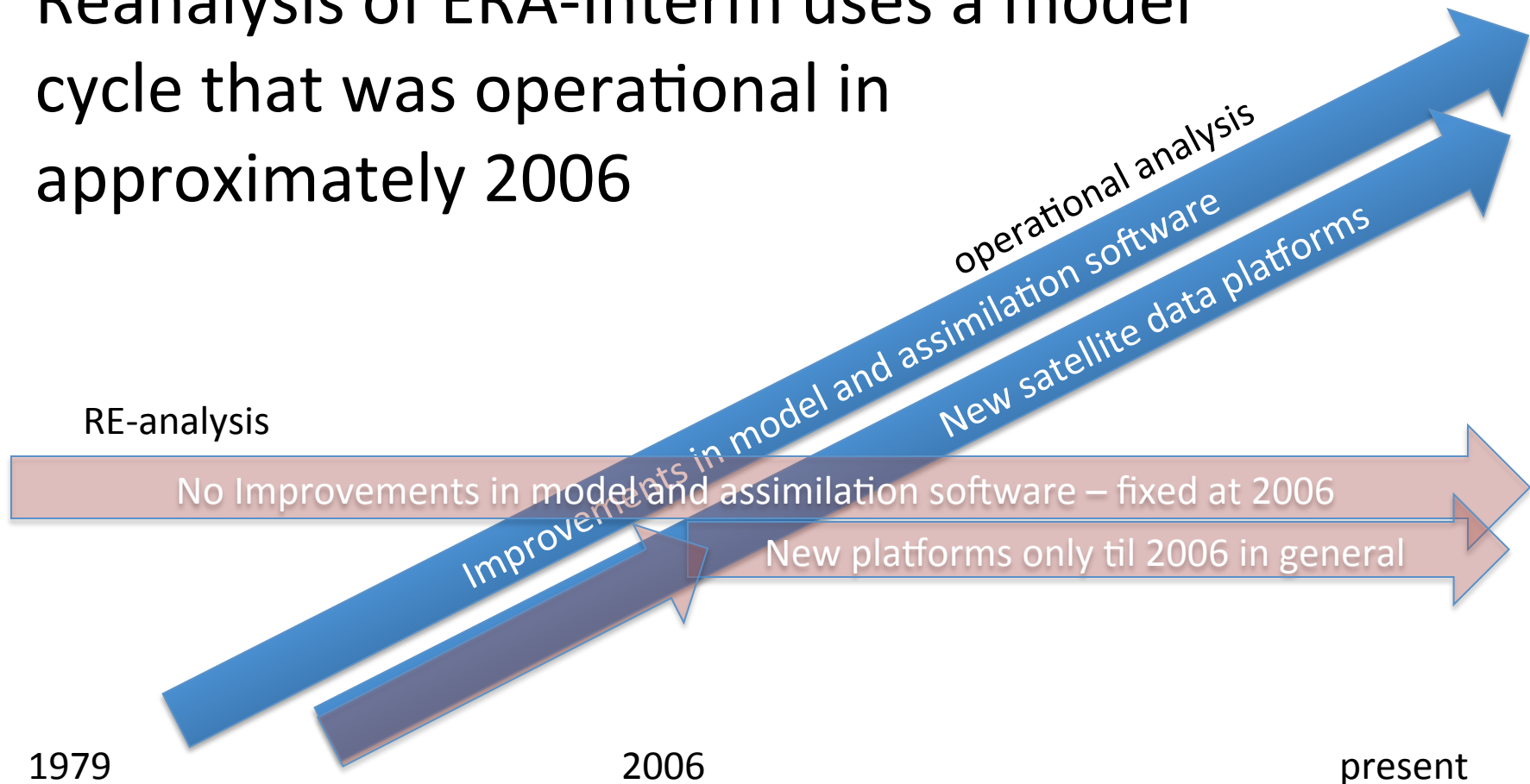
# VECTRI malaria simulations using FEWS ARC2 and ERA-Interim for Temperature

- Large uncertainty associated with the model – close to transmission limits
- Model indicates transmission would cease around 2006/2007 simply due to climate variability -



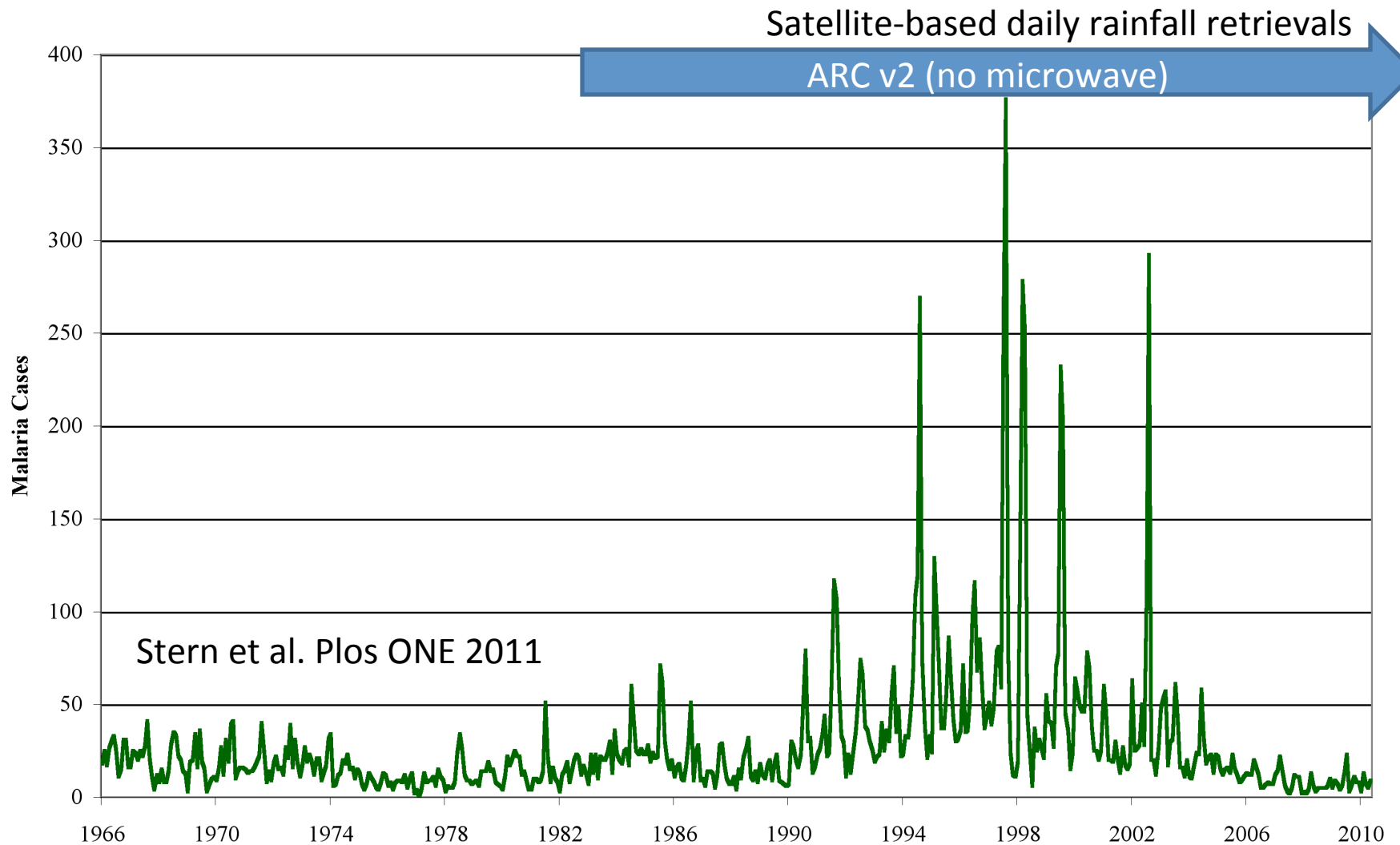
# To give you an idea

- Reanalysis of ERA-40 uses a model cycle that was operational in 2000
- Reanalysis of ERA-Interim uses a model cycle that was operational in approximately 2006





# Observed Cases for Kericho...



# Take home messages

- ▶ Analysis products are a useful supplement to observations
- ▶ Instantaneous fields are directly from the model analysis. Fluxes are from a short-range forecast.
- ▶ For recent case studies much better to use the analysis than reanalysis (higher resolution and better system)

[http://apps.ecmwf.int/datasets/data/interim\\_full\\_daily/](http://apps.ecmwf.int/datasets/data/interim_full_daily/)

- We will now try to download some fields using the reanalysis server
- and post-process them using cdo

