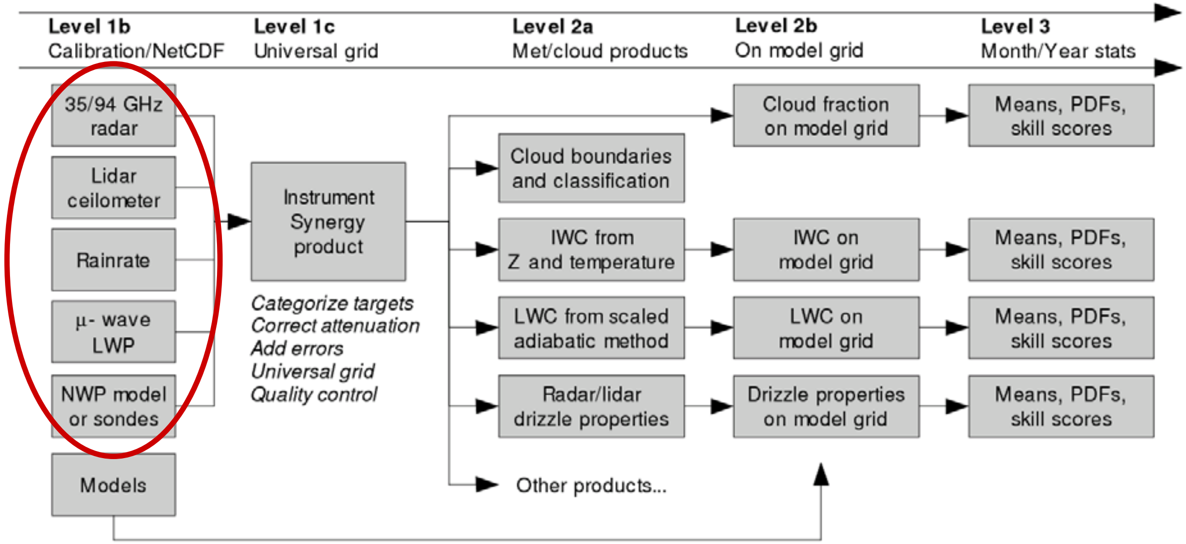
Towards an NWP-testbed

Ewan O'Connor and Robin Hogan
University of Reading, UK

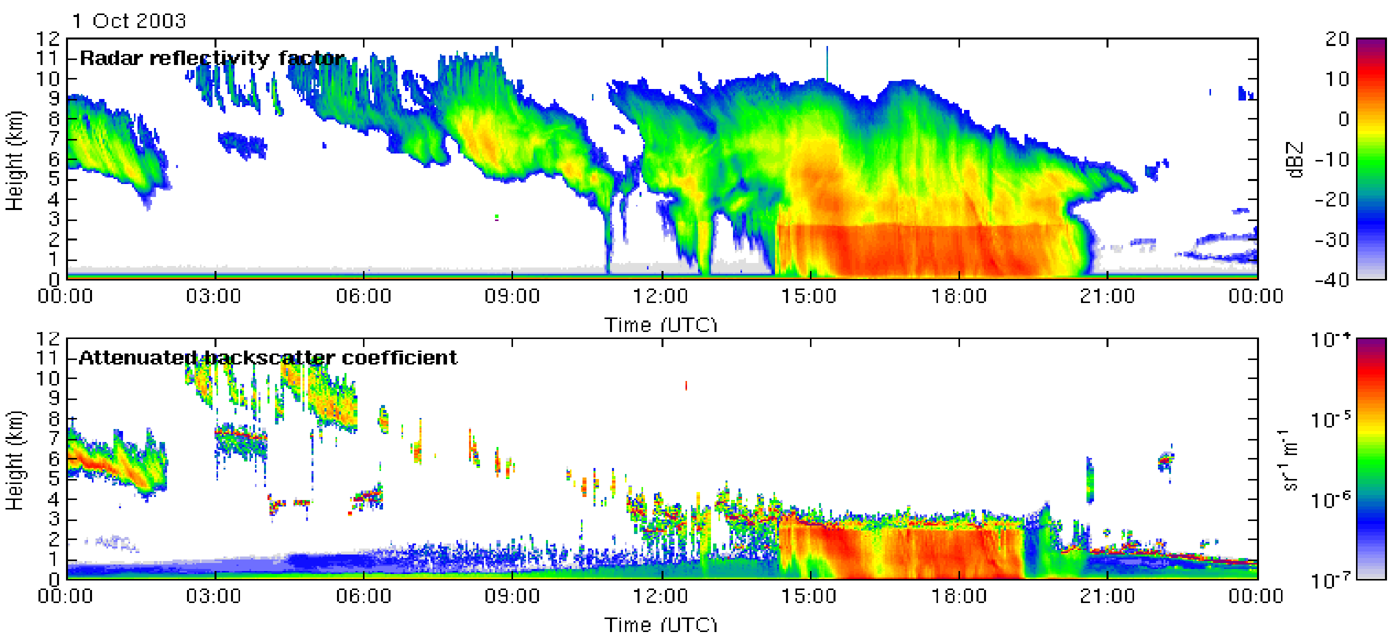
Overview

- Cloud schemes in NWP models are basically the same as in climate models, but easier to evaluate using ARM because:
  - NWP models are trying to simulate the actual weather observed
  - They are run every day
  - In Europe at least, NWP modelers are more interested in comparisons with ARM-like data than climate modelers (not true in US?)
- But can we use these comparisons to improve the physics?
  - Can compare different models which have different parameterizations
  - But each model uses different data assimilation system
  - Cleaner test if the setup is identical except one aspect of physics
  - SCM-testbed is the crucial addition to the NWP-testbed
- How do we set such a system up?
  - Start by interfacing Cloudnet processing with ARM products
  - Metrics: test both *bias* and *skill* (can only test bias of climate model)
  - Diurnal compositing to evaluate boundary-layer physics



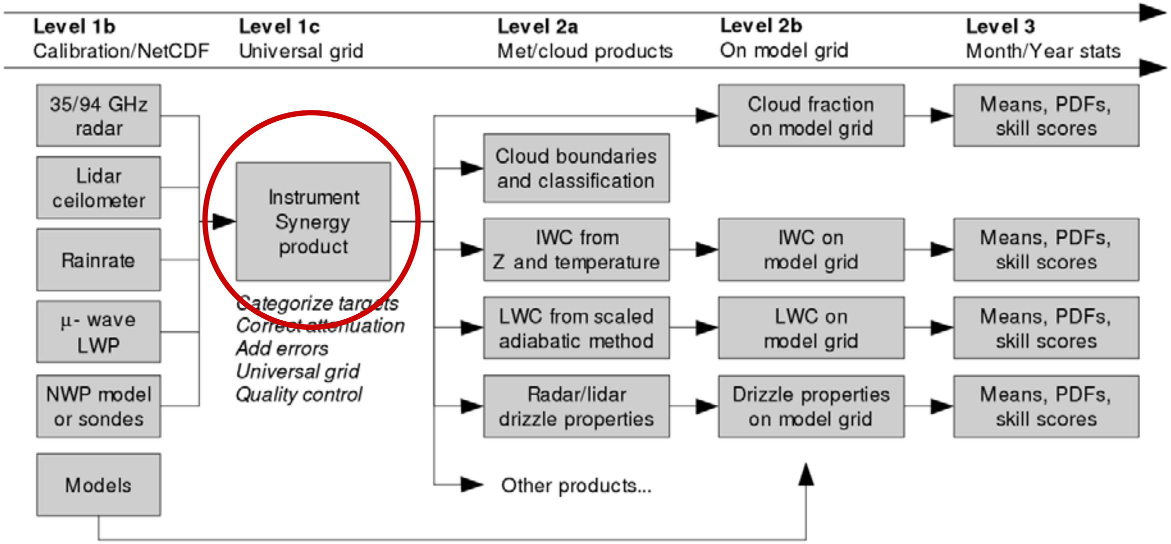
Level 1b

- Minimum instrument requirements at each site
  - Cloud radar, lidar, microwave radiometer, rain gauge, model or sondes



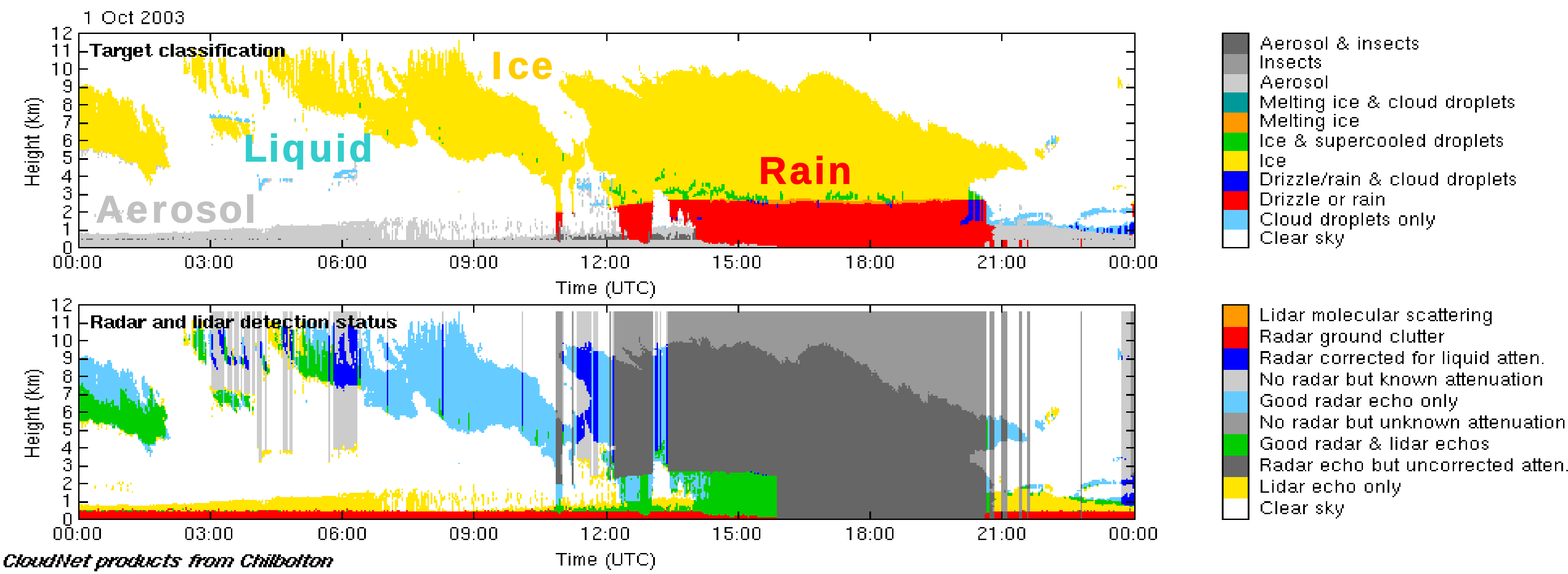
Radar

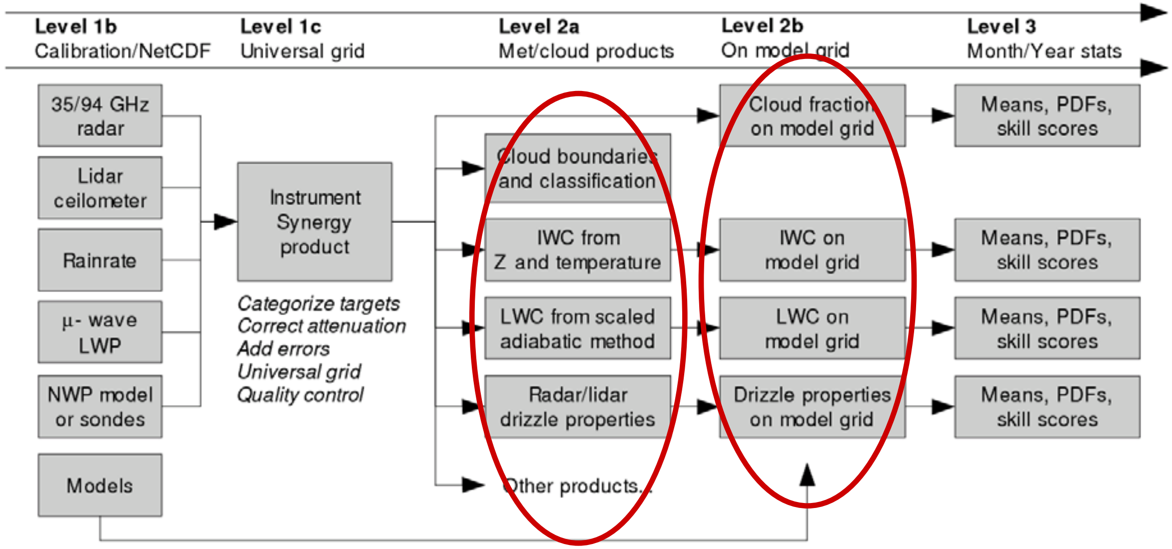
Lidar



Level 1c

- Instrument Synergy product
  - Example of target classification and data quality fields:

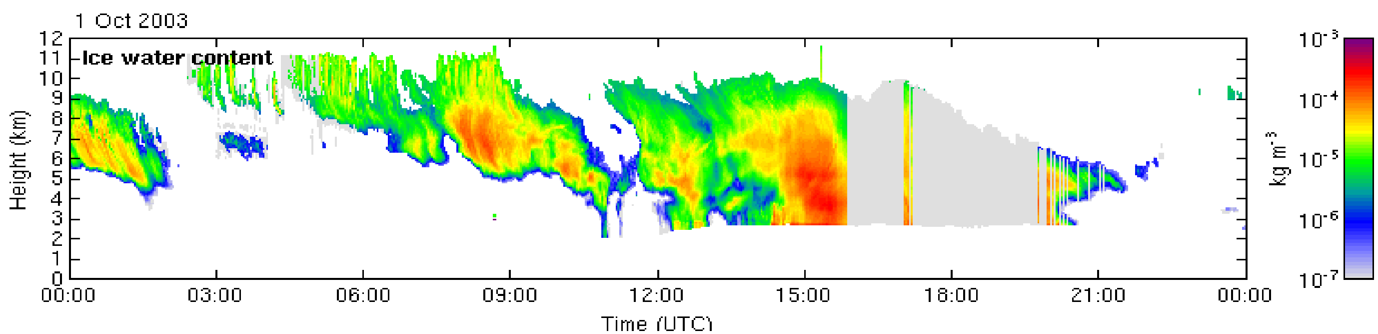




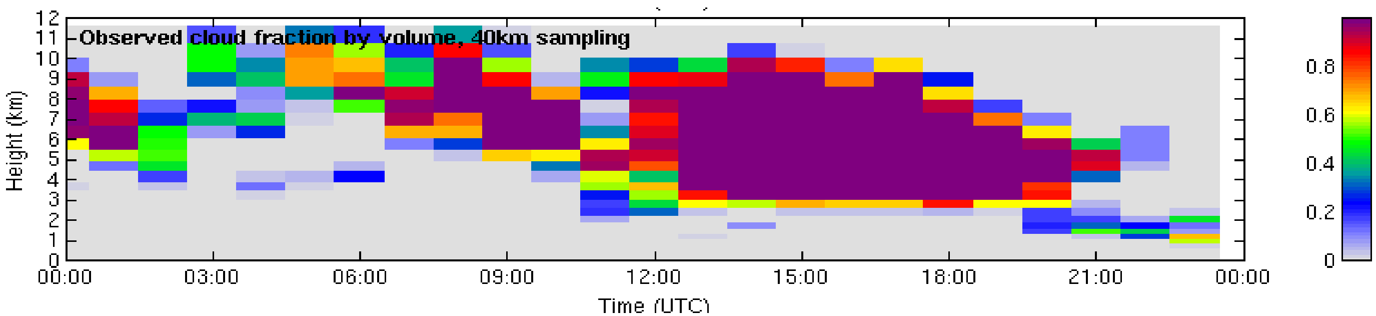
Cloudnet

Level 2a/2b

- Cloud products on (L2a) observational and (L2b) model grid
  - Water content and cloud fraction



L2a IWC on radar/lidar grid



L2b Cloud fraction on model grid

Cloud fraction

Chilbolton Observations

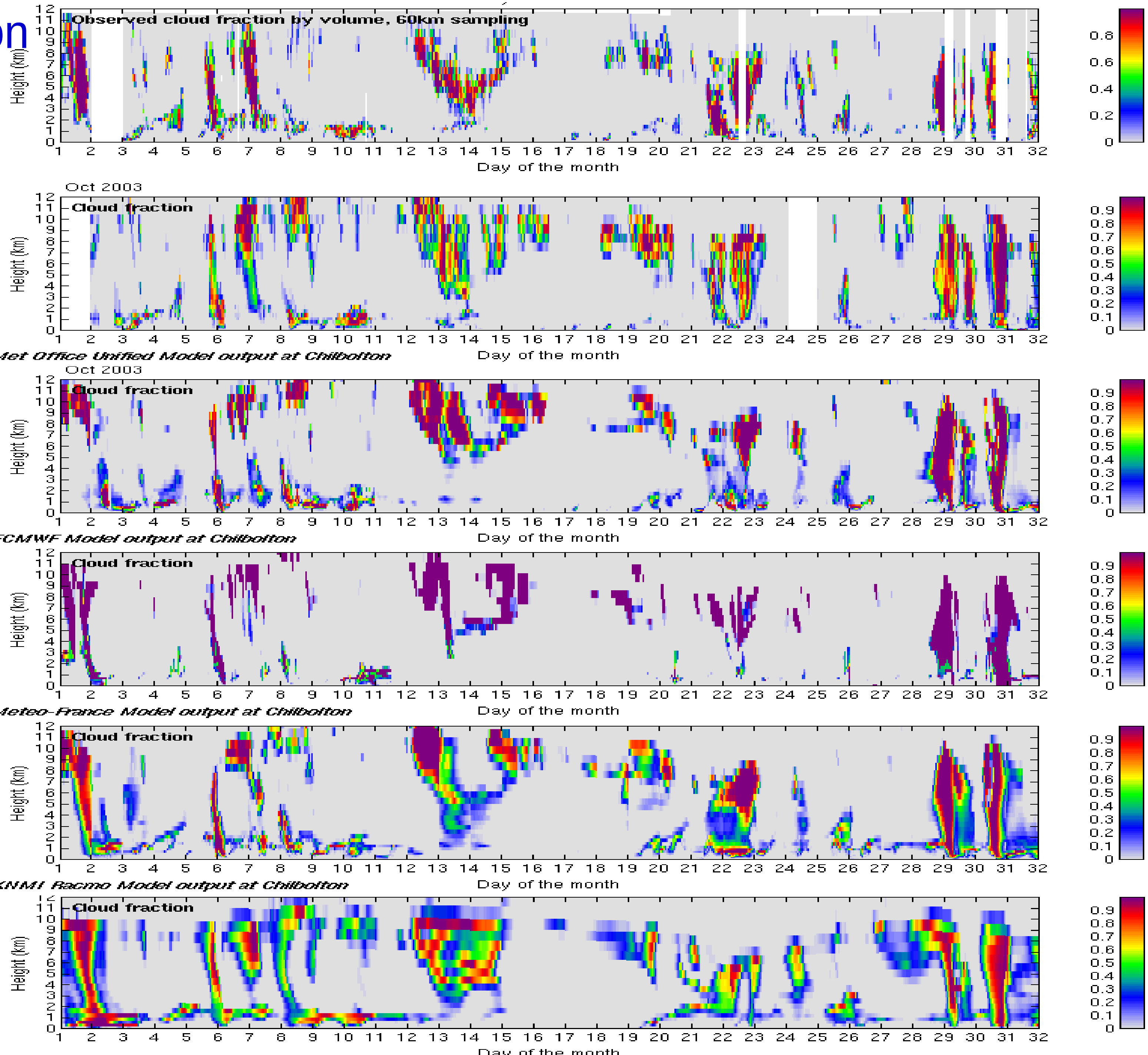
Met Office Mesoscale Model

ECMWF Global Model

Meteo-France ARPEGE Model

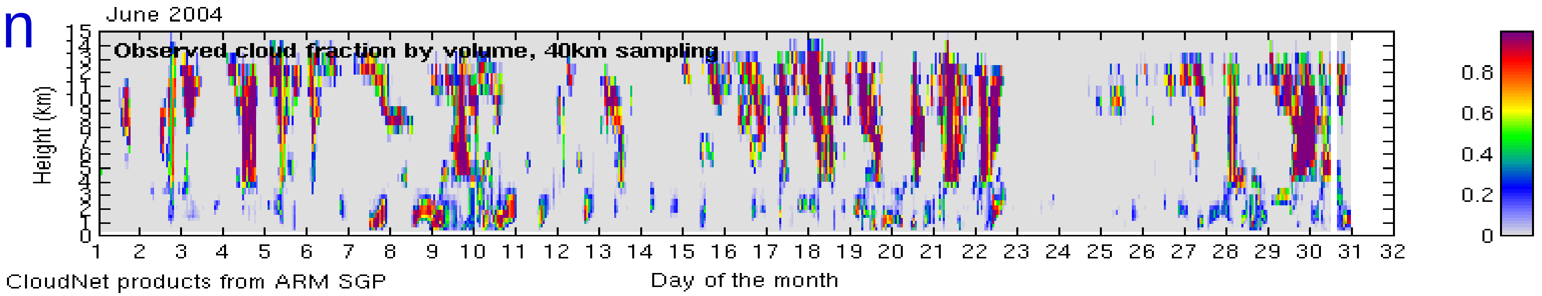
KNMI RACMO Model

Swedish RCA model

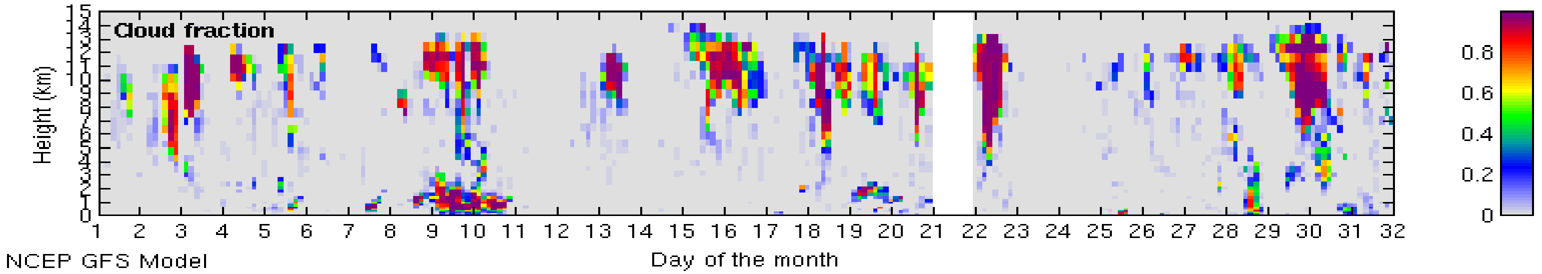


Cloud fraction

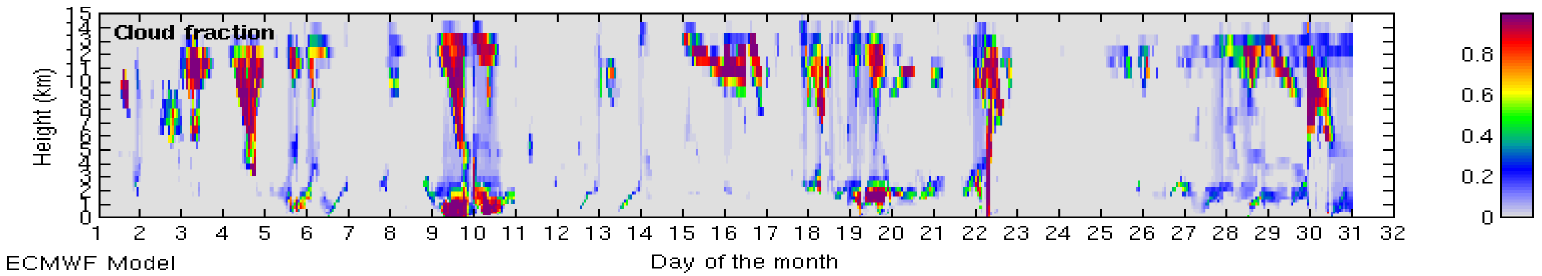
ARM SGP Observations



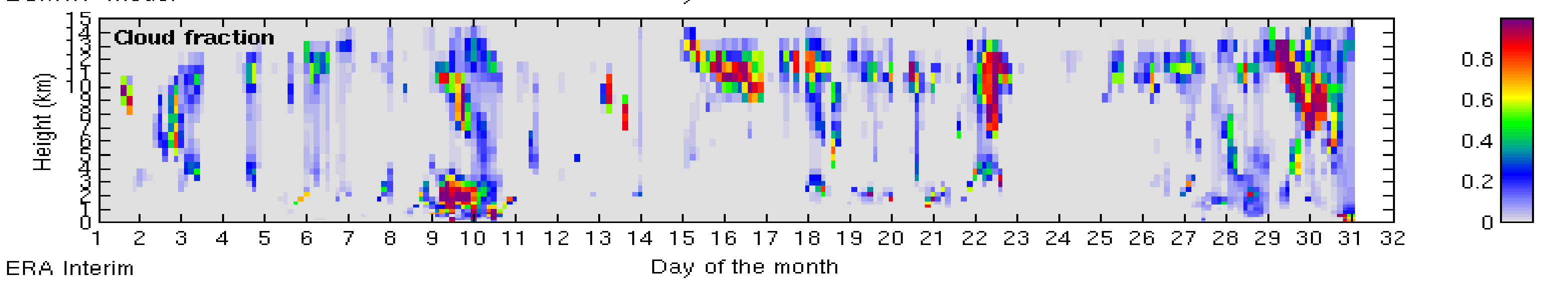
NCEP GFS model



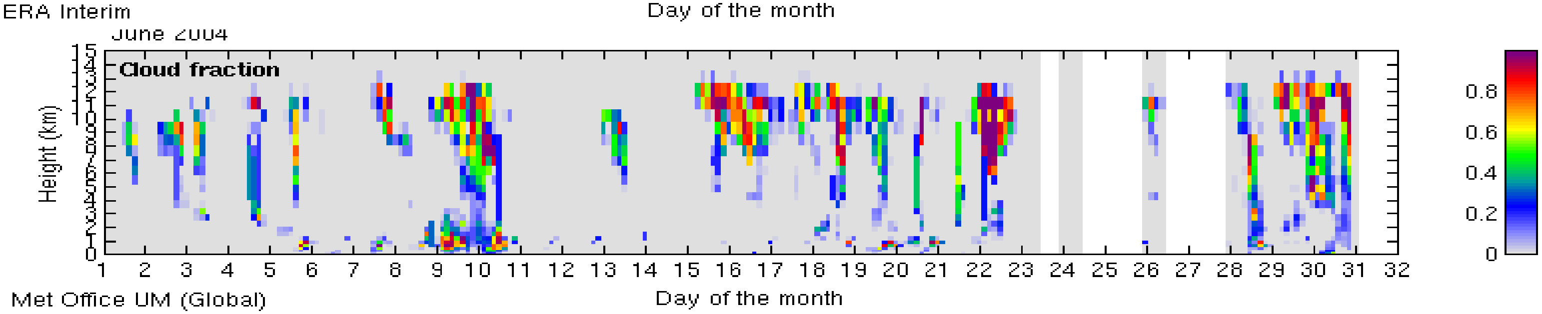
ECMWF model



ERA Interim Analyses

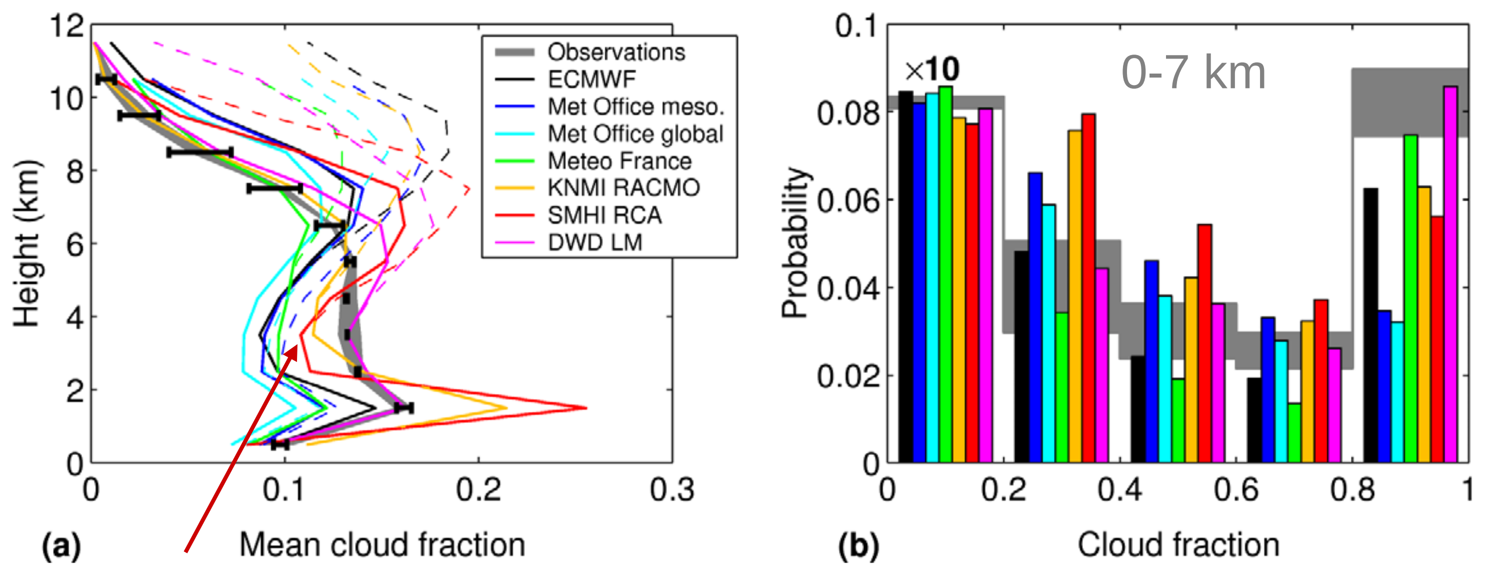


Met Office Global Model



Cloud fraction in 7 models

- Mean & PDF for 2004 for Chilbolton, Paris and Cabauw

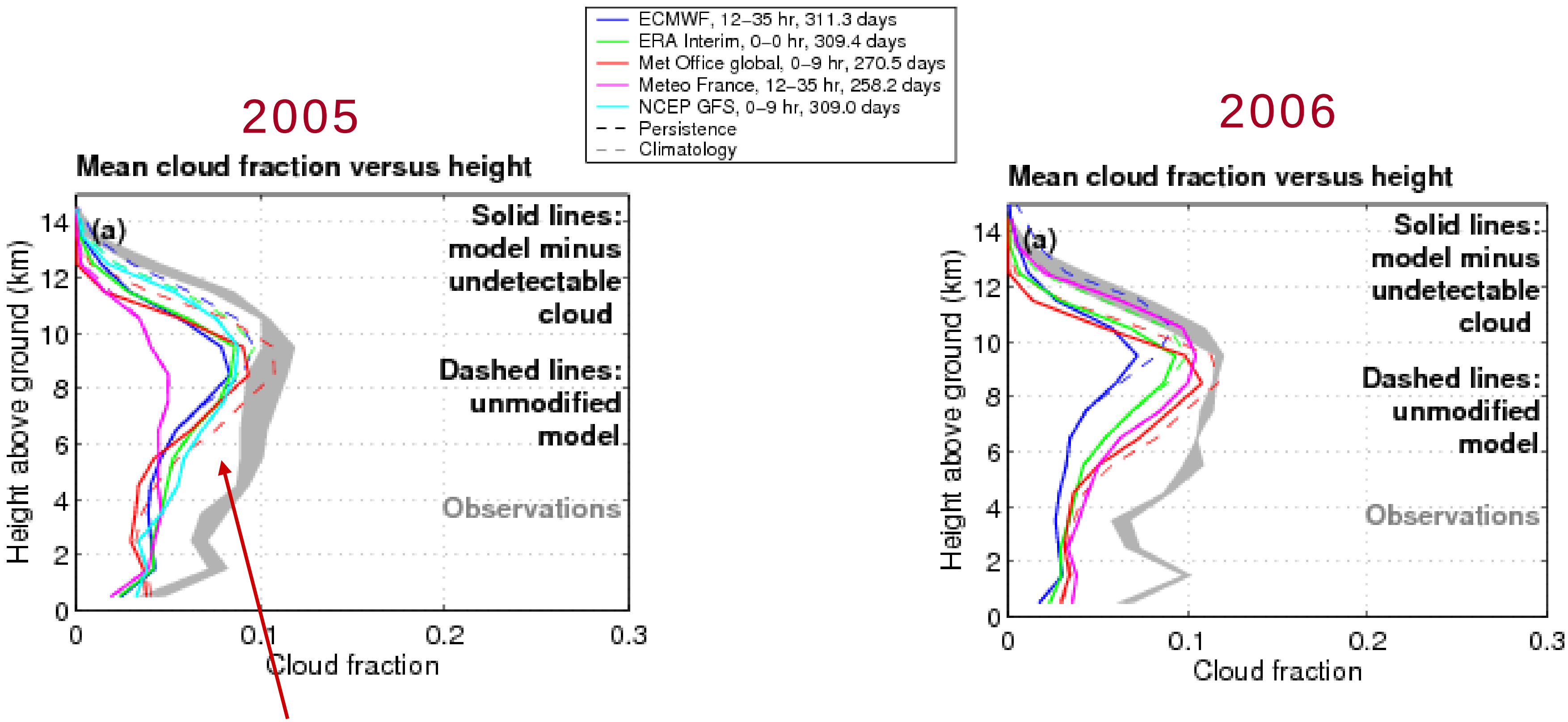


- All models except DWD underestimate mid-level cloud
- Some have separate “radiatively inactive” snow (ECMWF, DWD); Met Office has combined ice and snow but still underestimates cloud fraction
- Wide range of low cloud amounts in models
- Not enough overcast boxes, particularly in Met Office model

Illingworth et al. (BAMS 2007)

Cloud fraction in 5 models

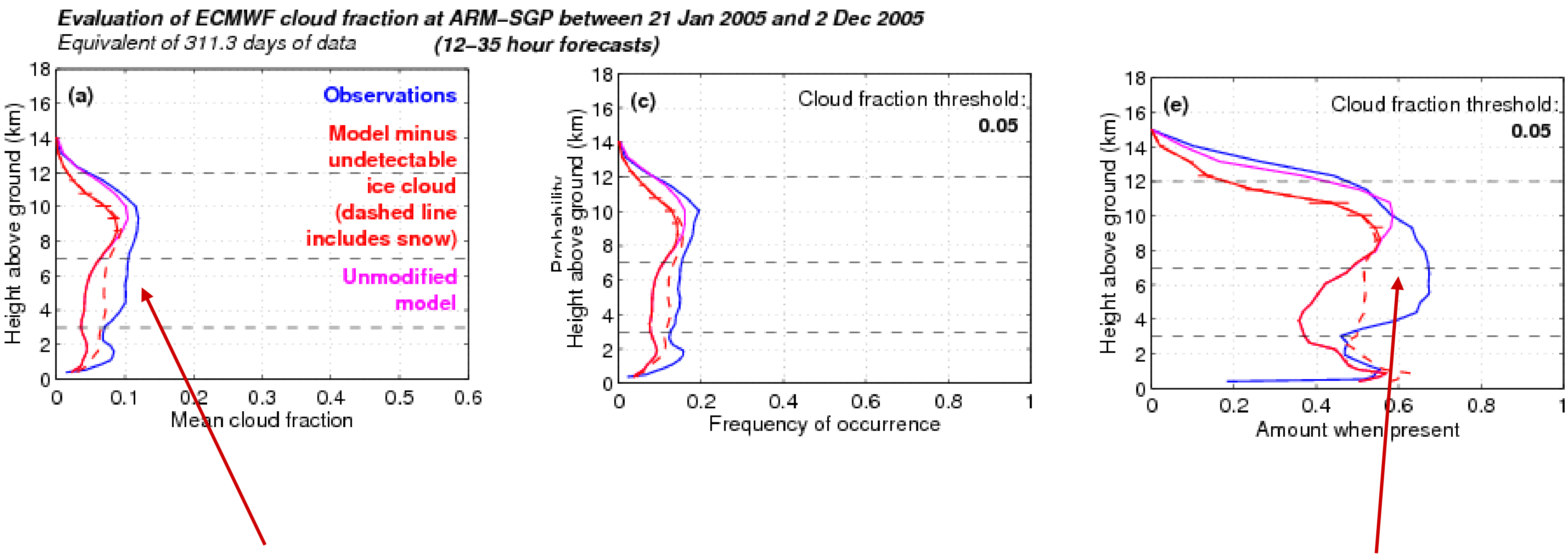
- Mean for ARM SGP



- All models again underestimate mid-level cloud
- Météo france shows improvement from 2005 to 2006

Cloud fraction components

- ECMWF model at ARM SGP for 2005

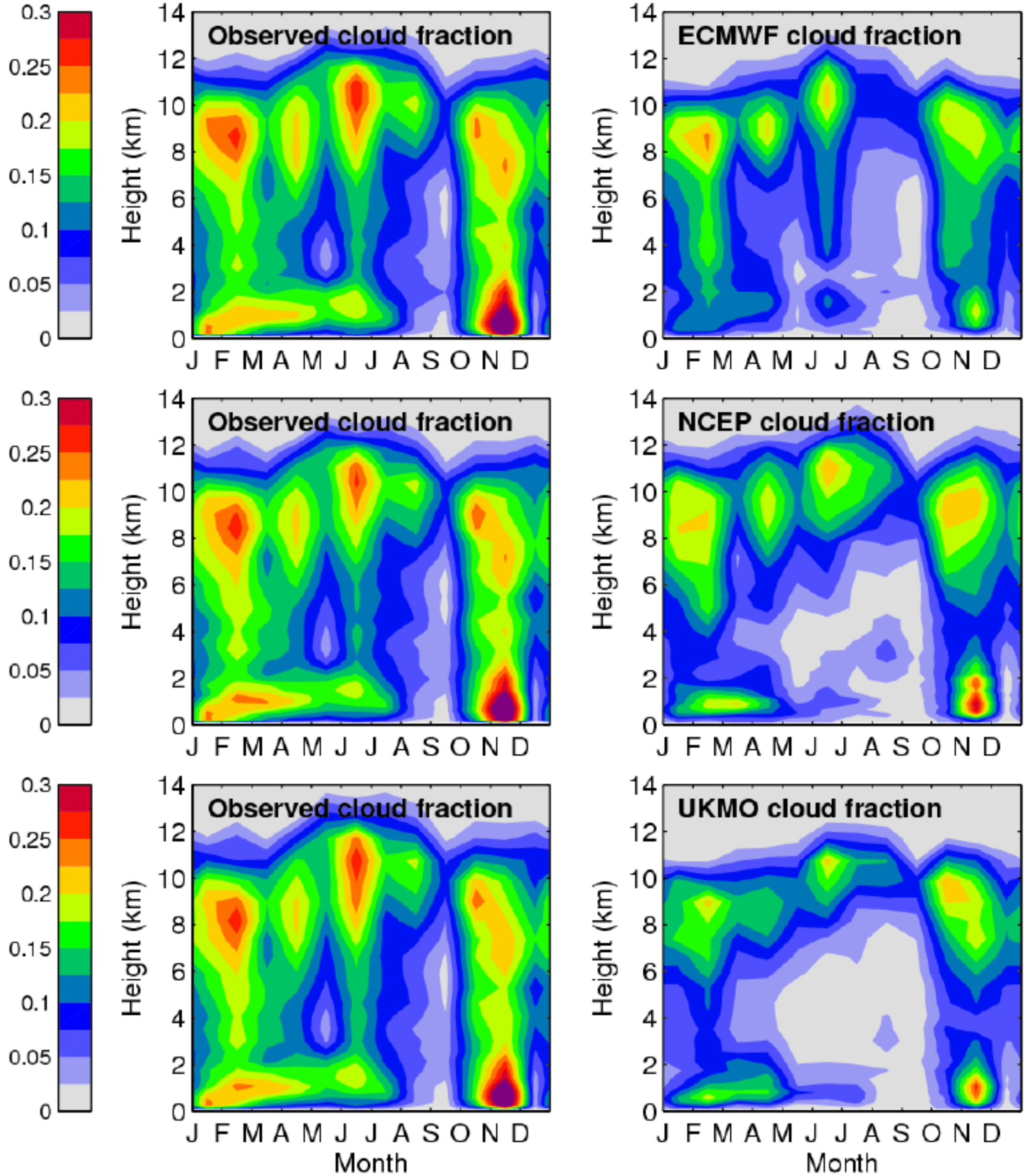


Mean cloud fraction underestimated. Improves slightly with the inclusion of snow.

Clouds are forecast often enough, when snow is included, except in BL.

Underestimate of the mid-level cloud fraction amounts, even when snow is included.

Seasonal variation

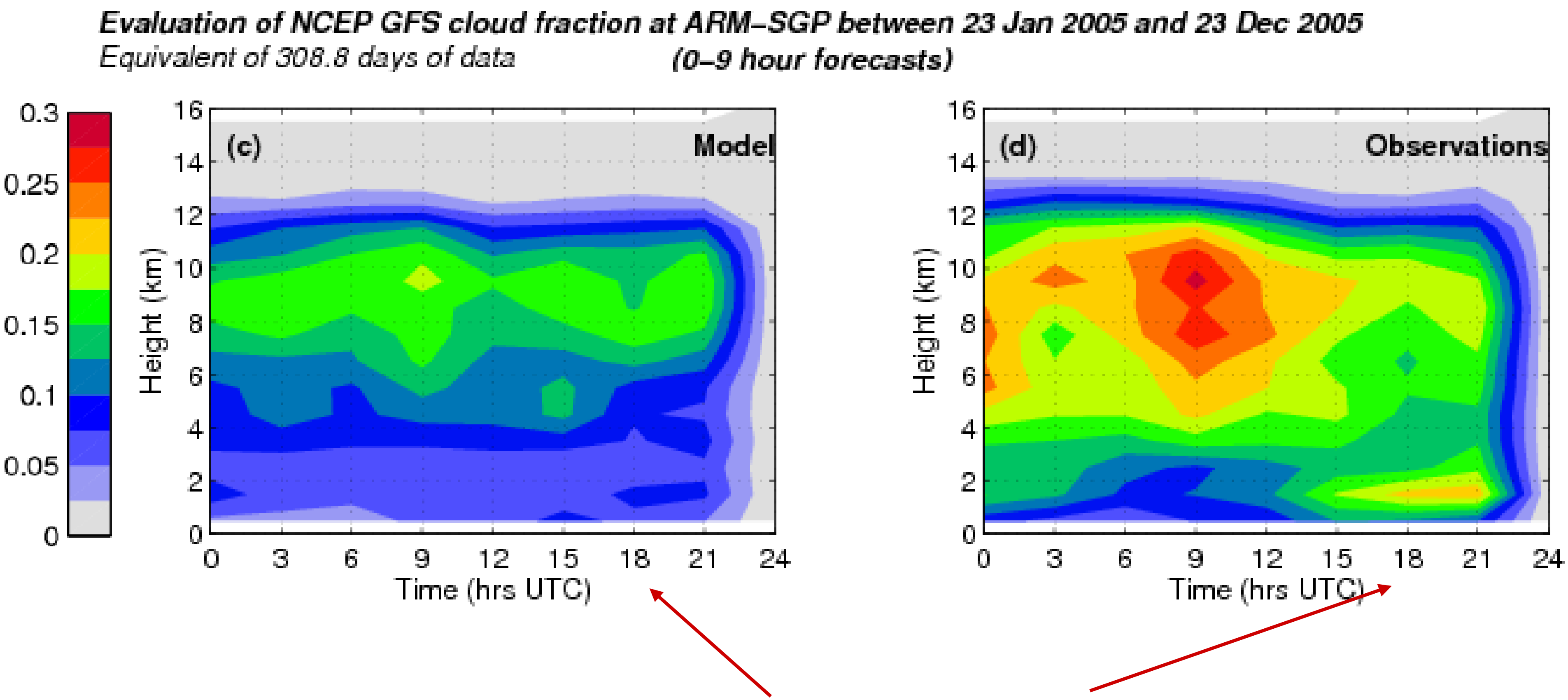


ECMWF

NCEP

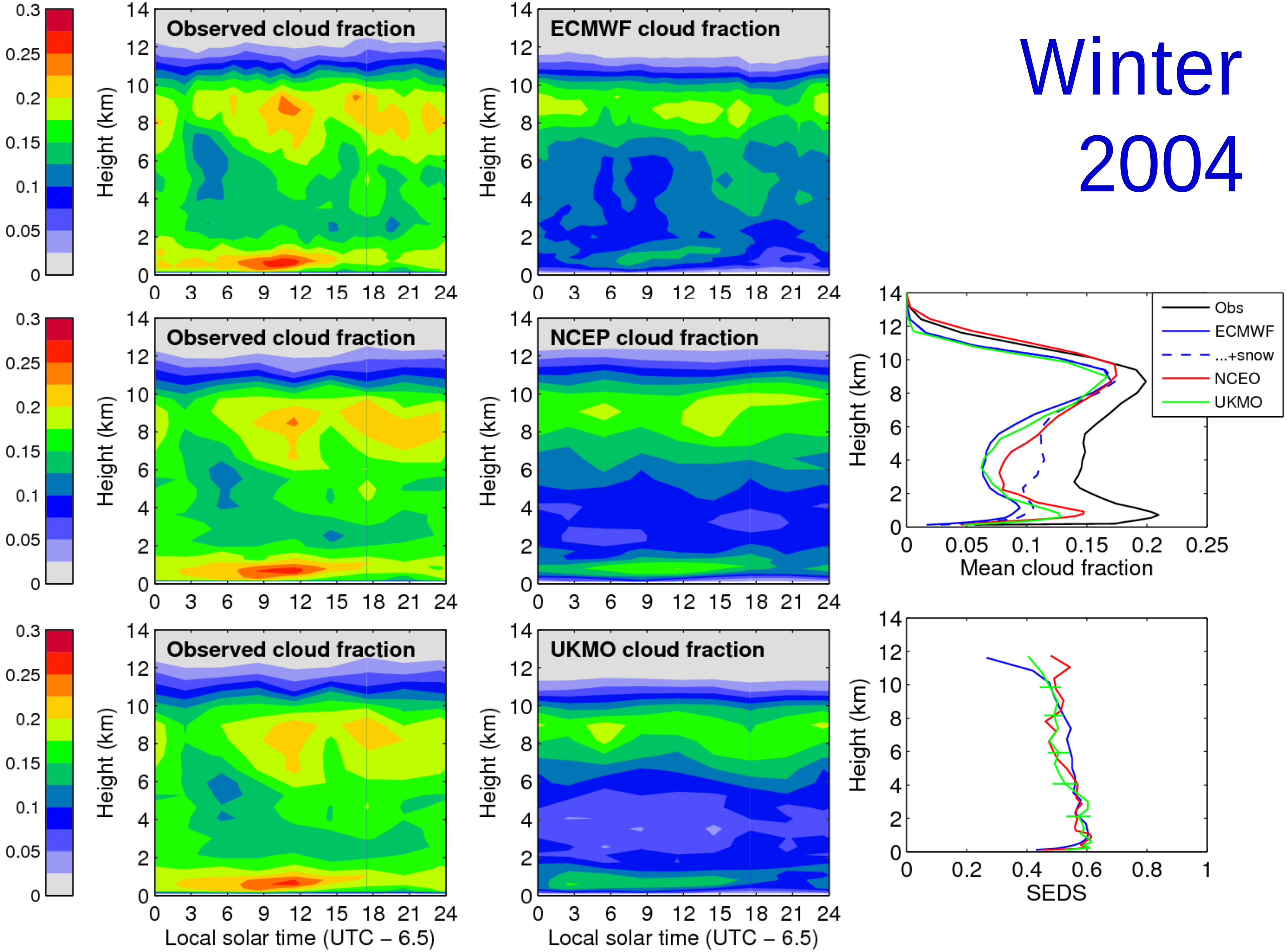
UK Met Office

Diurnal variation



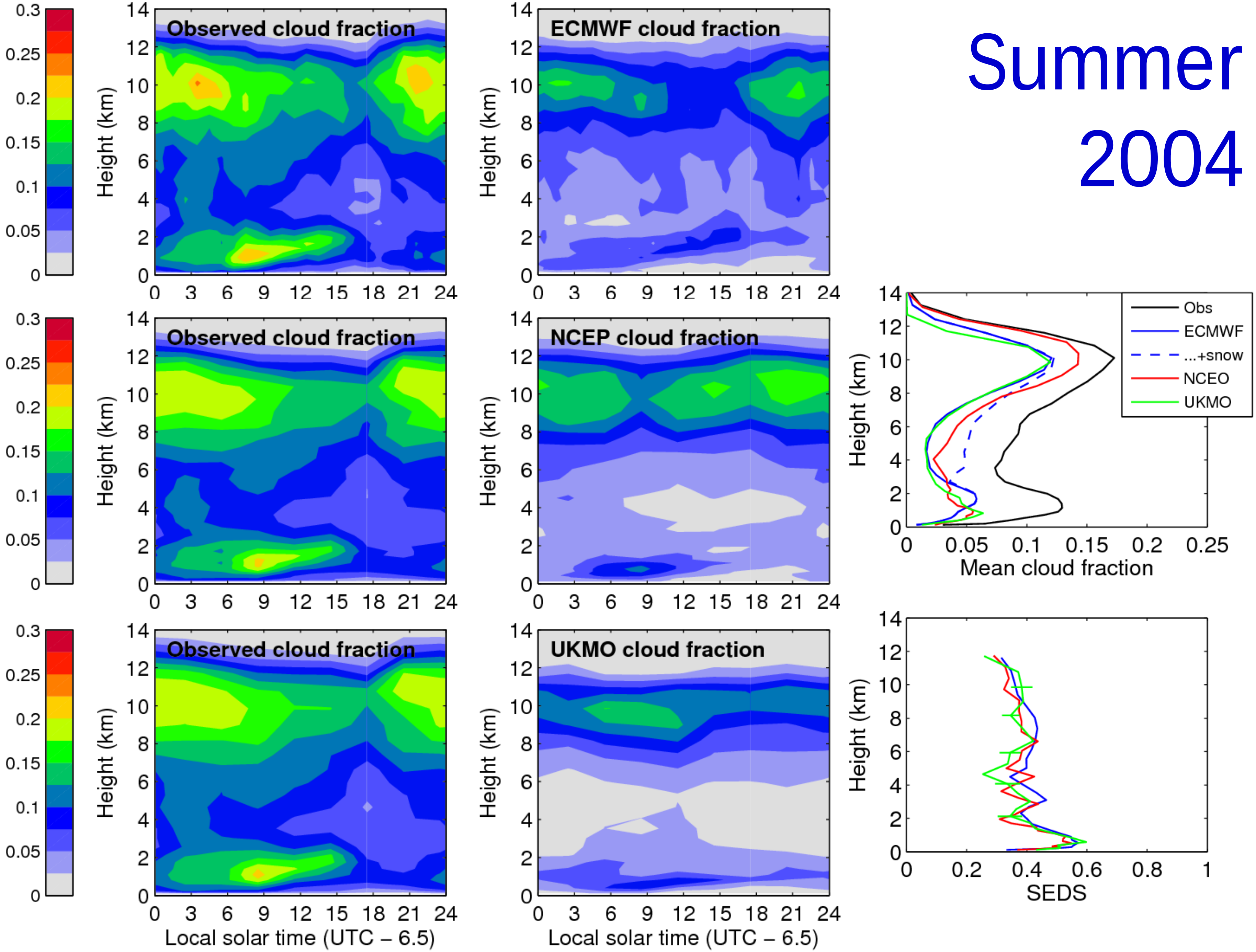
- Model cloud fraction always lower than observed.
- Not enough boundary layer cloud during the day.
- Can we simply scale the model cloud fraction?

Winter 2004

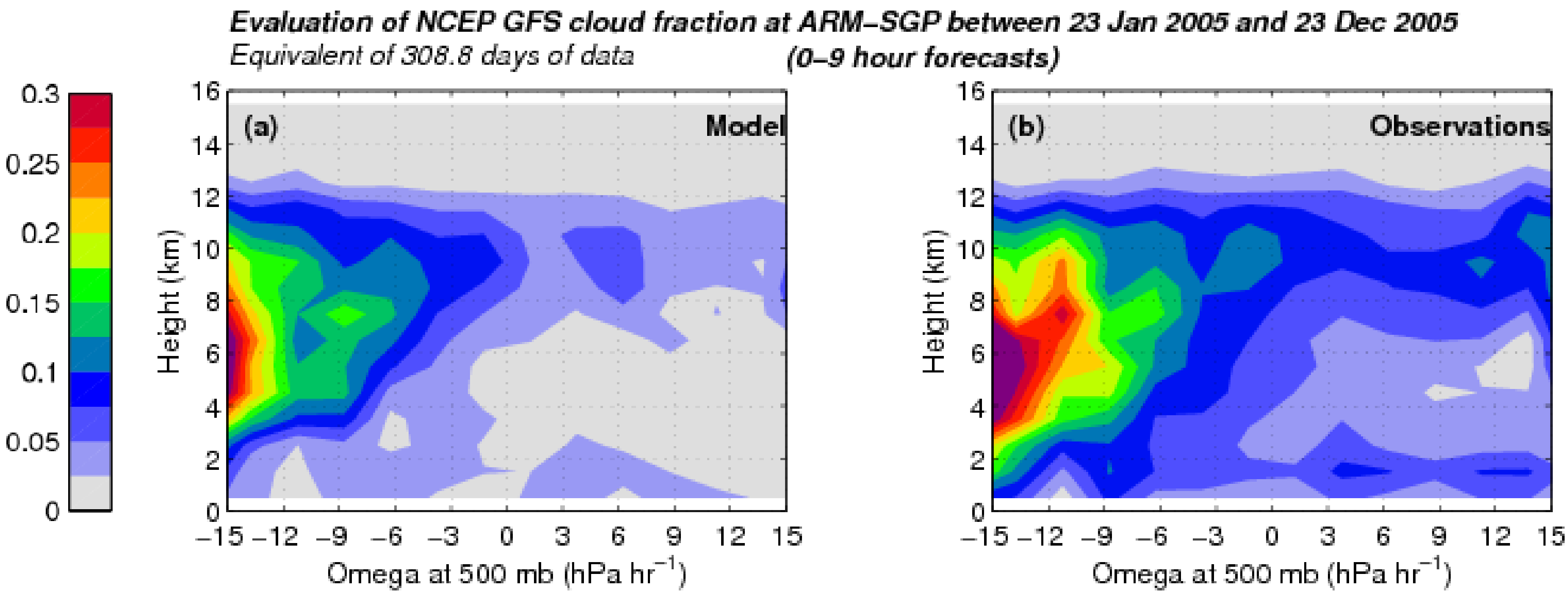


Observed cloud fraction
ECMWF cloud fraction
NCEP cloud fraction
UKMO cloud fraction
Height (km)
Local solar time (UTC – 6.5)
Obs
ECMWF
...+snow
NCEO
UKMO
Mean cloud fraction
SEDS

Summer 2004

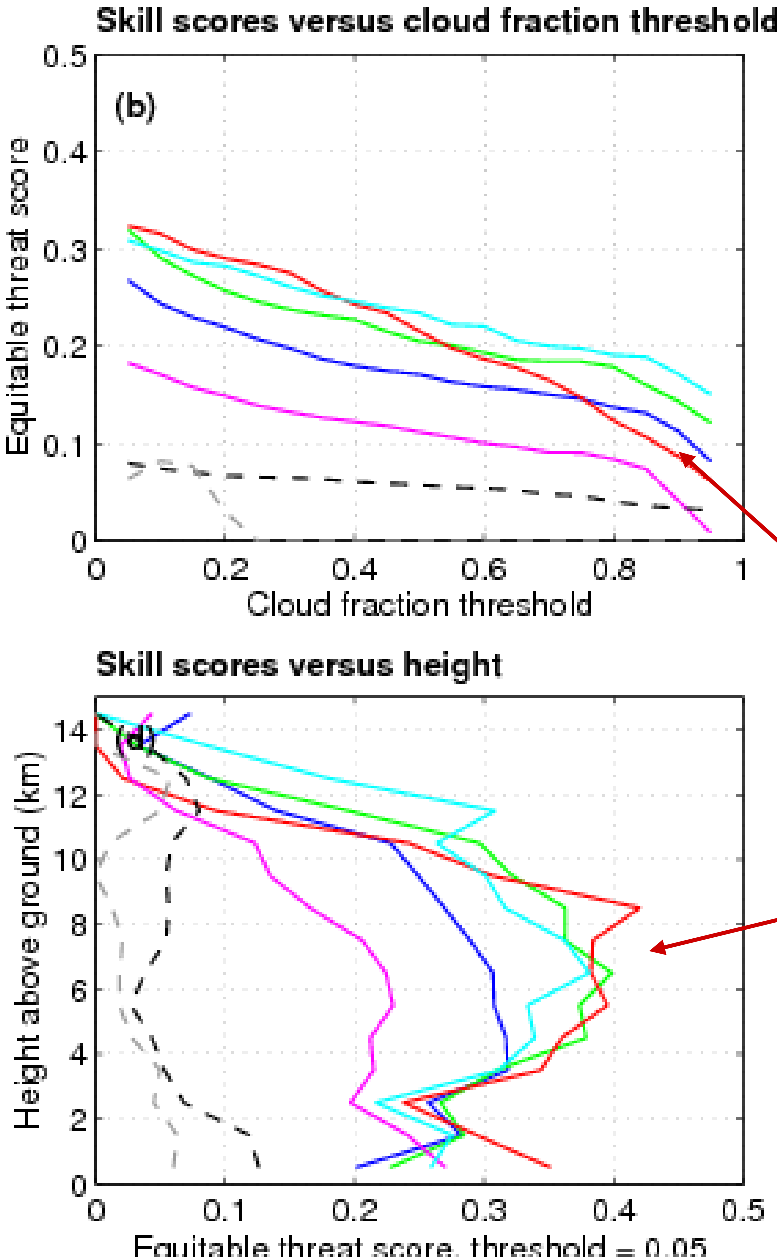

Observed cloud fraction
ECMWF cloud fraction
Observed cloud fraction
NCEP cloud fraction
Observed cloud fraction
UKMO cloud fraction
Height (km)
Local solar time (UTC – 6.5)
Obs
ECMWF
...+snow
NCEO
UKMO
Mean cloud fraction
SEDS

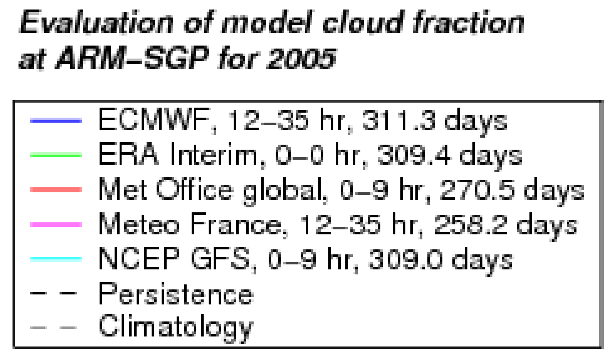
Omega at 500 mb



- Model cloud fraction always lower than observed.
- Not enough cloud in anticyclonic conditions, especially boundary layer cloud.
- Can we scale cloud fraction? Only in large-scale ascent.

Skill Scores



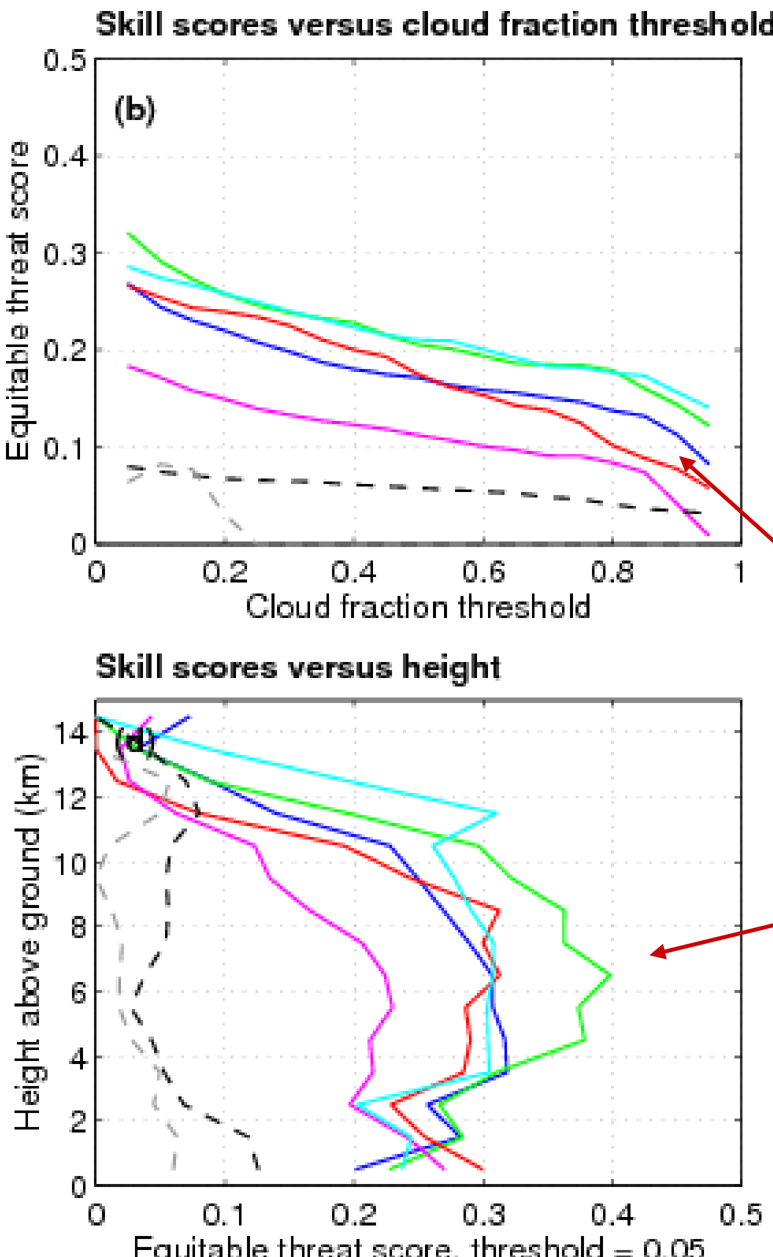


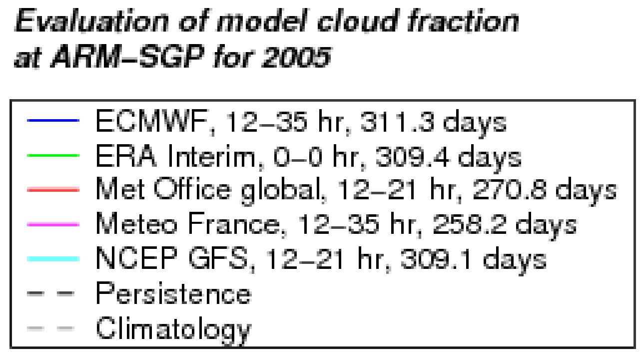
- Met Office Global model has much lower skill for high cloud fraction amounts.

- Most models show more skill in the mid-levels than in the BL.

NB. Not all models are shown with the same forecast leadtime!

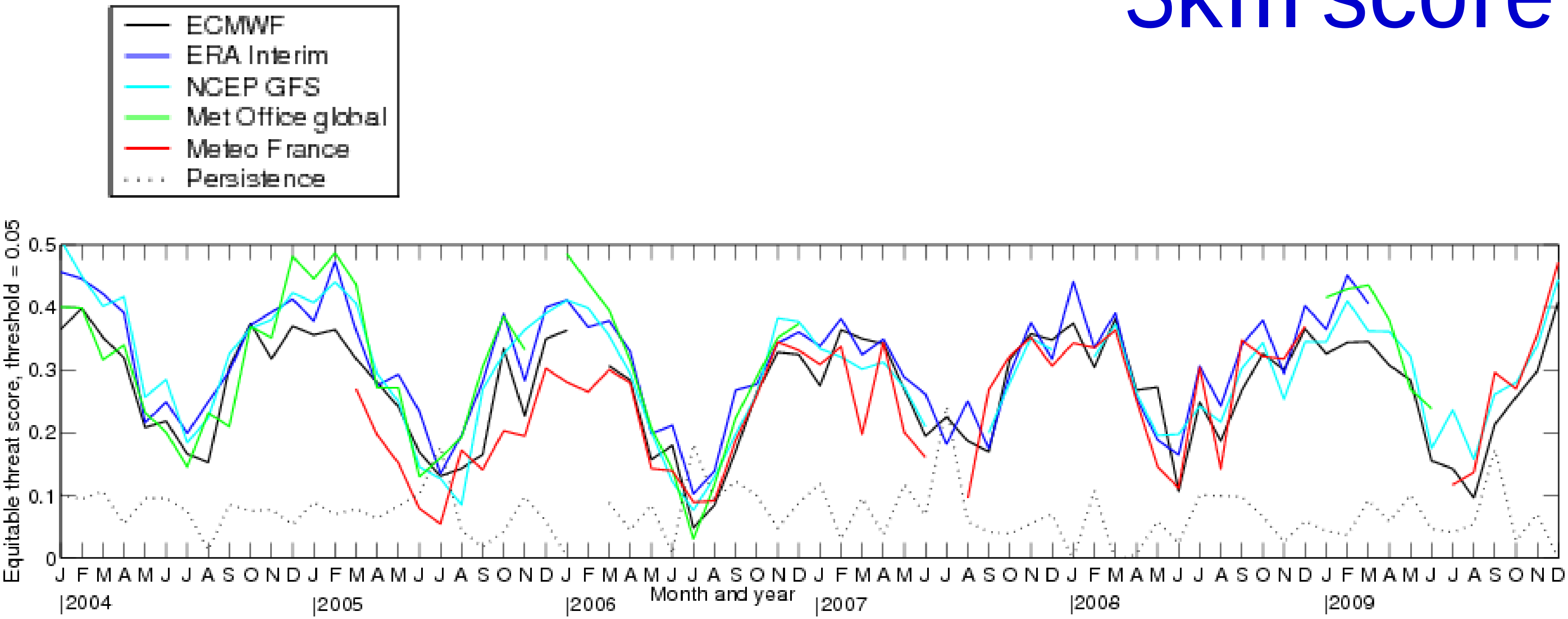
Skill Scores





- Met Office Global model has much lower skill for high cloud fraction amounts.
- Most models show more skill in the mid-levels than in the BL.

Skill score

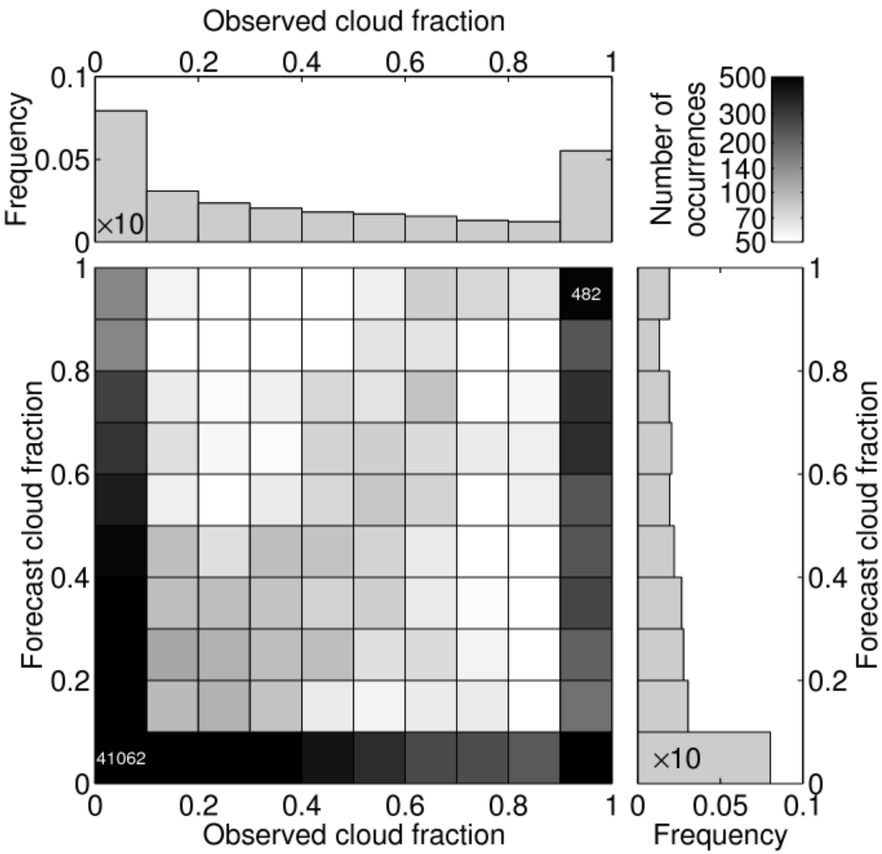


- Six years of cloud fraction evaluation over SGP
  - Clearly less skill in summer, often no better than persistence
  - ERA Interim Reanalysis no different to forecast
  - Any improvement in the cloud fraction forecast over time?
    - *For Météo France, yes..*

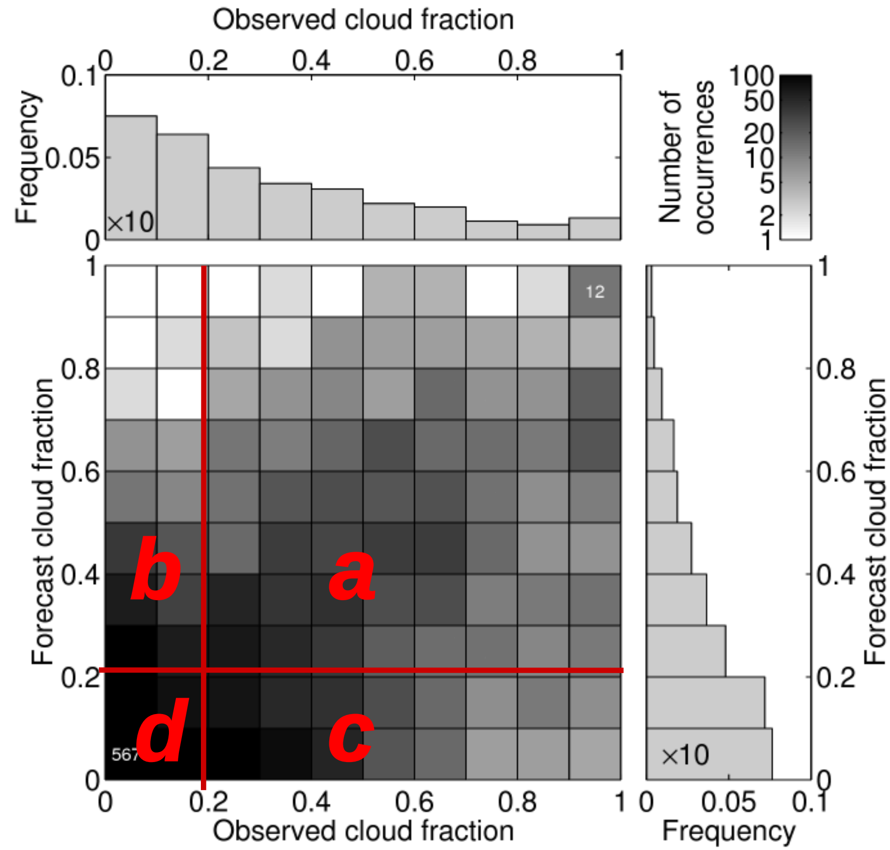
Summary and future work

- Six years of evaluation over SGP (extending to nine)
  - All models underestimate mid- and low-level cloud
  - Skill may be robustly quantified: less skill in summer
- Infrastructure to interface ARM and Cloudnet data has been tested with cloud fraction, IWC/LWC ongoing
  - So far Met Office, NCEP, ECMWF, Météo-France and ERA Interim processed.
  - Analyses do not show much improvement over NWP forecasts.
  - Next implement code at BNL, with other ARM sites and models.
  - Question: have cloud forecasts improved in 10 years?
- Next compare with results from SCM-testbed
  - We have the tools to quantify objectively improvements in both bias and skill with changed parameterizations in NWP models and SCMs.
  - Other metrics of performance or compositing methods required?

Joint PDFs of cloud fraction



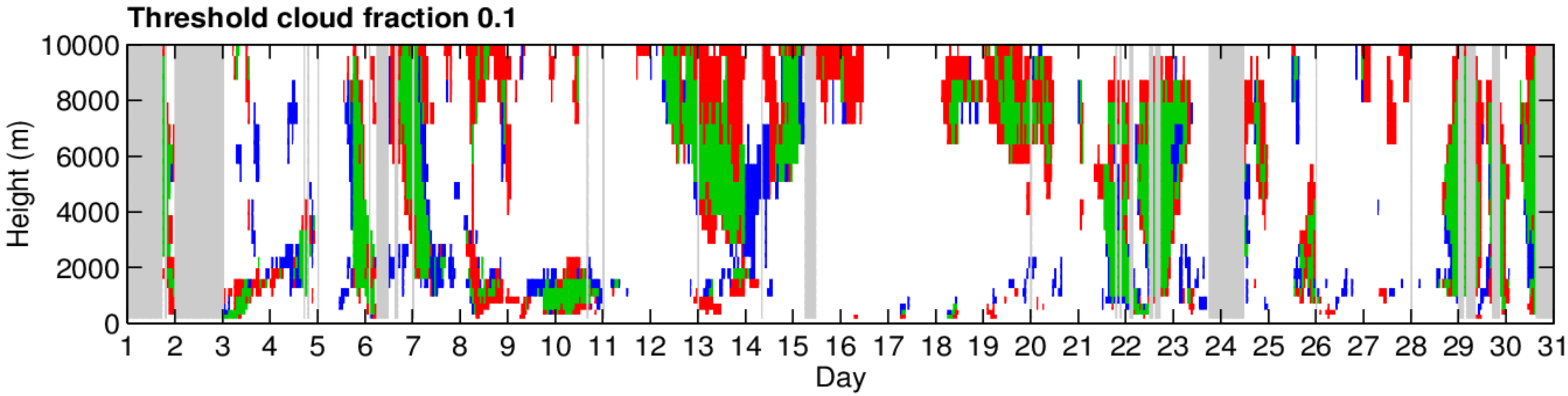
- Raw (1 hr) resolution
  - 1 year from Murgtal
  - DWD COSMO model



- 6-hr averaging

...or use a simple contingency table

Contingency tables



	Observed cloud	Observed clear-sky
Model cloud	*a*: Cloud hit	*b*: False alarm
Model clear-sky	*c*: Miss	*d*: Clear-sky hit

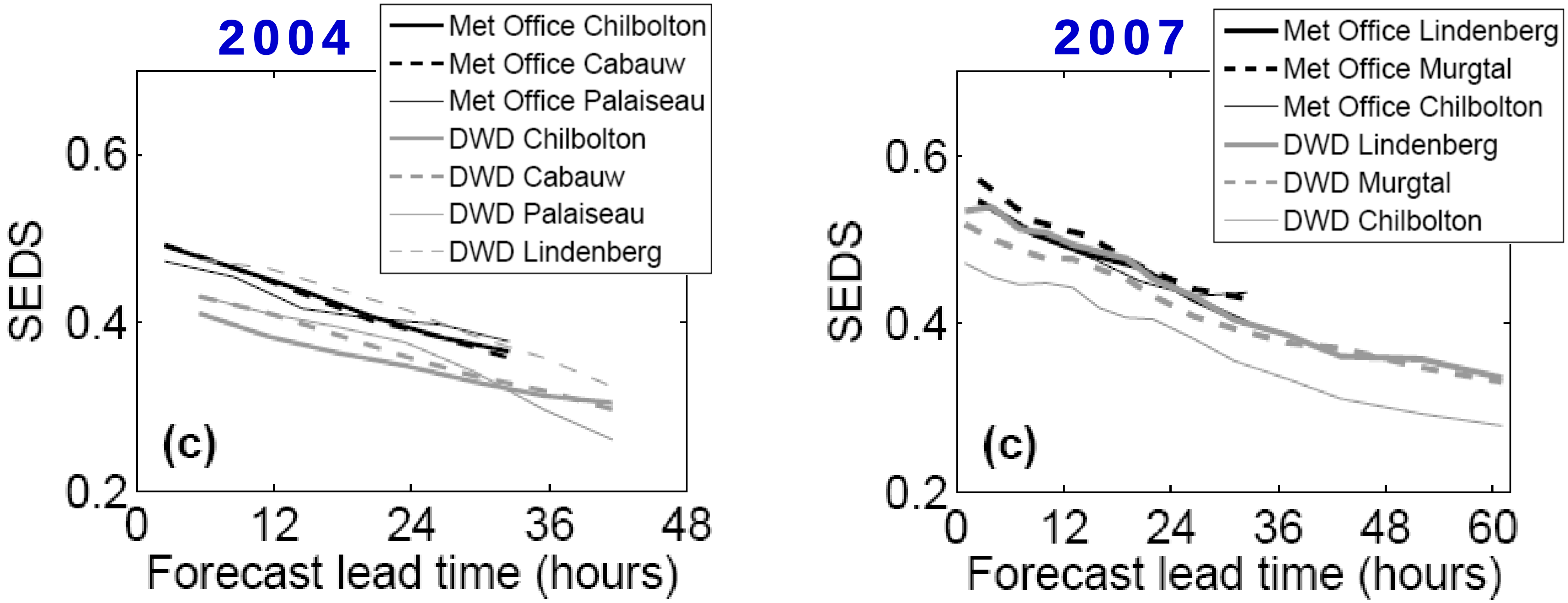
a = 7194	b = 4098
c = 4502	d = 41062

DWD model, Murgtal

For given set of observed events, only 2 degrees of freedom in all possible forecasts (e.g. *a* & *b*), because 2 quantities fixed:

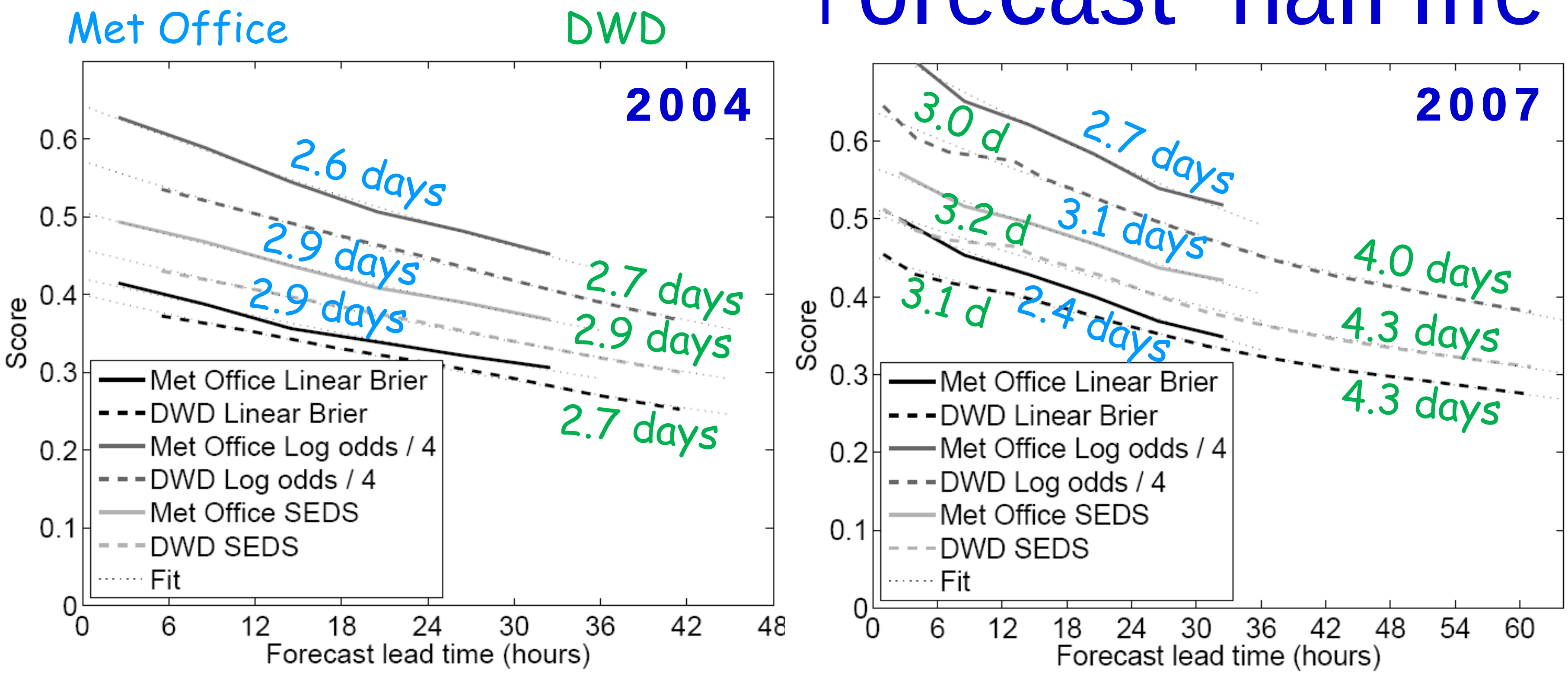
- Number of events that occurred $n = a + b + c + d$
- Base rate (observed frequency of occurrence) $p = (a + c)/n$

Skill versus lead time



- Only possible for UK Met Office 12-km model and German DWD 7-km model
  - *Steady decrease of skill with lead time*
  - *Both models appear to improve between 2004 and 2007*
- Generally, UK model best over UK, German best over Germany
  - *An exception is Murgtal in 2007 (Met Office model wins)*

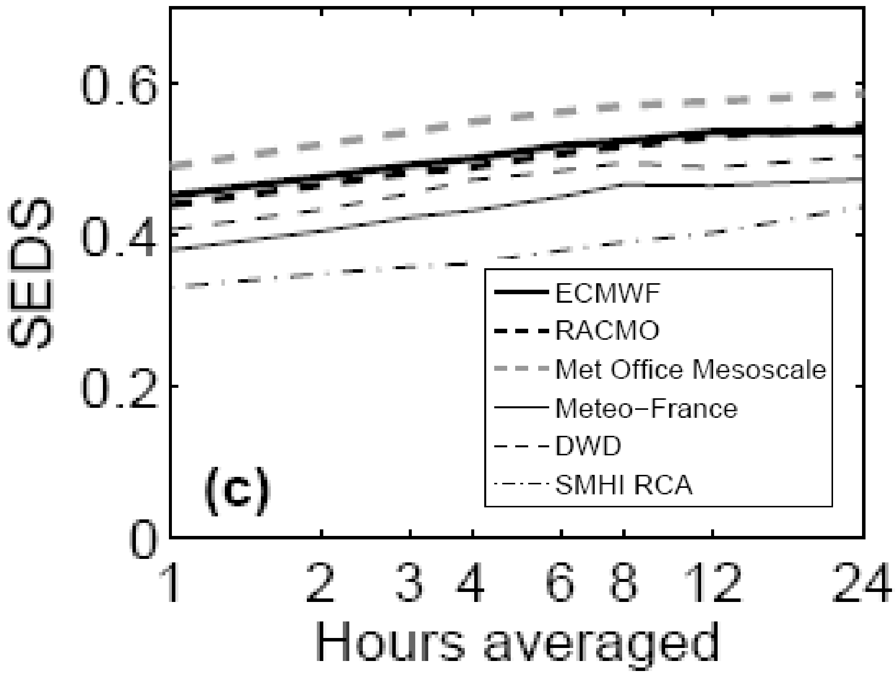
Forecast "half life"

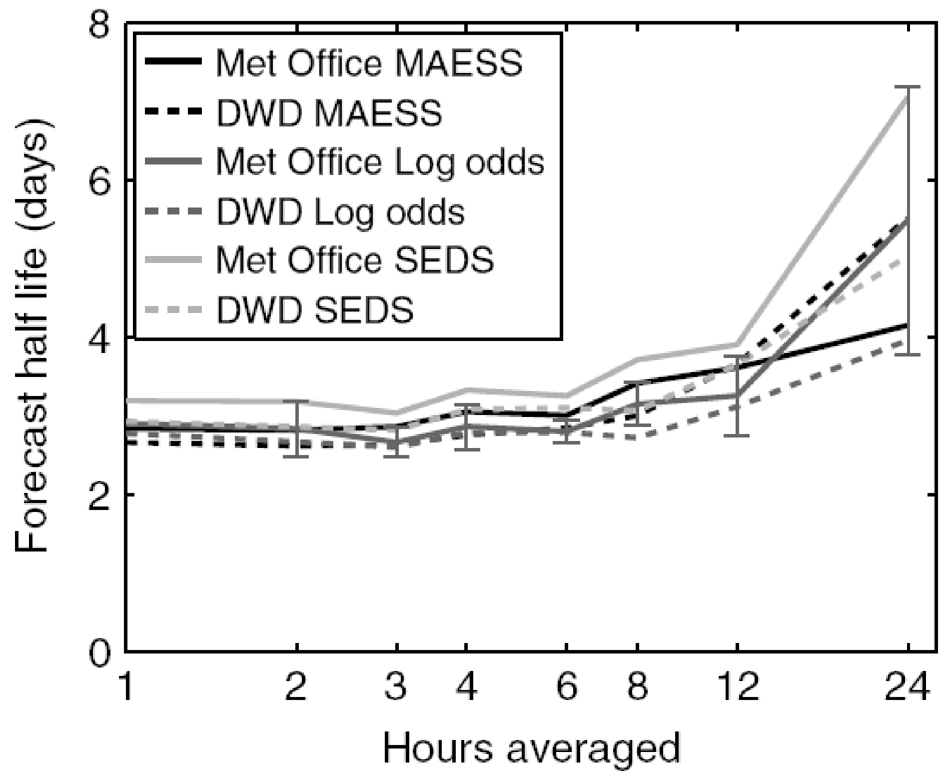


- Fit an inverse-exponential: $S(t) = S_0 \times 2^{-t/\tau_{1/2}}$
  - S_0 is the initial score and $\tau_{1/2}$ is the half-life
- Noticeably longer half-life fitted after 36 hours
  - Same thing found for Met Office rainfall forecast (Roberts 2008)
  - First timescale due to data assimilation and convective events
  - Second due to more predictable large-scale weather systems

Why is half-life less for clouds than pressure?

- Different spatial scales? Convection?
  - Average temporally before calculating skill scores:





  - Absolute score and half-life increase with number of hours averaged