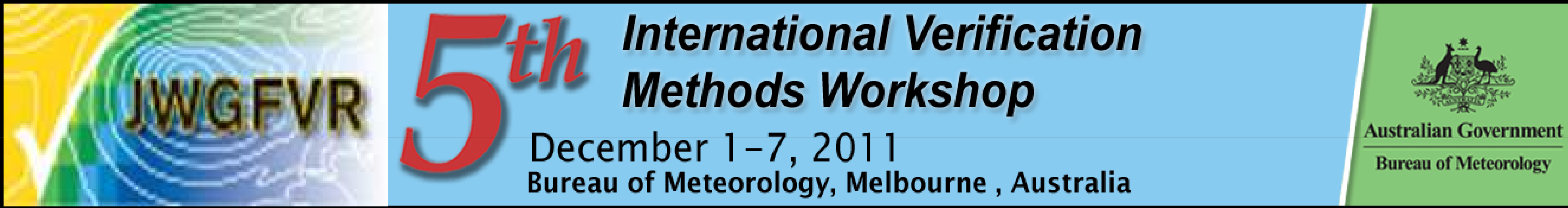
# Review JWGFVR 5th Verification workshop & SRNWP-V programme

Clive Wilson –Expert Team on diagnostics, validation & verification

With thanks to Marion Mittermaier, Rachel North, Matthew Trueman,Ric Crocker

34th EWGLAM/19th SRNWP meetings –Helsinki 8-11  October 2012

# Contents




- WWRP/JWGFVR Cloud verification recommendations
- SRNWP-V programme
    - Progress & update
    - WebEx ET meeting – June 2012
    - Final reports

# $5^{th}$ verification methods workshop Melbourne, 5-7 December 2011

**Marion Mittermaier**

- http://cawcr.gov.au/events/verif2011/
- 5 keynote presentations on 5 key topics

1. **Ensembles and probabilistic forecasts**: *Verification of ensemble forecasts: a guided tour through a zoo of skill metrics* – A. Weigel
2. **Seasonal and Climate**: *Verification of seasonal forecasts* – A. Watkins
3. **Aviation specific verification**: *Recent developments in Aviation Verification at the UK Met Office* – P. Gill
4. **User-oriented systems and topics**: *Uncertainty Forecasts and the End-User* – S. Jocelyn
5. **Tropical Cyclones and high-impact weather**: *Verification of tropical cyclone forecasts* – B. Brown

# 5th verification methods workshop Melbourne, 5-7 December 2011

**Marion Mittermaier**

- Special verification issue of Meteorological Applications. *[Deadline was 15 September 2012.]*
- Introductory paper to special issue by Ebert *et al.* on "**Progress and challenges in forecast verification**"
- Key areas where progress has been made include:
    - Improved reporting
    - Wider use of diagnostic verification
    - Development of new scores and techniques for difficult problems, and
    - Evaluation of forecasts for applications using meteorological information

# 5th verification methods workshop Melbourne, 5-7 December 2011

**Marion Mittermaier**

## Many interesting challenges remain:

- Improvements in methods to verify high-resolution ensembles
- Seamless predictions spanning multiple space and time scales
- Multivariate forecasts.

## More work is needed to:

- make best use of new observations
- investigate links with data assimilation, and
- develop better intuitive “verification products” for end-users

WORLD METEOROLOGICAL ORGANIZATION

WORLD WEATHER RESEARCH PROGRAMME

WWRP 2012 - 1

# RECOMMENDED METHODS

# FOR EVALUATING

# CLOUD AND RELATED PARAMETERS

March 2012

WWRP/WGNE Joint Working Group on Forecast Verification Research (JWGFVR)

**http://www.wmo.int/pages/prog/arep/wwrp/new/documents/WWRP_2012_1_web.pdf**

# Cloud verification -summary recommendations

- User-orientated:
  - total cloud cover and cloud base height
    - Low, medium, high desirable
    - Spatial bias – eg satellite cloud mask
- More generally use satellite cloud – but not analyses at short range –model contamination
- model-oriented :
  - Simulated cf observed radiances
  - profiles

# Summary recommendations contd.

- Verify against:
    - Gridded observations and profiles
    - Remote-sensed satellite, lidar
    - Surface station observations
        - If auto and manual – do not mix
        - Automated cloud base for low thresholds (aviation)
- Model impact/sensitivity
        - Use cloud radar & lidar
- Compare against climatology , persistence forecasts
- Include 95%, median & IQ range for model assessments

Table 5 - A summary of metrics for evaluating categorical predictions

| *Metric* | ***TCA, Low, Medium, High*** | ***CBH*** | ***Cloud profile*** | ***LWC, IWC*** | ***Radiances, Brightness Temperature*** |
|---|---|---|---|---|---|
| Frequency Bias (FB) | *** | *** | * | * | * |
| Symmetrical Extreme Dependency Score (SEDS) (Hogan *et al.*, 2009) | *** | *** | * | * | * |
| Odds or log-odds ratio (OR) | ** | ** | * | * | * |
| Hanssen Kuipers Score (HK), Pierce Skill Score (PSS), Kuipers Skill Score (KSS) | ** | ** | * | * | * |
| Probability Of Detection (POD), Hit Rate (HR) | *** | *** | * | * | * |
| Probability Of False Detection (PODF), False Alarm Rate (F) | *** | *** | * | * | * |
| Heidke Skill Score (HSS) | ** | ** | * | * | * |
| Gilbert Skill Score (GSS), Equitable Threat Score (ETS) | ** | ** | * | * | * |
| False Alarm Ratio (FAR) | *** | *** | * | * | * |
| Odds Ratio Skill Score (ORSS) | ** | ** | * | * | * |
| Proportion Correct (PC) | ** | ** | * | * | * |
| Threat Score (TS), Critical Success Index (CSI) | * | * | * | * | * |
| Gerrity Skill Score (GSS) for multi-category forecasts (see description in Annex B) | *** | *** | * | * | * |

**http://www.wmo.int/pages/prog/arep/wwrp/new/documents/WWRP_2012_1_web.pdf**

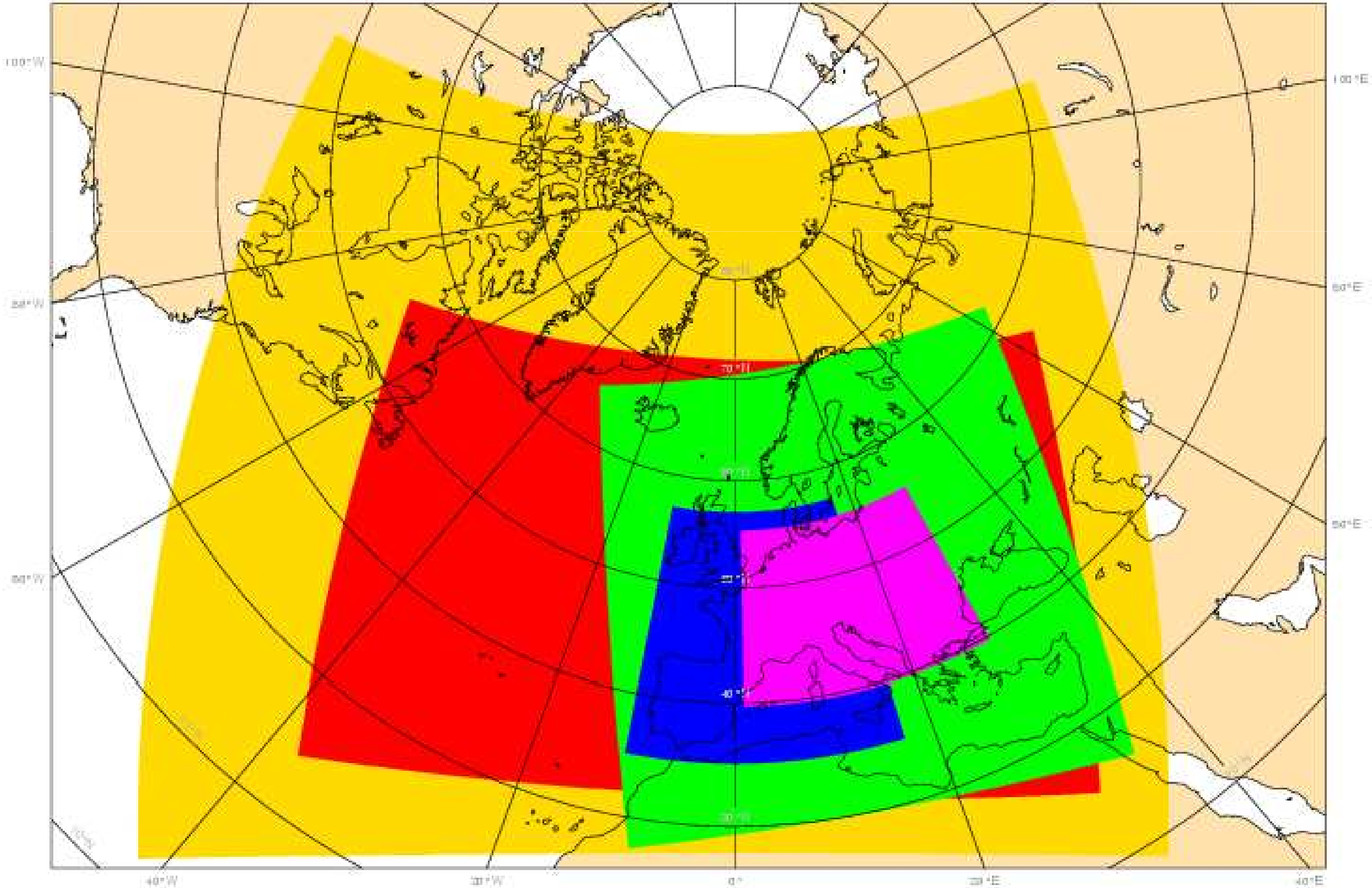
# SRNWP-V Phase 2 Programme progress in red

Met Office

EUMETNET

- Continue & expand comparison
  - Longer more robust results – up-to-date, publication of results on EUMETNET Portal
  - higher resolution of future operational models
    - AROME and ARPEGE results processed since start of 2011
  - SEEPS scores calculated over common domain
- Additional products verified
  - Cloud amount/base
    - truth ? Auto/manual different biases
    - Distribution approach
  - Wind gust –validation underway, also distribution

# Domains of 5 consortia reference models



15km

12km

7km

10km

5 km

(10km)

Hirlam UM COSMO ALADIN Aladin-Lace

| **Parameter** | **Scores** |
|---|---|
| Mean sea level pressure | mean bias and root mean square errors |
| 2m temperature | Bias, rmse |
| 2m relative humidity | Bias, rmse |
| 10m winds | mean bias speed error and root mean square vector wind error |
| 6 hourly total precipitation | equitable threat score and frequency bias for 0.5, 1.0 and 4.0 mm $6h^{-1}$ |

# Key

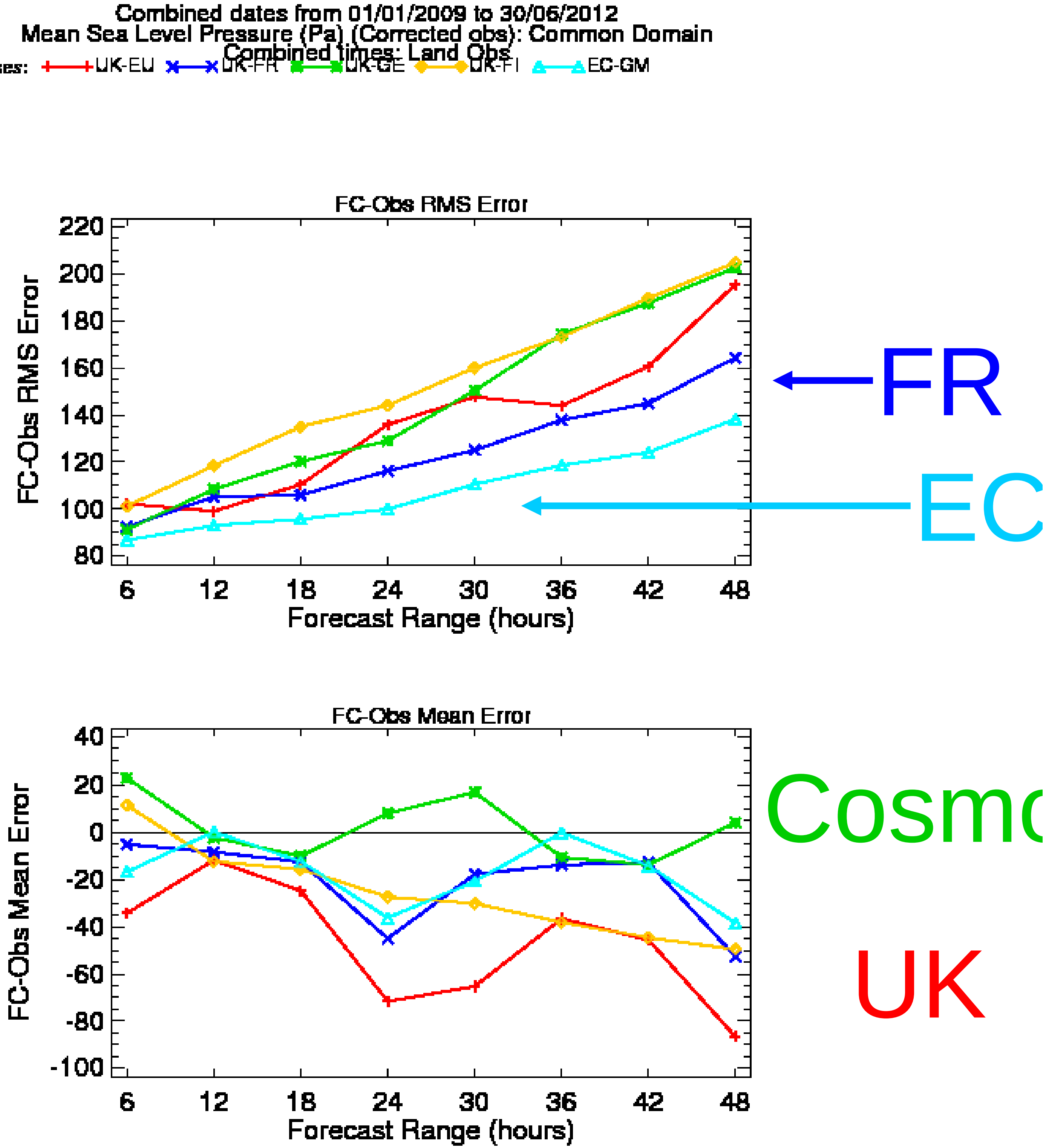
| Model | Label |
|---|---|
| Hirlam reference run by FMI | UK-FI |
| Aladin-France run by Meteo-France | UK-FR |
| COSMO Europe run by DWD | UK-GE |
| The North Atlantic European configuration of the Unified Model run by the Met Office | UK-EU |
| *Aladin-Czech (LACE) run by CHMI* | *UK-LC* |
| ECMWF high resolution global model | EC-GM |

Comparison over ALADIN-France domain unless otherwise stated

Met Office

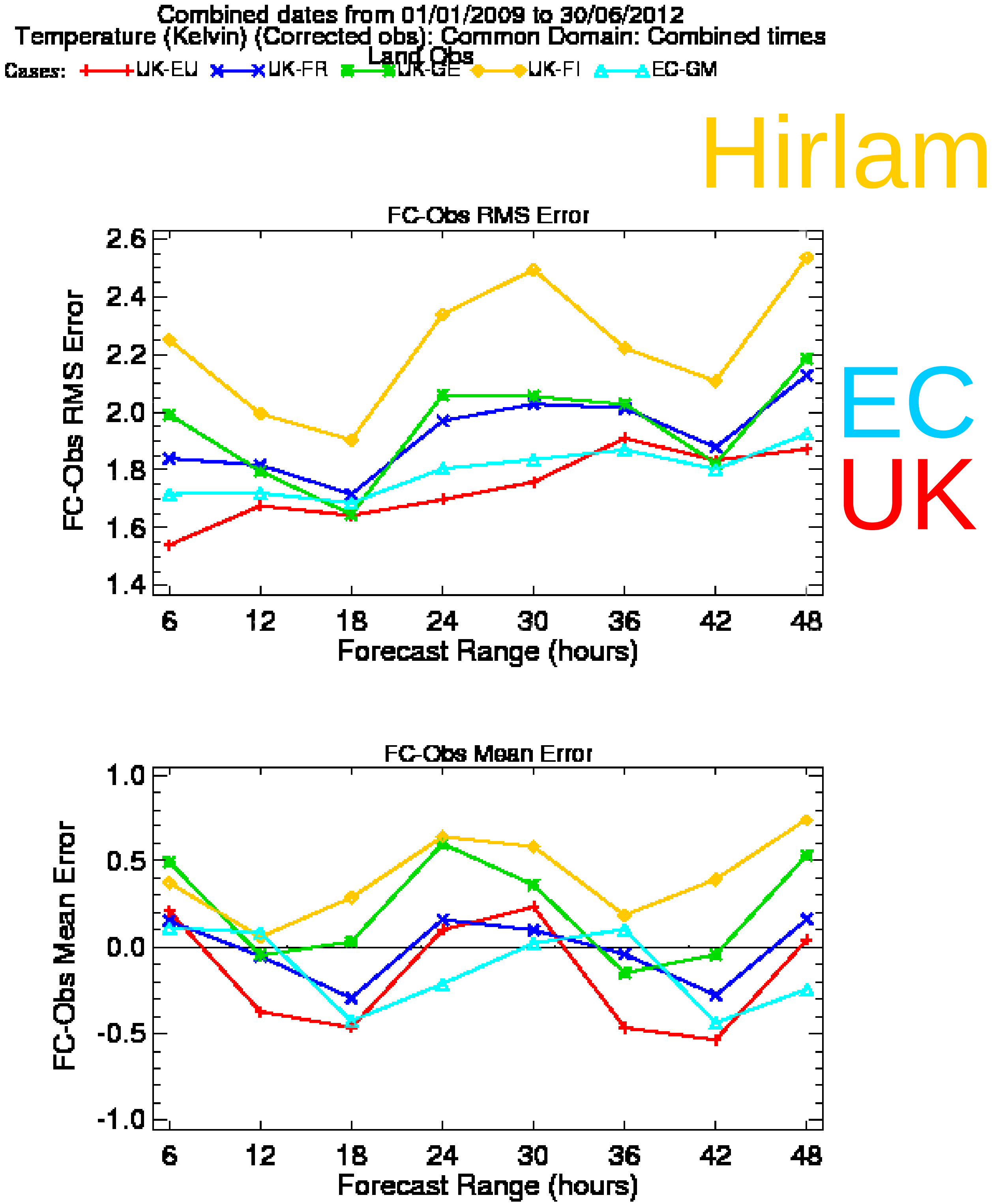
# Sea level Pressure

3.5 year mean 01/2009-06/2012

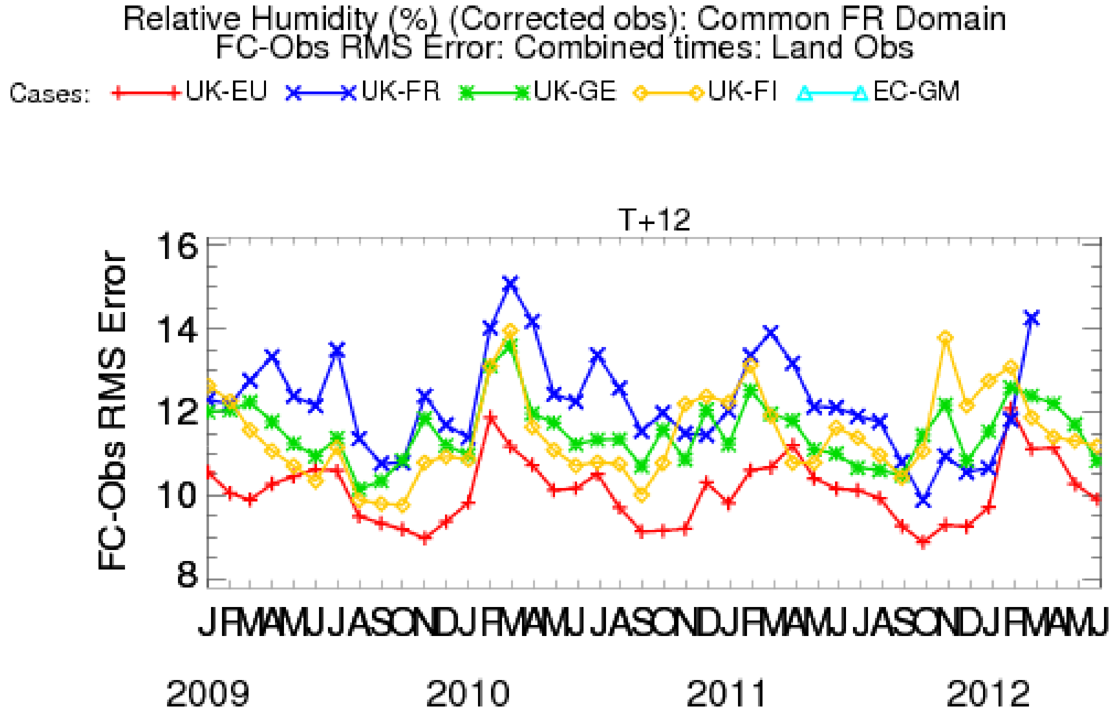


# Screen temperature

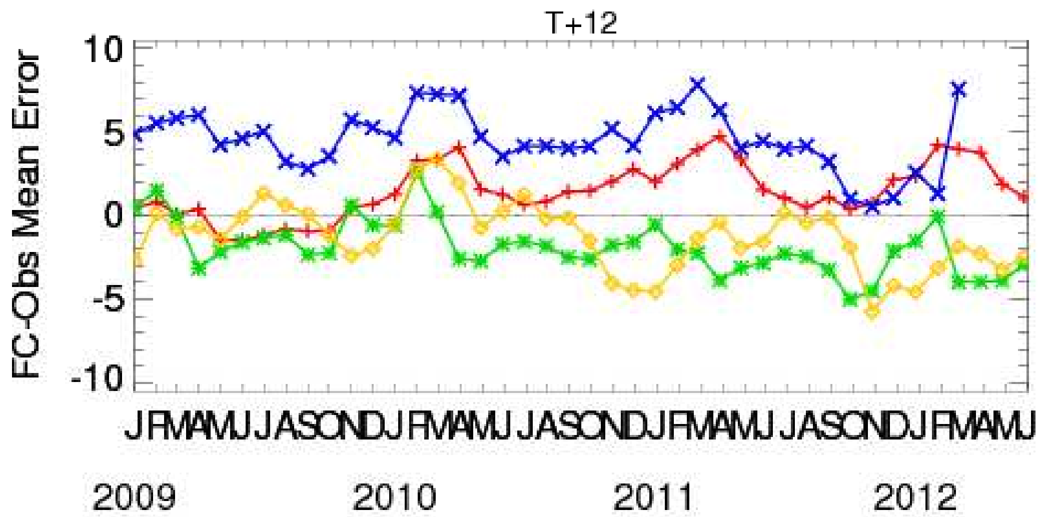
3.5 year mean
01/2009-06/2012



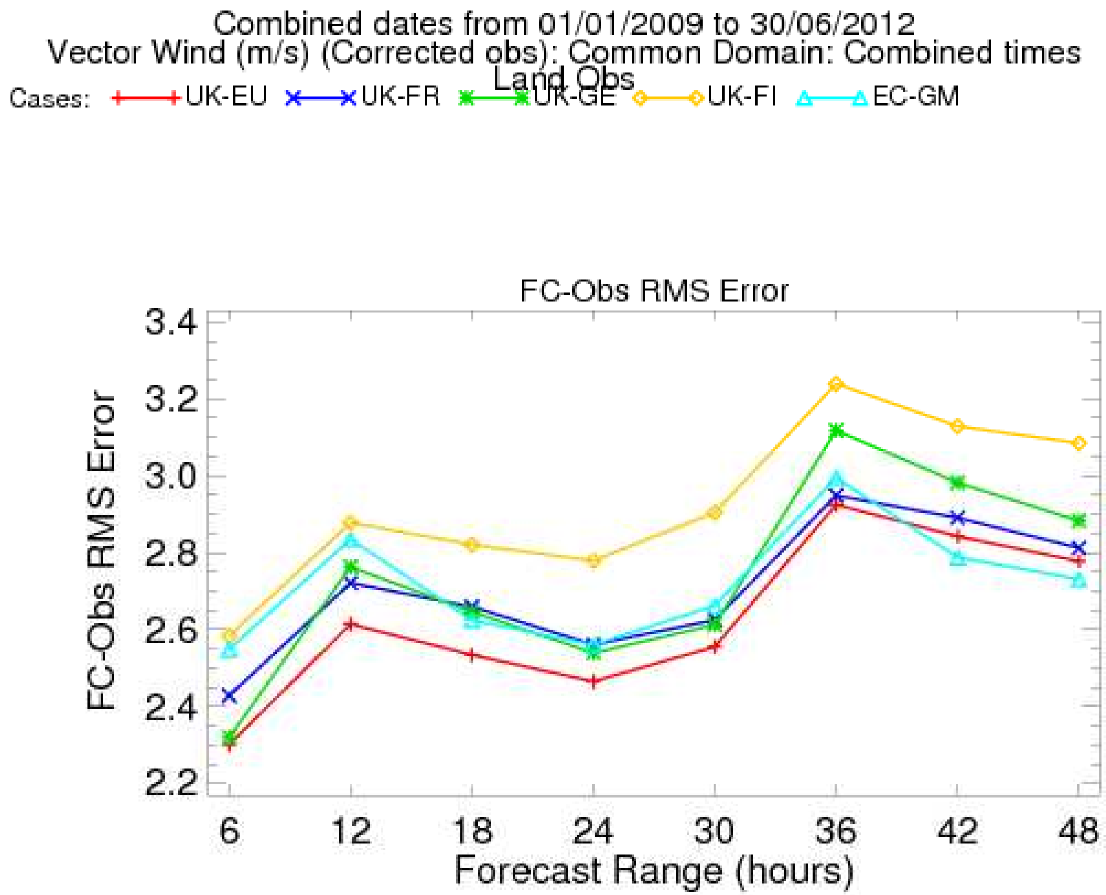
# Relative humidity




NB screen temperature and humidity assimilated in UK-NAE



Met Office



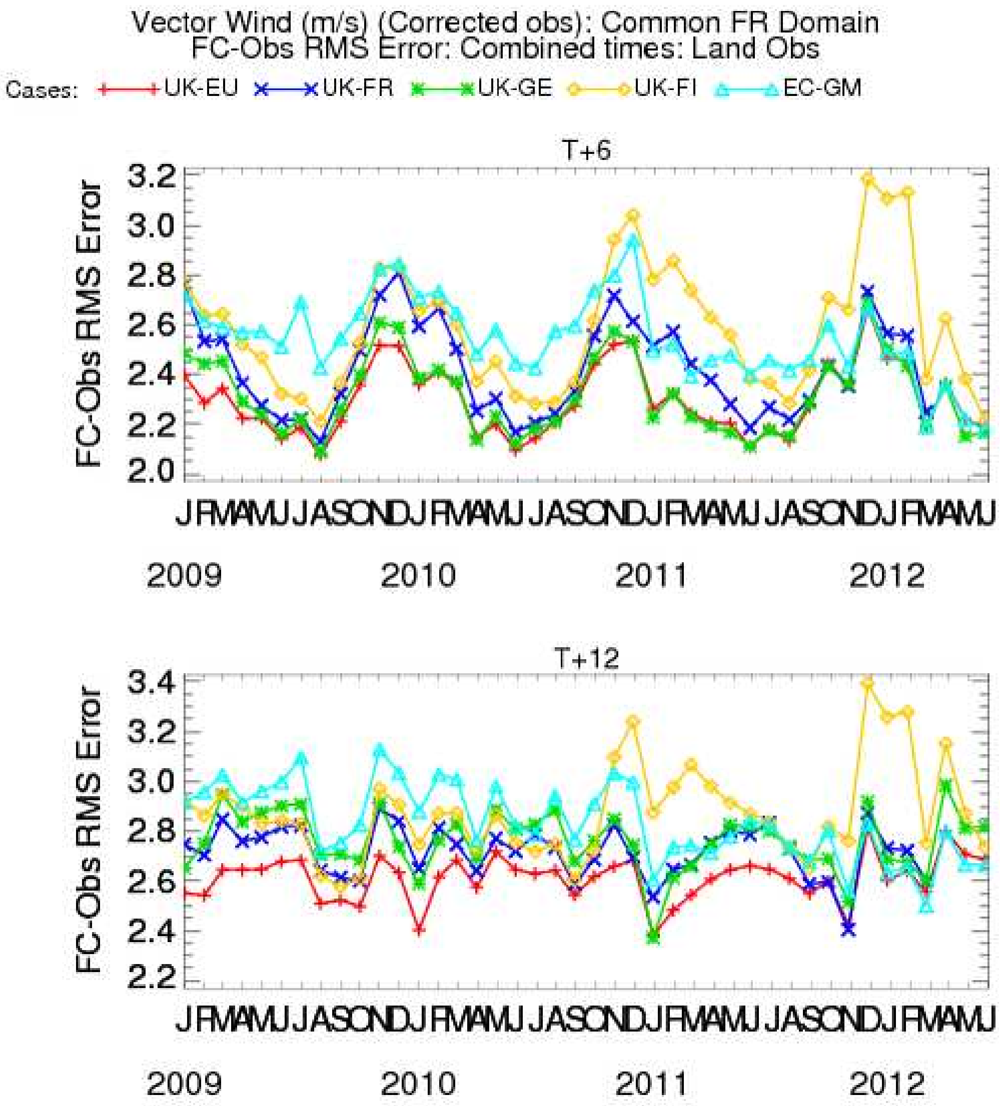

# 10m vector wind rmse

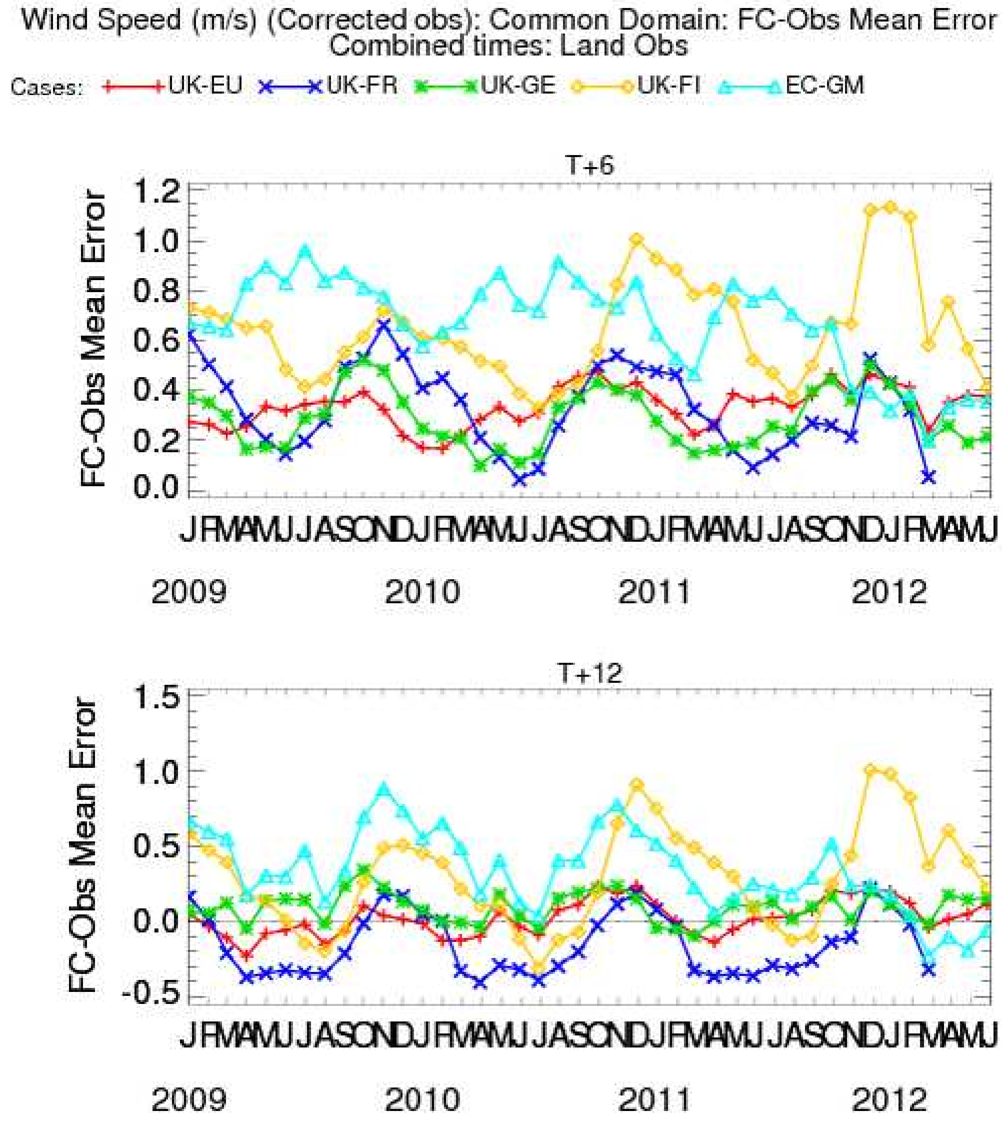
3.5 years 01/2009-06/2012

Regional models better than global especially shorter range

# Vector wind error

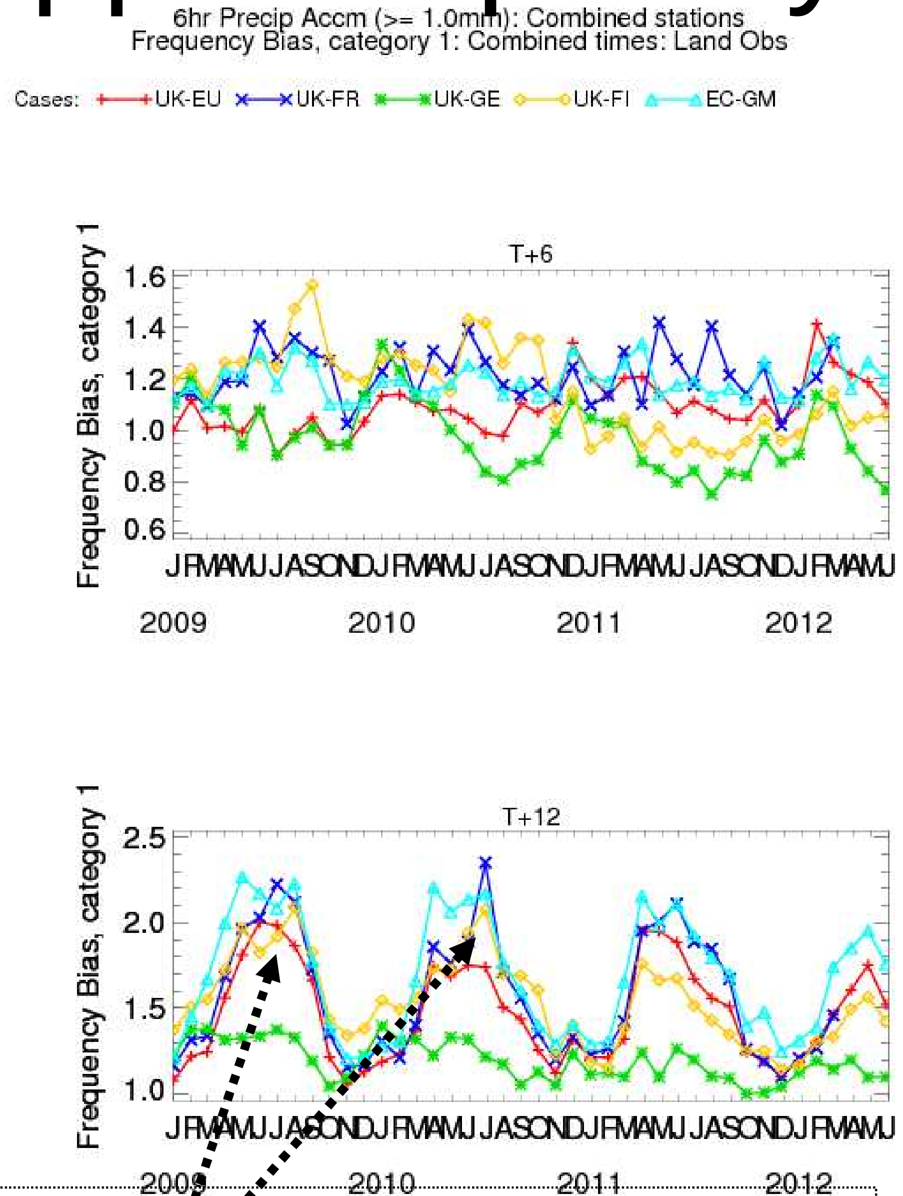
# Speed Bias

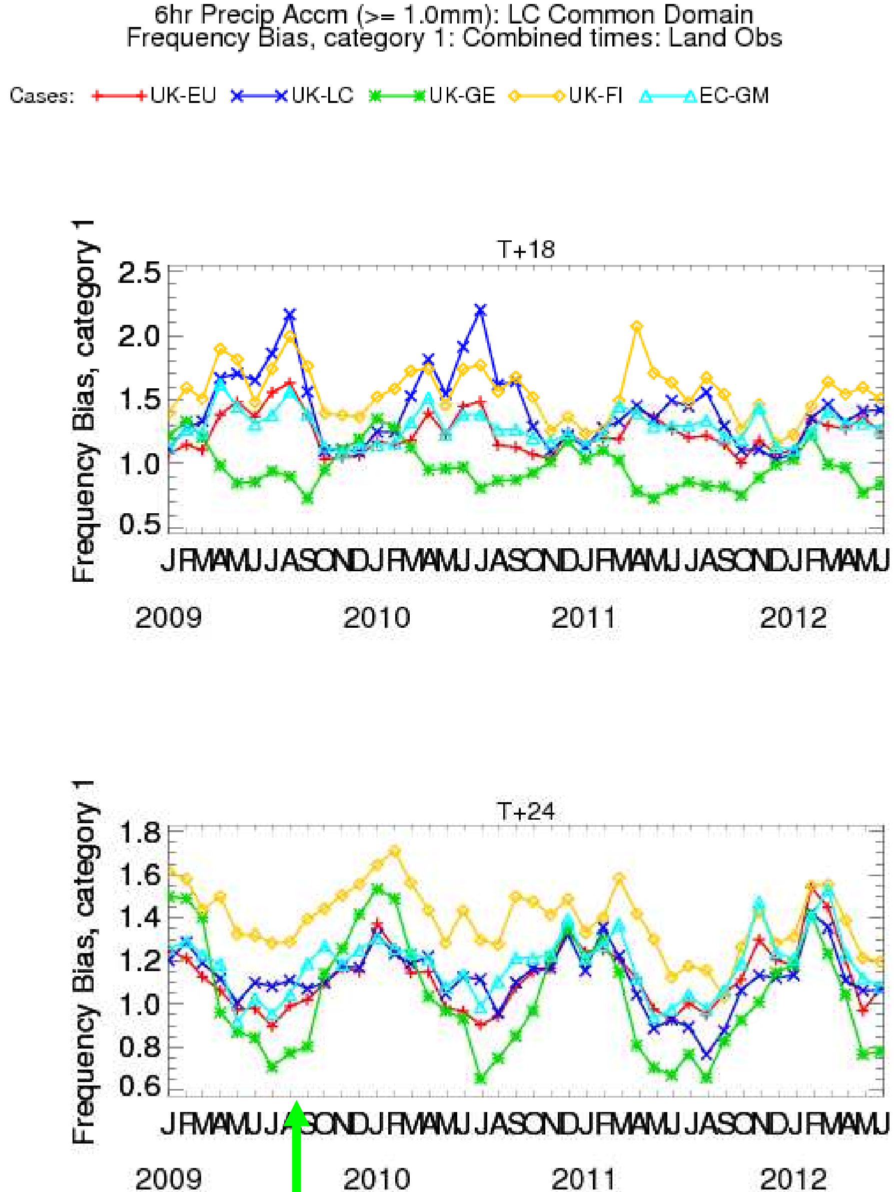



Models closer in 2012, Hirlam relatively poorer since 2011

# ppn frequency bias >1mm/6h




Summer day bias, convection

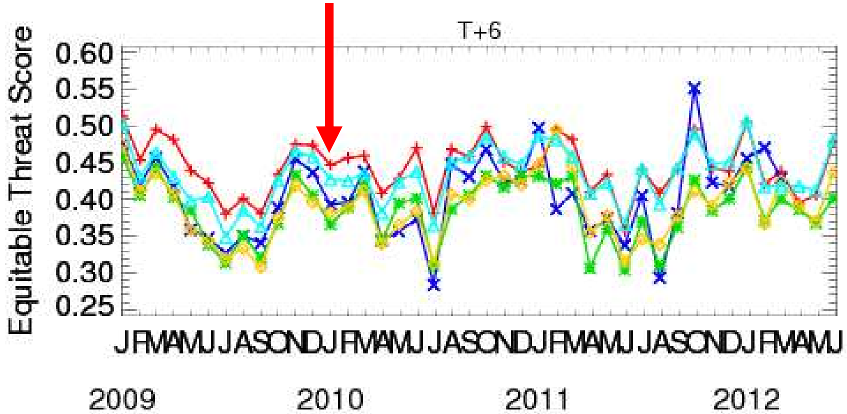
Night negative bias

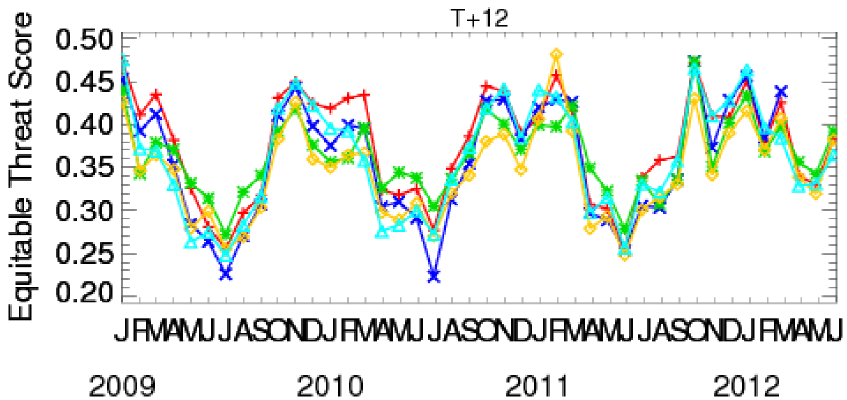
# ppn ETS >1mm/6h

6hr Precip Accm (>= 1.0mm): Combined stations: Equitable Threat Score
Combined times: Land Obs

Cases: UK-EU UK-FR UK-GE UK-FI EC-GM

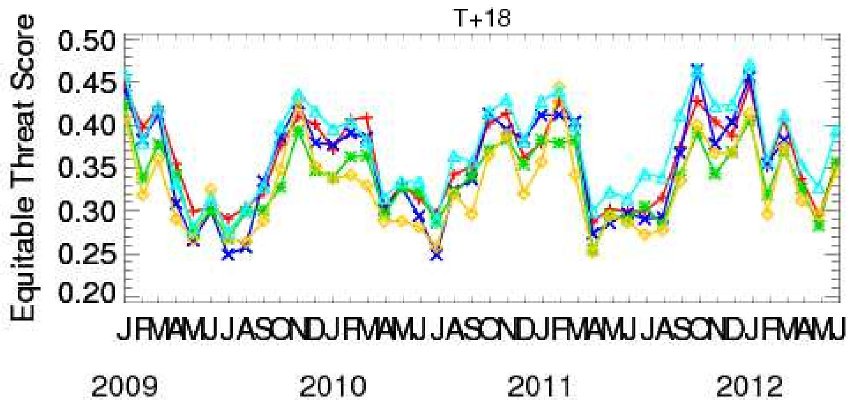
DA – radar/cloud

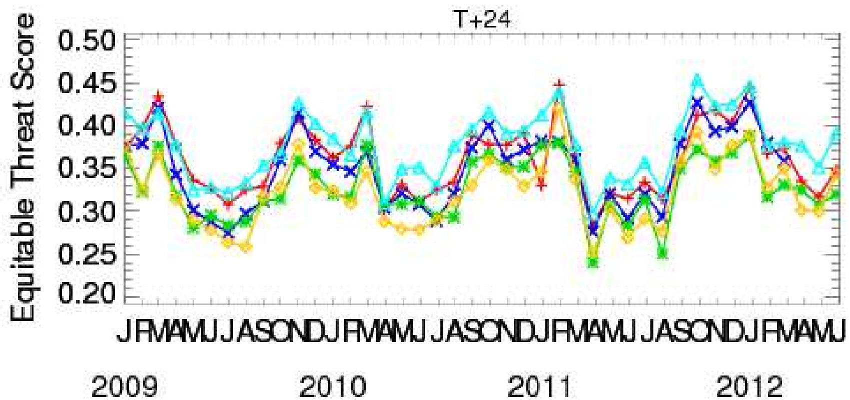


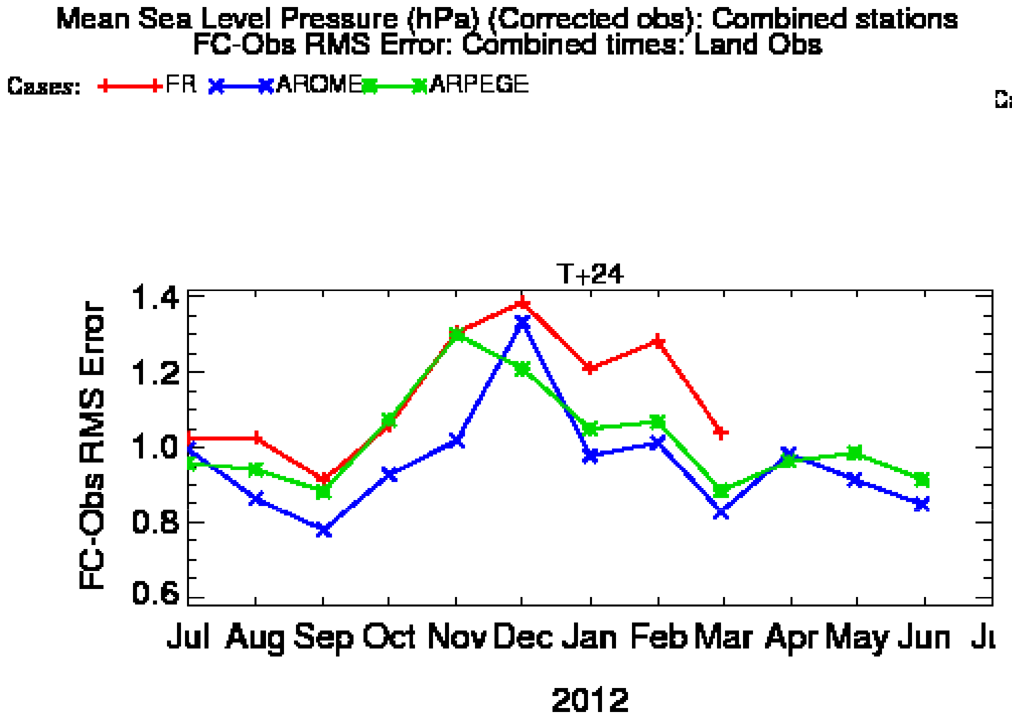



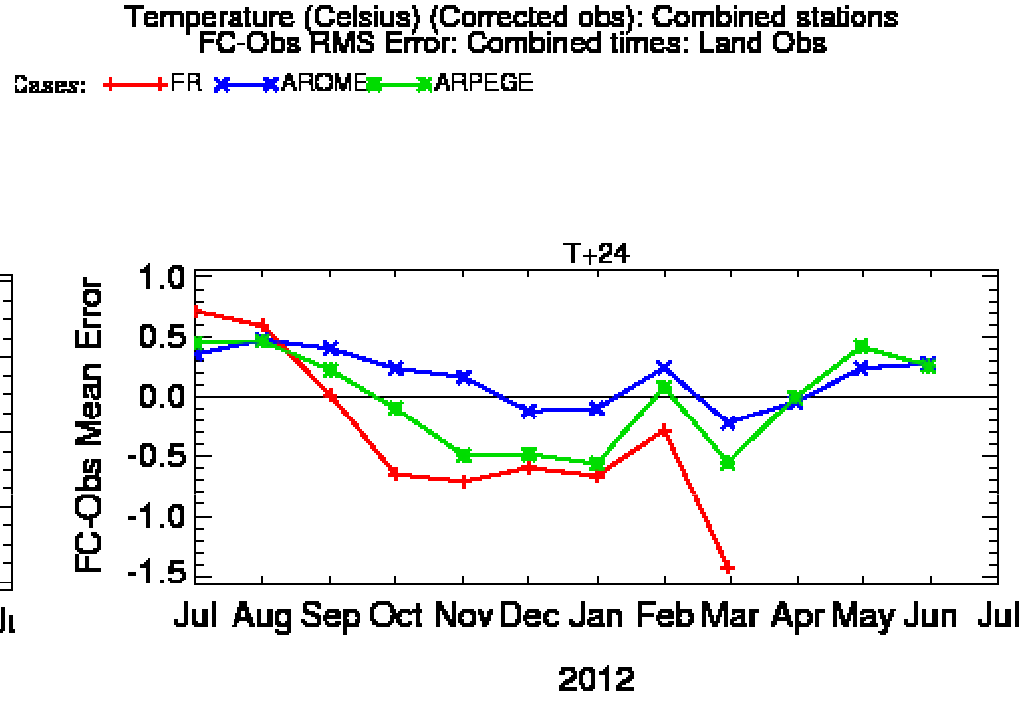

6hr Precip Accm (>= 1.0mm): Combined stations: Equitable Threat Score
Combined times: Land Obs

Cases: UK-EU UK-FR UK-GE UK-FI EC-GM

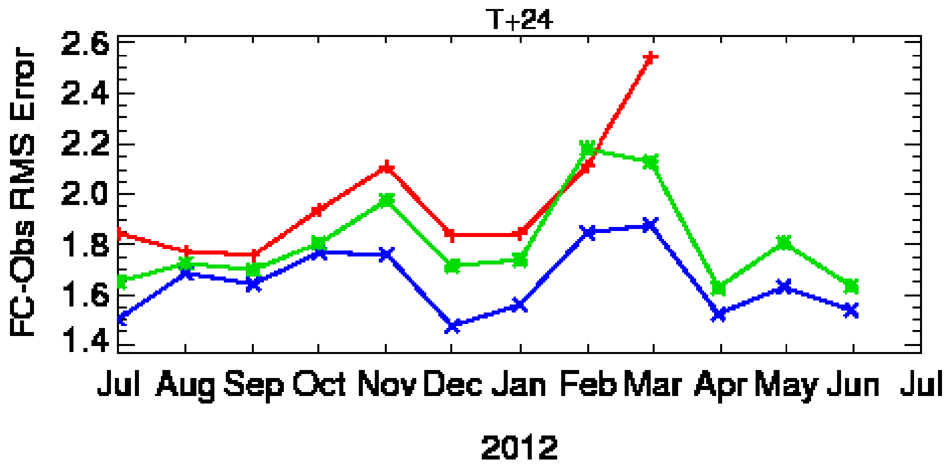





# AROME v ARPEGE v ALADIN

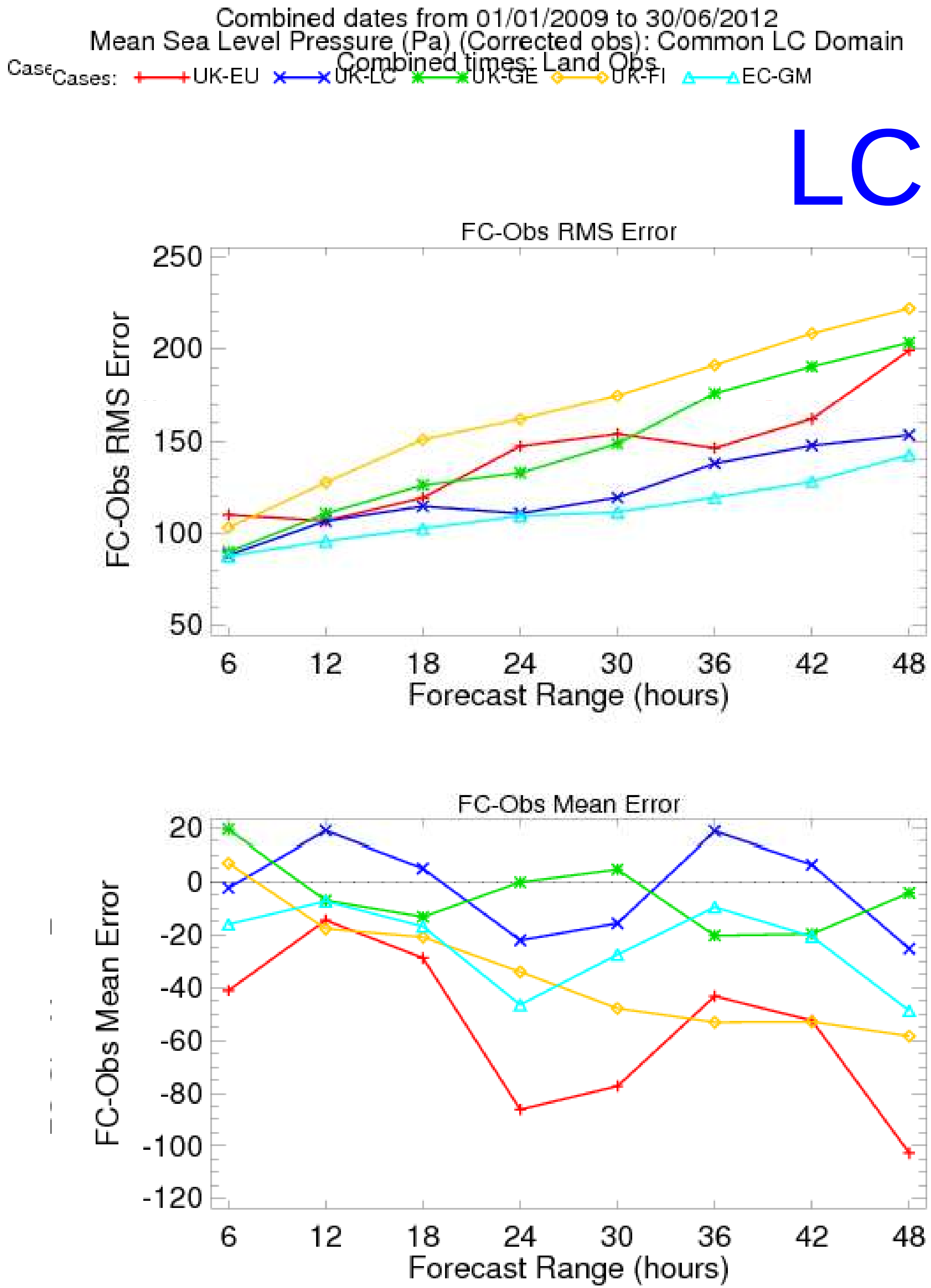


Met Office

# Sea level Pressure

# ALADIN-LACE domain

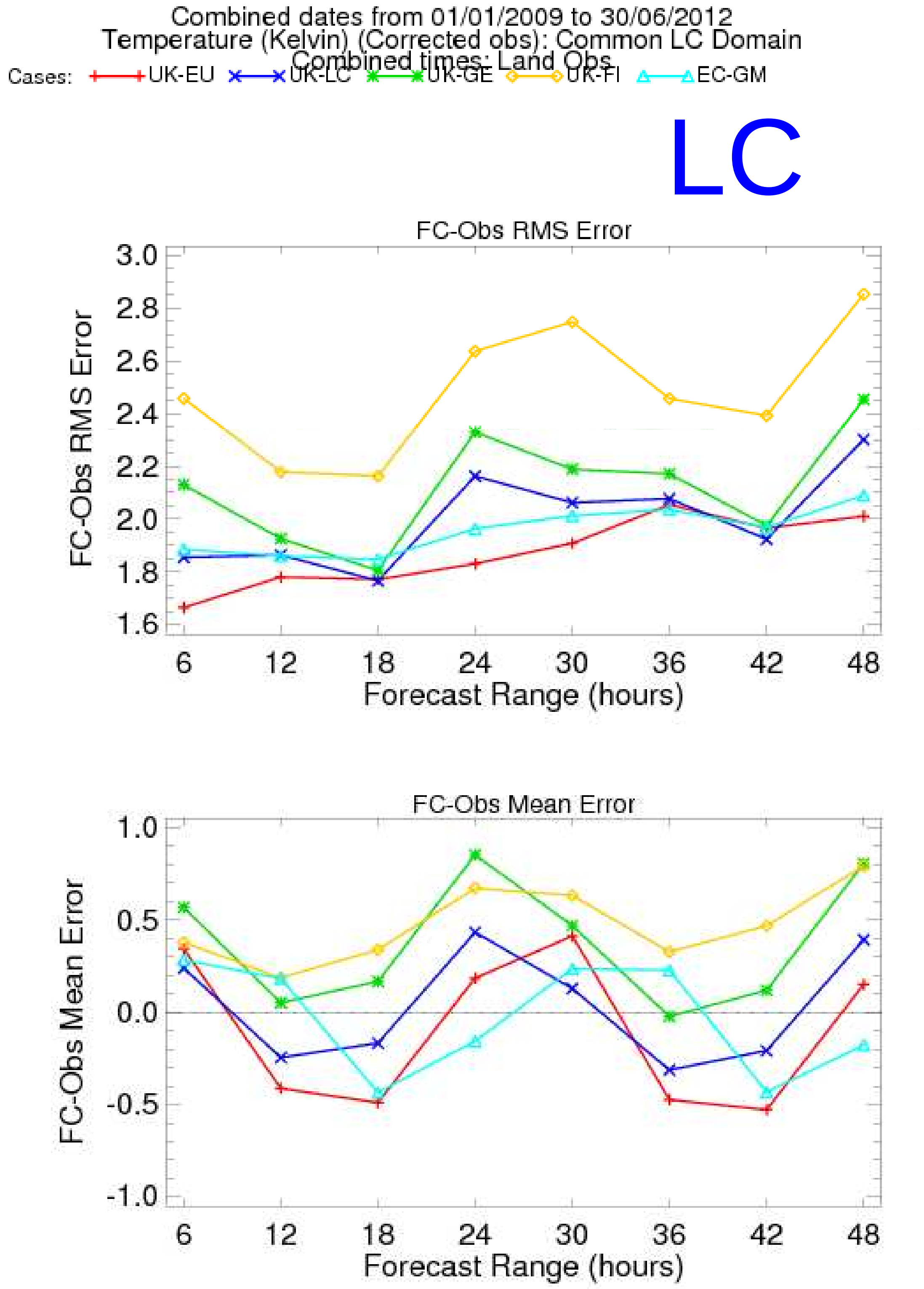
Mean 01/2009-06/2012



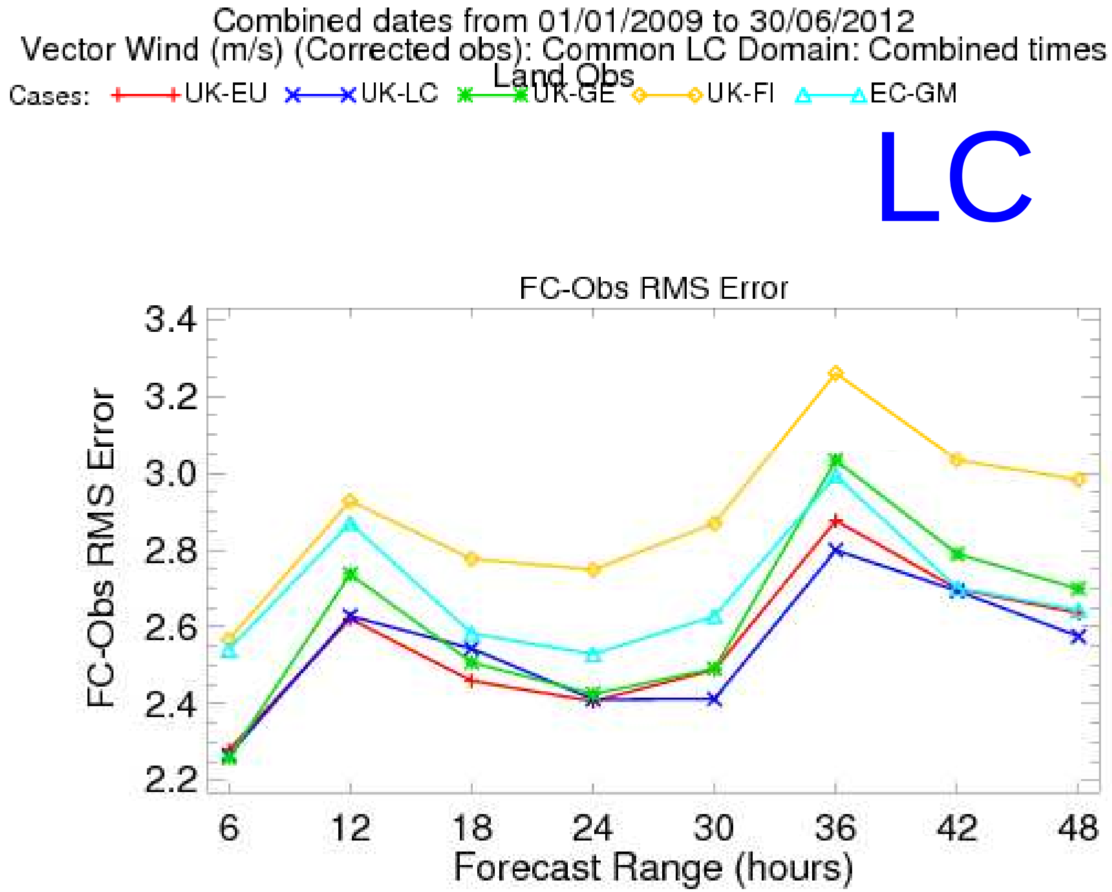
Screen temperature

ALADIN-LACE domain

Mean 01/2009-06/2012



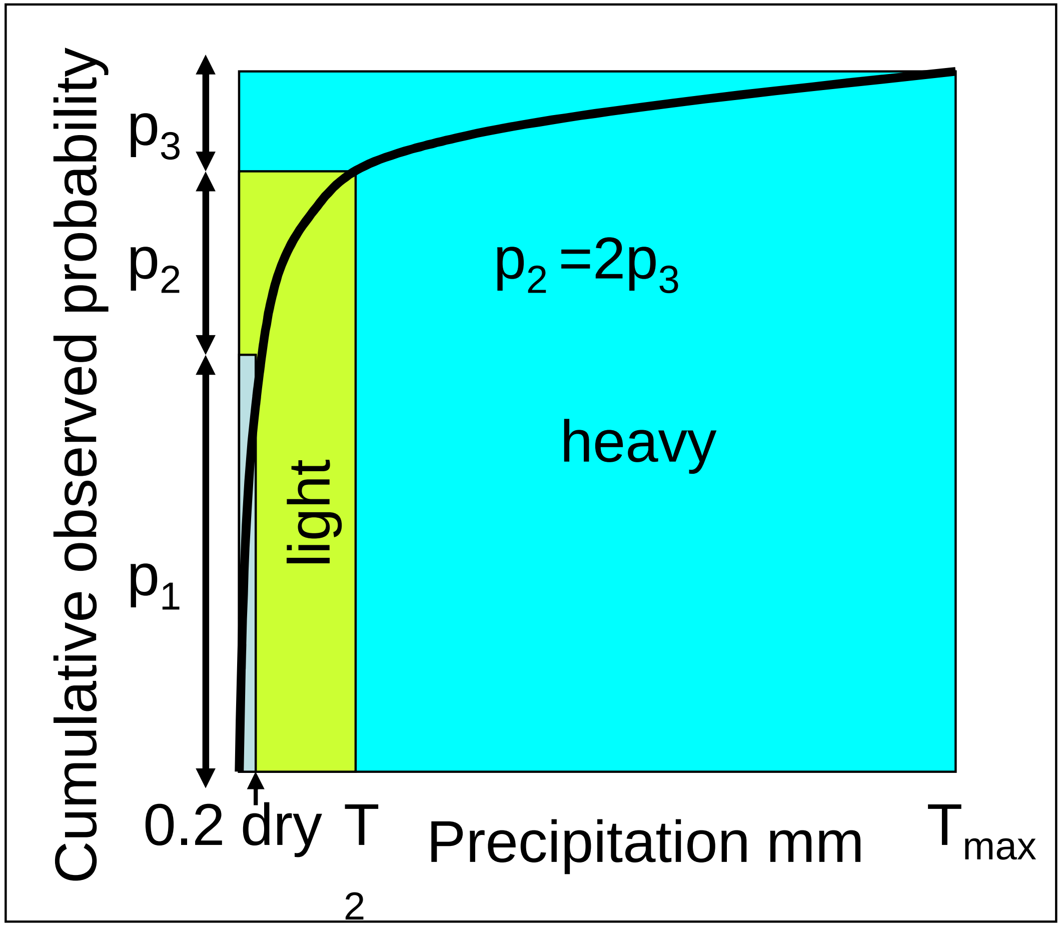
# 10m vector wind rmse ALADIN-LACE domain



# SEEPS=Stable Equitable Error in Probability Space

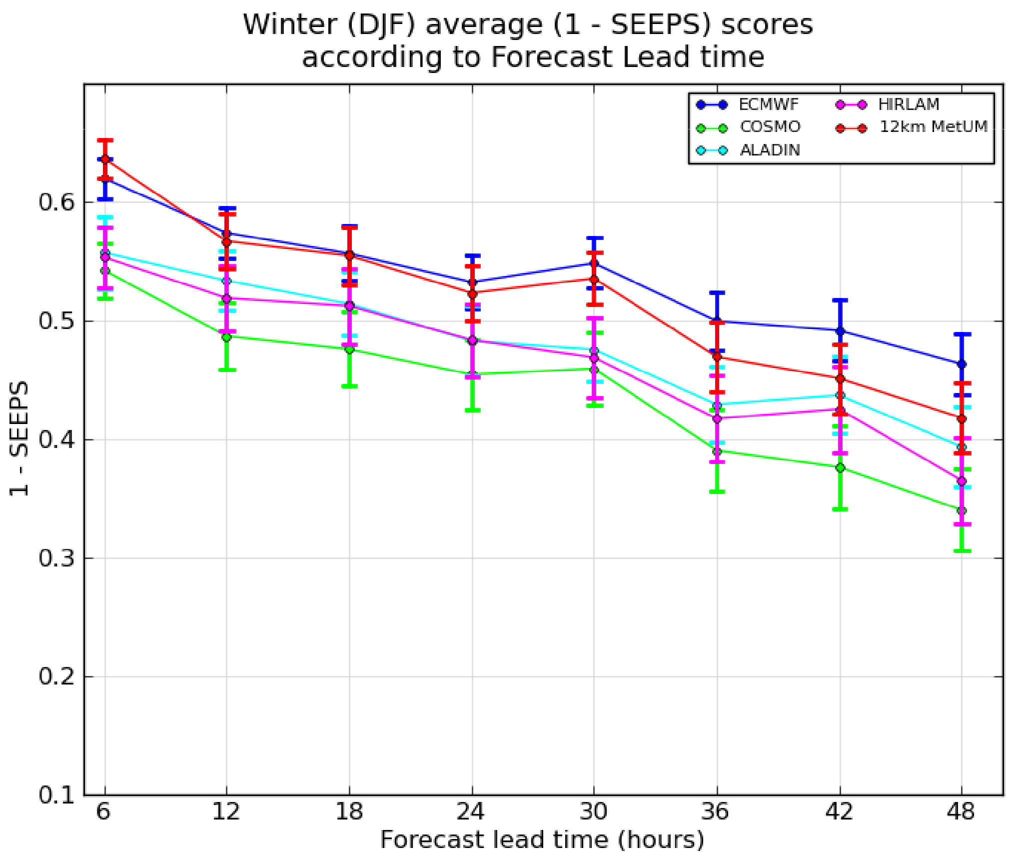
Rodwell et al, 2010, QJRMS 136

- Dry, light , heavy based on observed climatology (24h) at station – $p_1$ , $p_2$ , $p_3$
- Contingency table probabilities based on these categories
- Scoring matrix – stable, equitable
    - SEEPS=0 (perfect) , =1 ( no skill - , eg constant)
- Now applying to 6h accumulations in SRNWP-V
    - 6h climatology (courtesy Mark Rodwell)



# Winter (DJF) 2009-2012 data $3^2/_3$ winters

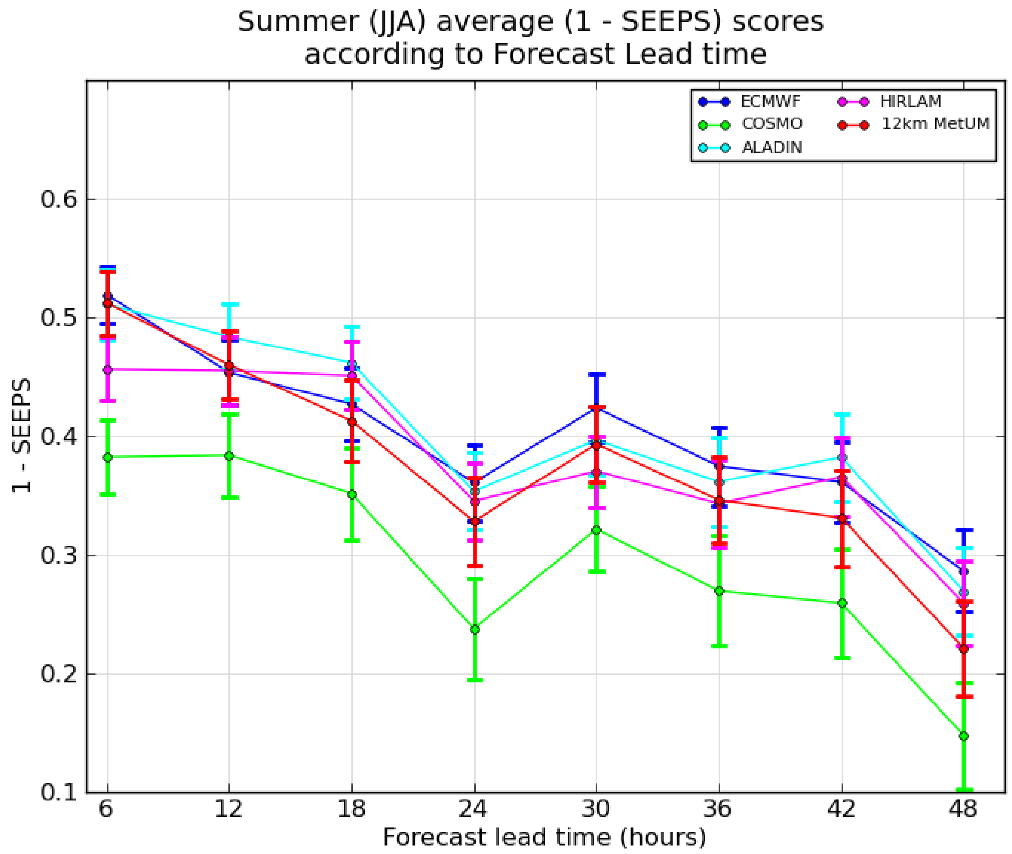
- Higher skill than summer
- EC best at longer range
- 3 groupings
    - EC/UM
    - Aladin/Hirlam
    - COSMO



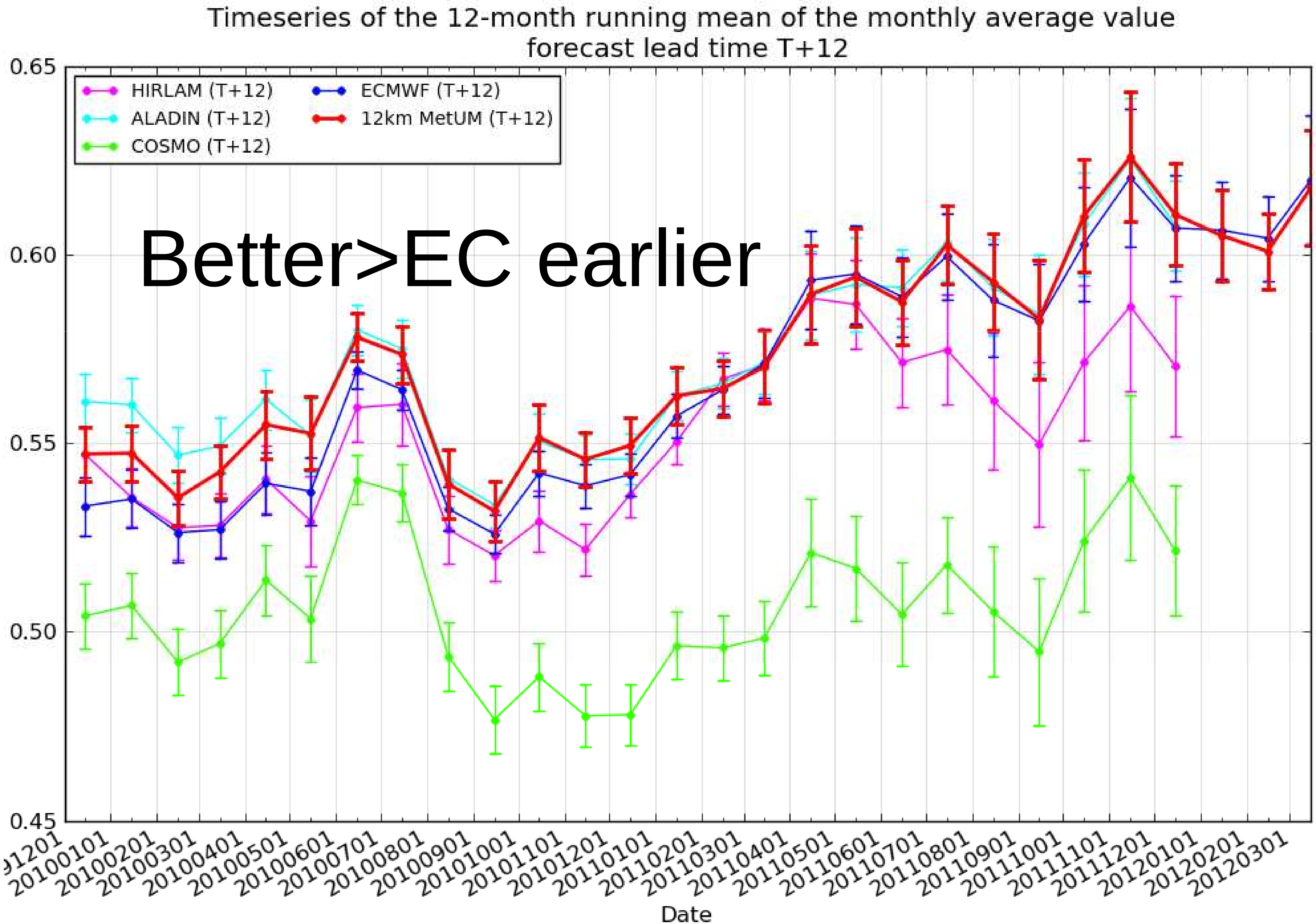

Error bars 70% confidence intervals

# Summer (JJA) 2009-2011 data
# 3 summers

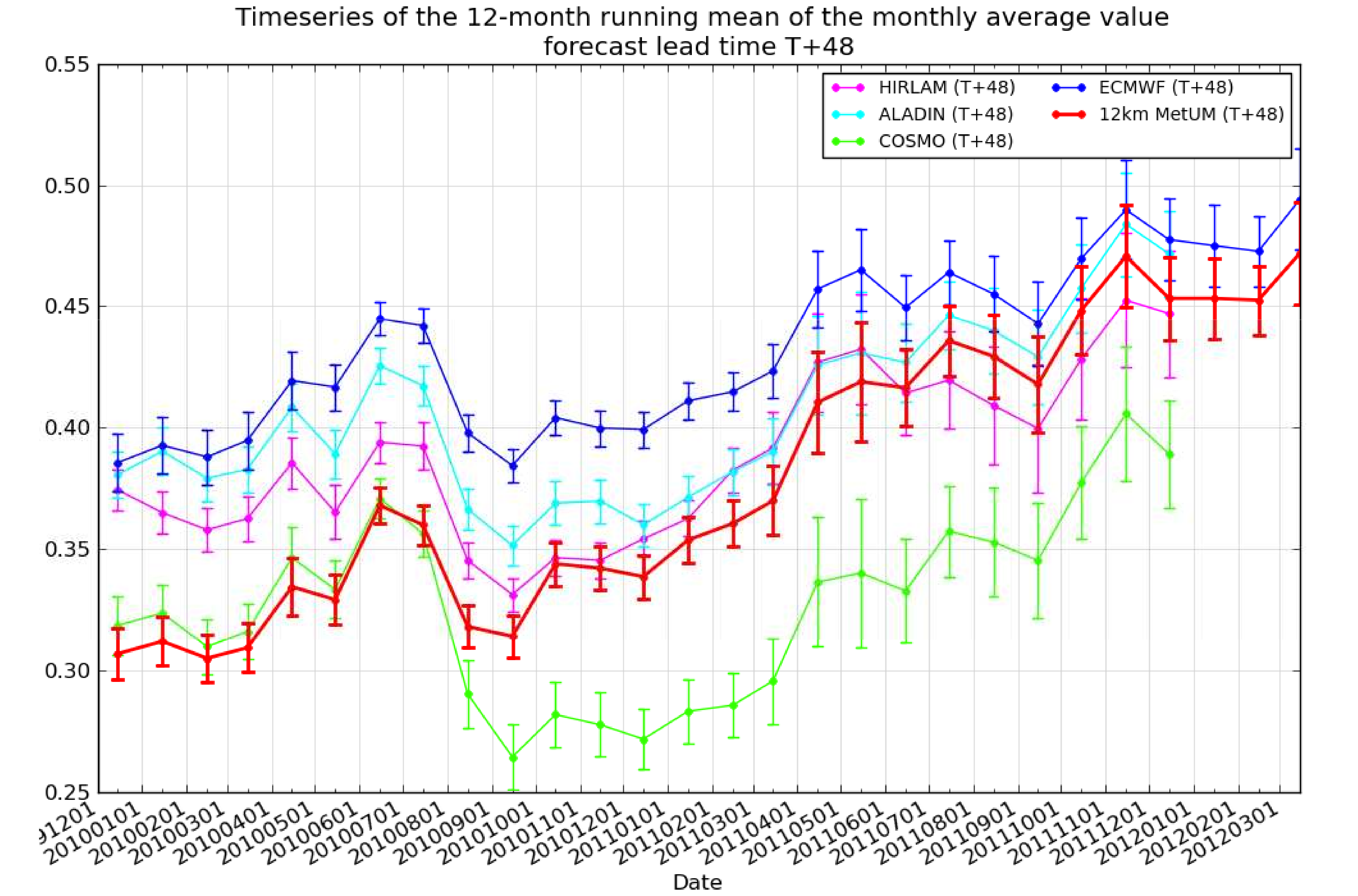
- Lower skill ( ≈ -0.1) than winter
- 2 groupings
  - EC/UM/Aladin/Hirlam
  - COSMO
  - Aladin >EC
- Dip in skill at T+24 (evening-night)
- Lack of showers persisting into late evening




Error bars 70% confidence intervals



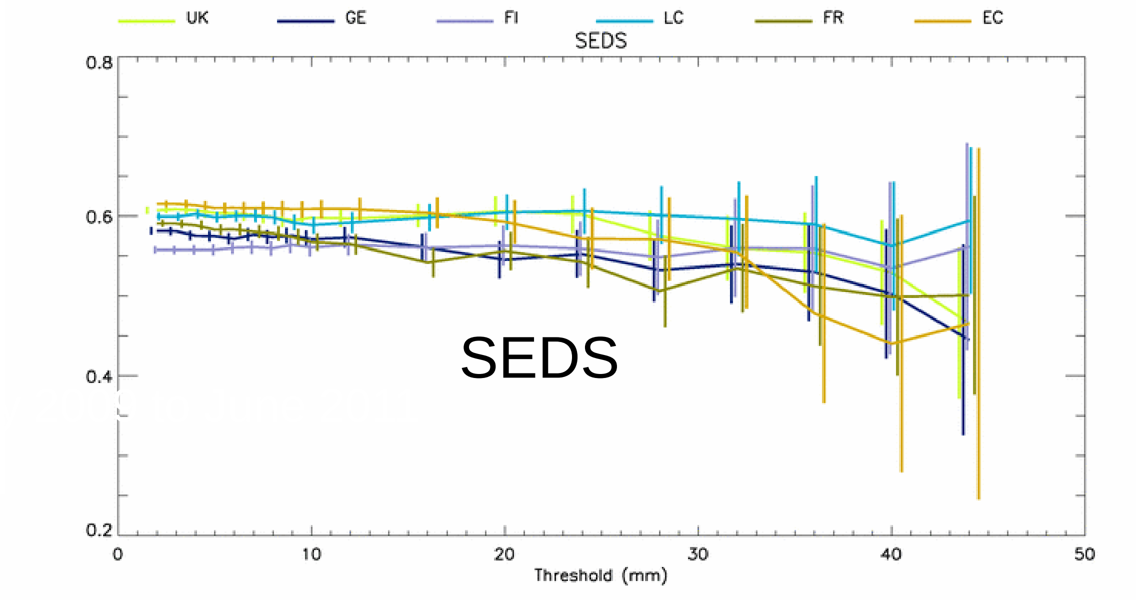

All models show improving trend

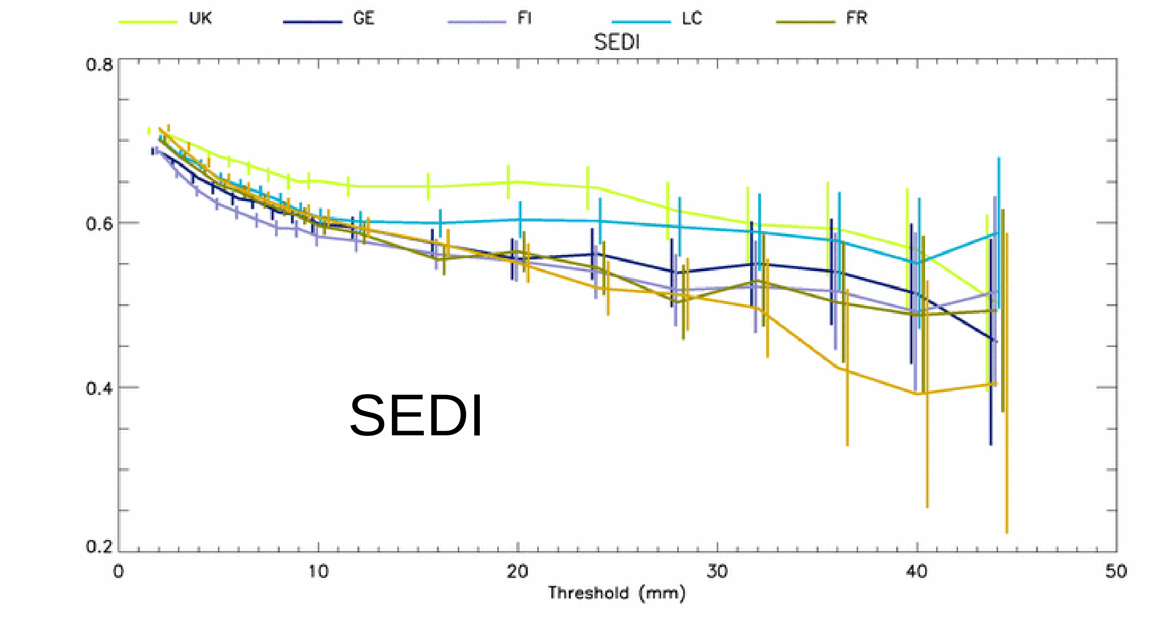

Timeseries of the 12-month running mean of the monthly average value
forecast lead time T+48
HIRLAM (T+48)
ALADIN (T+48)
COSMO (T+48)
ECMWF (T+48)
12km MetUM (T+48)
0.55
0.50
0.45
0.40
0.35
0.30
0.25
20100101
20100201
20100301
20100401
20100501
20100601
20100701
20100801
20100901
20101001
20101101
20101201
20110101
20110201
20110301
20110401
20110501
20110601
20110701
20110801
20110901
20111001
20111101
20111201
20120101
20120201
20120301
Date

# Extremal dependency scores

$$\text{SEDS} = \frac{\log(q) - \log(H)}{\log(p) + \log(H)} \qquad \text{SEDI} = \frac{\log(F) - \log(H) - \log(1-F) + \log(1-H)}{\log(F) + \log(H) + \log(1-F) + \log(1-H)}$$

Jan 2009 – June 2011



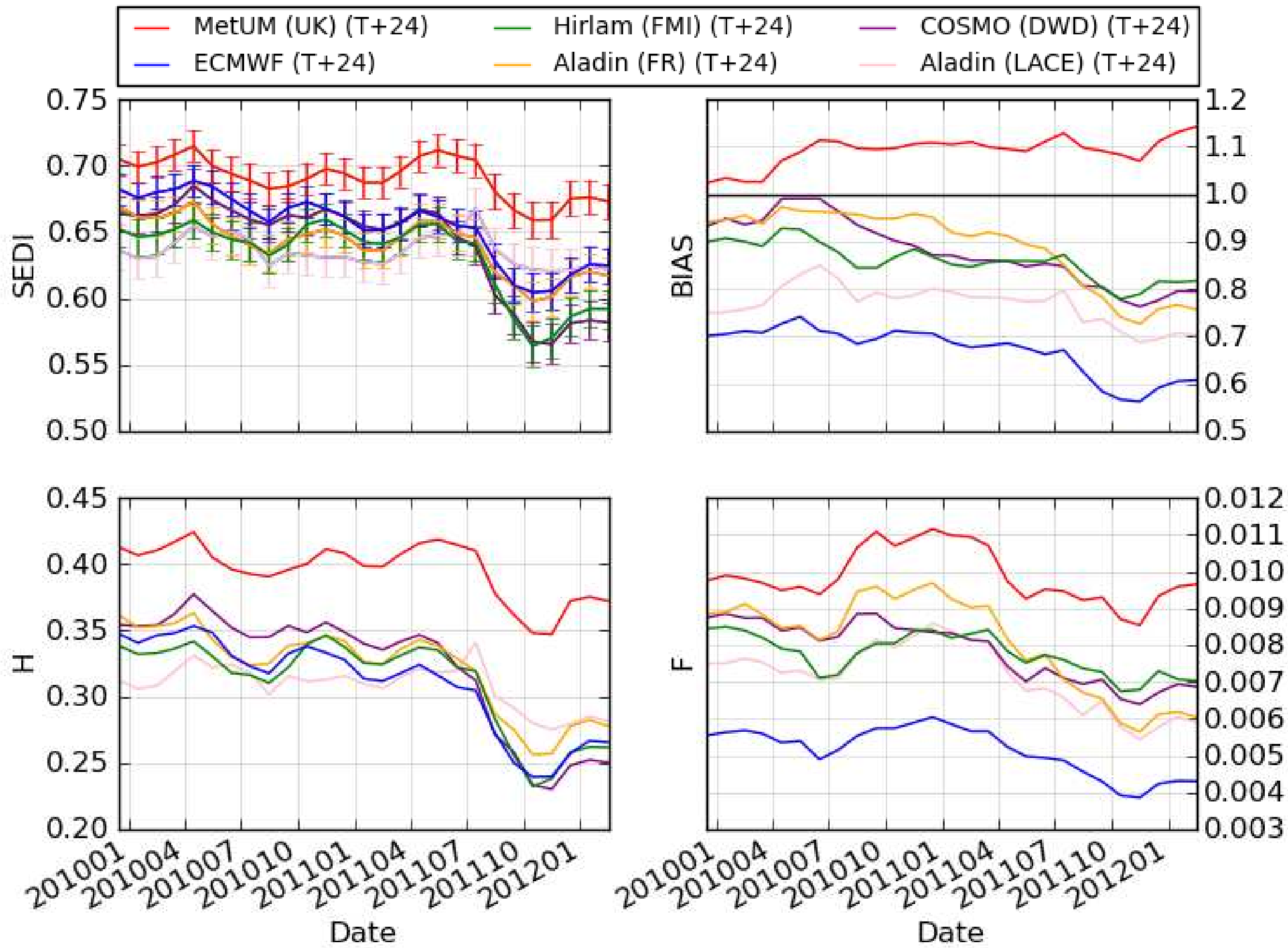




Matthew Trueman

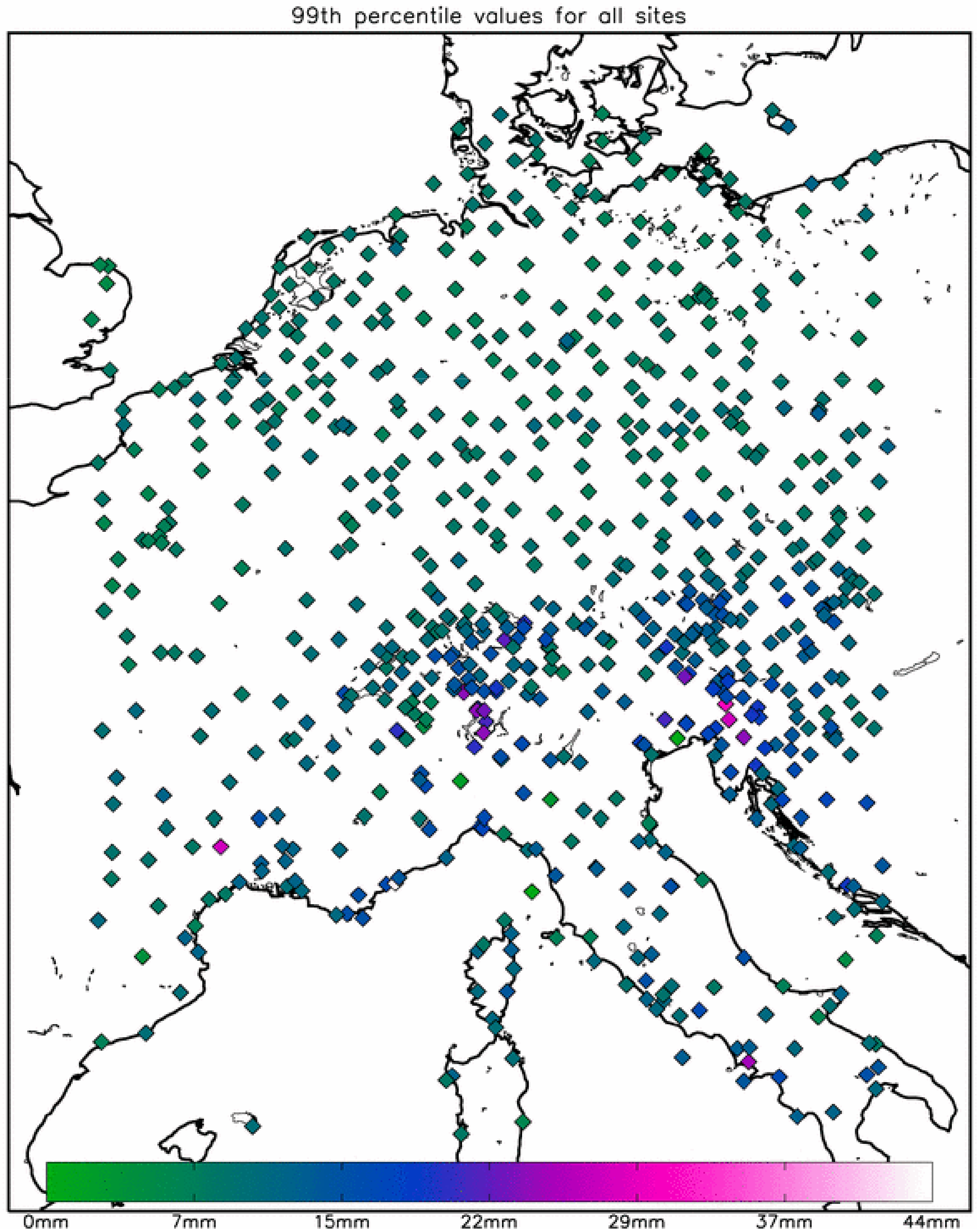
# Extremal dependency scores

$$\text{SEDI} = \frac{\log(F) - \log(H) - \log(1-F) + \log(1-H)}{\log(F) + \log(H) + \log(1-F) + \log(1-H}$$

- 99th percentile threshold
- LAMS have better bias than ECMWF
- EC under-forecast extremes
- H, hit rate LAMS better




Matthew Trueman


99th percentile values for all sites
0mm
7mm
15mm
22mm
29mm
37mm
44mm

# ND3: Spatial & scale selective verification of precipitation and cloud

- Verify against
  - Gridded analyses- ECMWF, Meteo-France, Met Office (UK only)
    - Other national gridded sets ?
  - High resolution radar (5 min,1-2 km) -UK
  - OPERA radar composite QC / gauge bias correction

- Methods
  - Fractional skill (Roberts & Lean)
  - Upscaling
  - Intensity scale (Casati)
  - Structure,amplitude,location (SAL) (Wernli et al)
  - Contiguous rain areas (Ebert & McBride)

Met Office

Fuzzy verification

using SEVERI cloud mask as truth binary 0 or 1

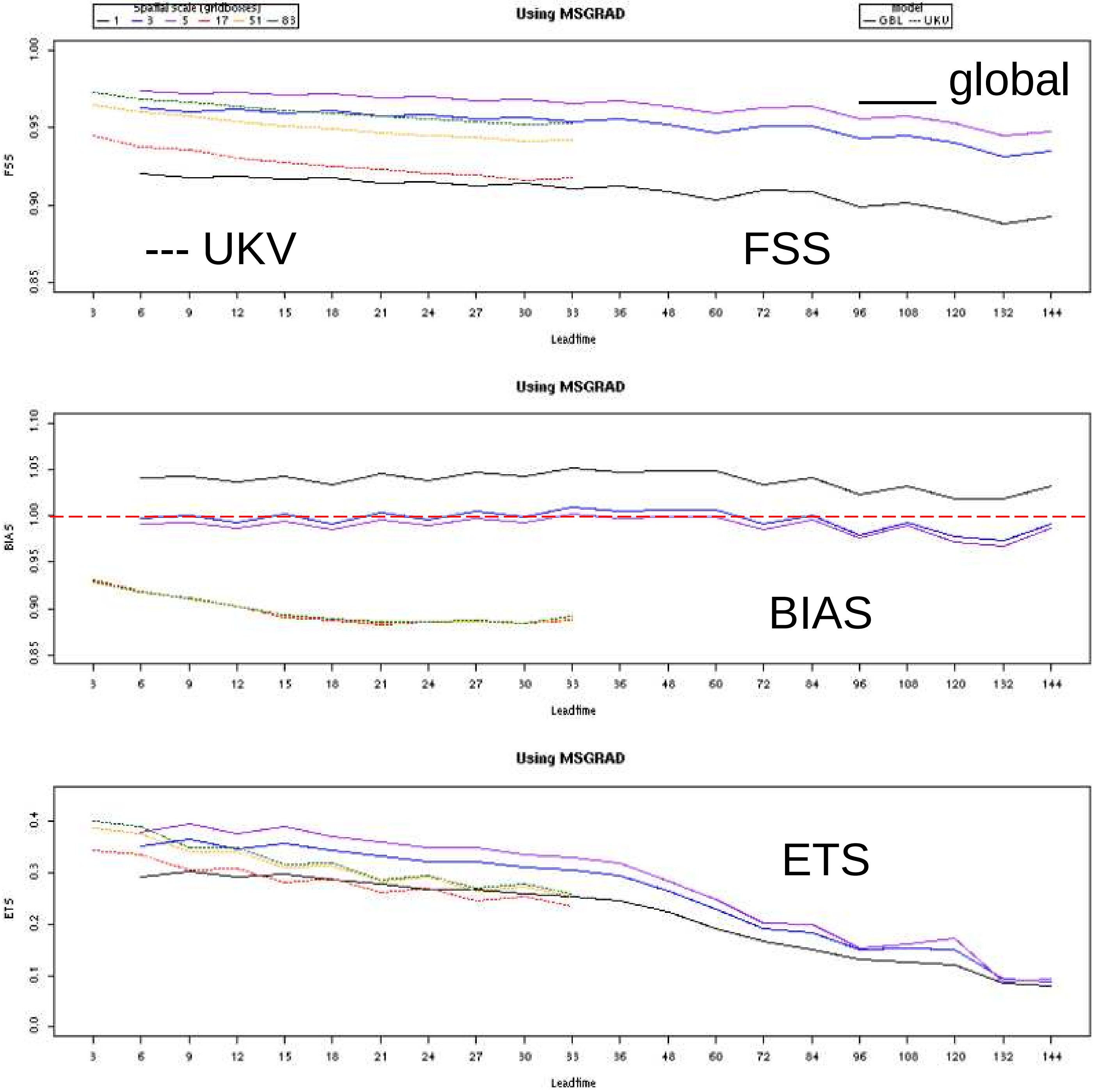
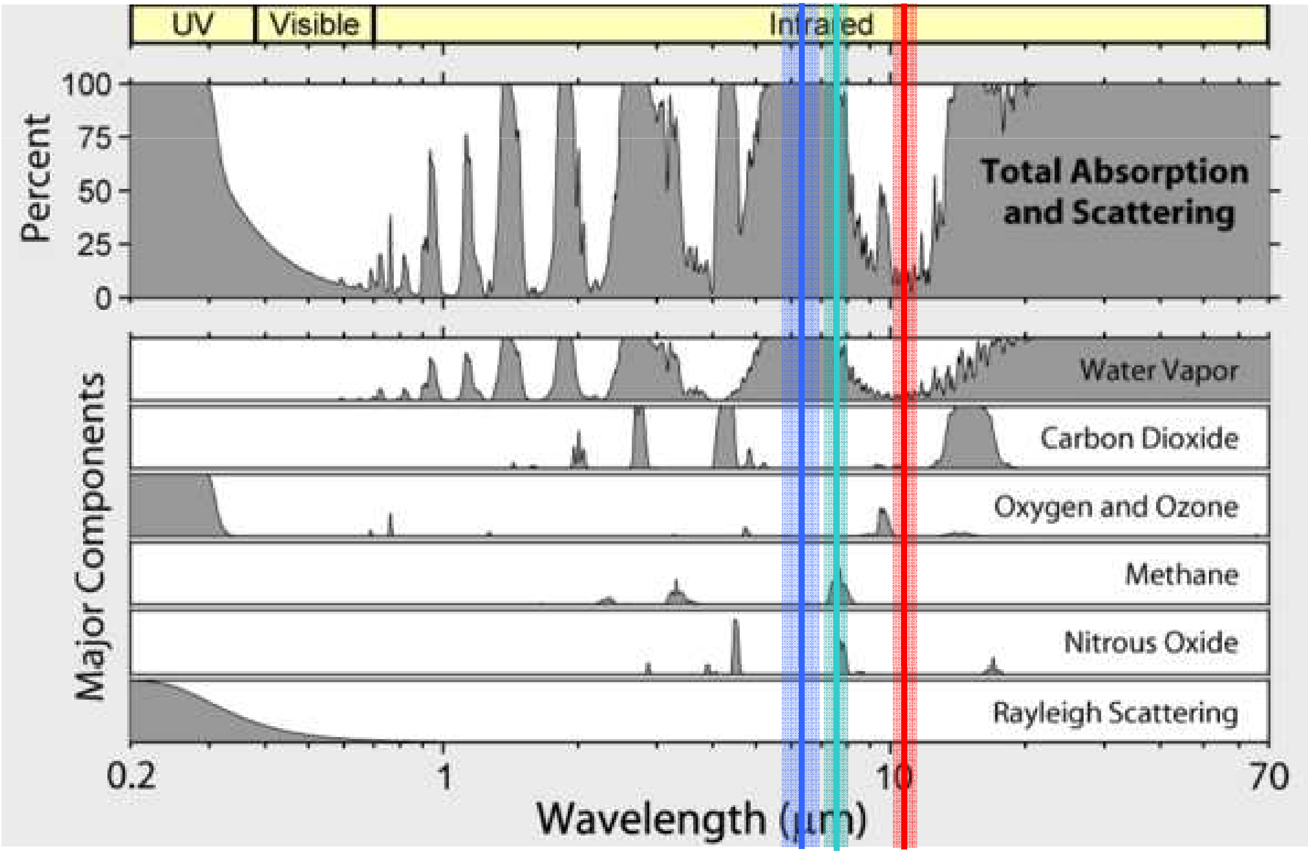



Figure 1 - FSS, bias and ETS scores for a cloud threshold >0.375 when verified against the MSGRAD cloud mask data. All spatial scales of the GM and UKV models are displayed.

# Simulating images to compare to SEVIRI[1]

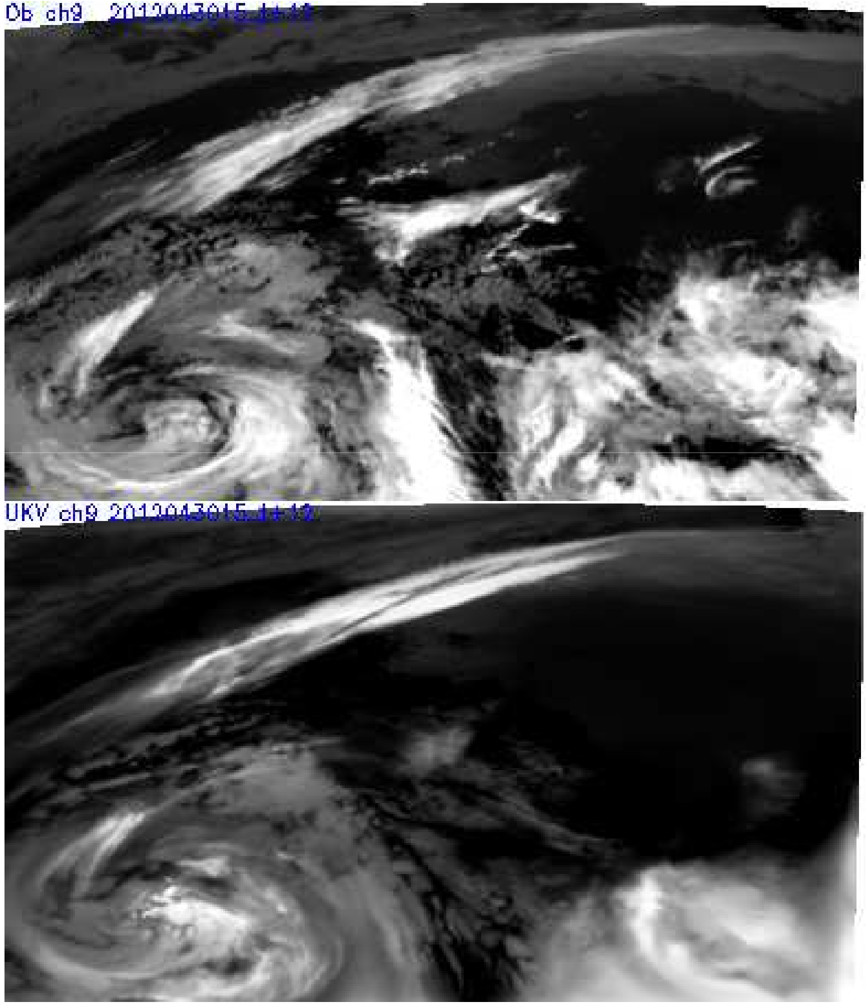




Ch05 6.2 μm - upper water vapour sensitive

Ch06 7.3 μm

Ch09 10.8 μm- detects cloud and surface

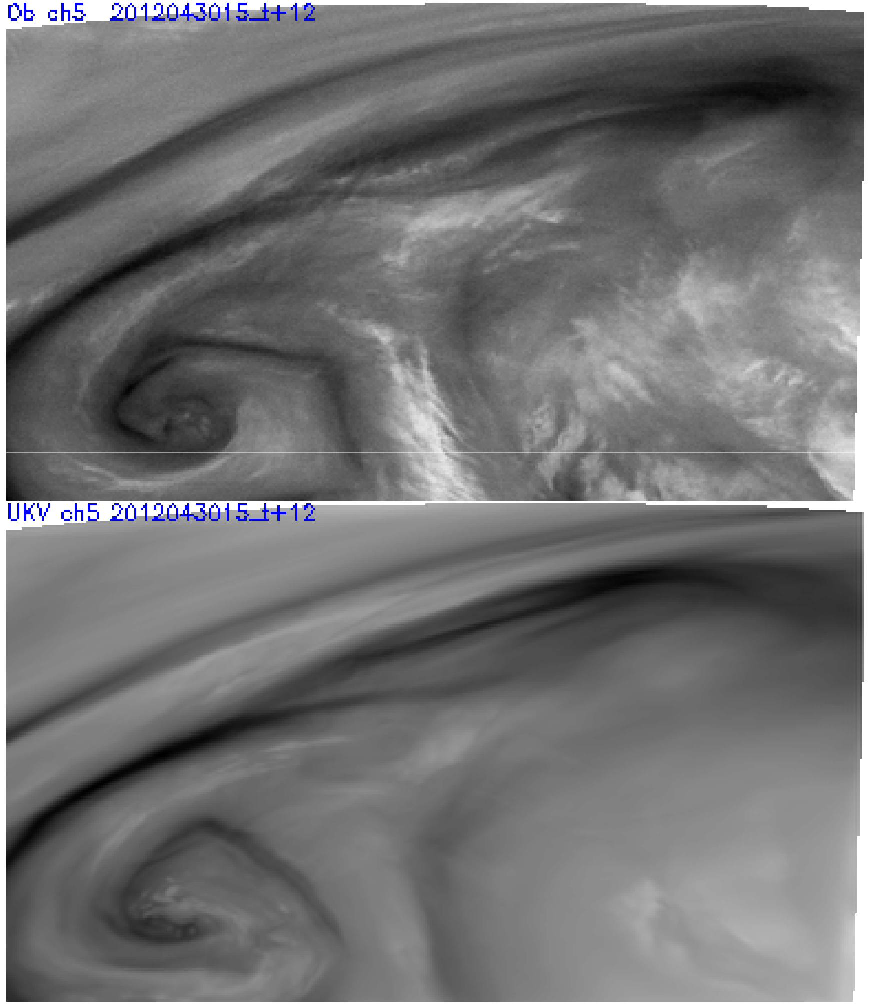
1. Spinning Enhanced Visible and Infrared Imager




SEVIRI

UKV model

T+12

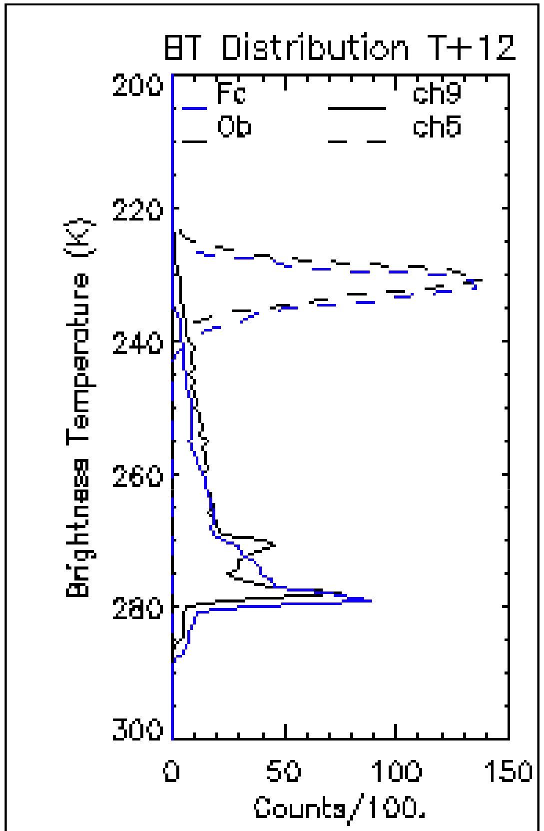



ch5

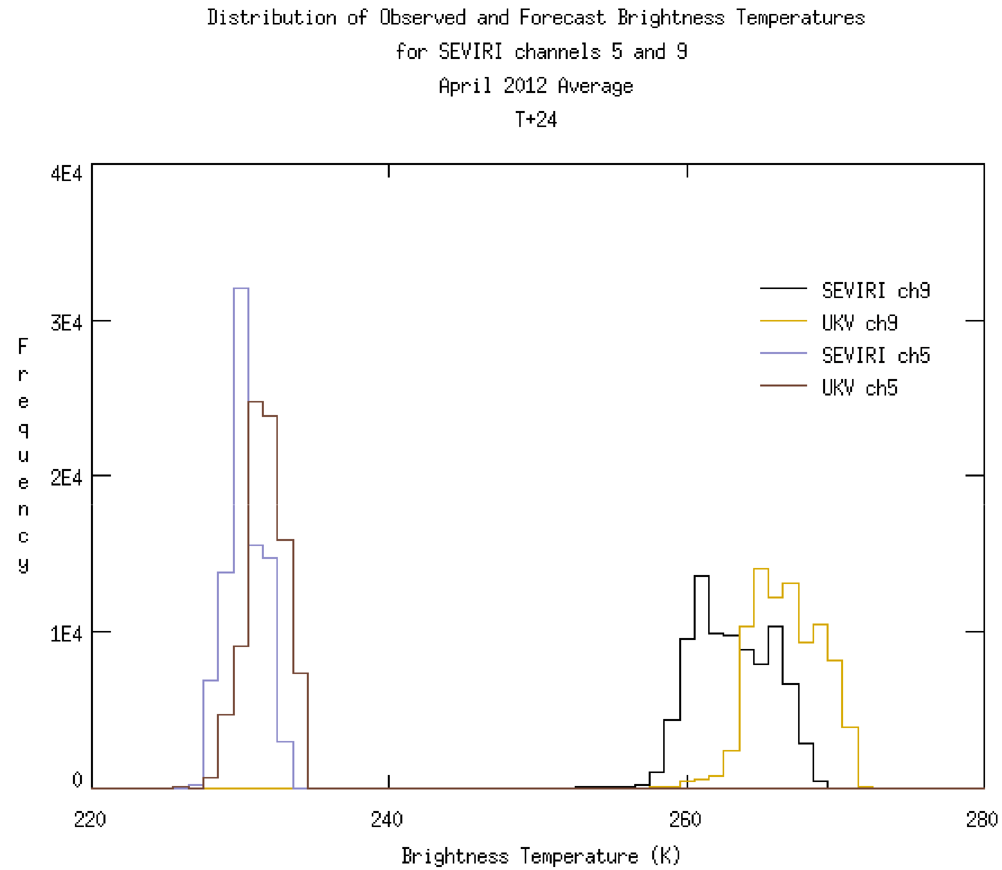
ch9

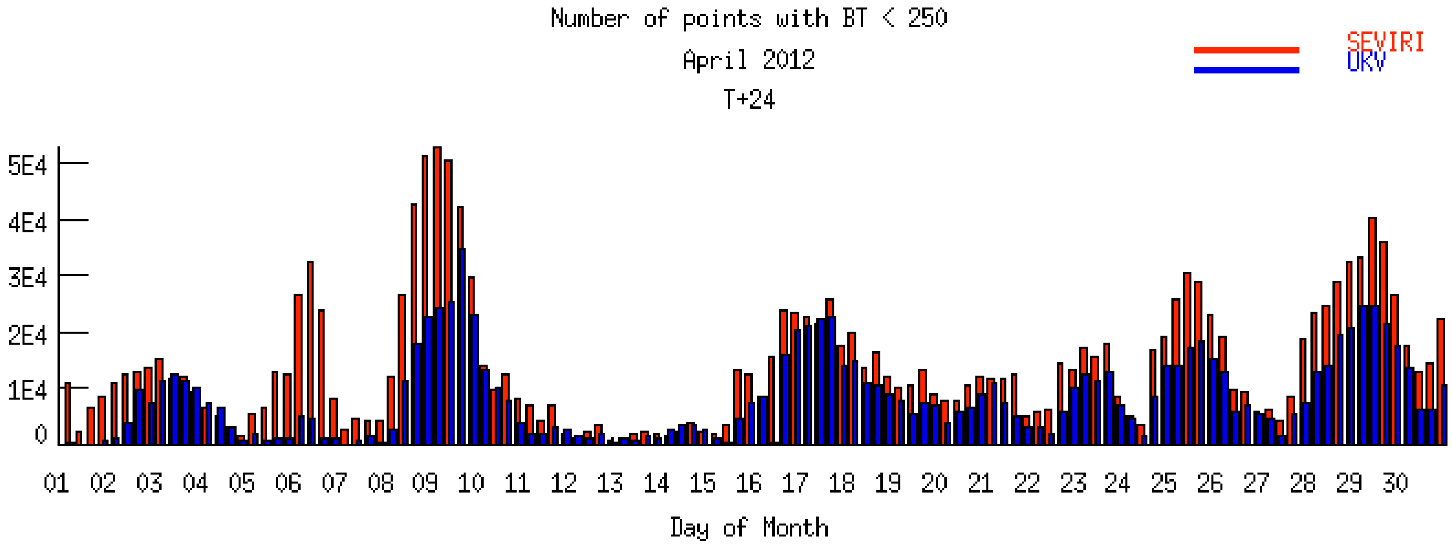
15UTC 30th April 2012

Bob Tubbs



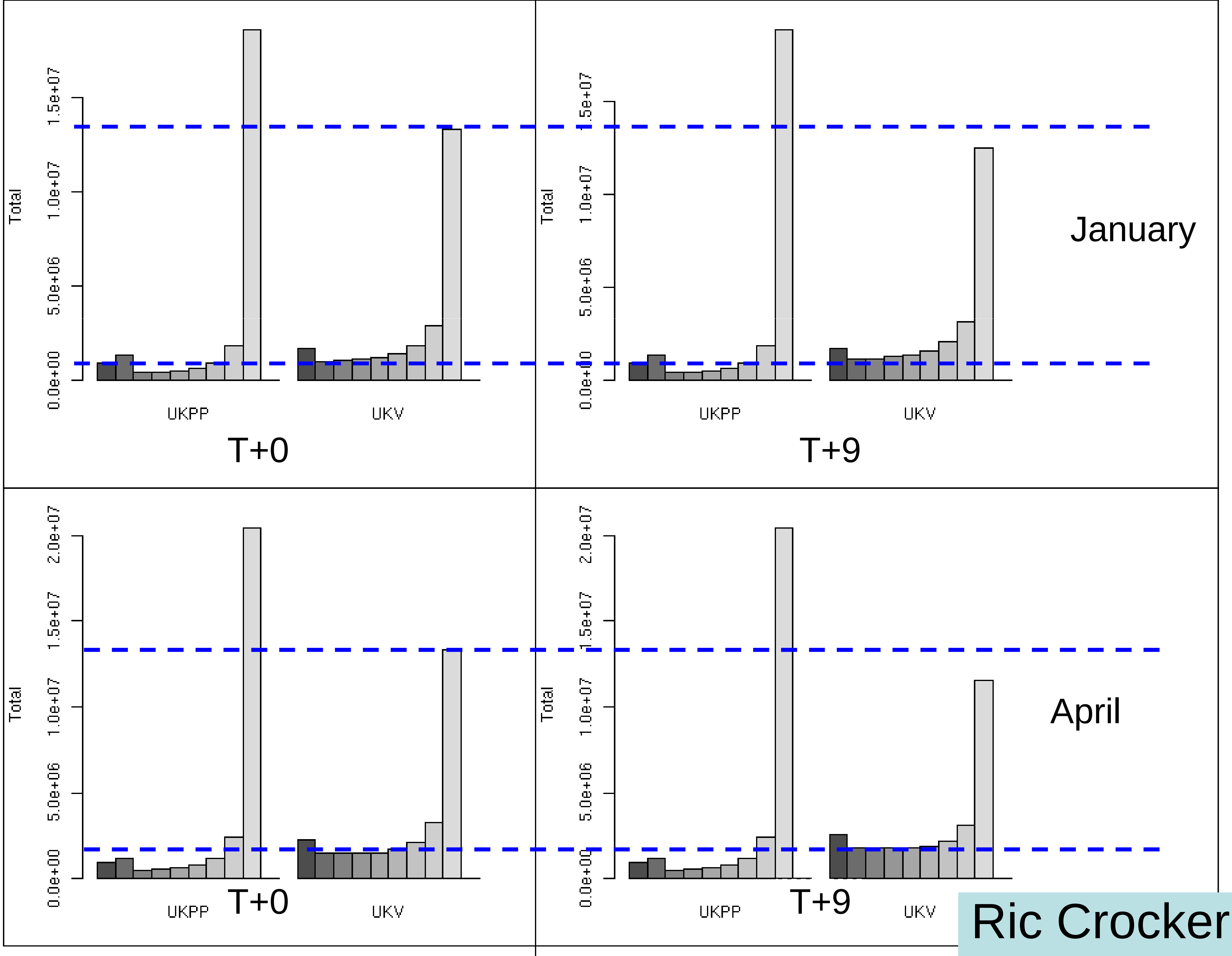

15UTC 30th April 2012




Ric Crocker

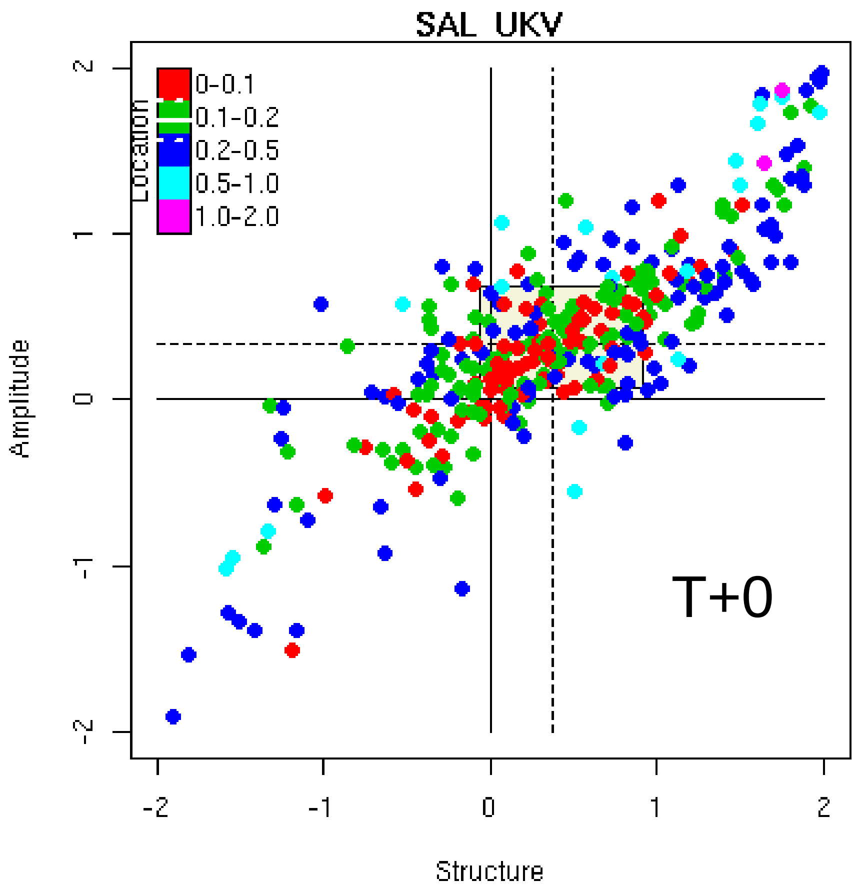
E

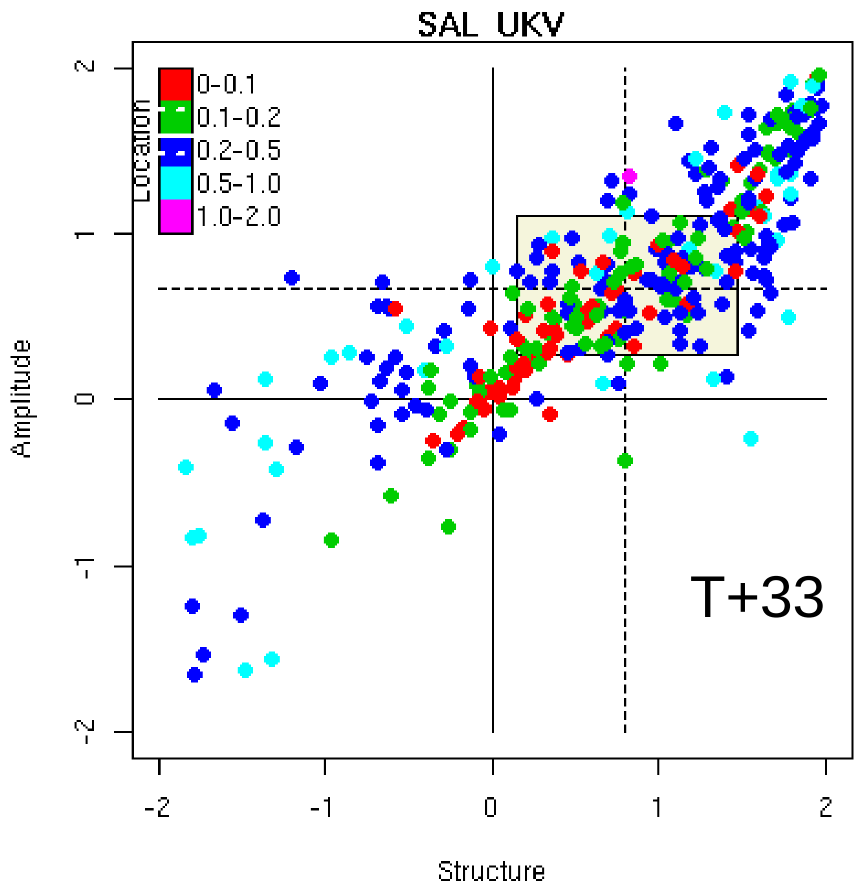

January
April
T+0
T+9
T+0
T+9
Total
UKPP
UKV
0.0e+00
5.0e+06
1.0e+07
1.5e+07
2.0e+07

Ric Crocker

# SAL for clear area (negative cloud threshold). UKV model compared to cloud analysis

- Positive S,A – too little cloud
- Increase with forecast – even less and bigger holes

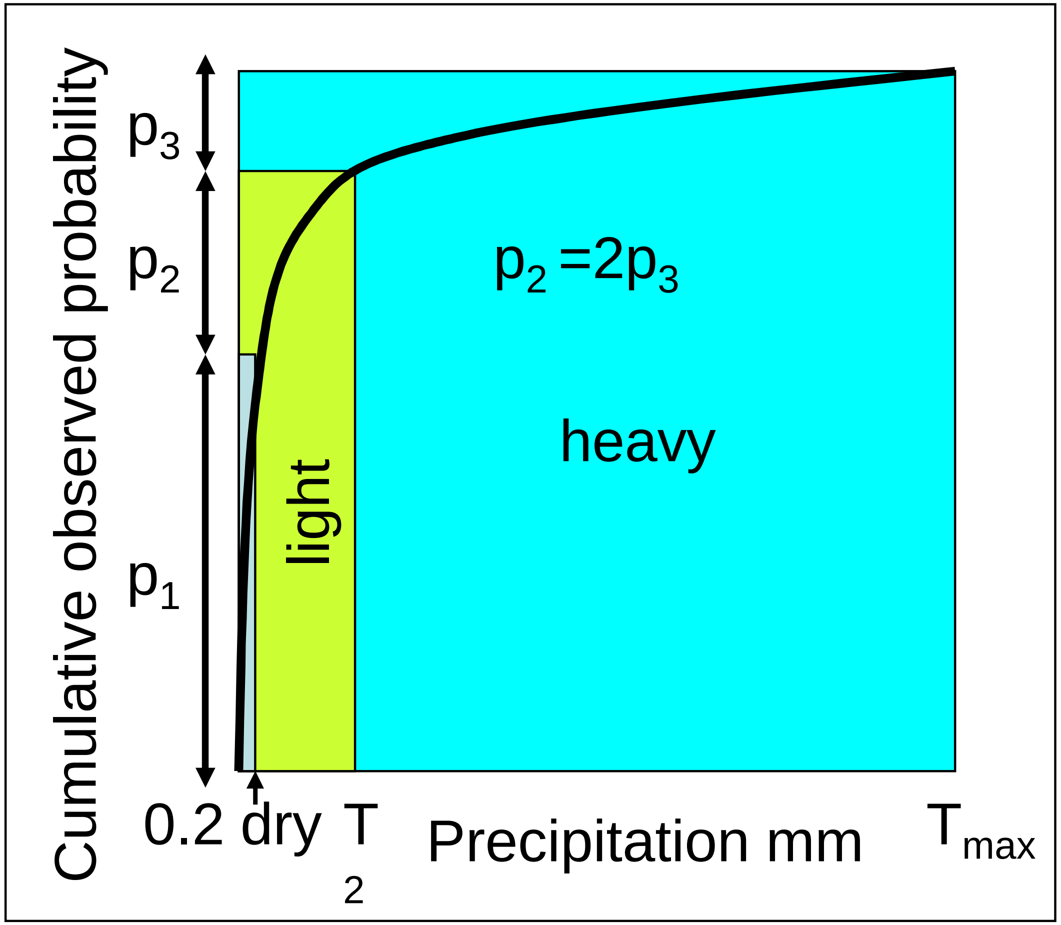


# Other deliverables

Met Office

- ND4 Inclusion of severe/high impact weather verification
  - Extreme dependency scores being applied to precipitation and wind forecasts from models
- Deliverable ND5:
  - Full documentation of the methods used in the intercomparison.
  - Newer spatial methods code to be portable.


Cumulative observed probability
p3
p2
p1
p2 =2p3
heavy
light
0.2 dry
T2
Precipitation mm
Tmax

# Thanks - Questions