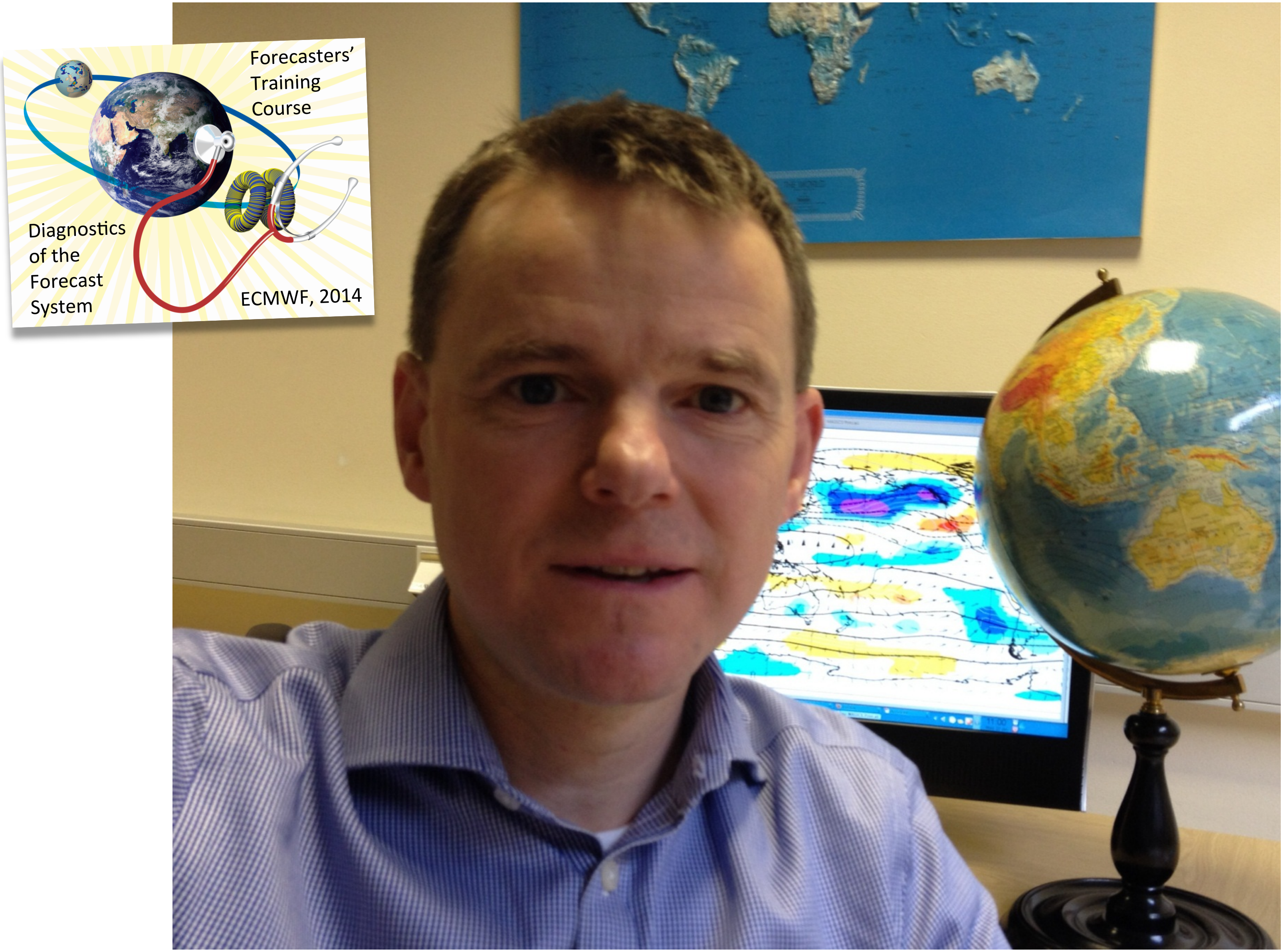
Model Errors & Diagnostic Tools: Mark Rodwell



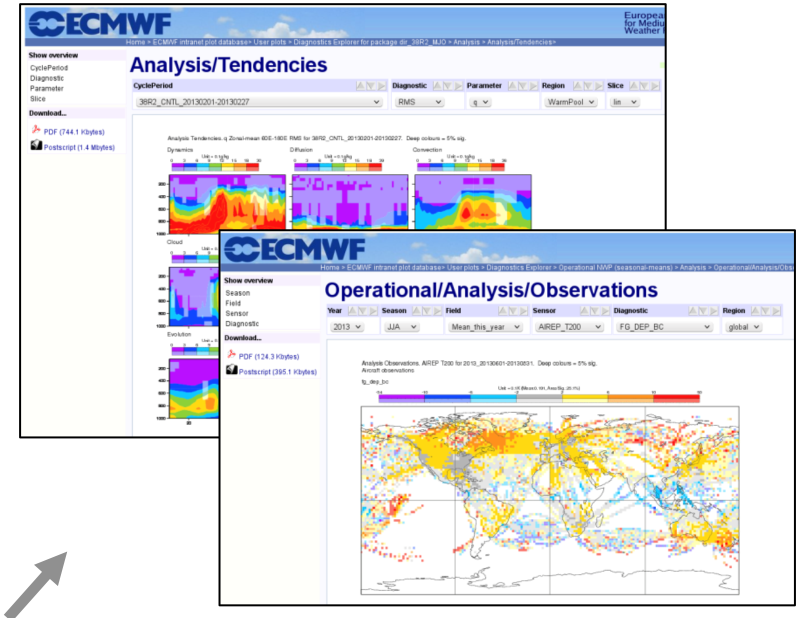
Diagnostics in weather prediction

Which model is best & why?

What are the common errors?

Do we have the correct uncertainty?

"Diagnostics Explorer"



Home > ECMWF intranet plot database> User plots > Diagnostics Explorer for package dir_38R2_MJO > Analysis > Analysis/Tendencies>

Analysis/Tendencies

CyclePeriod: 38R2_CNTL_20130201-20130227
Diagnostic: RMS
Parameter: q
Region: WarmPool
Slice: lin

Analysis Tendencies.q Zonal-mean 80E-180E RMS for 38R2_CNTL_20130201-20130227. Deep colours = 5% sig.

Home > ECMWF intranet plot database> User plots > Diagnostics Explorer > Operational NWP (seasonal-means) > Analysis > Operational/Analysis/Obs

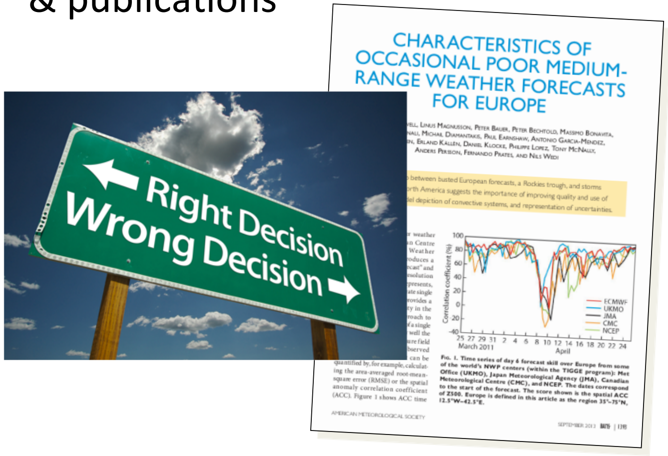
Operational/Analysis/Observations

Year: 2013
Season: JJA
Field: Mean_this_year
Sensor: AIREP_T200
Diagnostic: FG_DEP_BC
Region: global

Analysis Observations. AIREP T200 for 2013_20130601-20130831. Deep colours = 5% sig.
Aircraft observations
fg_dep_bc

Smart & easy

Development decisions
& publications



CHARACTERISTICS OF OCCASIONAL POOR MEDIUM-RANGE WEATHER FORECASTS FOR EUROPE

Linus Magnusson, Peter Bauer, Peter Bechtold, Massimo Bonavita,
Michail Diamantakis, Paul Earnshaw, Antonio Garcia-Mendez,
Erland Källén, Daniel Klocke, Philippe Lopez, Tony McNally,
Anders Persson, Fernando Prates, and Nils Wedi

between busted European forecasts, a Rockies trough, and storms
orth America suggests the importance of improving quality and use of
del depiction of convective systems, and representation of uncertainties.

quantified by, for example, calculating the area-averaged root-mean-square error (RMSE) or the spatial anomaly correlation coefficient (ACC). Figure 1 shows ACC time

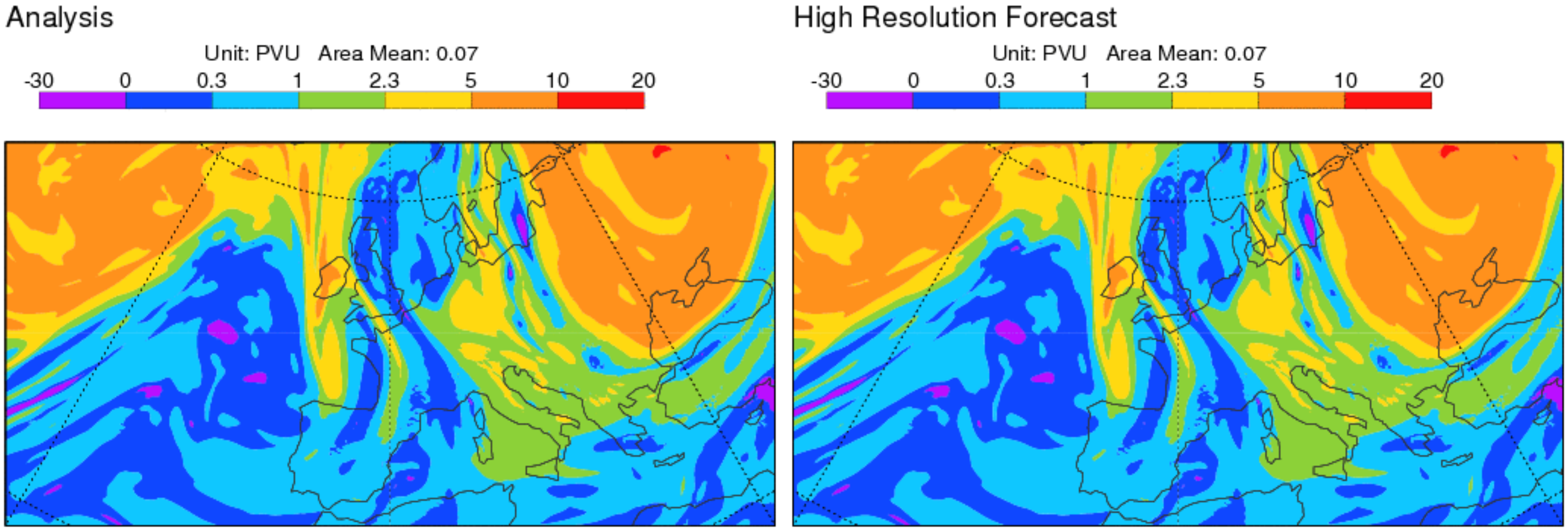
Fig. 1. Time series of day 6 forecast skill over Europe from some of the world's NWP centers (within the TIGGE program): Met Office (UKMO), Japan Meteorological Agency (JMA), Canadian Meteorological Centre (CMC), and NCEP. The dates correspond to the start of the forecast. The score shown is the spatial ACC of Z500. Europe is defined in this article as the region 35°–75°N, 12.5°W–42.5°E.

AMERICAN METEOROLOGICAL SOCIETY

SEPTEMBER 2013 BAMS | 1393

Deterministic forecasting (initial conditions)

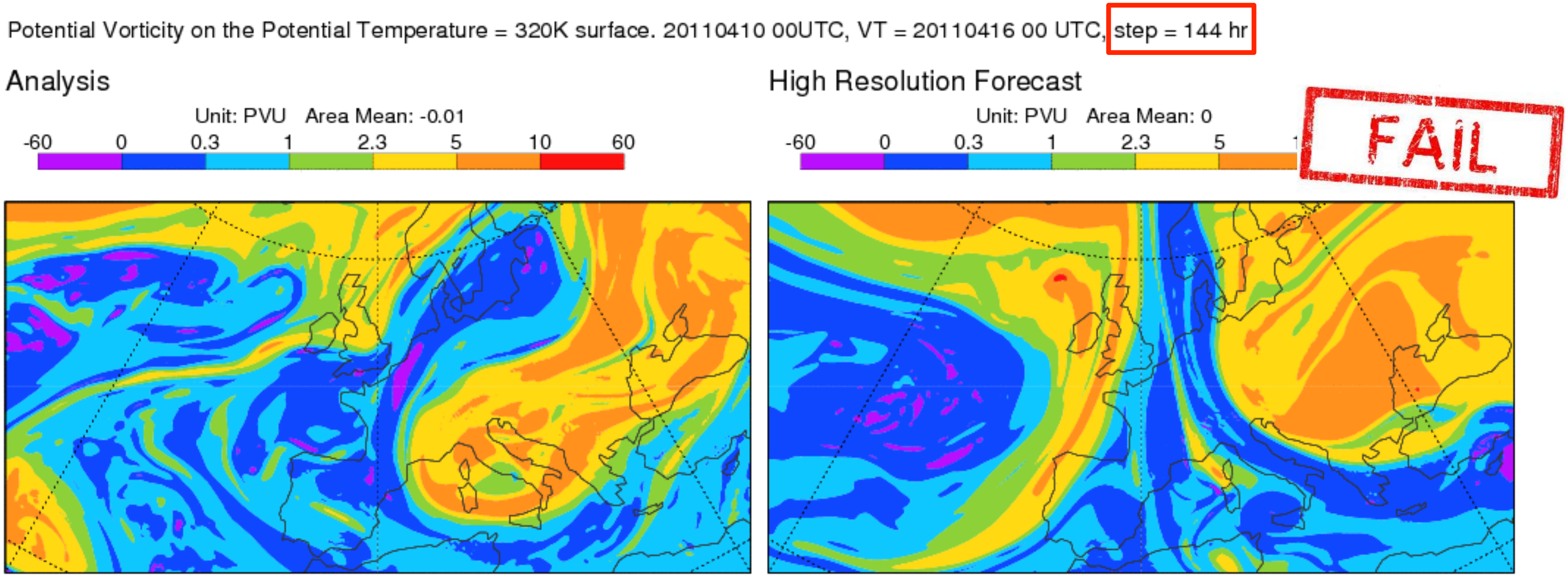
Potential Vorticity on the Potential Temperature = 320K surface. 20110410 00UTC, VT = 20110410 00 UTC, step = 000 hr



Analysis

High Resolution Forecast

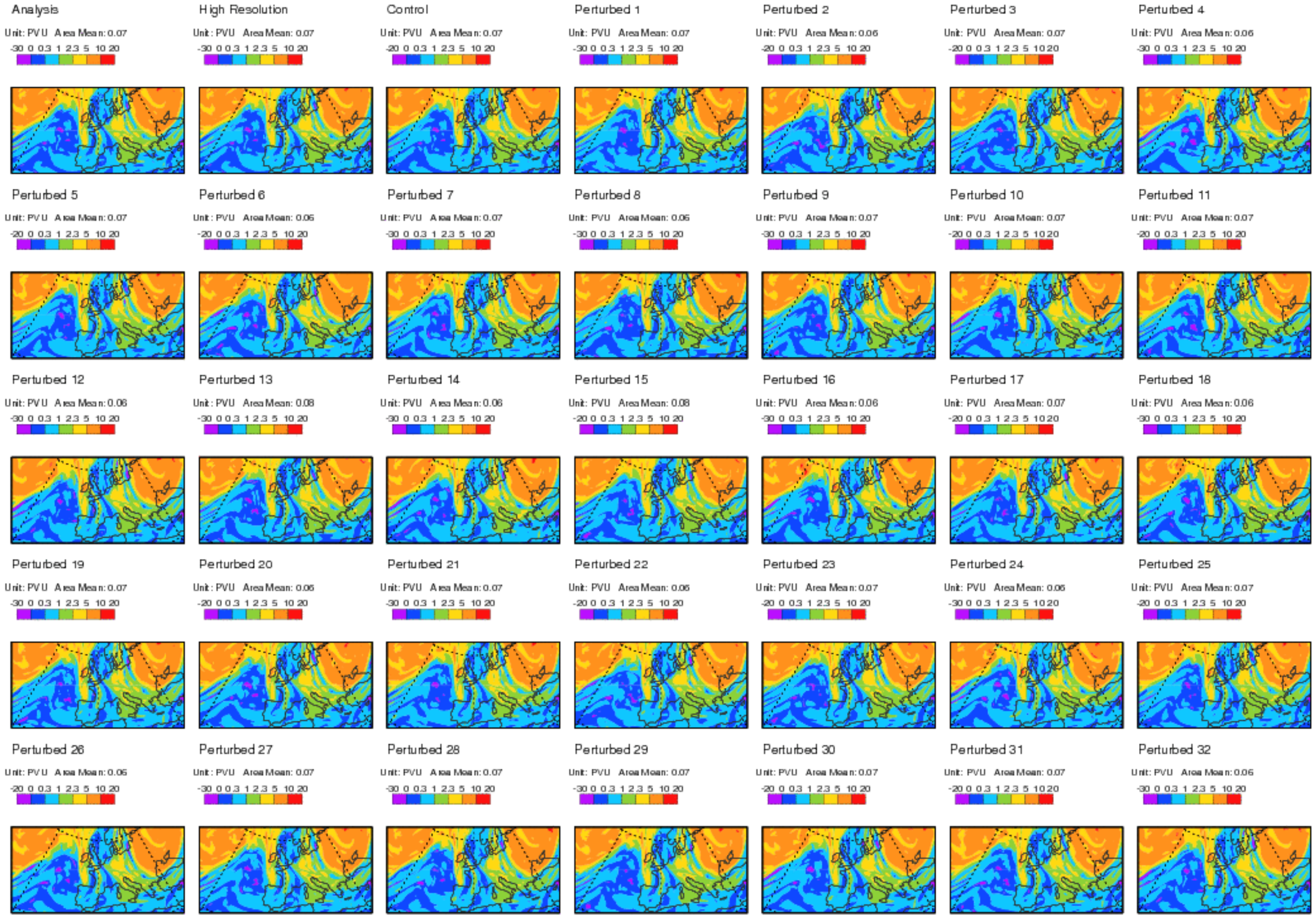
Deterministic forecasting (flow evolution to day-6)



It is difficult, by day-6, to disentangle model error from the natural growth of initial condition uncertainty (chaos)

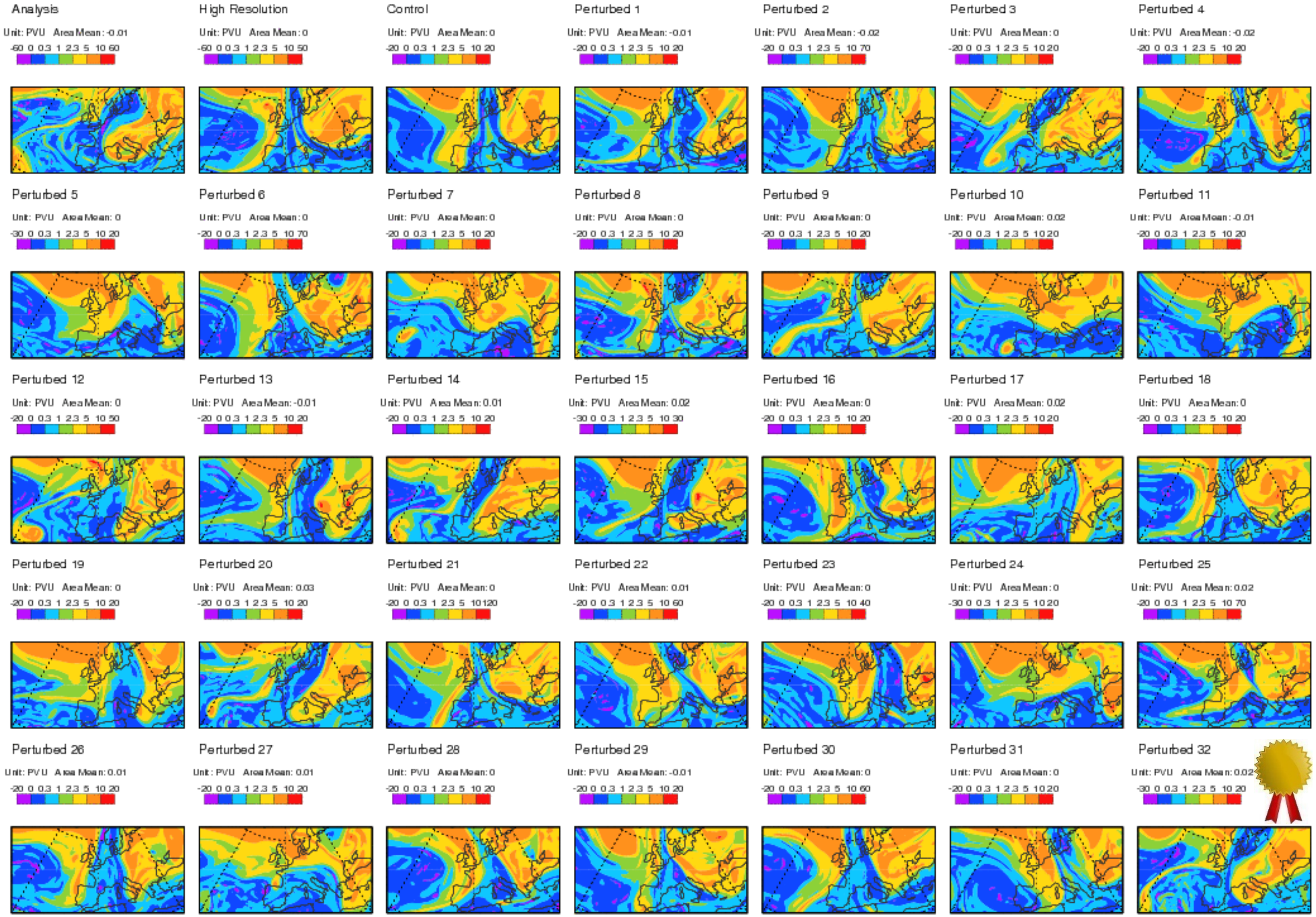
Ensemble forecasting (initial conditions)

Potential Vorticity on the Potential Temperature = 320K surface. 20110410 00UTC, VT = 20110410 00 UTC, step = 000 hr

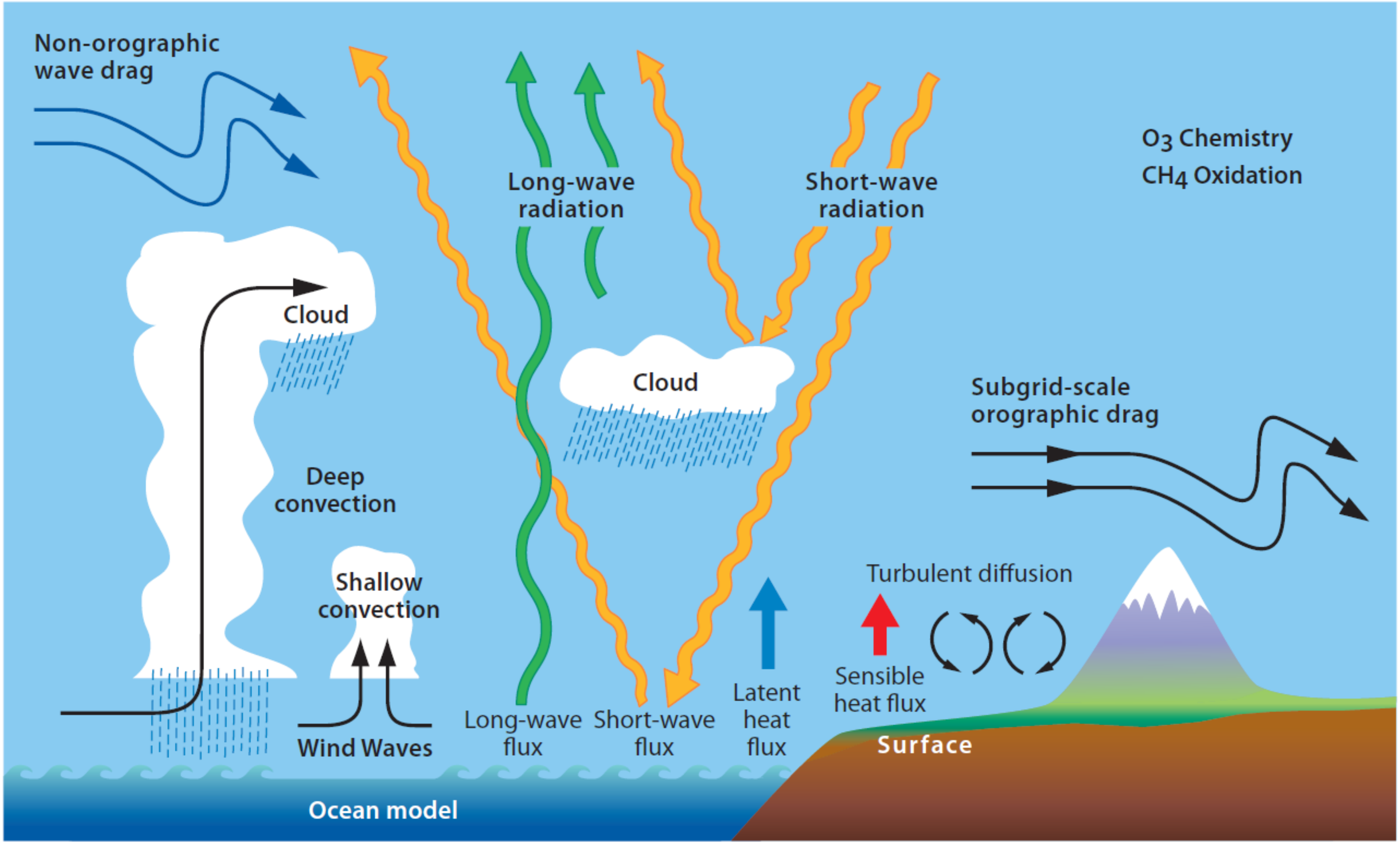


Ensemble forecasting (flow evolution to day-6)

Potential Vorticity on the Potential Temperature = 320K surface. 20110410 00UTC, VT = 20110416 00 UTC, step = 144 hr



The complexity of present-day model physics

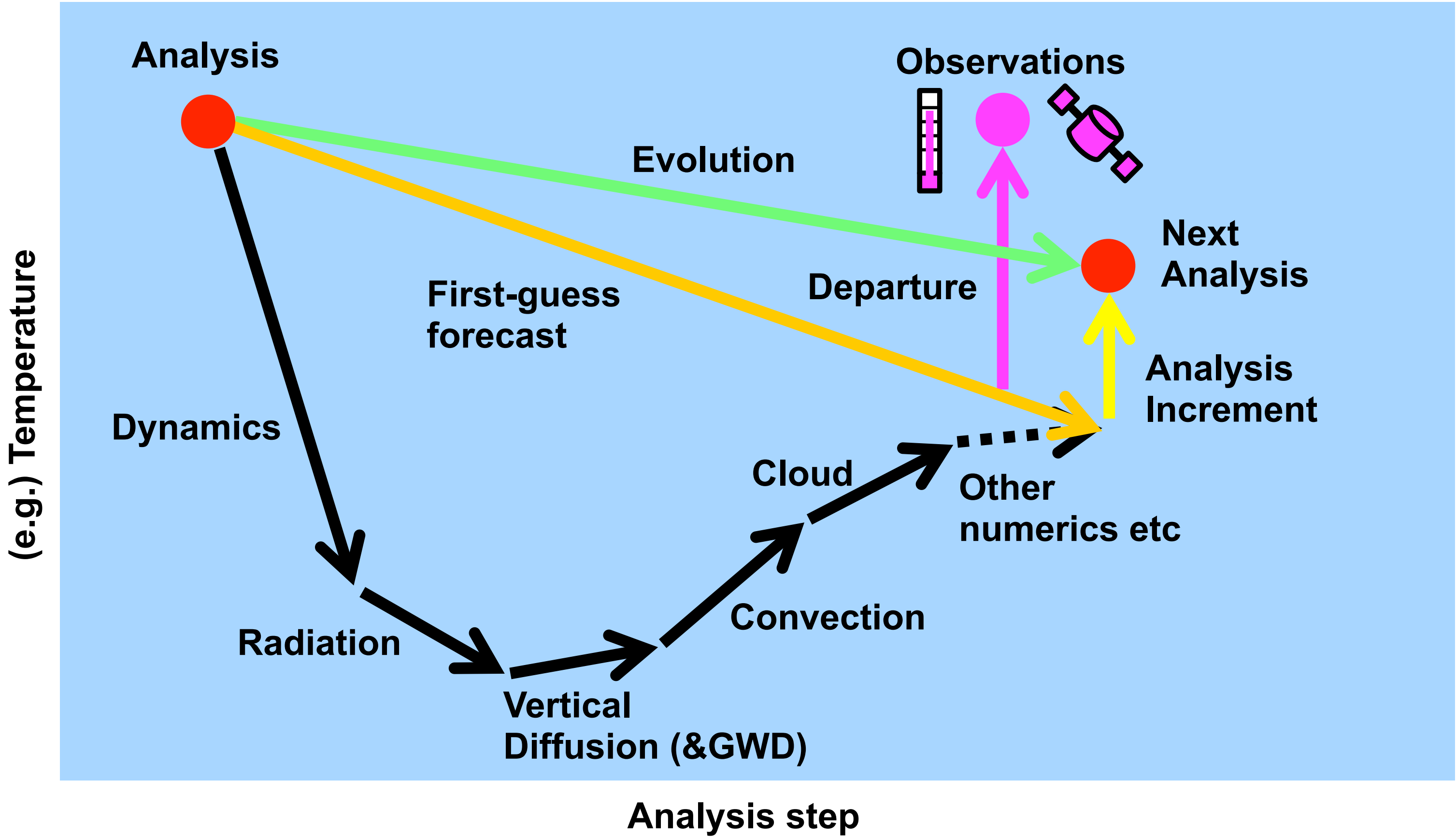


The complexity of today's models, with numerous interactions between physical processes and the resolved flow (including teleconnections), can make it very difficult to isolate the offending process(es). Single column and LES models can help, but these do not take into account the evolution of the resolved flow.

Figure from Peter Bechtold

Diagnosis of analysis & deterministic model error

Schematic of the data assimilation process – a diagnostic perspective



An approach to over-coming the issues of predictability and model complexity: Look at initial process tendencies

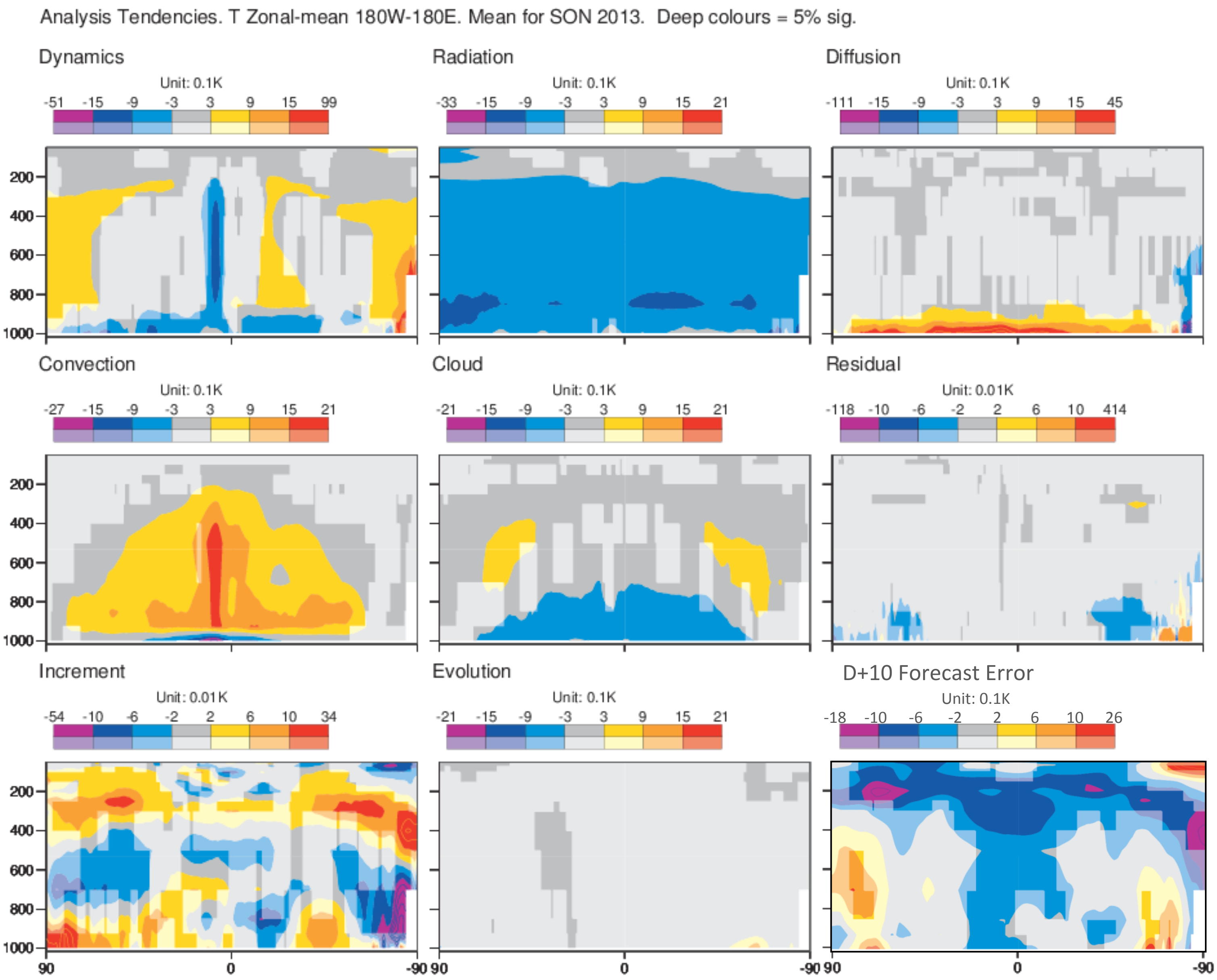
First-guess = sum of all processes.

Analysis increment corrects first-guess error, and draws next analysis closer to observations.

Relationship between increment and individual process tendencies can help identify key errors.

Approach first discussed by Klinker and Sardeshmukh (1992). Refined by Rodwell and Palmer (2007)

Initial temperature tendencies and D+10 error



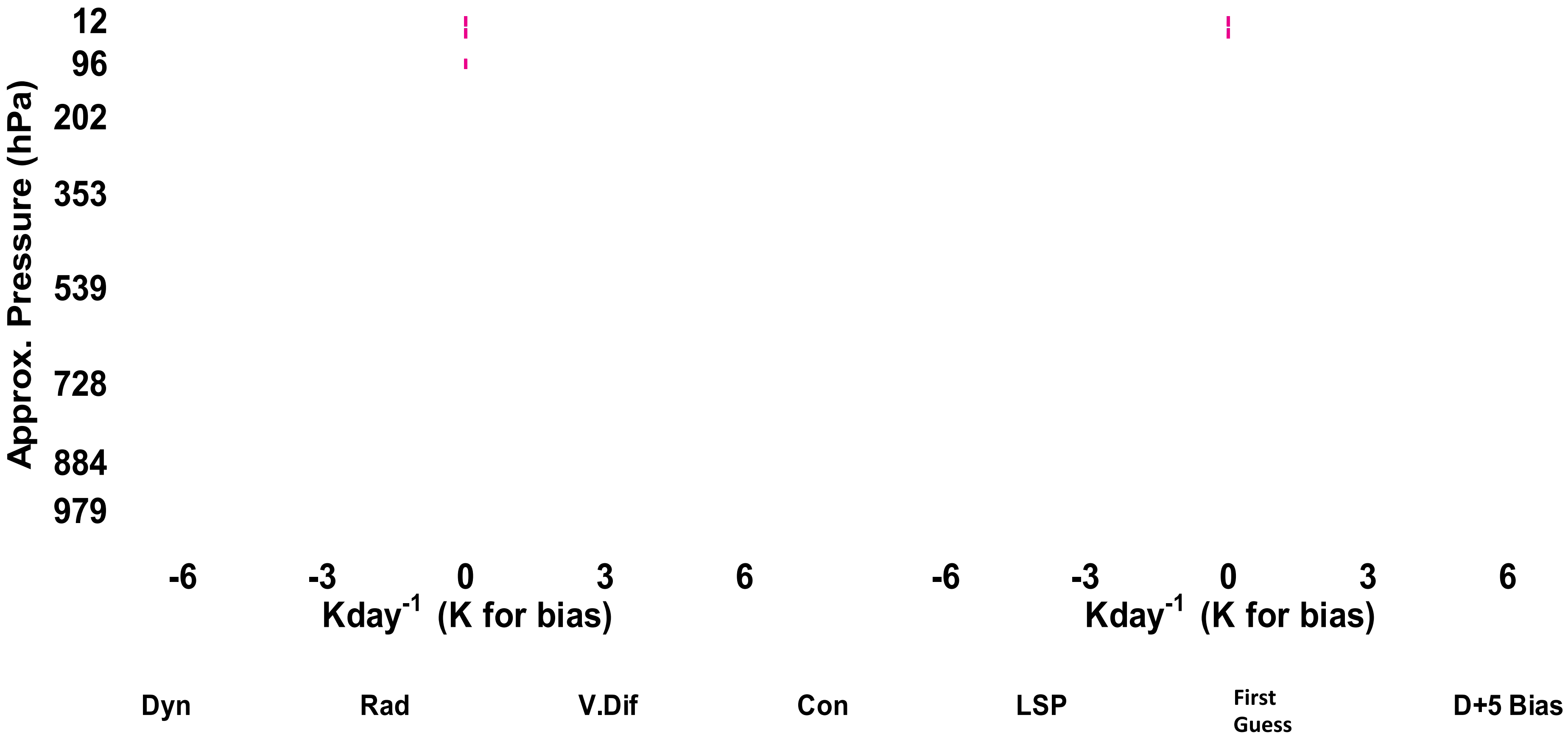
Strong upper-tropospheric increments (where radiation is not balanced by dynamics)

Error grows x10 by D+10 (due to poorly constrained humidities?)

Note that increment and residual plotted with smaller contour interval. D+10 error also has different interval.

1st example: Method questions 12K warming

Temperature tendency profiles over the Amazon (300-320°E, 20°S-0°N)

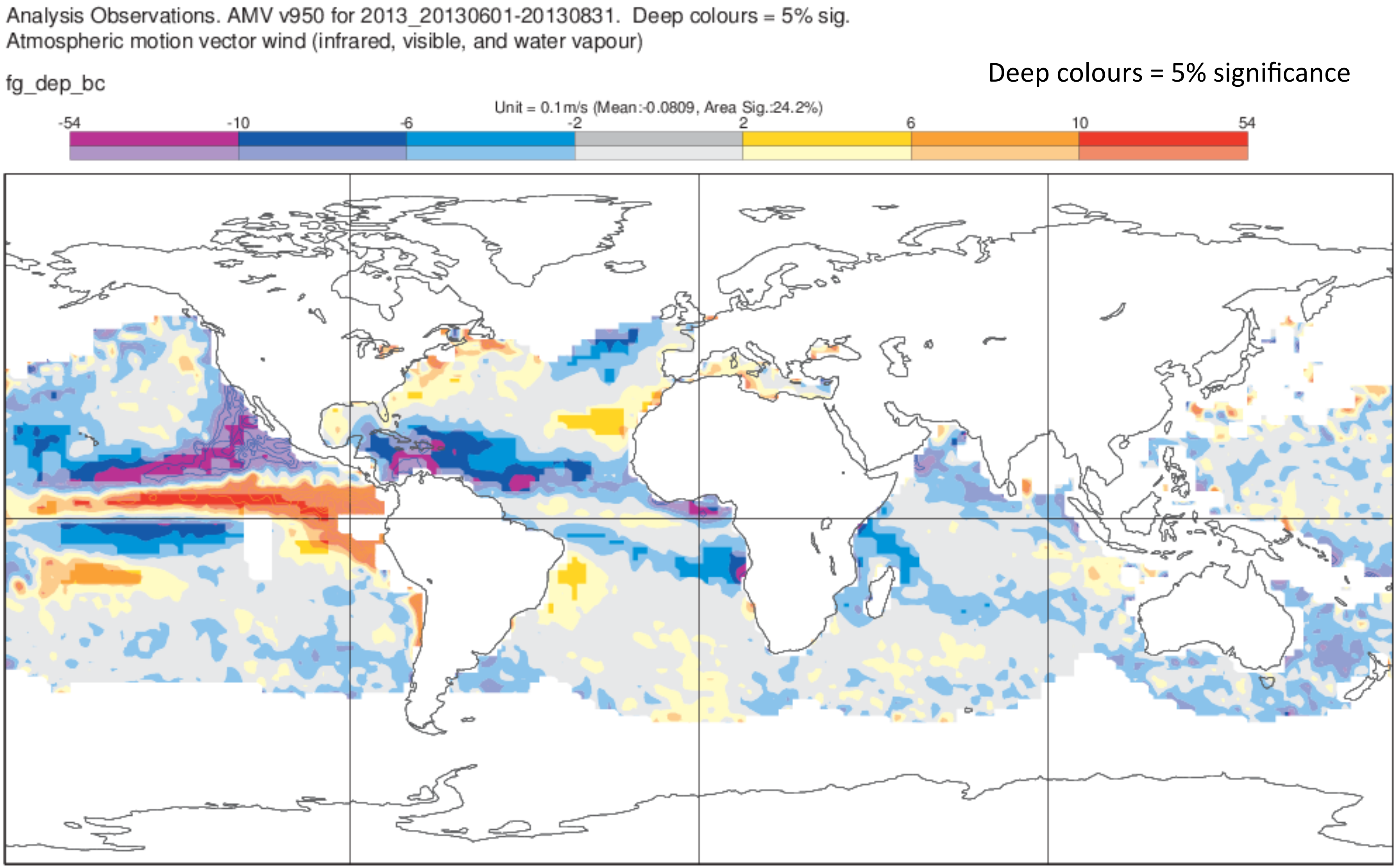


Mean first guess tendency, red, (the sum of all processes) is 'quite small': A reference value for the realism of the model's physics

Greatly increased first-guess tendency: Perturbation leads to poorer physics. Reject this perturbation from climate ensemble?

Rodwell and Palmer (2007). 6hr tendencies. 31 days (January 2005) X 4 forecasts per day. 70% conf.int. T159, L60,1800s.

2013 JJA Mean FG Departure AMV v950



Analysis Observations. AMV v950 for 2013_20130601-20130831. Deep colours = 5% sig.
Atmospheric motion vector wind (infrared, visible, and water vapour)

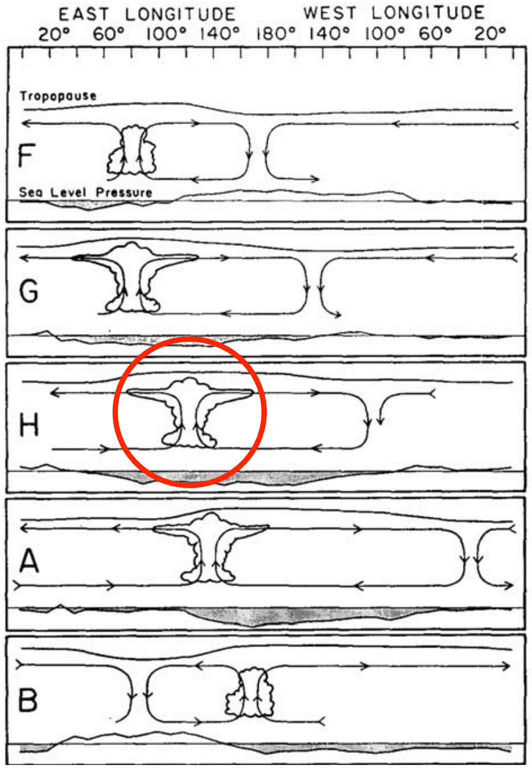
fg_dep_bc

Unit = 0.1m/s (Mean:-0.0809, Area Sig.:24.2%)

Deep colours = 5% significance

Sometimes the increments (or departures) may reflect observation issues

Mean zonal wind tendency (60-180°E) during MJO

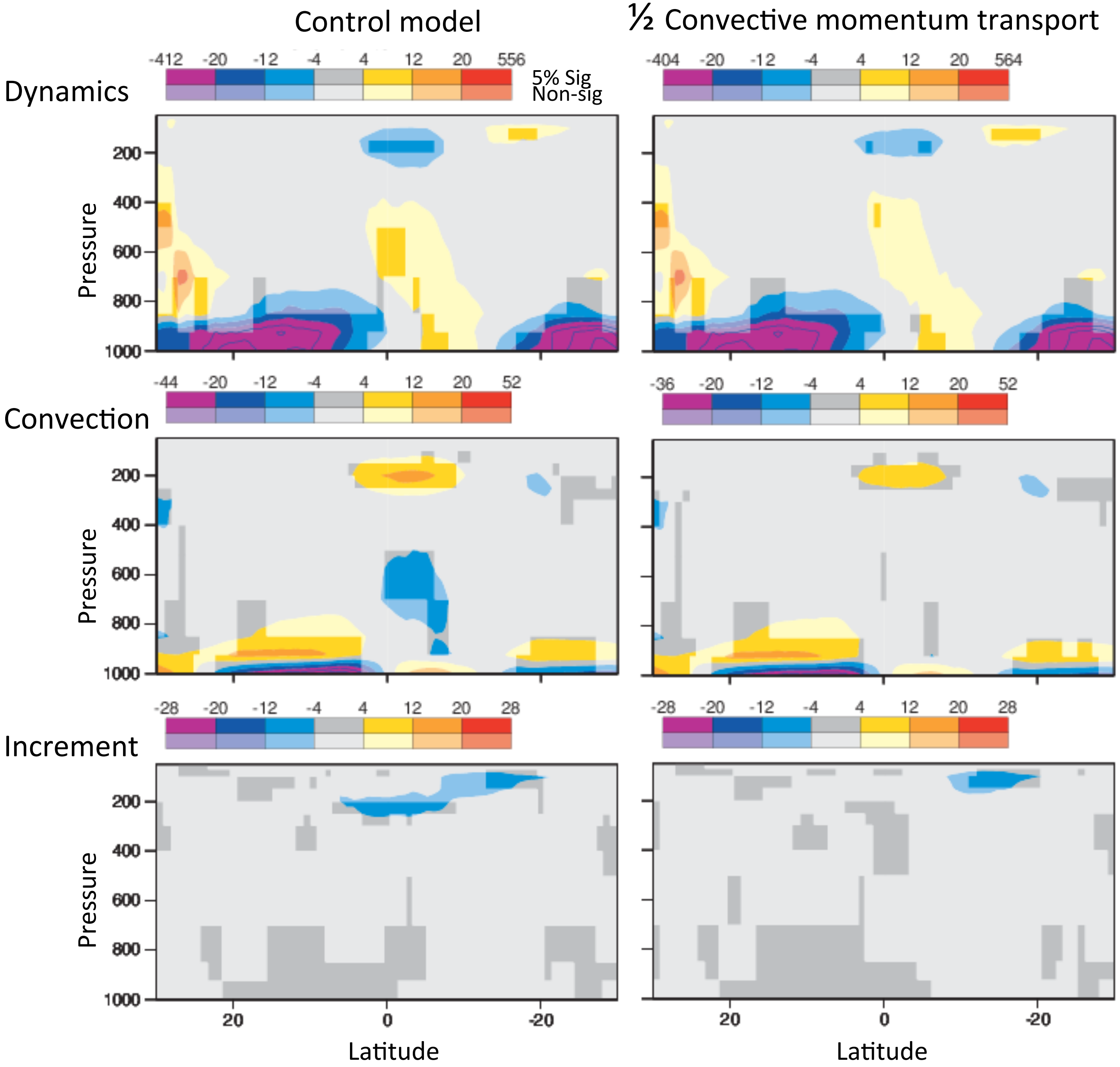


From Madden and Julian (1972)

Period : 20130201-27 (MJO convection active over warm-pool)

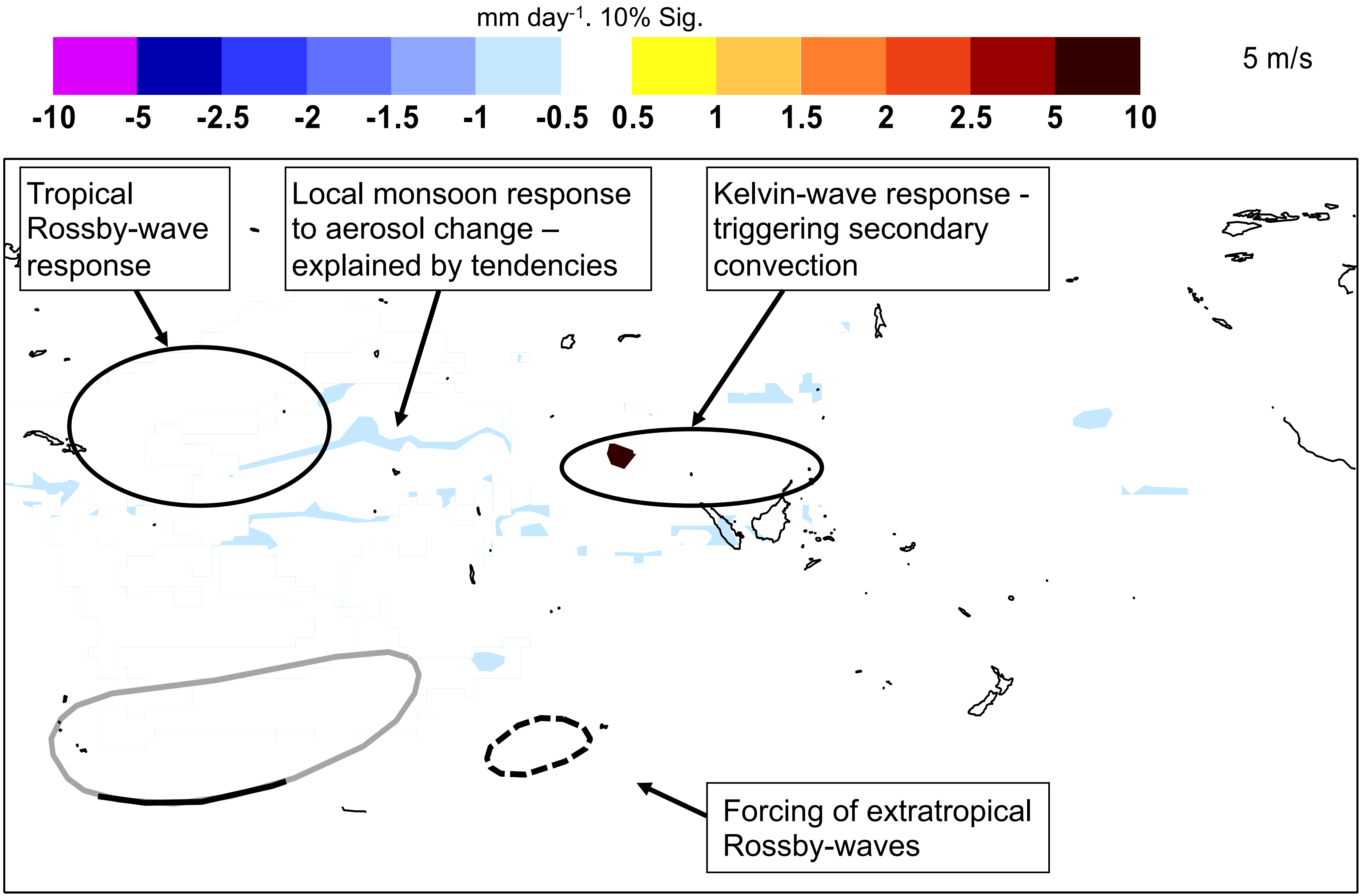
Better balance with dynamics when convective momentum transport is halved

Work with Peter Bechtold, Anton Beljaars , Jian Ling, Philippe Lopez, Frederic Vitart & Chidong Zhang



Model climate response to Sahara aerosol change

Precipitation, 850hPa winds and 500hPa heights



Tropical Rossby-wave response

Local monsoon response to aerosol change – explained by tendencies

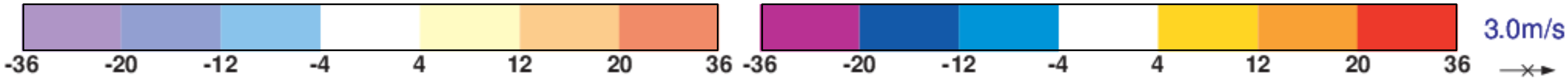
Kelvin-wave response - triggering secondary convection

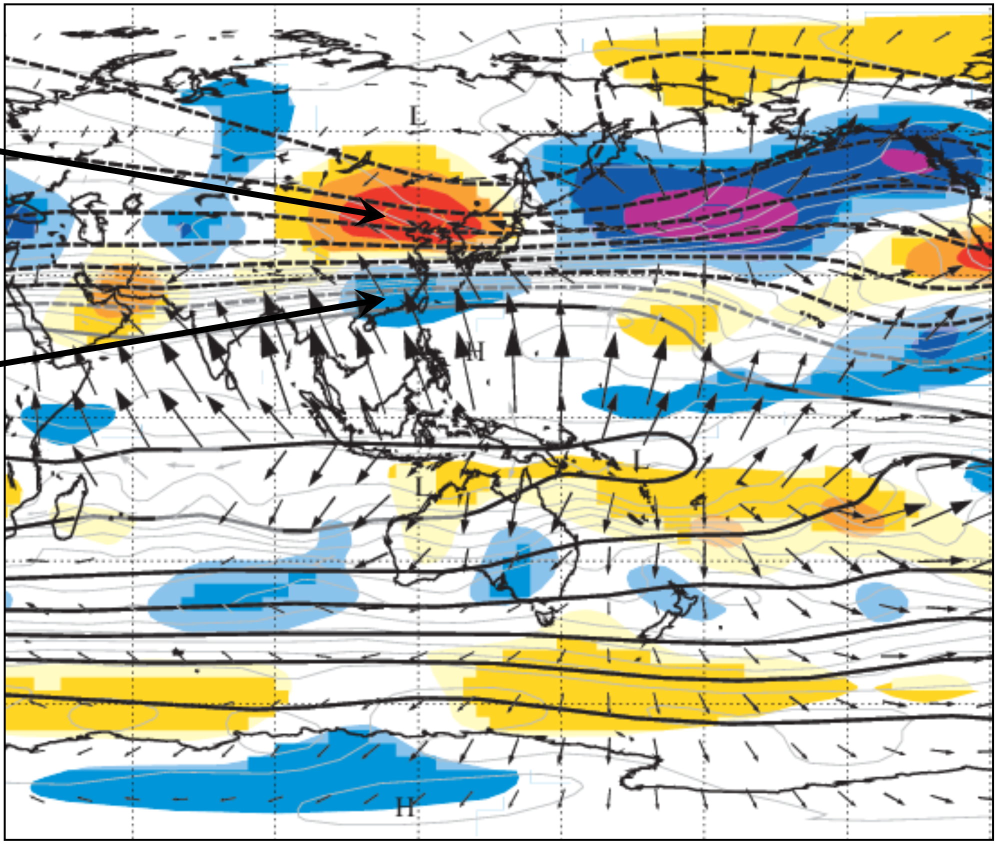
Forcing of extratropical Rossby-waves

Rodwell and Jung (2008). JJA season response to (primarily) a reduction in aerosol over the Sahara

‘Stretching’ and vorticity advection from Tropics

Rossby Wave Source: shading unit = 10^{-11} s^{-2} . Streamfunction: contour interval = $2x10^7$ m^2s^{-1}. Divergent wind vectors





‘Stretching’

Advection

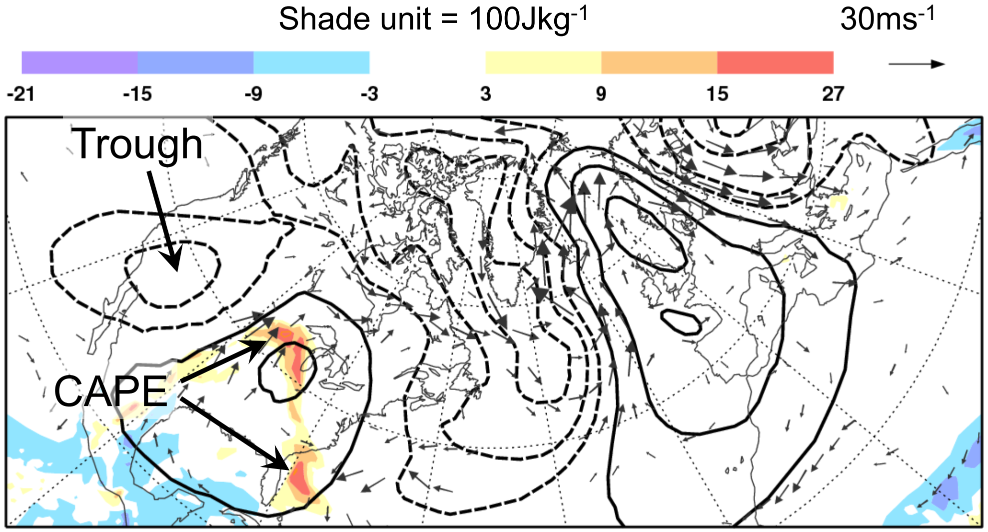
Mean errors in stretching and advection account for ⅓ to ½ of RMSE of vorticity forcing at day-1

Reducing in this mean error should improve prediction of stormtracks

All integrated 100-300hPa, DJF 2012/13

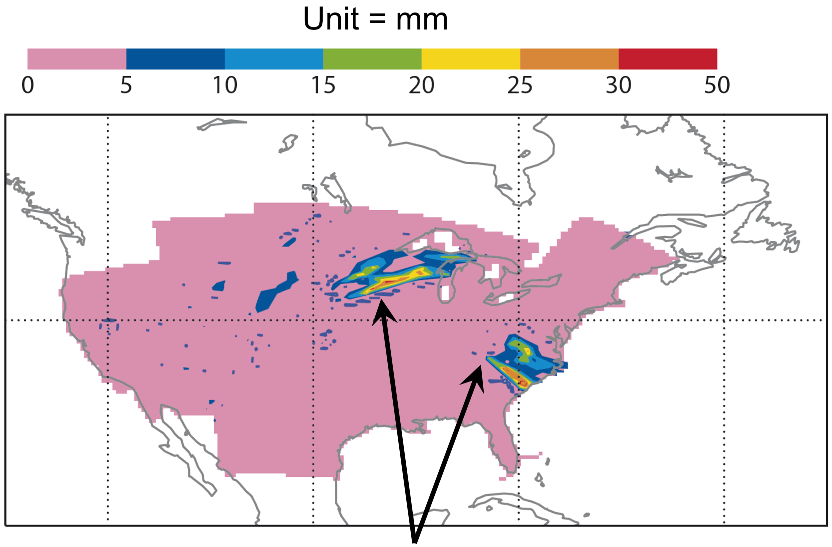
10 April Rockies trough with CAPE & MCS ahead

Z500 and CAPE anomaly



Contour Interval = 100m

12-hr Radar-observed precipitation

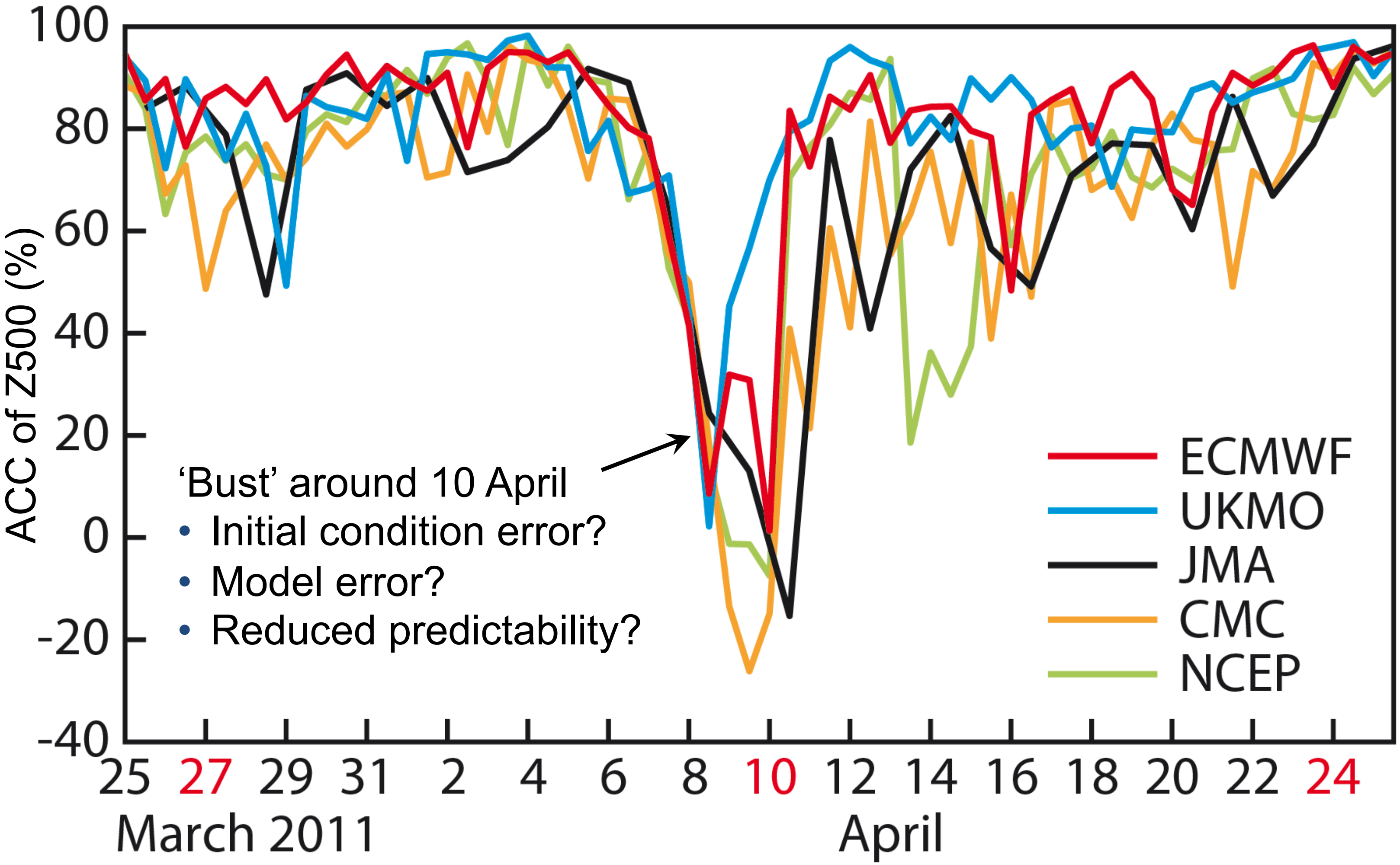


Mesoscale convective systems (MCSs)

Rodwell et al. (2013)

Z500 at 0UTC, CAPE at 6UTC (T+6), 12hr ‘NEXRAD’ Radar precipitation accumulated to 9 UTC

Skill of single forecasts (Europe, leadtime = 6 days)



‘Bust’ around 10 April

- Initial condition error?
- Model error?
- Reduced predictability?

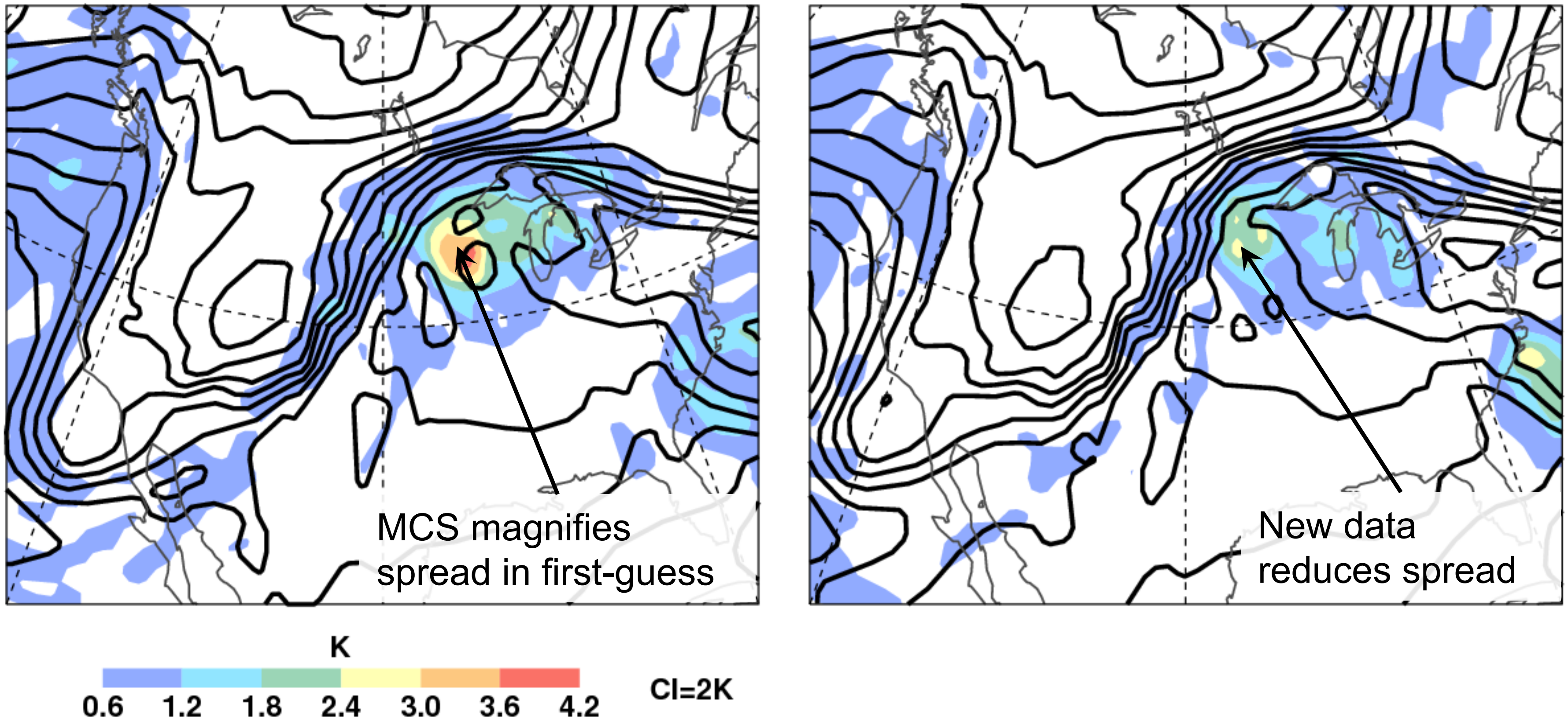
Score is the spatial Anomaly Correlation Coefficient (ACC)x100 for 500 hPa geopotential height (Z500) over Europe (12.5°W –42.5°E, 35°N–75°N). The date shown is the forecast start date

Ensemble of data assimilations, EDA

10 April T200 mean & spread

First-guesses T+12hr

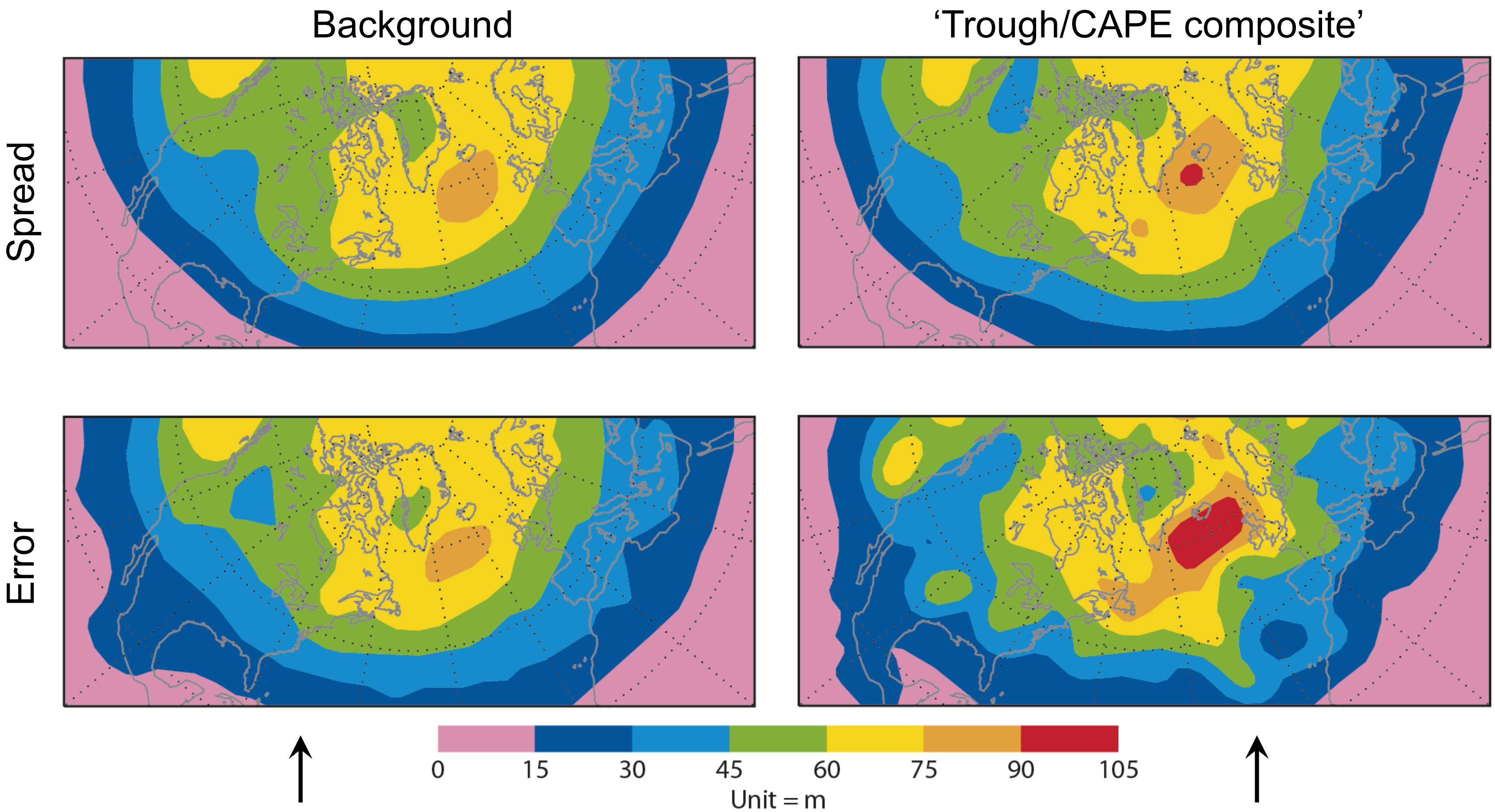
Analyses


MCS magnifies spread in first-guess

New data reduces spread

The ensemble of first-guess forecasts develops spread over the first 12 hours associated with uncertainties in the prediction of a mesoscale convective system. The incorporation of new observations by the ensemble of data assimilations results in a contraction of the spread. Key question: Is the final analysis spread too large or too small to correctly reflect the predictability of the subsequent flow? Data: Temperature at 200 hPa from 10-member EDA, valid at 6UTC.

Composite ensemble spread & error (Z500 at day 6)



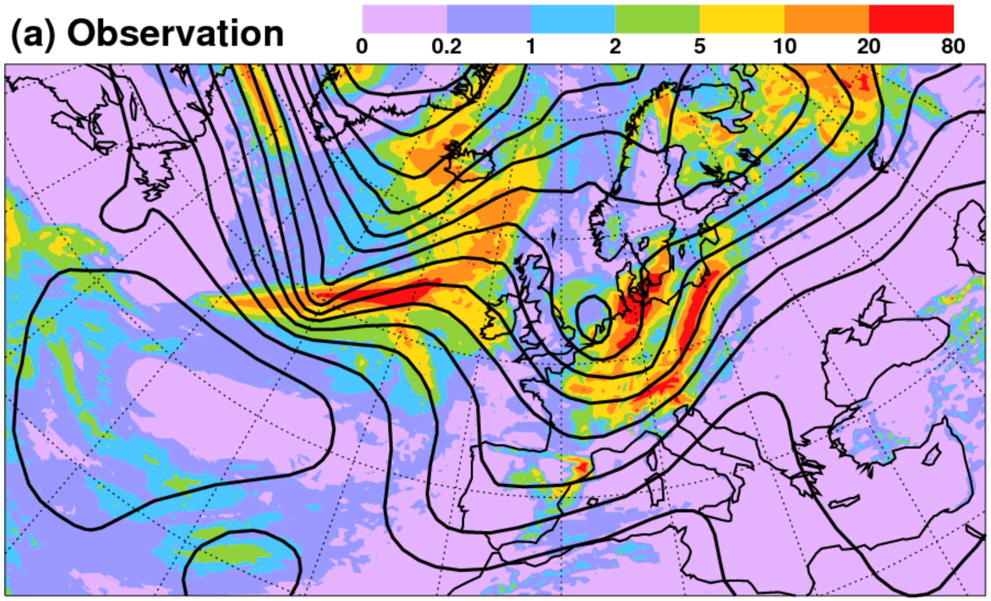
Error ≈ spread (system 'reliable' in the mean). *e.g.* stormtrack

30% increased error. Spread not fully predicting the reduced predictability?

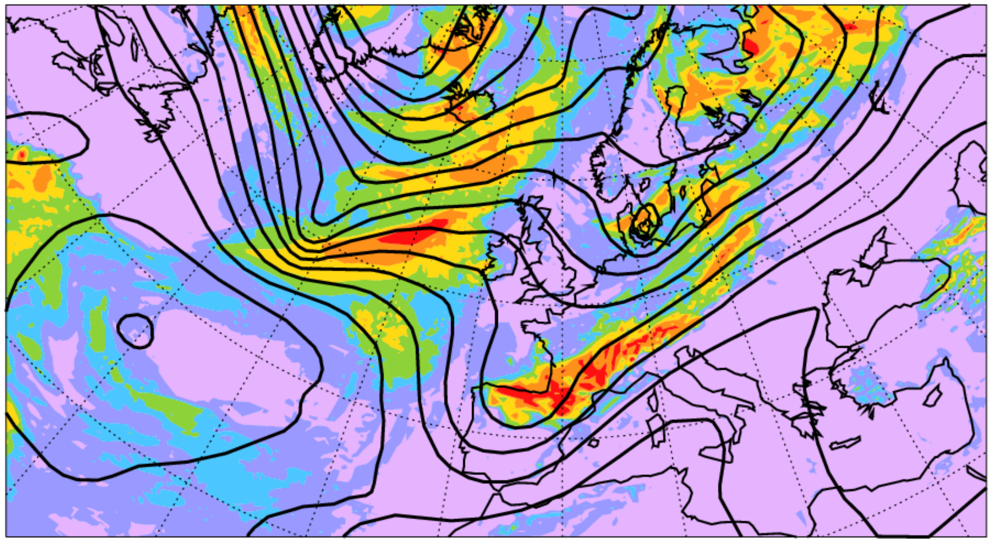
Composite over all 84 events 10 November 2010 – 20 March 2012 (0 or 12UTC) with a strong trough over the Rockies and positive CAPE ahead. 'Error' is RMSE of ensemble-mean (dominated by random component), 'Spread' is ensemble standard deviation, scaled for finite ensemble

Z500 and Precipitation: 23/08/2008

(a) Observation

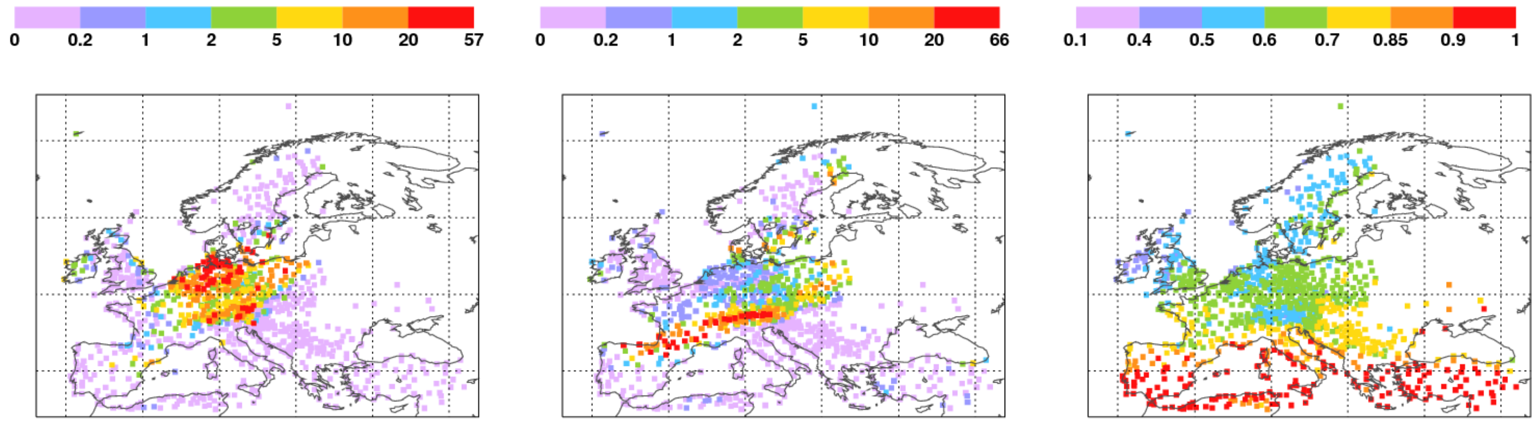


(b) Forecast D+4



Contours: Z500
Shading: Precipitation (24hr accumulation)

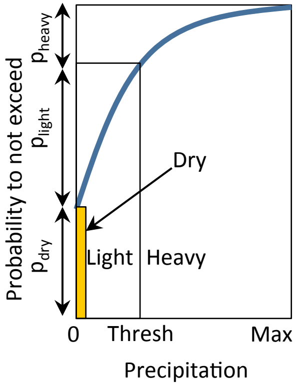
Calculation of ‘SEEPS’ precipitation score

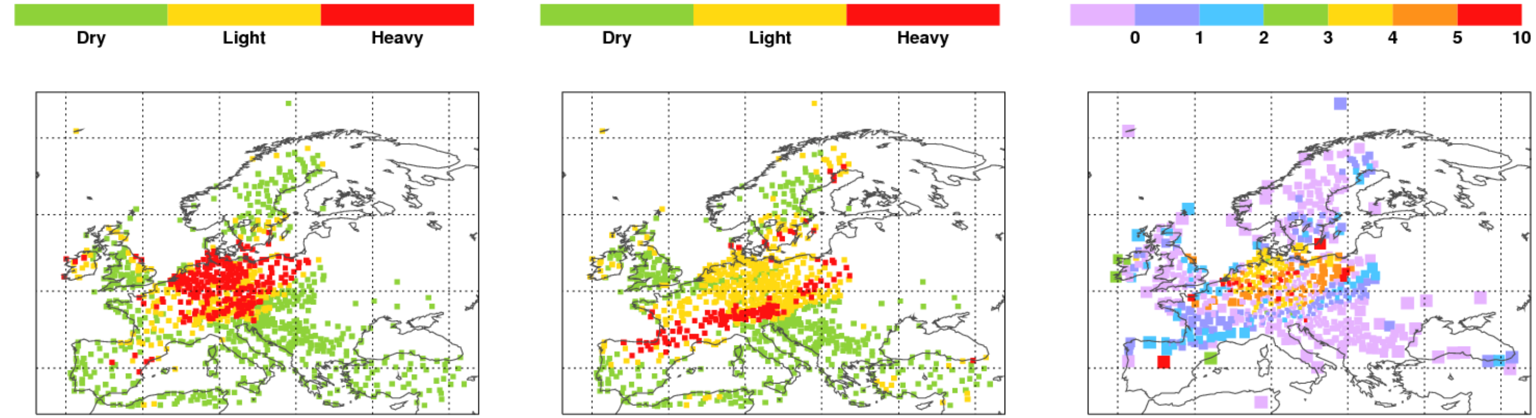


(a) Observation

(b) Forecast

(c) Probability Dry





(d) Observed Category

(e) Forecast Category

(f) SEEPS

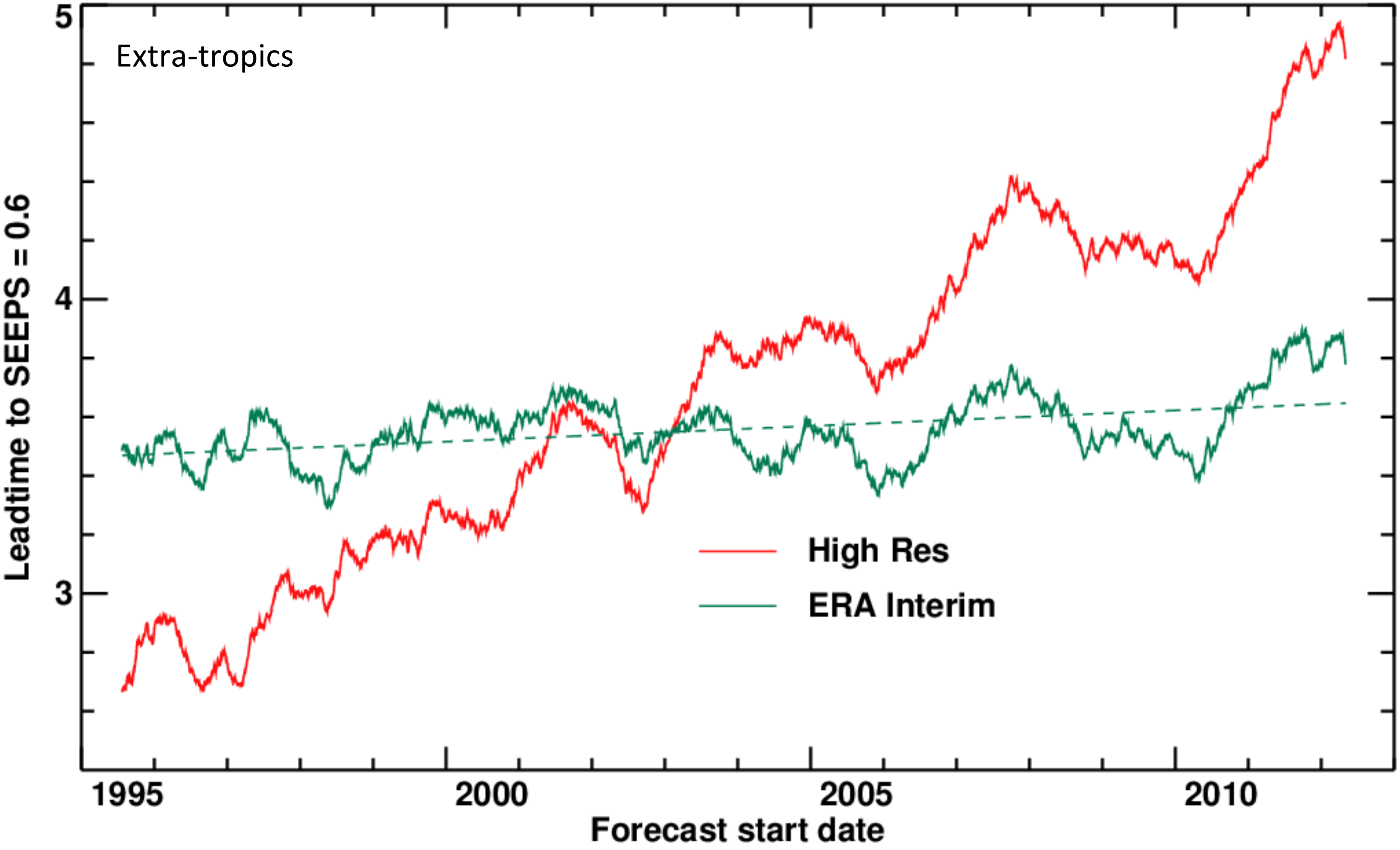
‘1-to-1’ relationship with actual error

Score is equitable & inhibits hedging

Rodwell et al. (2010)

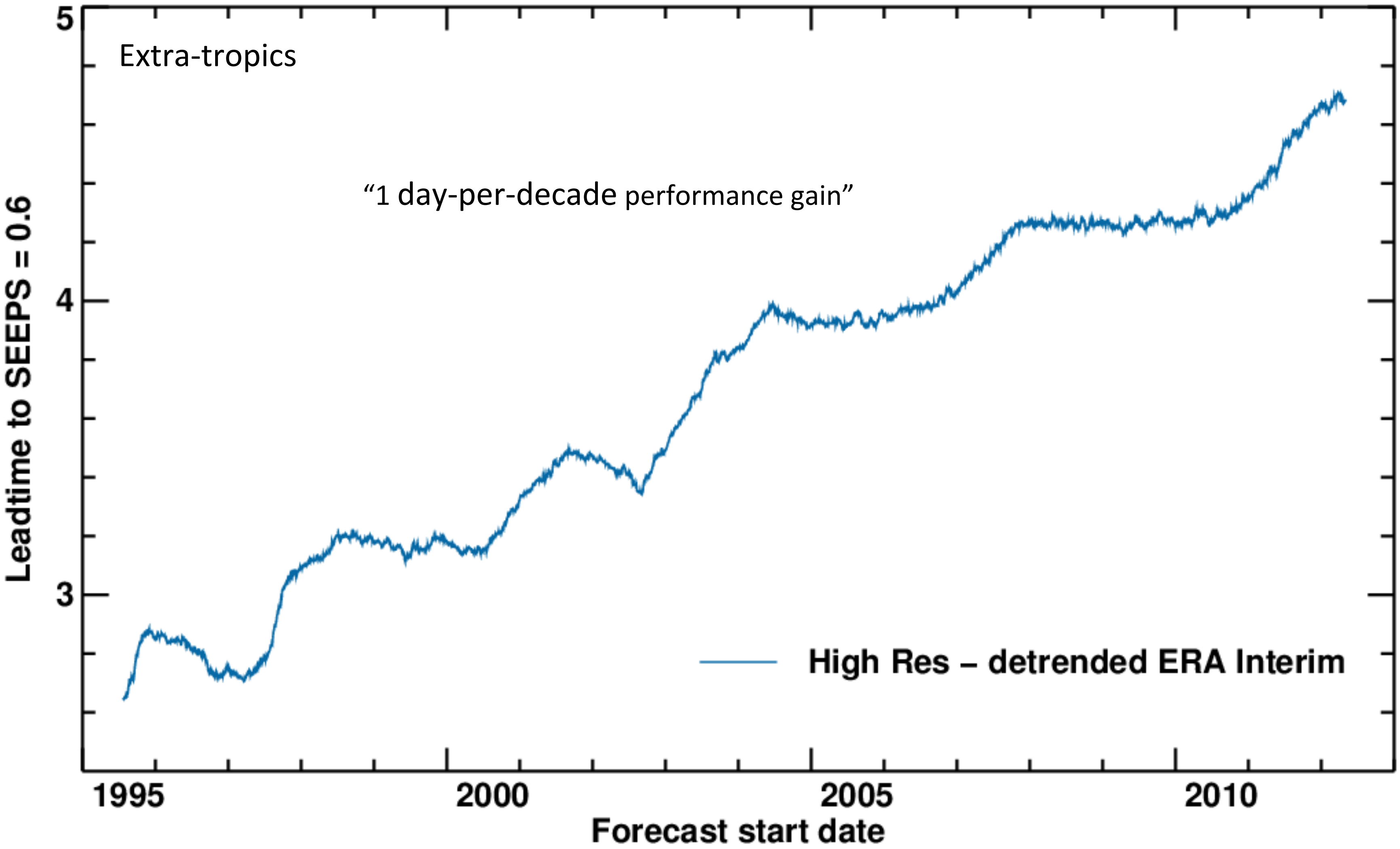
Failure to predict heavy precipitation ahead of Low over northern Europe, too much frontal precipitation to the south. A station’s climatology is used to define threshold between ‘Light’ and ‘Heavy’. In (f), the box size indicates a station’s relative contribution to the area-mean score. Forecast is day 4 24-hr accumulation on 20080823

Score identifies trends & model improvements



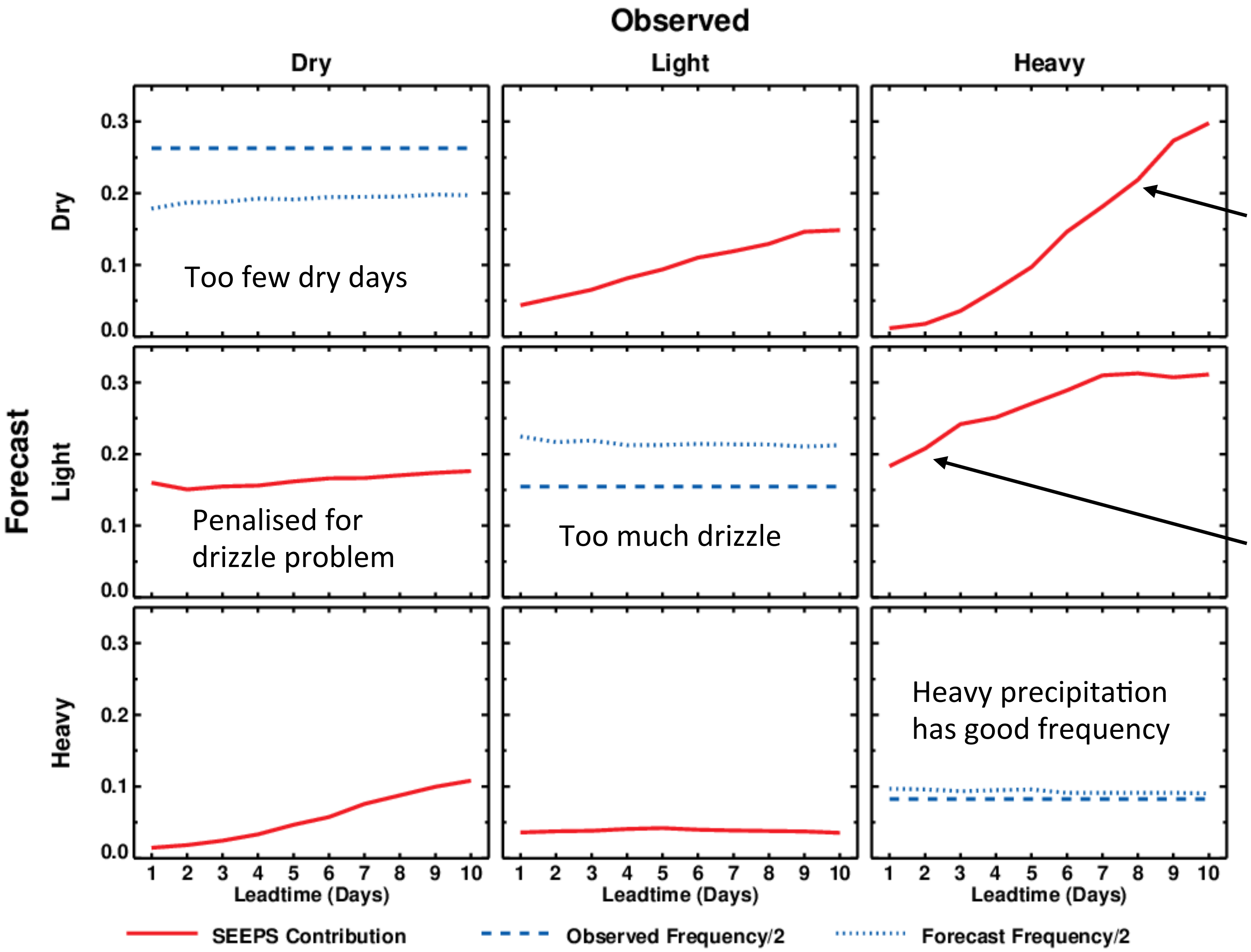
Lead-times have a 365-day running-mean applied. Clear trends in (single) high-resolution forecast skill. A strong argument for continuing to monitor deterministic forecasts of precipitation. The slight rend in lead-time for the (fixed system) ERA Interim forecasts must be due to improvements in the observing system.

Precipitation forecast performance trend



Lead-times have a 365-day running-mean applied.

Category frequency (blue) & SEEPS error (red)



Too few dry days

Penalised for drizzle problem

Too much drizzle

Heavy precipitation has good frequency

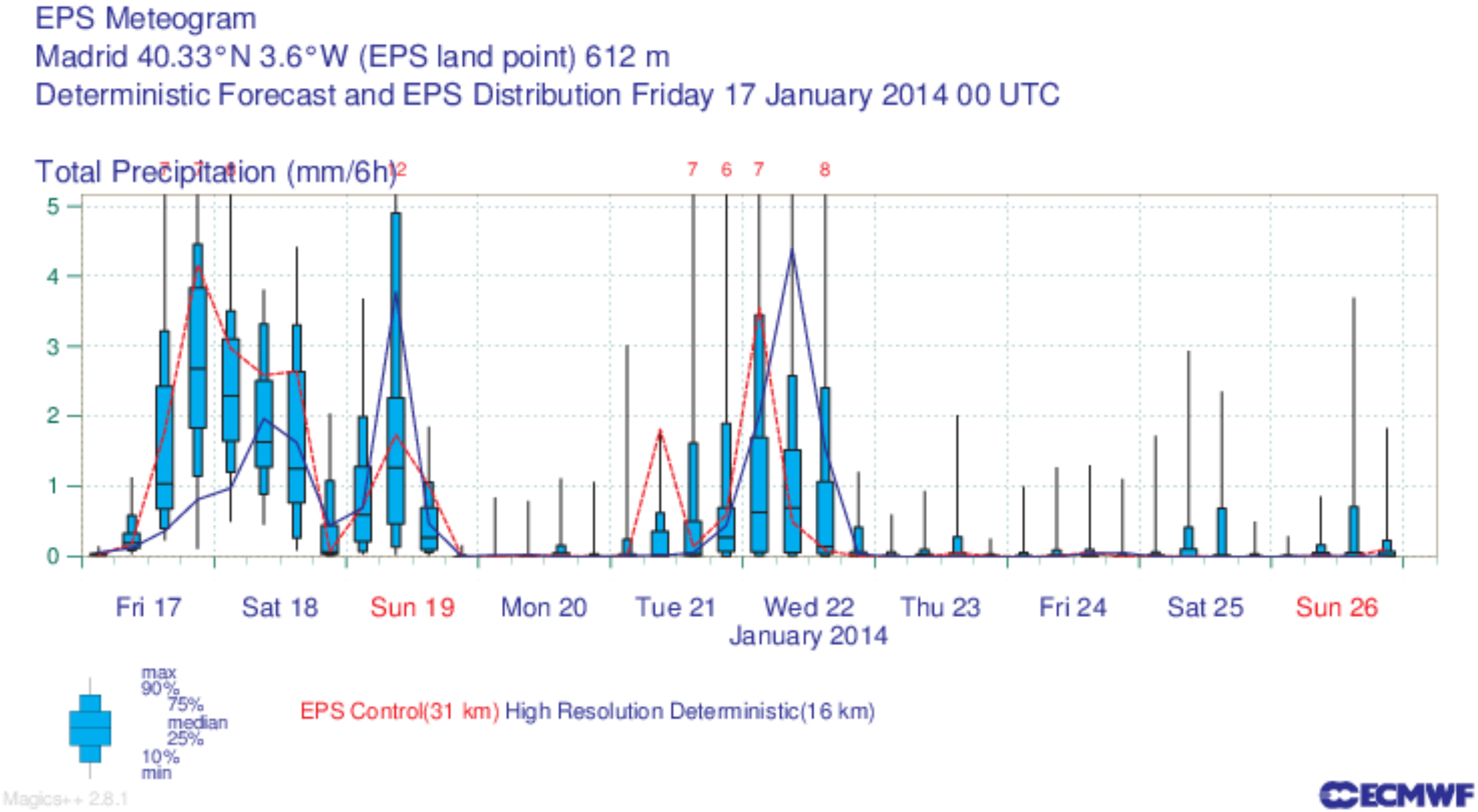
2-category errors rare at short lead-times. High penalties later-on.

1-category errors more common at short lead-times. Can we improve these?

DJF 2009/10. 24-hr precipitation accumulations, Extra-tropics
Diagonal panels show category frequencies (÷2)
Off-diagonal panels show SEEPS contributions for each error-type

Rodwell et al (2011), North et al (2013)

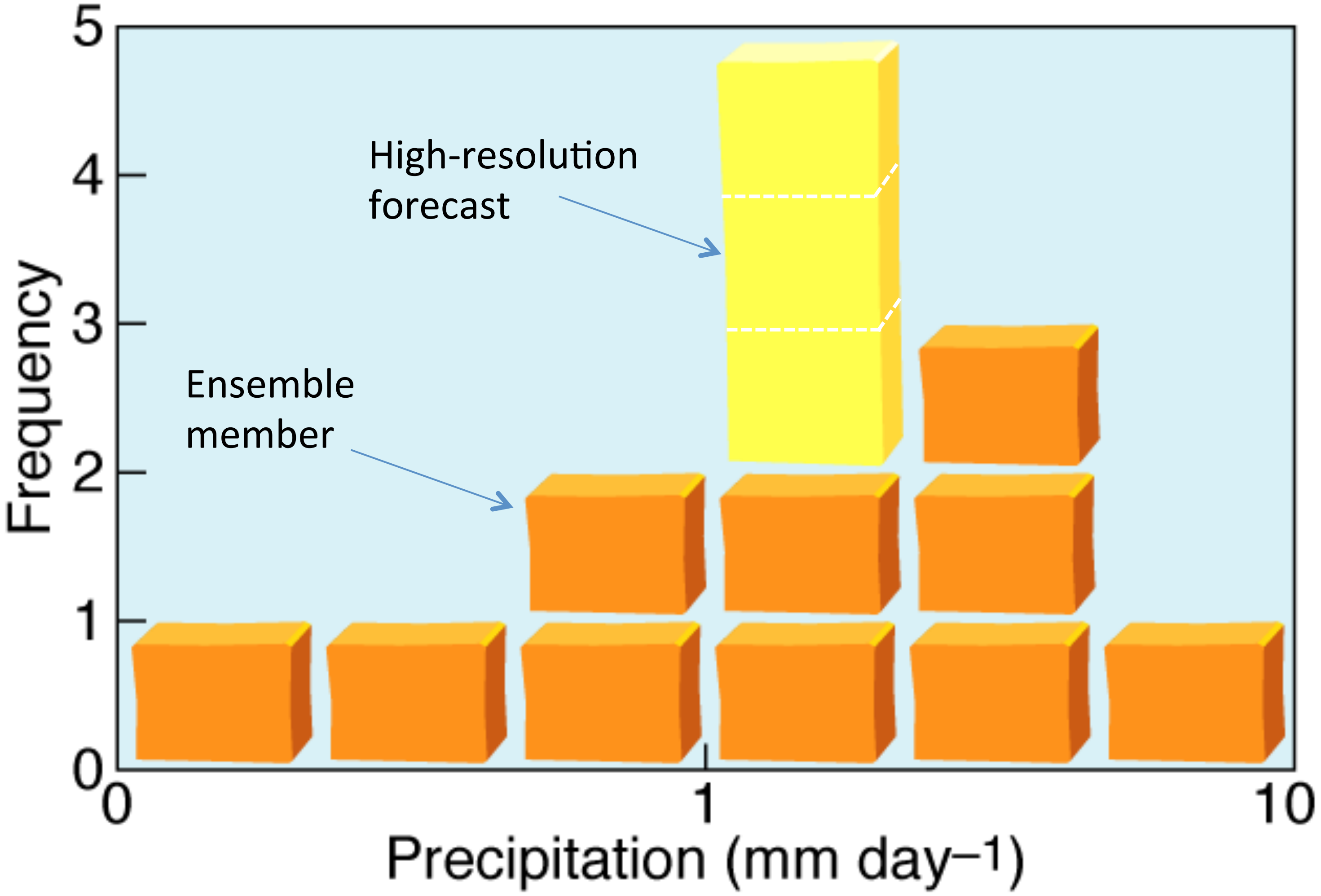
Ensemble and high-resolution information



EPS Meteogram
Madrid 40.33°N 3.6°W (EPS land point) 612 m
Deterministic Forecast and EPS Distribution Friday 17 January 2014 00 UTC

Total Precipitation (mm/6h)

Combined Prediction System: Methodology



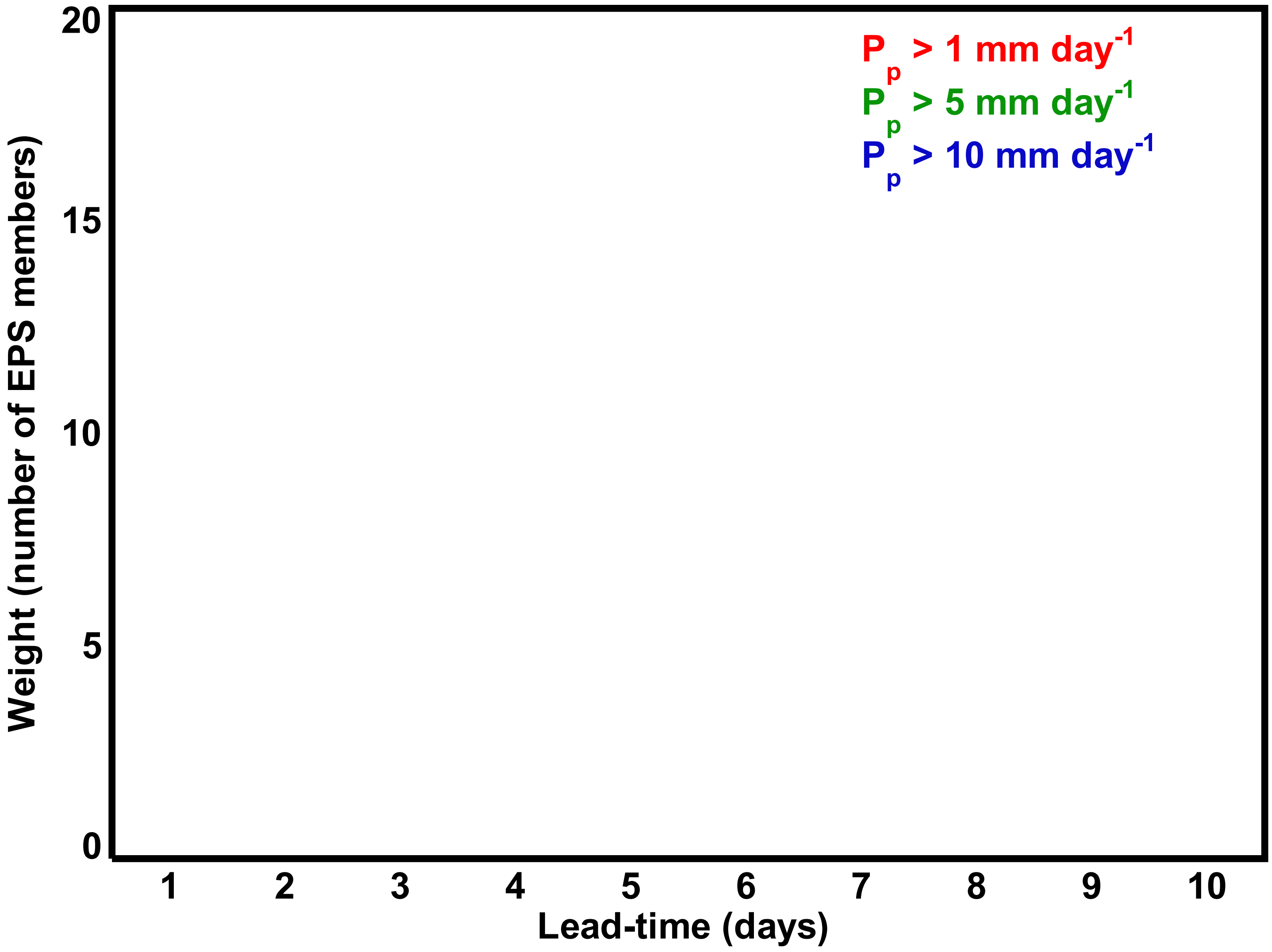
In the example, $\text{weight}_{HRES}=3$ and the probability of 1mm precipitation = 9/13

In the real case, find weight_{HRES} that maximises (*e.g.*) Brier Skill Score or Ignorance score

Can do analytically by solving $\partial BSS/\partial w_{HRES} = 0$

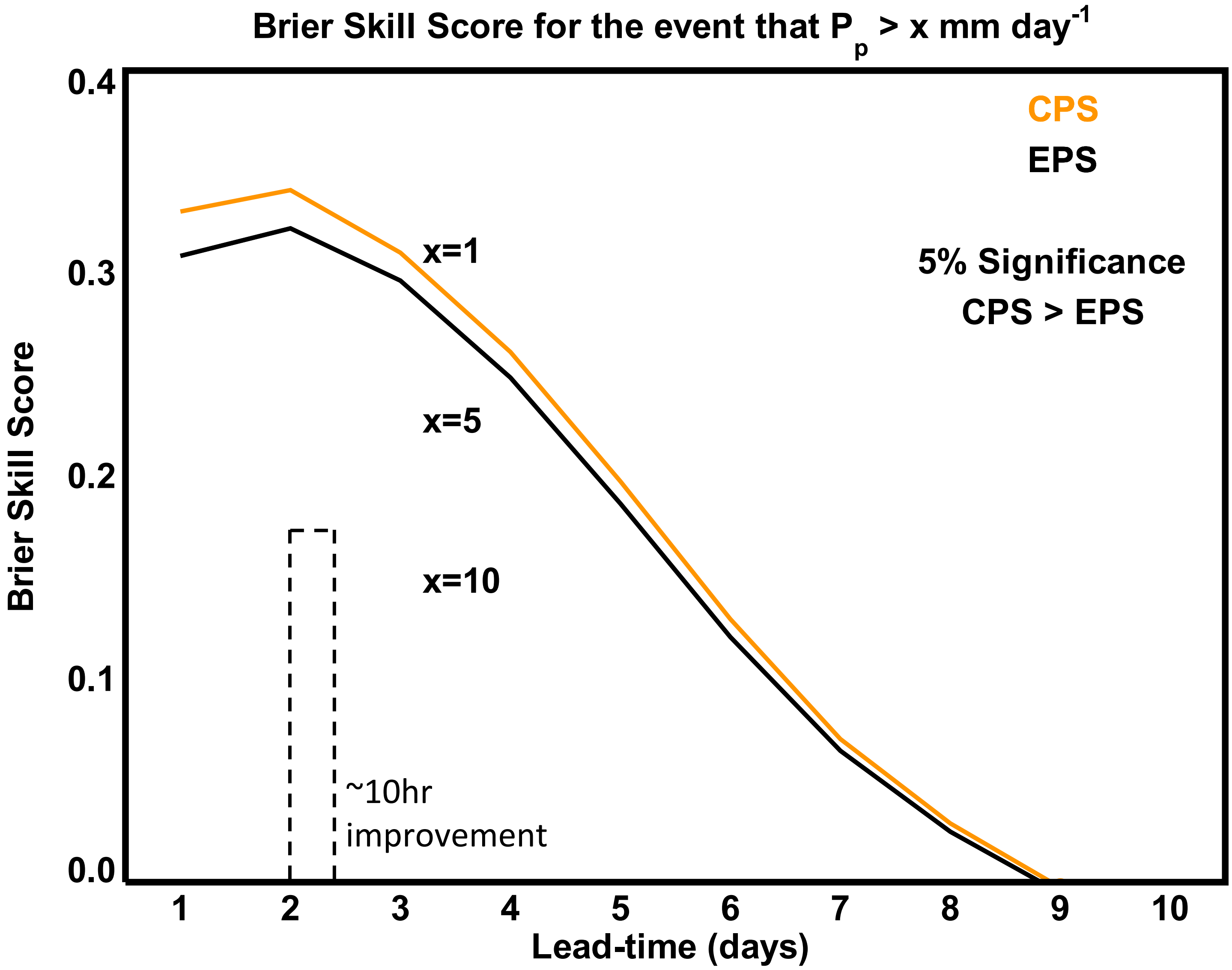
Rodwell ECMWF Newsletter 106 (2006)

The weight to give the high-resolution system



At short lead-times, high-resolution system is very valuable. At longer lead-times weight →1. Based on years 2001-2005

Combined system is more skilful



Results are cross-validated so no artificial inflation of skill. Based on years 2001-2005

M.J. Rodwell, RMetS theWeatherClub

Then & Now

It's often sunny this time of year – shall we go hiking?

I'll keep the weekend free

Weatherman says it'll be sunny on Saturday

I'll run to the station and buy the train tickets

Weathergirl says Saturday doesn't look so good now!

Let's hope for the best!

What a lovely time …
I actually think the countryside does look best in the rain …

No comment!

MON

Weather app says there's an 80% chance of a dry scenario this weekend

Let's keep it free for mountain biking!

THU

Its 50:50 on Saturday due to local showers, but 90% chance dry on Sunday

Let's aim for Sunday, and invite friends over for lunch on Saturday

FRI

Nearly 100% probability it'll be sunny on Sunday!

I'll book the train online for Sunday

SUN

Awesome time in the hills

3.6K Likes 127 Comments

Summary

- Diagnostic issues: Chaos & complexity
- Deterministic model error: Initial tendencies within data assimilation cycle
  - Assessing models (*e.g.* Perturbed climate ensemble)
  - Identifying errors (Upper-tropospheric cold bias, convective momentum transport)
- Ensemble distribution: Spread & error of EDA (observation-space) & ENS
  - Quantifying flow-dependent uncertainty (What is an ‘ensemble bust’?)
  - Key processes that magnify uncertainty (*e.g.* MCSs, baroclinic instabilities, *etc.*)
  - Key initial condition errors & sub-grid-scale physics uncertainty
- Understanding the circulation
  - Targeting and monitoring key sources of predictability (*e.g.* Rossby wave source)
- Diagnostics for users
  - The deterministic and probabilistic weaknesses of the forecast system
  - Exploiting information from high-resolution and ensemble (combined approaches)