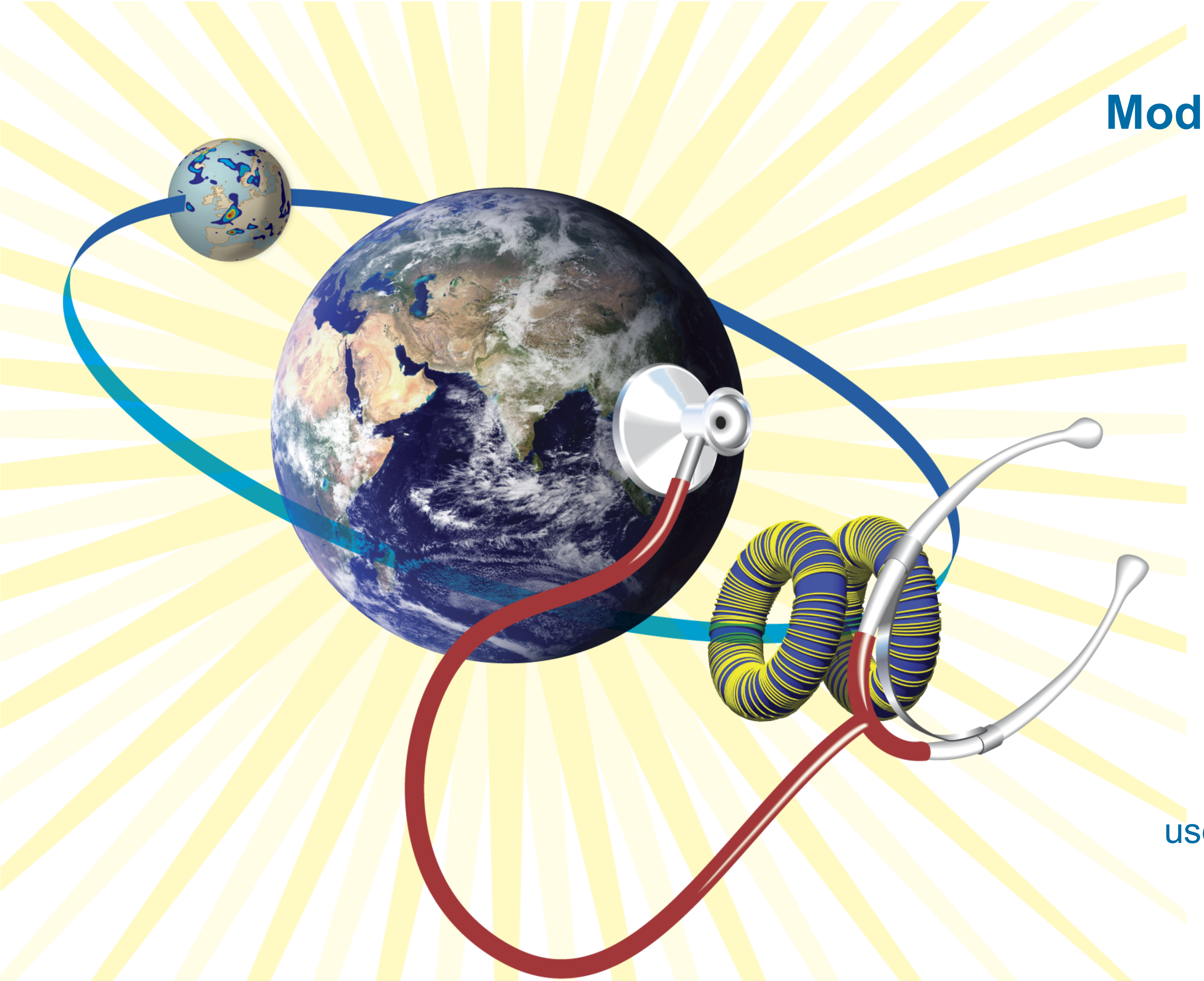


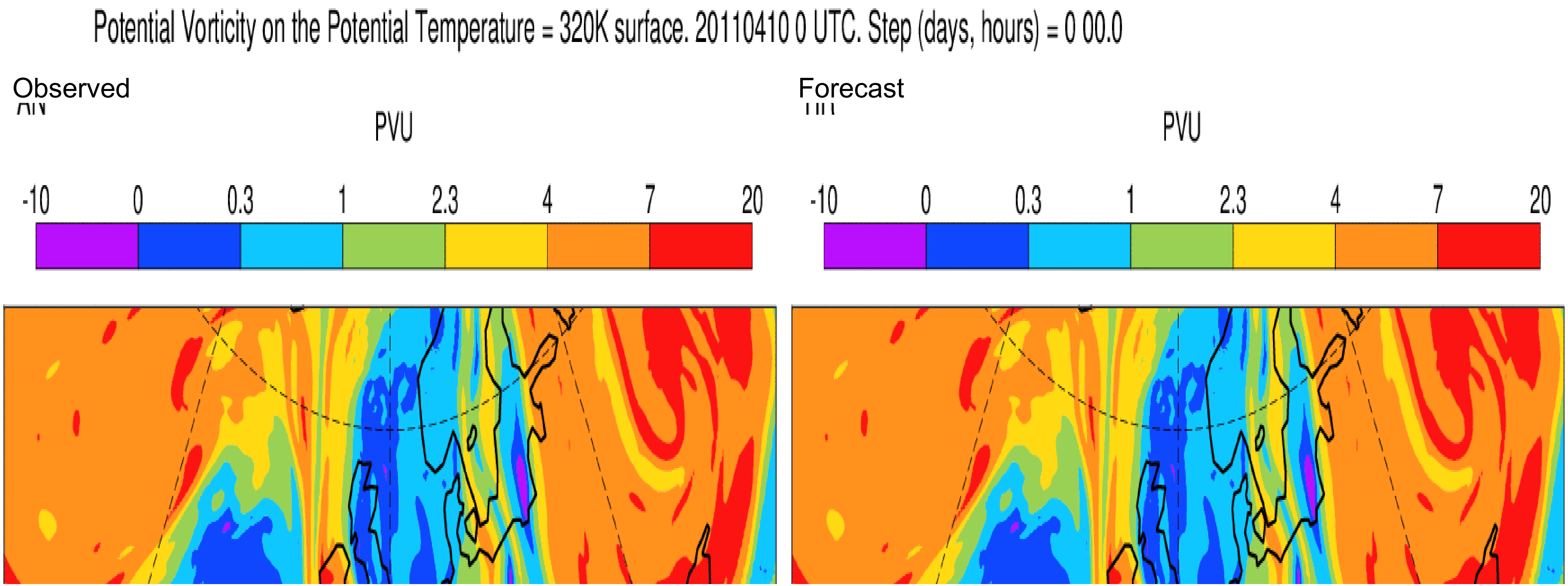
Model Errors & Diagnostic Tools

Mark Rodwell

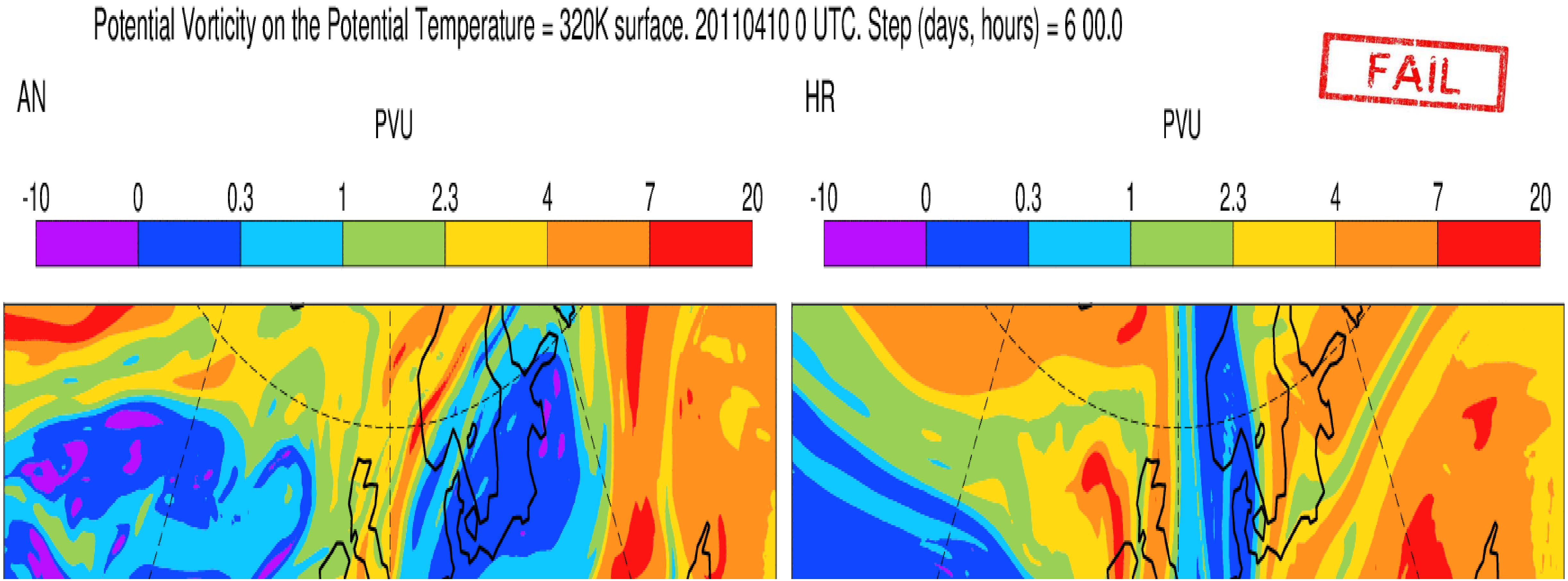
ECMWF training course on the
use and interpretation of ECMWF products

3 February 2017, ECMWF Reading

Animation of a very poor medium-range single forecast



Animation of a very poor medium-range single forecast

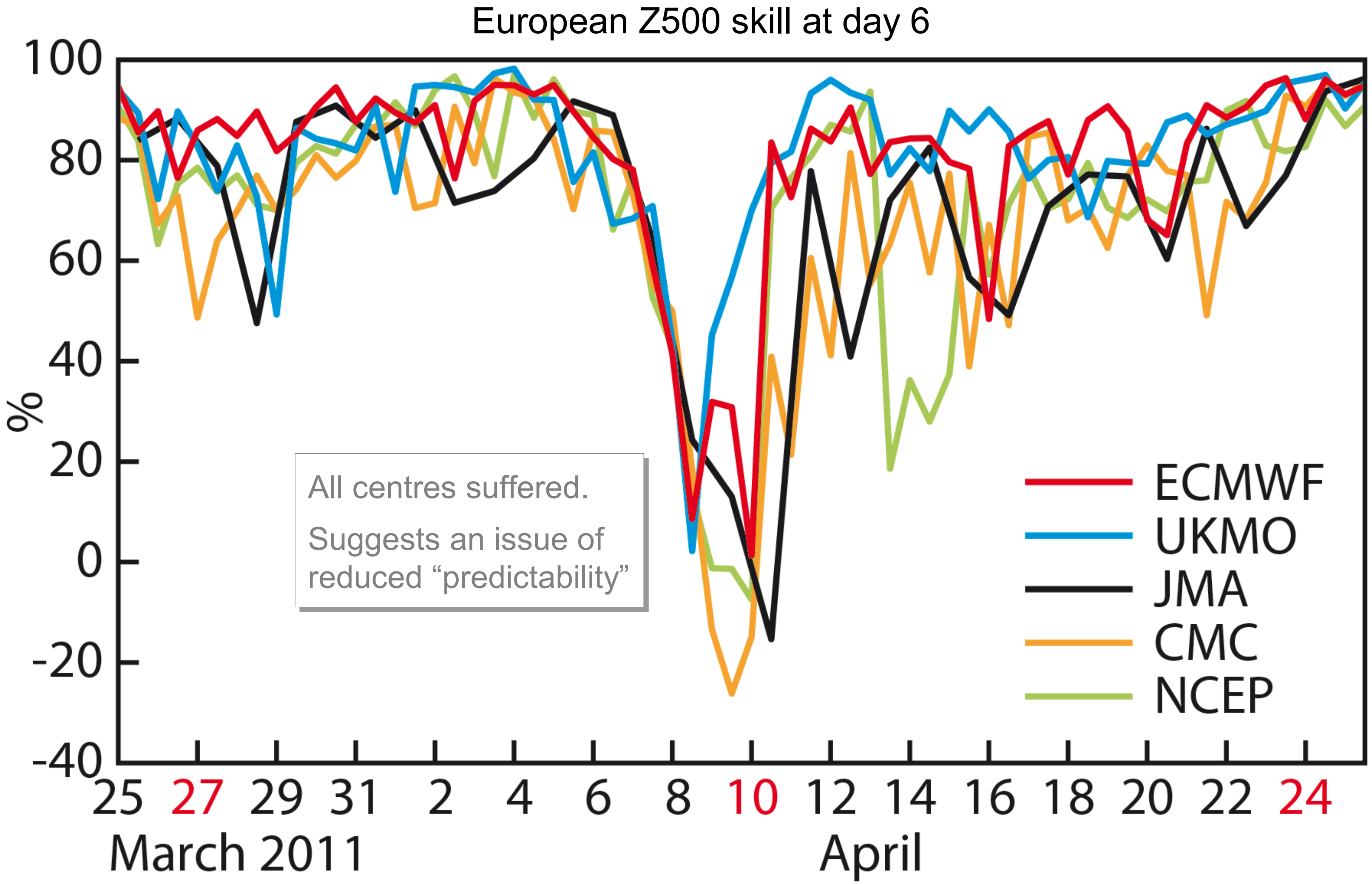


We see the mixing of air masses. The eventual block (high pressure) over Northern Europe is not well predicted

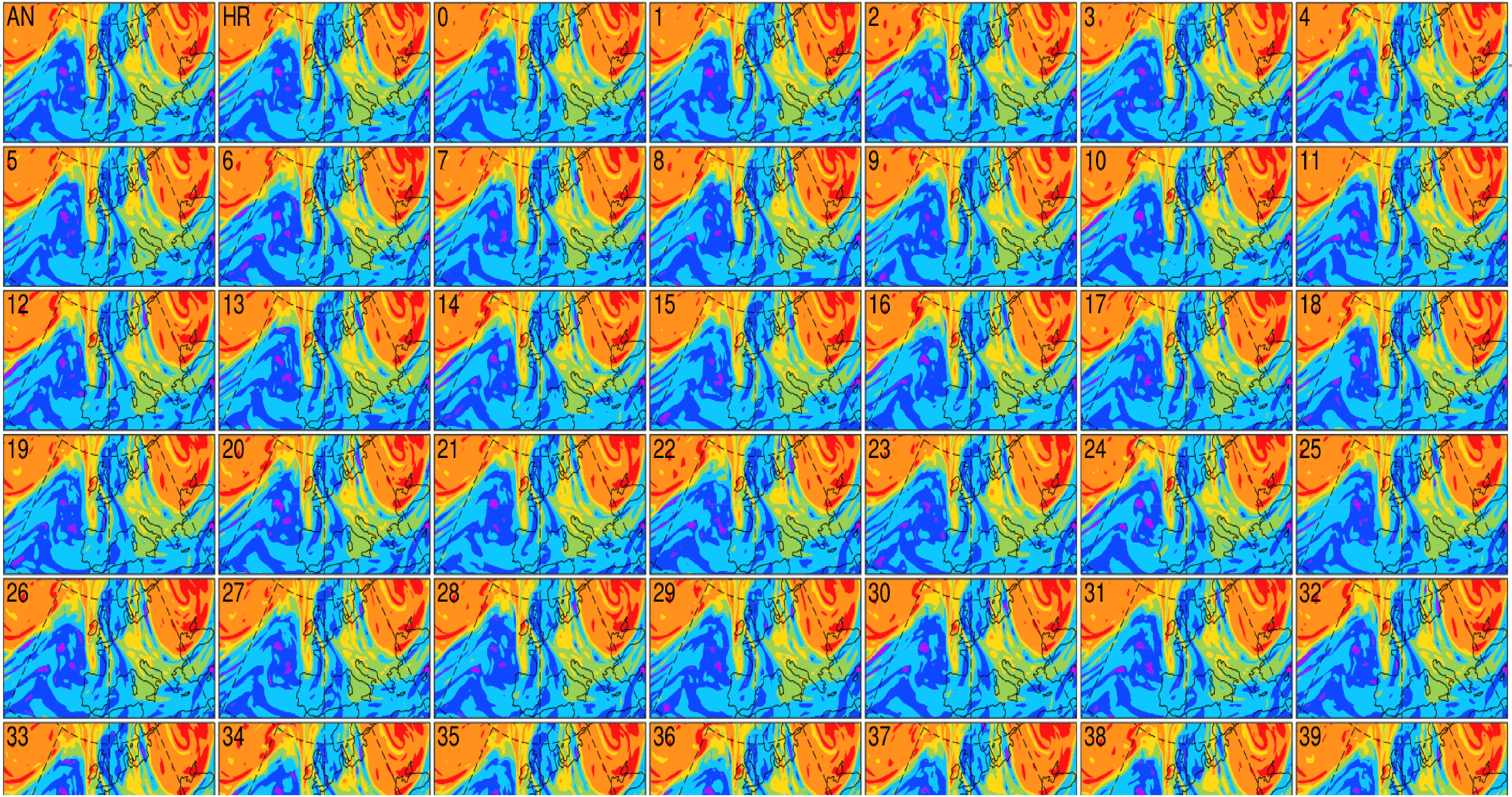
With a single forecast, it is easy to quantify the error (pointwise differences, pattern correlations etc.)

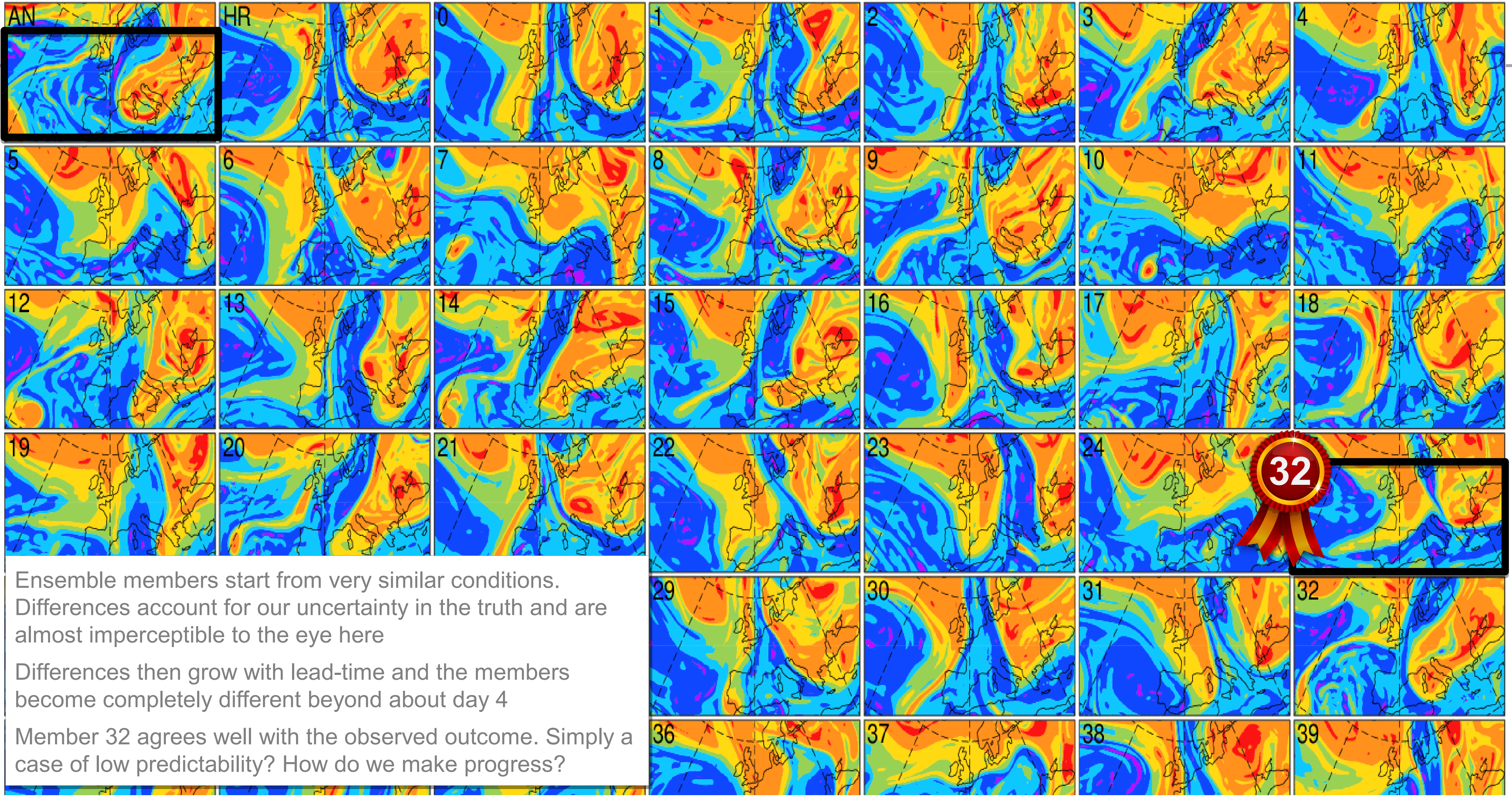
All forecast centres suffered

Rodwell et al, 2013, BAMS



Spatial Anomaly Correlation Coefficient for 500 hPa geopotential height in [12.5°W –42.5°E, 35°N–75°N]. Date is forecast start

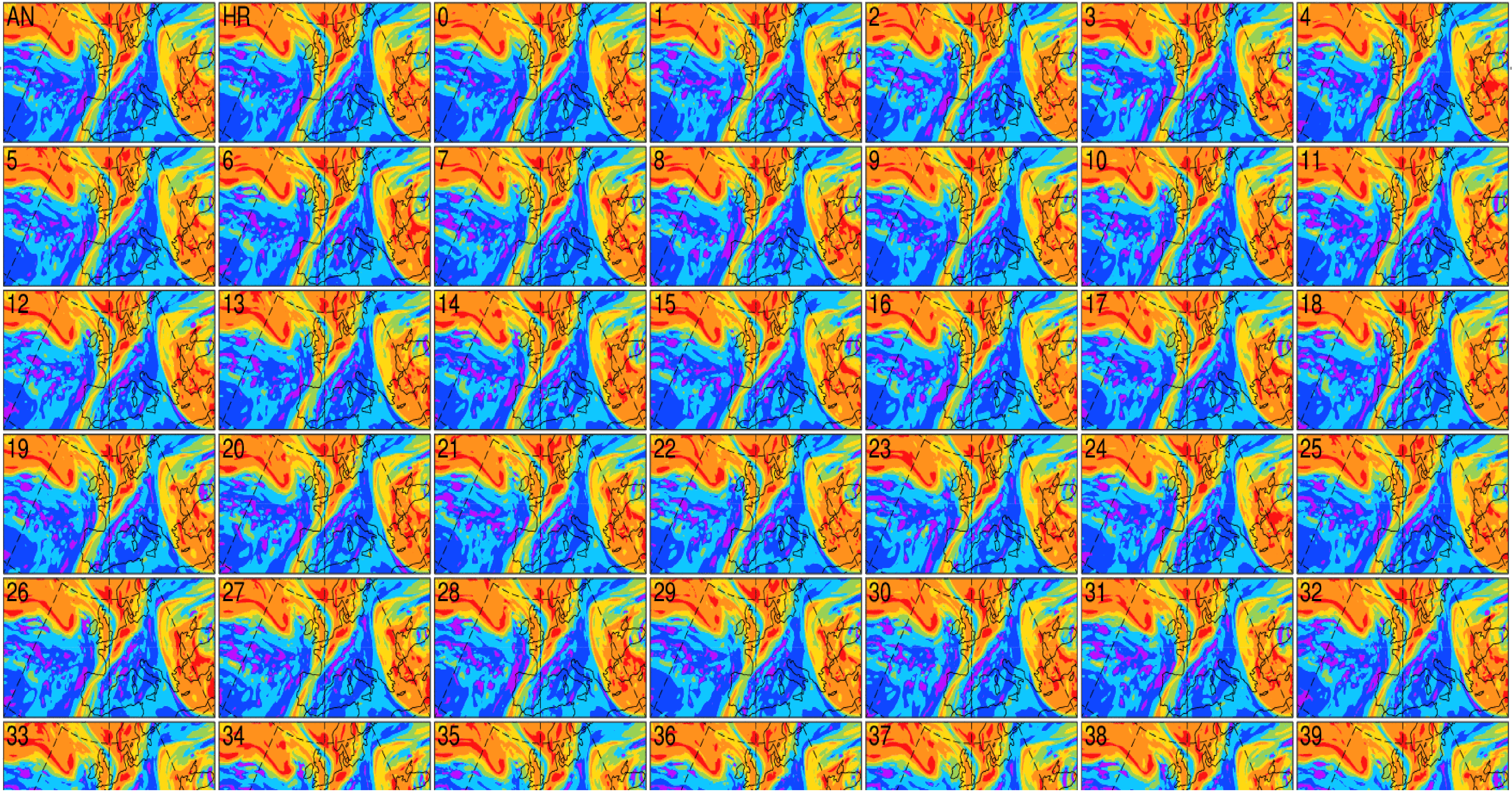

AN
HR
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39

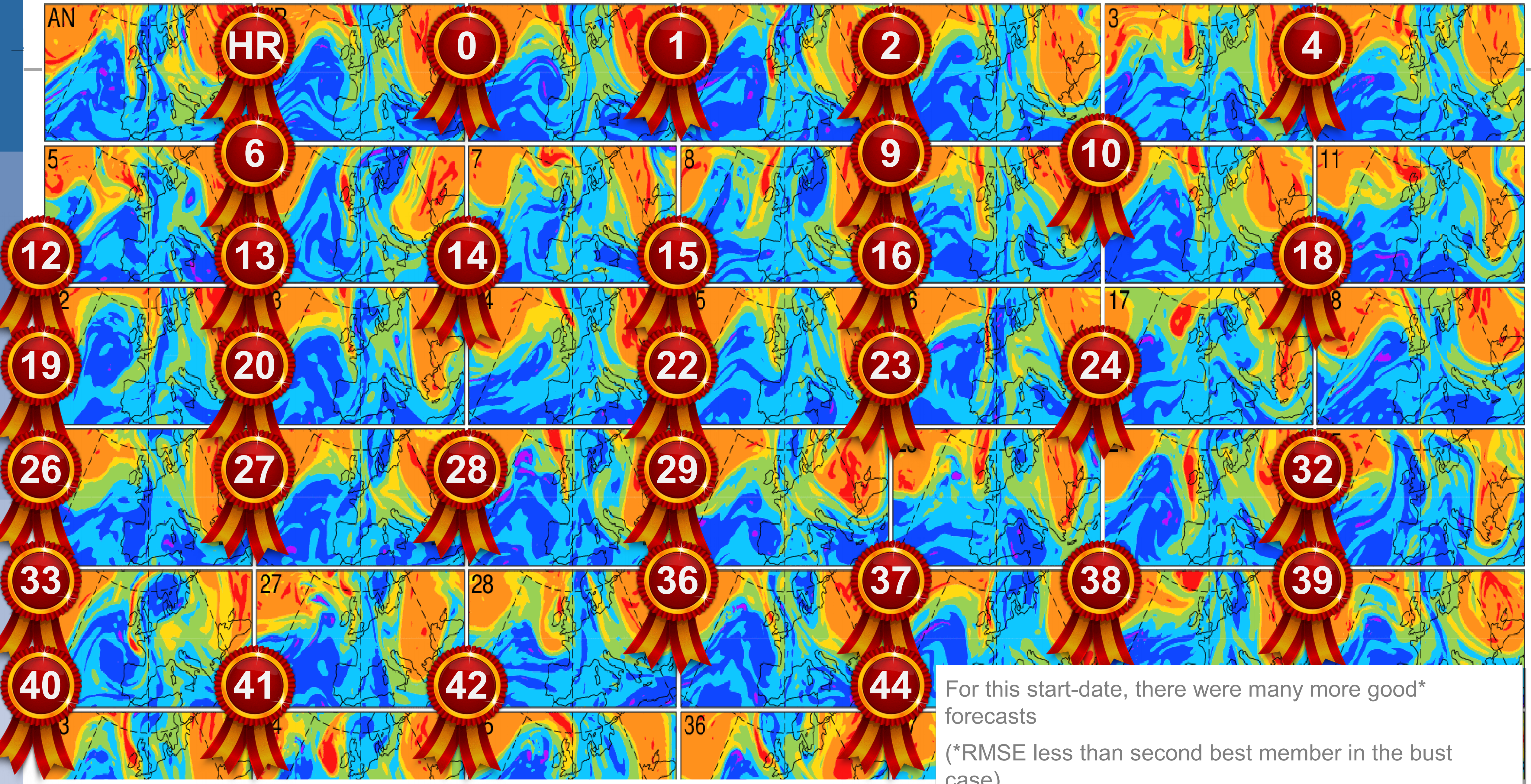


Ensemble members start from very similar conditions. Differences account for our uncertainty in the truth and are almost imperceptible to the eye here

Differences then grow with lead-time and the members become completely different beyond about day 4

Member 32 agrees well with the observed outcome. Simply a case of low predictability? How do we make progress?

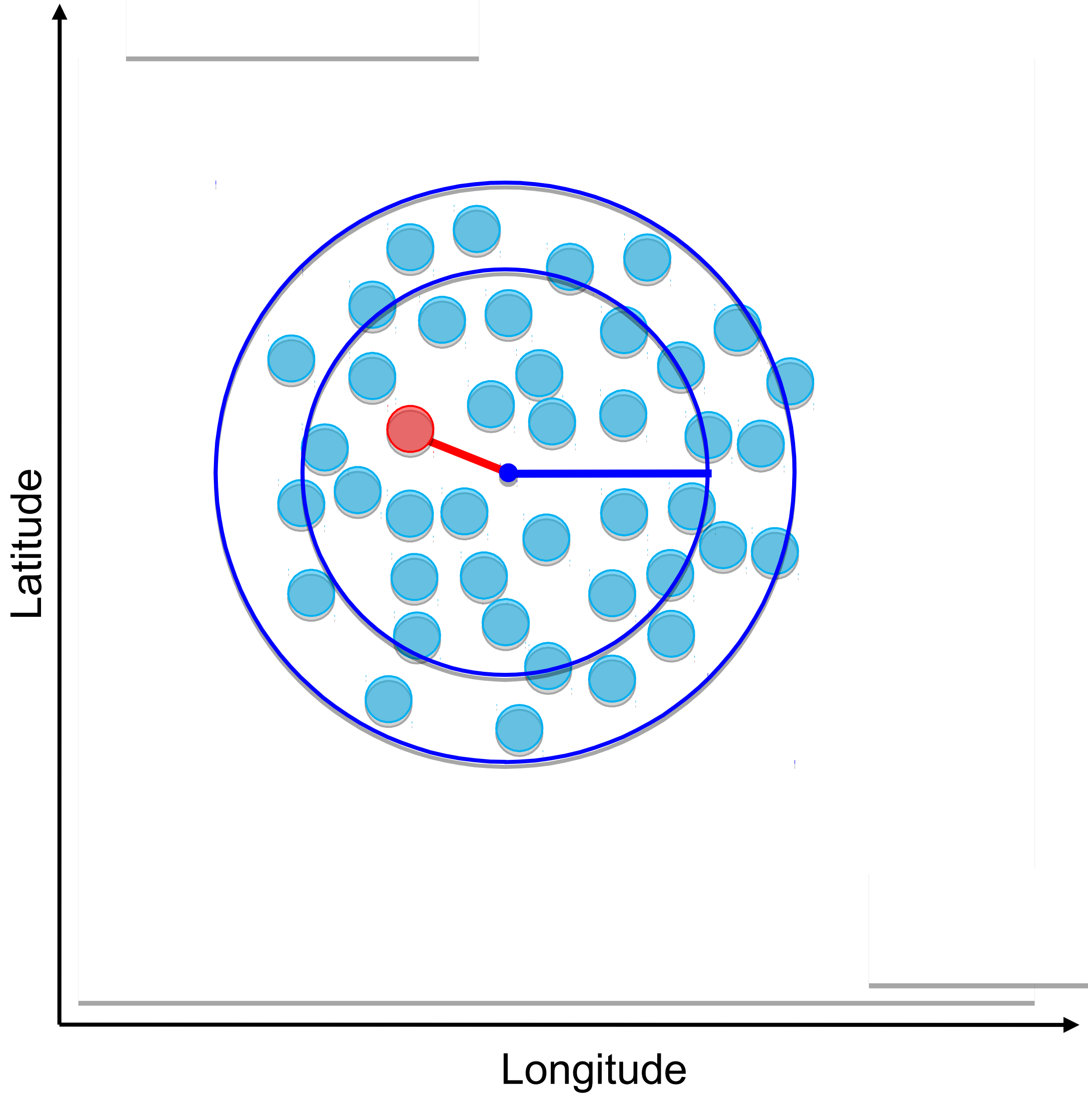

AN
HR
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39



For this start-date, there were many more good* forecasts

(*RMSE less than second best member in the bust case)

Reliability and Sharpness – Example based on forecast of storm location



In a "**reliable**" forecast system, the truth can be considered as another ensemble member

Reliability is very useful: if we predict an event with probability 70%, it will happen with frequency 70%

A testable consequence of reliability is that:

average Error = average Spread

(averaged over many forecast start dates)

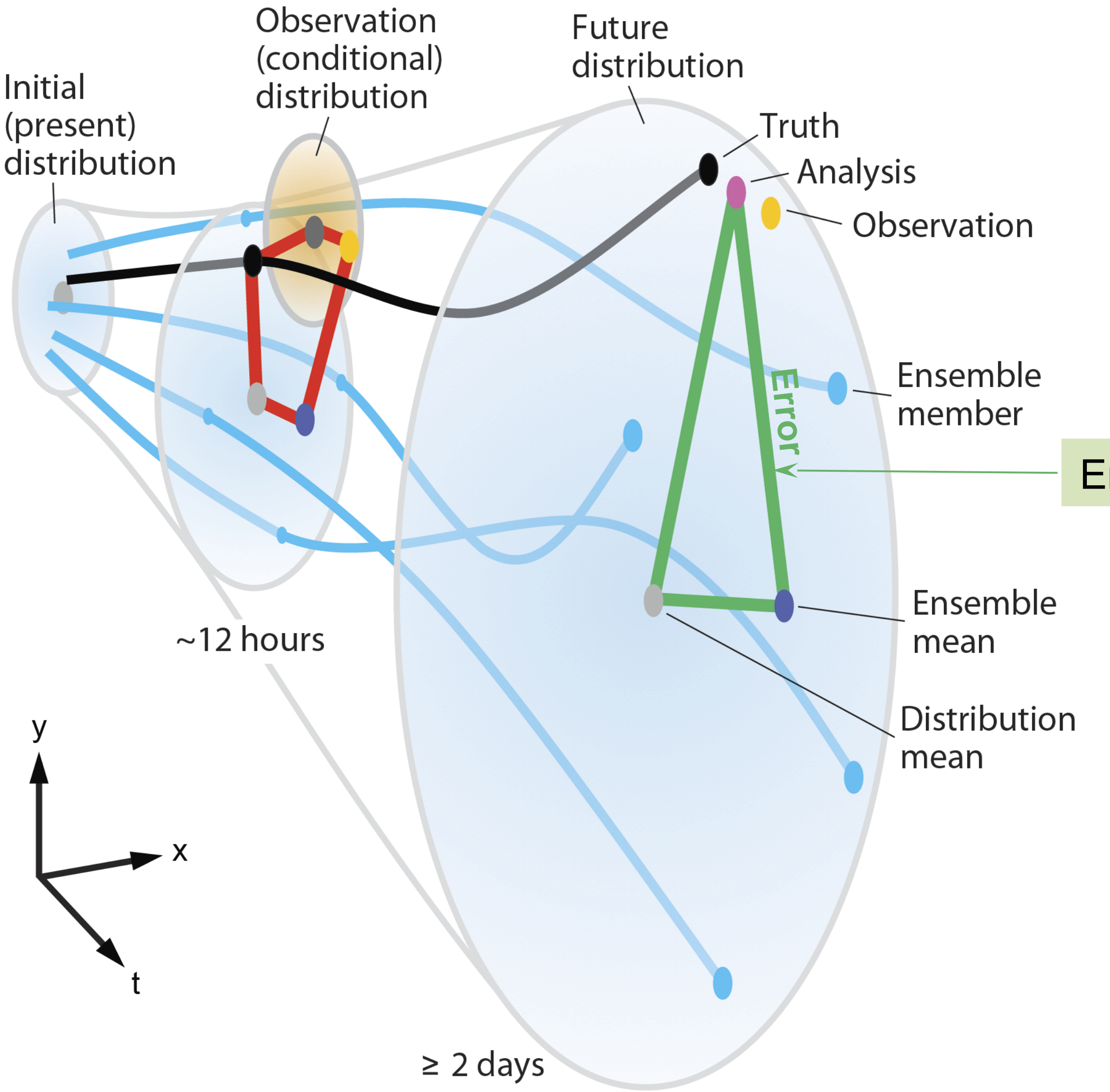
Given we had a reliable system, progress would be …

Predicting "**sharper**" (tighter) distributions **while retaining reliability**

(A more predictable day should also have a sharper distribution)

Reliability in ensemble forecasting

Adapted from Rodwell et al. (2015) QJRMS



$\text{Error}^2 = \text{EnsVar} + \text{Residual}$

Error ≈ Spread

(when averaged over enough start dates - a consequence of statistical "reliability")

Progress: While maintaining this relationship, we wish to reduce both error and spread (i.e. "shrink diagram in x,y")

Proper scores (Brier, CRPS, Ignorance etc.) help ensure that development decisions lead us in this direction

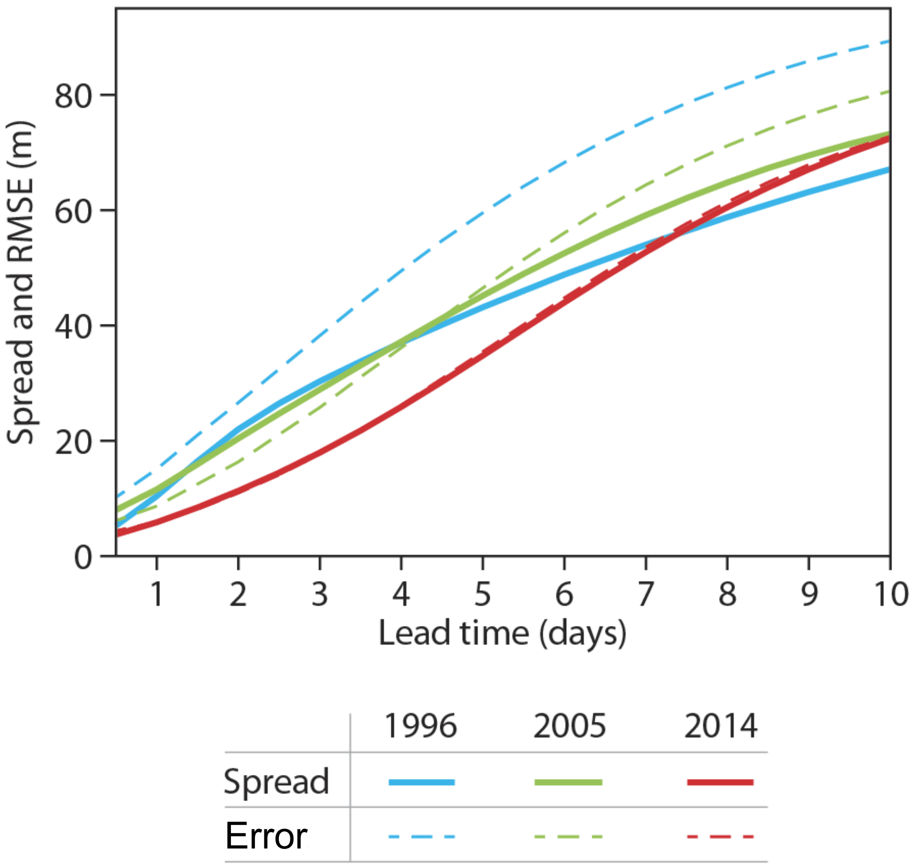
(Cross-terms on squaring have zero expectation. EnsVar is scaled variance to account for finite ensemble-size)

Ensemble spread and error

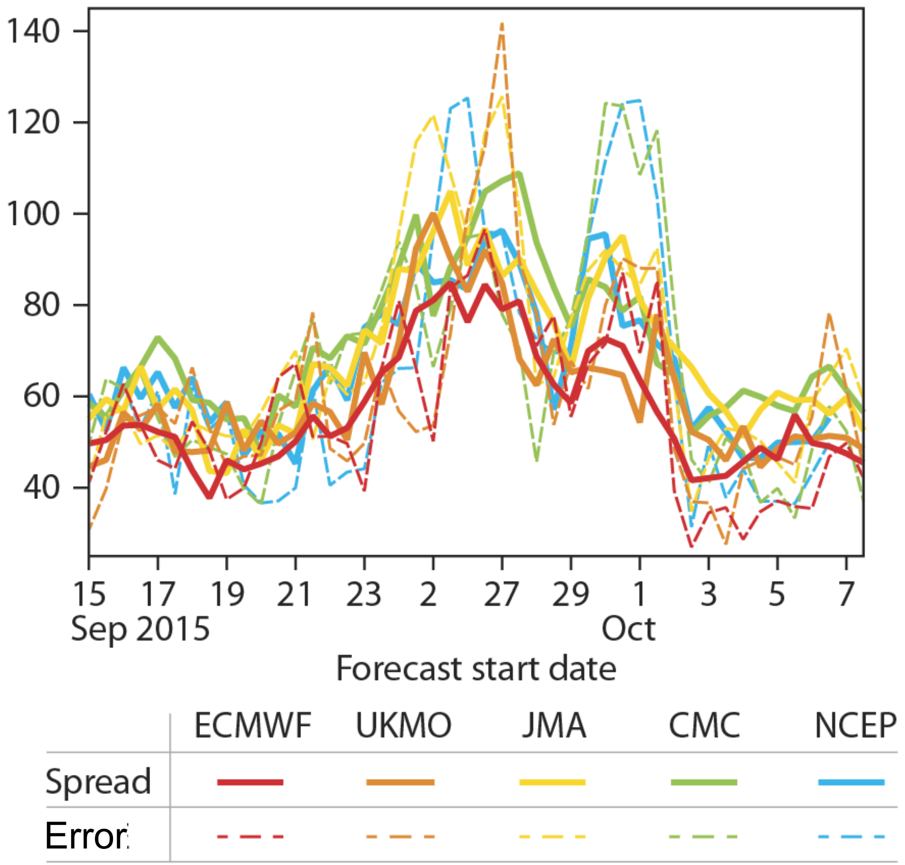
Z500

Rodwell 2016, ECMWF Newsletter

Northern Hemisphere, annual mean



Europe, day 6



Overall Error and Spread have reduced and come into alignment; due to better observations, initial conditions, forecast model and better representation of uncertainty

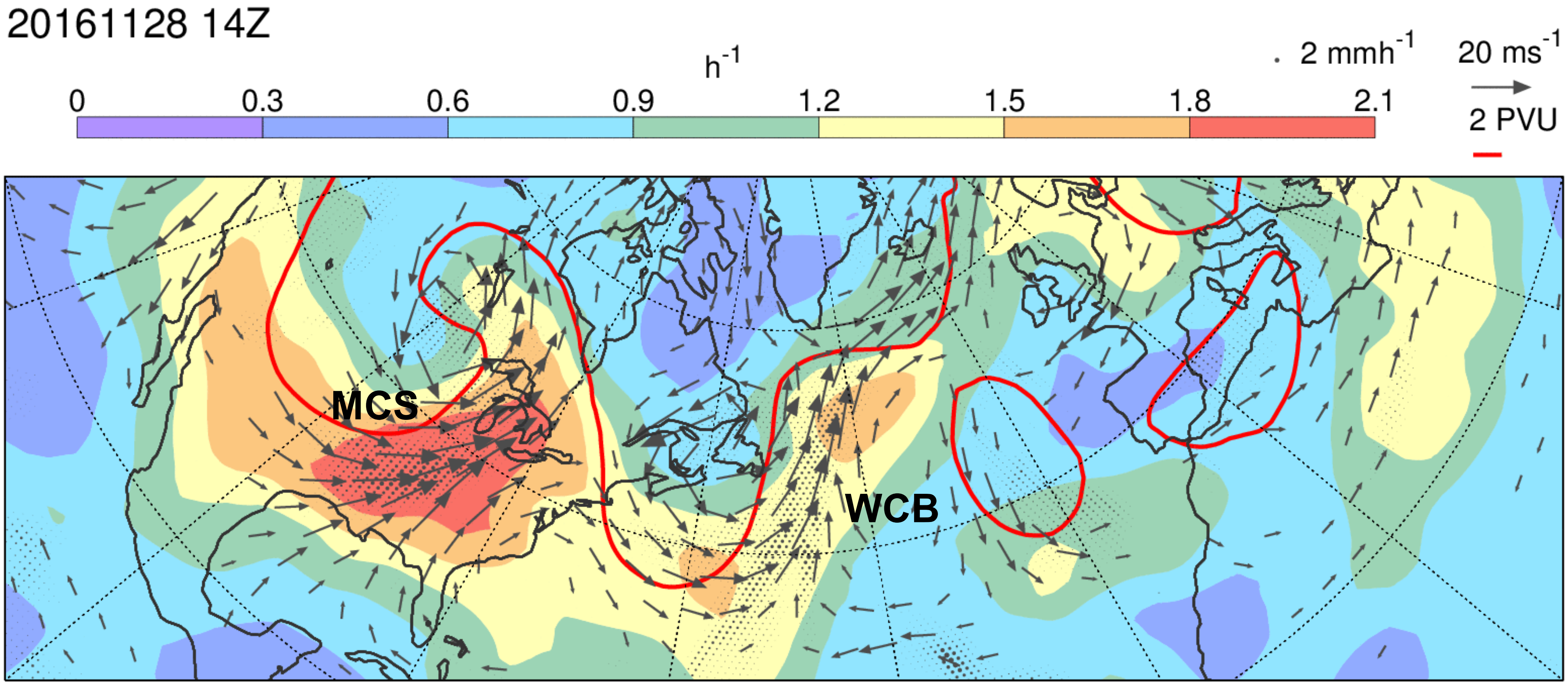
… but uncertainty varies from day-to-day. The real reason we make ensemble forecasts. What causes this, and how can we evaluate it in our forecasts?

To make progress, we must avoid too much chaos, and look at the growth of uncertainty at very short lead-times

500 hPa geopotential height (Z500). “Error” is RMS of ensemble-mean error
Spread = ensemble standard deviation (scaled to take account of finite ensemble size)

Animation of “instantaneous” growth of uncertainty

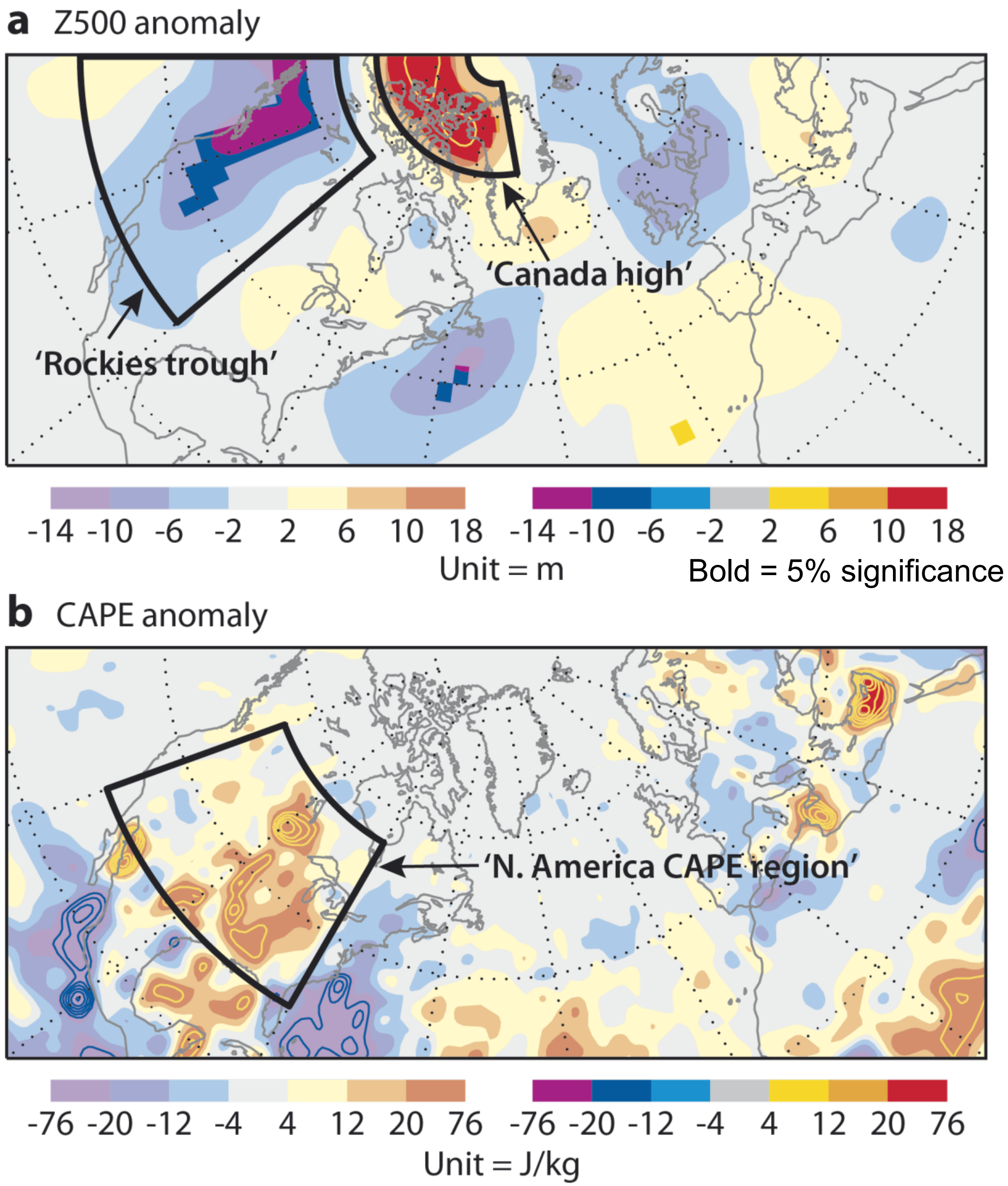
Instantaneous growth of uncertainty (in the background forecasts of the ensemble of data assimilations) for PV315 (shaded). Also shown are control forecast PV=2 on 315K (red contour) and v850 (vectors), and ensemble-mean precipitation (dots; size indicates rate). All with 1d running mean applied.



A lot of “instantaneous” growth of uncertainty is associated with cyclogenesis, warm conveyor-belts, and meso-scale convection (*e.g.* over the US). How do we evaluate our representation of this growth of uncertainty?

Average initial conditions of 584 single forecast "busts" over Europe at day 6

Rodwell et al, 2013, BAMS



Trough over the Rocky mountains, with high convective potential ahead

Conducive to the formation of mesoscale convection

Can average over such cases to evaluate flow-dependent reliability and thus our model uncertainty

(Subsequent evaluation requires independent data to avoid misleading results)

'CAPE' = Convective Available Potential Energy

Meso-scale convection over Kansas

Systems grow to typically 500km in scale, with embedded convective cells and tornados

The Jetstream and meso-scale convection: “The piano string and hammer”

54 cases

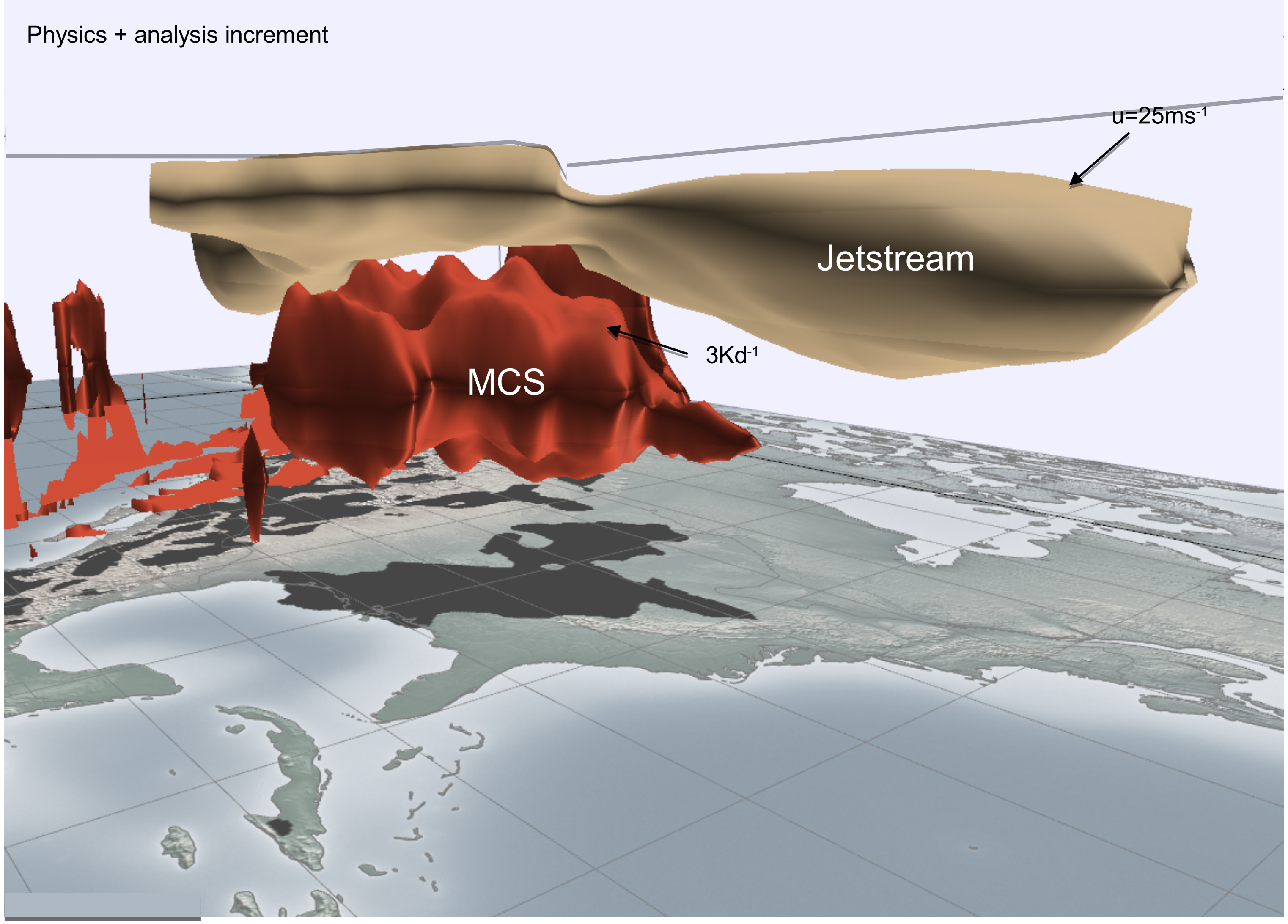
Met3D: Marc Rautenhaus

If we don’t hit the string hard enough, the wave in the string will be too weak

If we hit the string at the wrong time, the wave will arrive over Europe at the wrong time

We do not know when to press the key (mesoscale convection itself involves chaotic uncertainty)

What we want is that the ensemble members generate such convection with the “right” uncertainty



Reliability in ensemble data assimilation

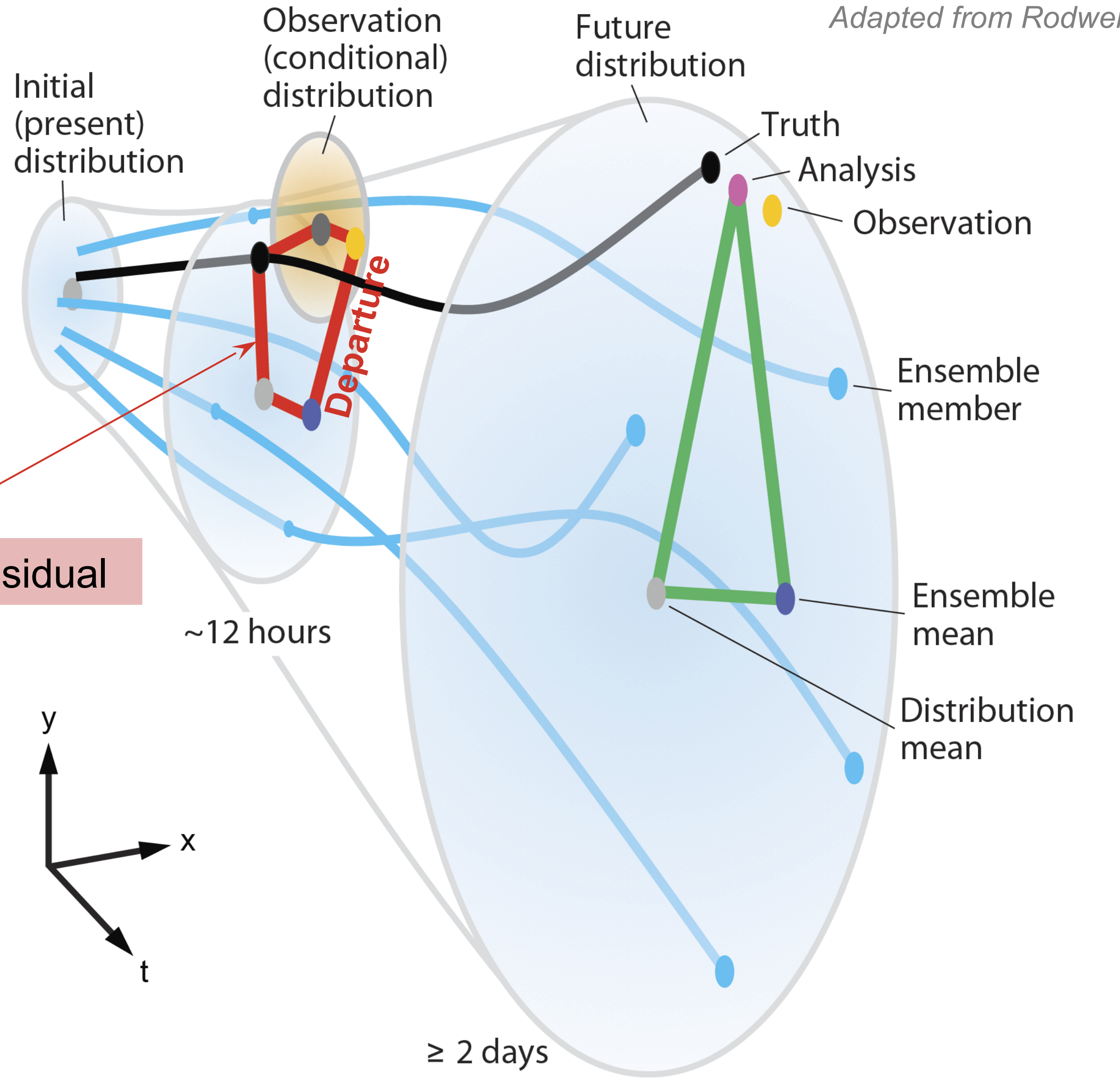
Chaos makes it difficult to identify problems in the medium-range using the spread-error relationship

Go to much shorter lead-times – within ensemble data assimilation process

Need to take account for observation error

Obtain diagnostic equation to evaluate "instantaneous" growth of uncertainty

Adapted from Rodwell et al. (2015) QJRMS



Depar^2 = Bias^2 + EnsVar + ObsUnc^2 + Residual

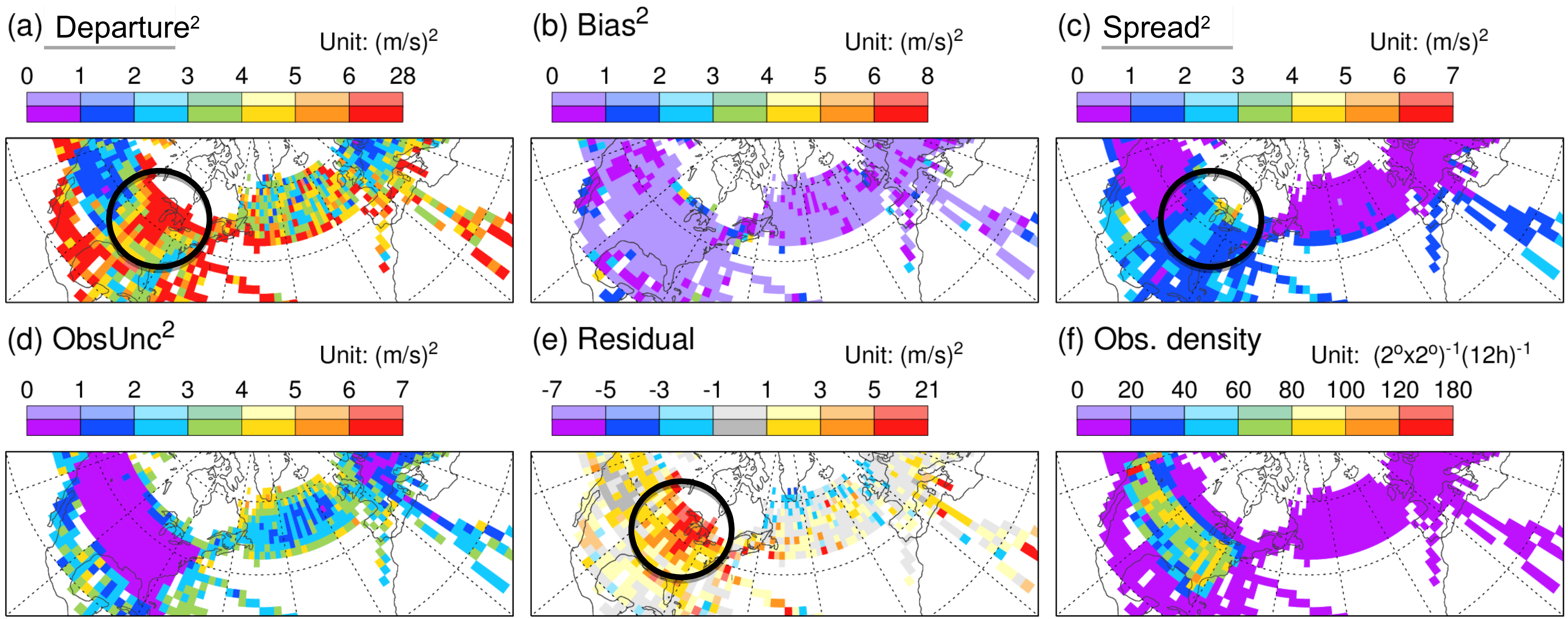
(Cross-terms on squaring have zero expectation. EnsVar is scaled variance to account for finite ensemble-size)

Evaluating model uncertainty in upper-level winds during “Rocky trough” situations

54 cases

Relative to aircraft west-east wind observations at 200hPa (±15)

Rodwell 2016, ECMWF Newsletter



Ensemble spread highlights enhanced uncertainty around the Great Lakes

Large errors ensue

Errors are relative to observations that are also uncertain but, even if we take this into account, there appears to be too little spread (and model uncertainty) in this flow situation

$\text{Depar}^2 = \text{Bias}^2 + \text{Spread}^2 + \text{ObsUnc}^2 + \text{Residual}$

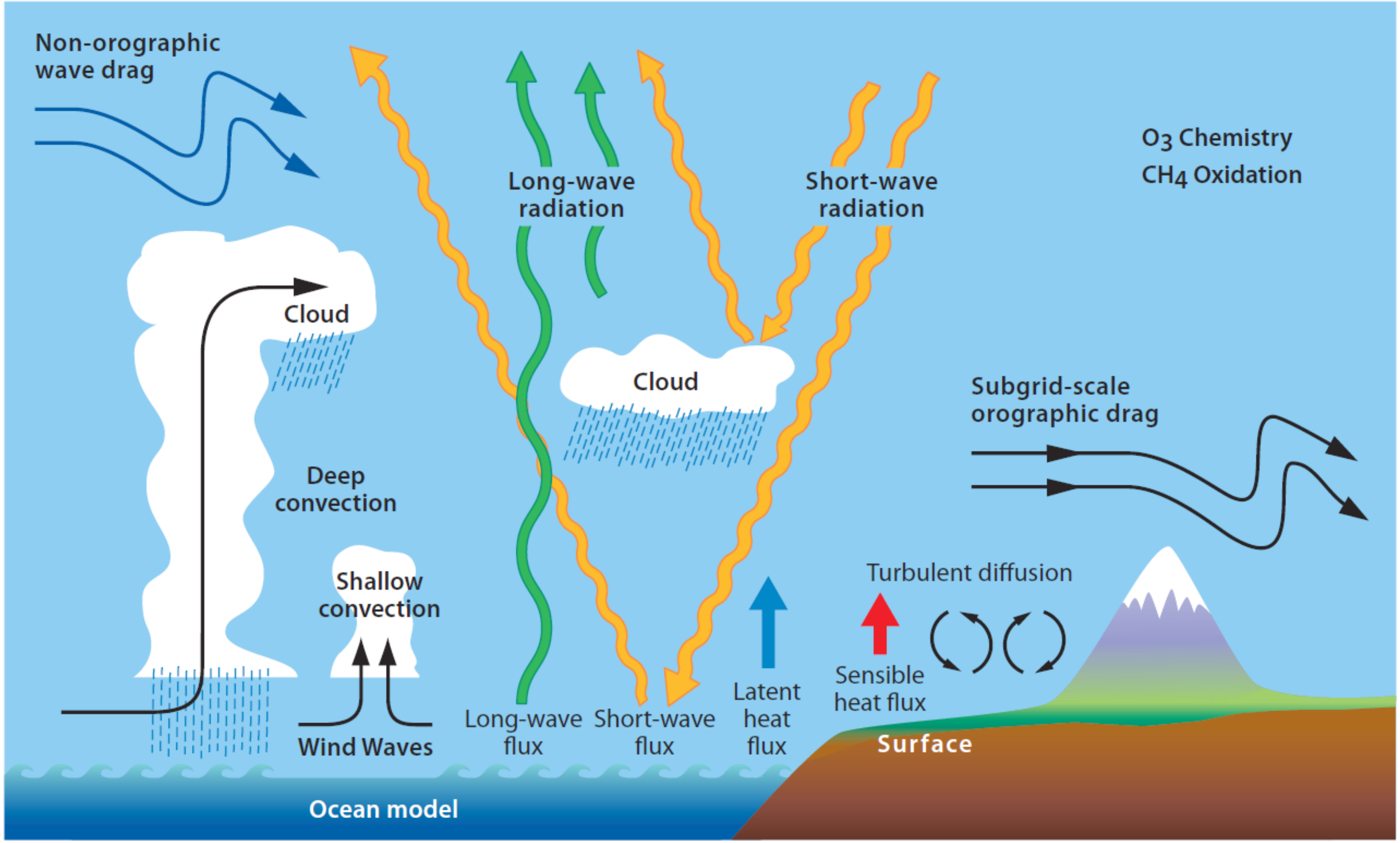
Reliability $\Rightarrow \mathbb{E}[\text{Residual}]=0$

The complexity of present-day model physics

Figure from Peter Bechtold

Ideally, we wish to identify deficiencies at the process level. Again, this should be easier at short timescales since interactions between physical processes and the resolved flow (including teleconnections) are minimised.

Single column and LES models can help, but these do not take into account the evolution of the resolved flow.



mark.rodwell@ecmwf.int

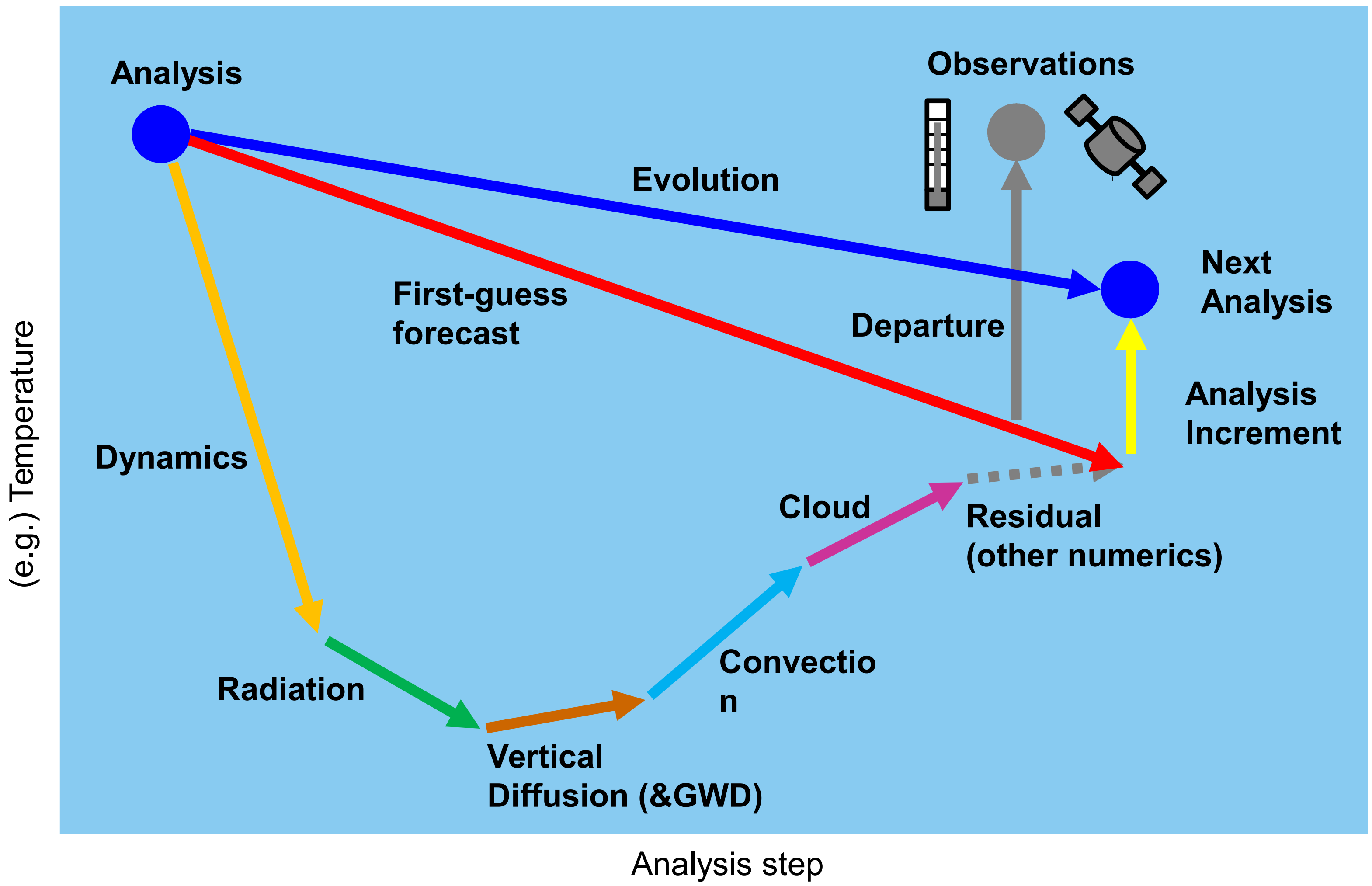
The Initial Tendency approach to diagnosing model error

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

First-guess = sum of all processes

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

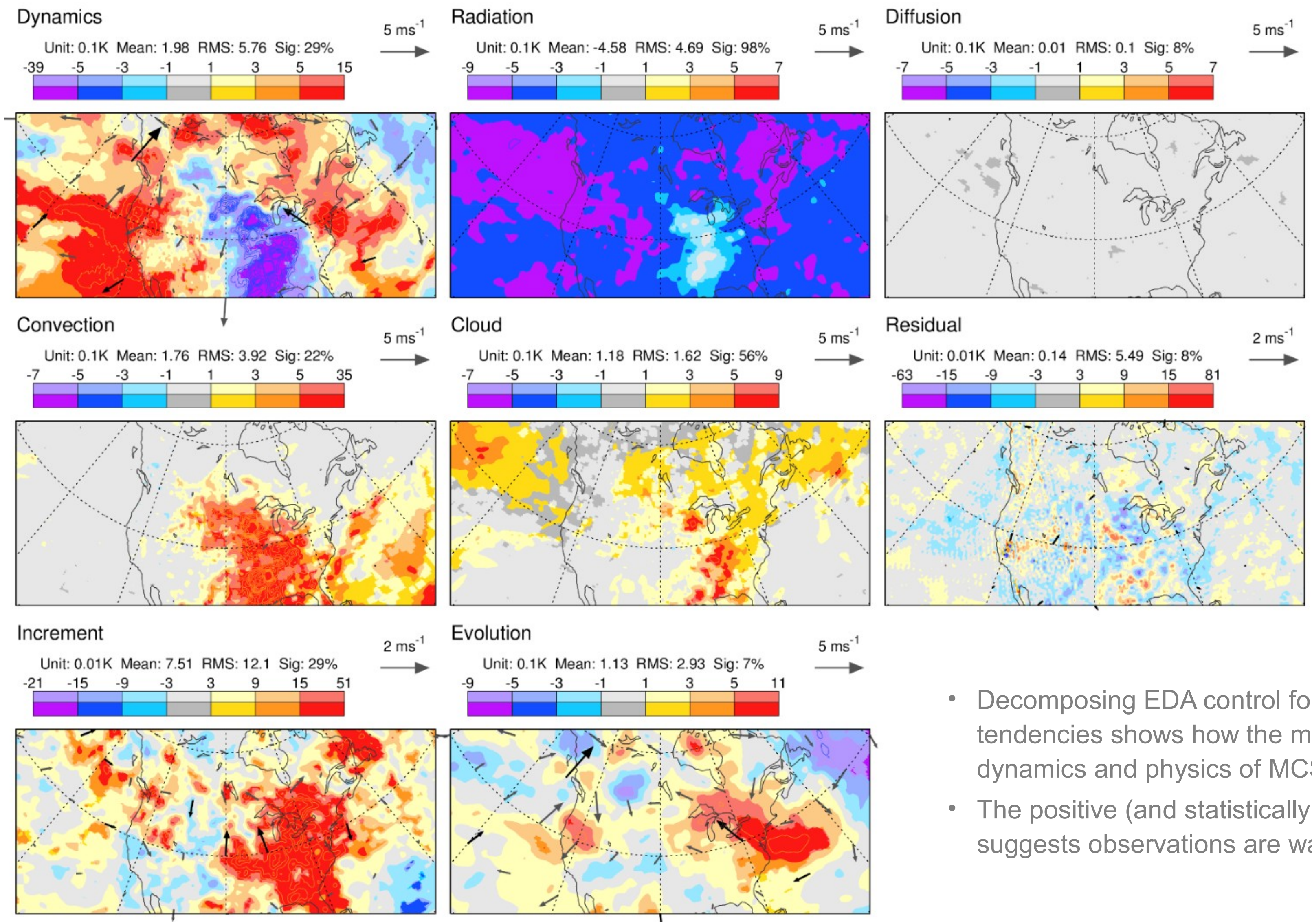
Schematic of the data assimilation process – a diagnostic perspective



"Initial Tendency" approach discussed by Klinker & Sardeshmukh (1992). Refined by Rodwell & Palmer (2007)

Initial tendency budget from control forecast: Trough/CAPE comp.

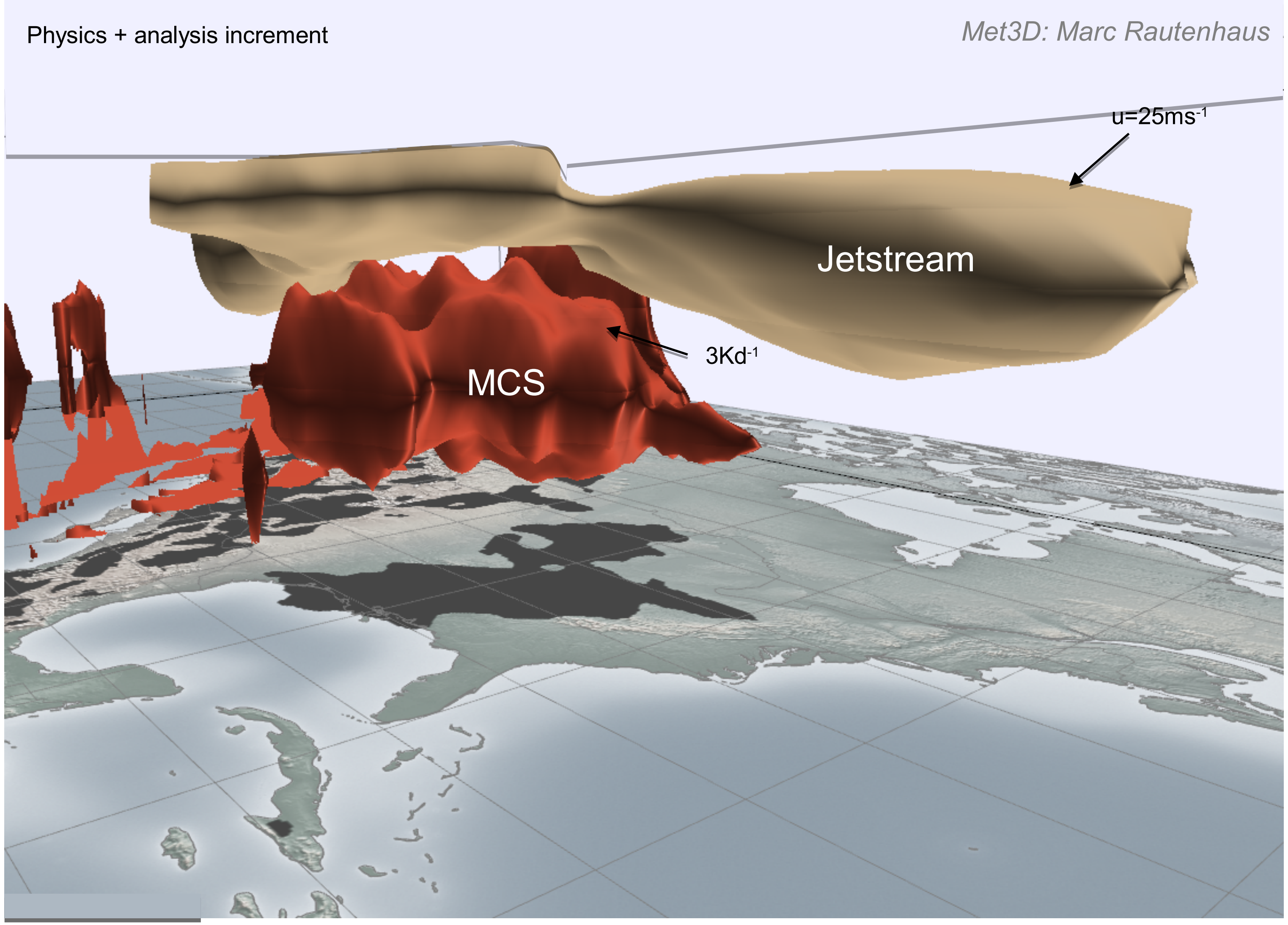
T300, 54 cases



- Decomposing EDA control forecast into process tendencies shows how the model represents dynamics and physics of MCS
- The positive (and statistically significant) increment suggests observations are warmer than the model

Process tendencies accumulated over 12hr background, the analysis increment, and evolution of the flow

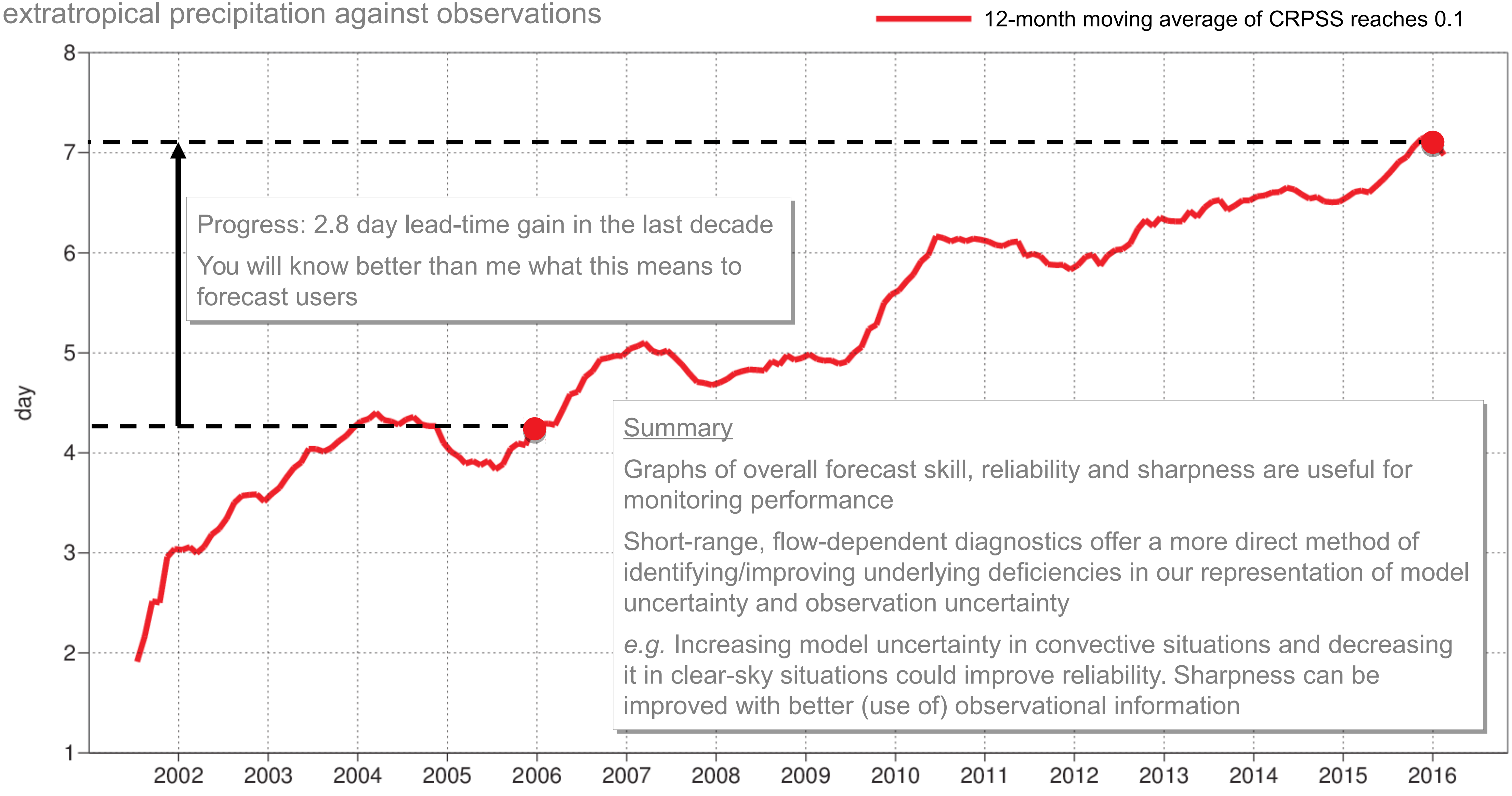
MCS – Jetstream interaction (composite)



- *Increments emphasize model systematic error: MCS does not interact enough with Jetstream*
- *Also need to strengthen stochastic physics to increase background variance?*

Trend in probabilistic forecast performance & Summary

CRPSS, extratropical precipitation against observations



Summary

Graphs of overall forecast skill, reliability and sharpness are useful for monitoring performance

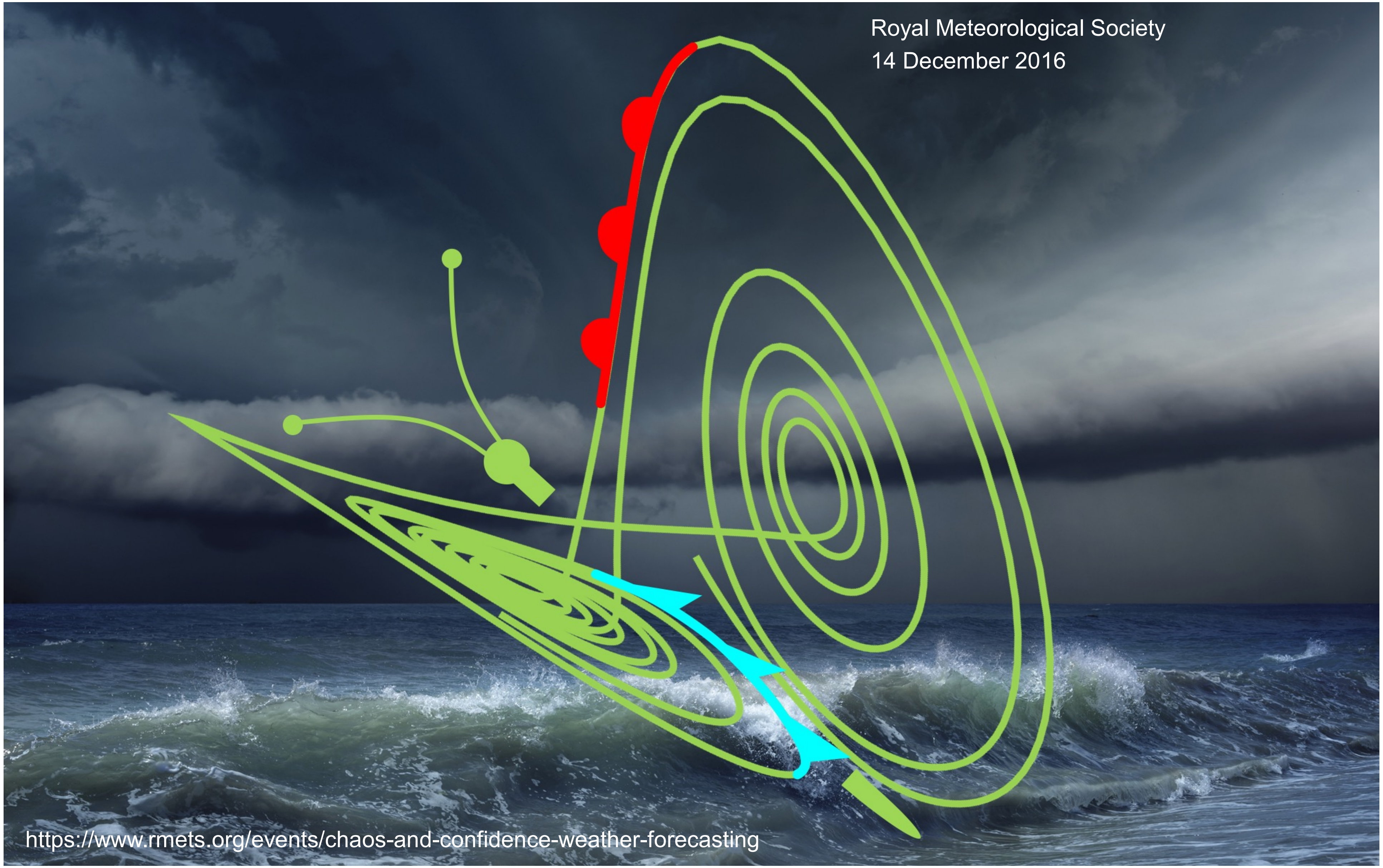
Short-range, flow-dependent diagnostics offer a more direct method of identifying/improving underlying deficiencies in our representation of model uncertainty and observation uncertainty

e.g. Increasing model uncertainty in convective situations and decreasing it in clear-sky situations could improve reliability. Sharpness can be improved with better (use of) observational information

Thank you

Extra slides

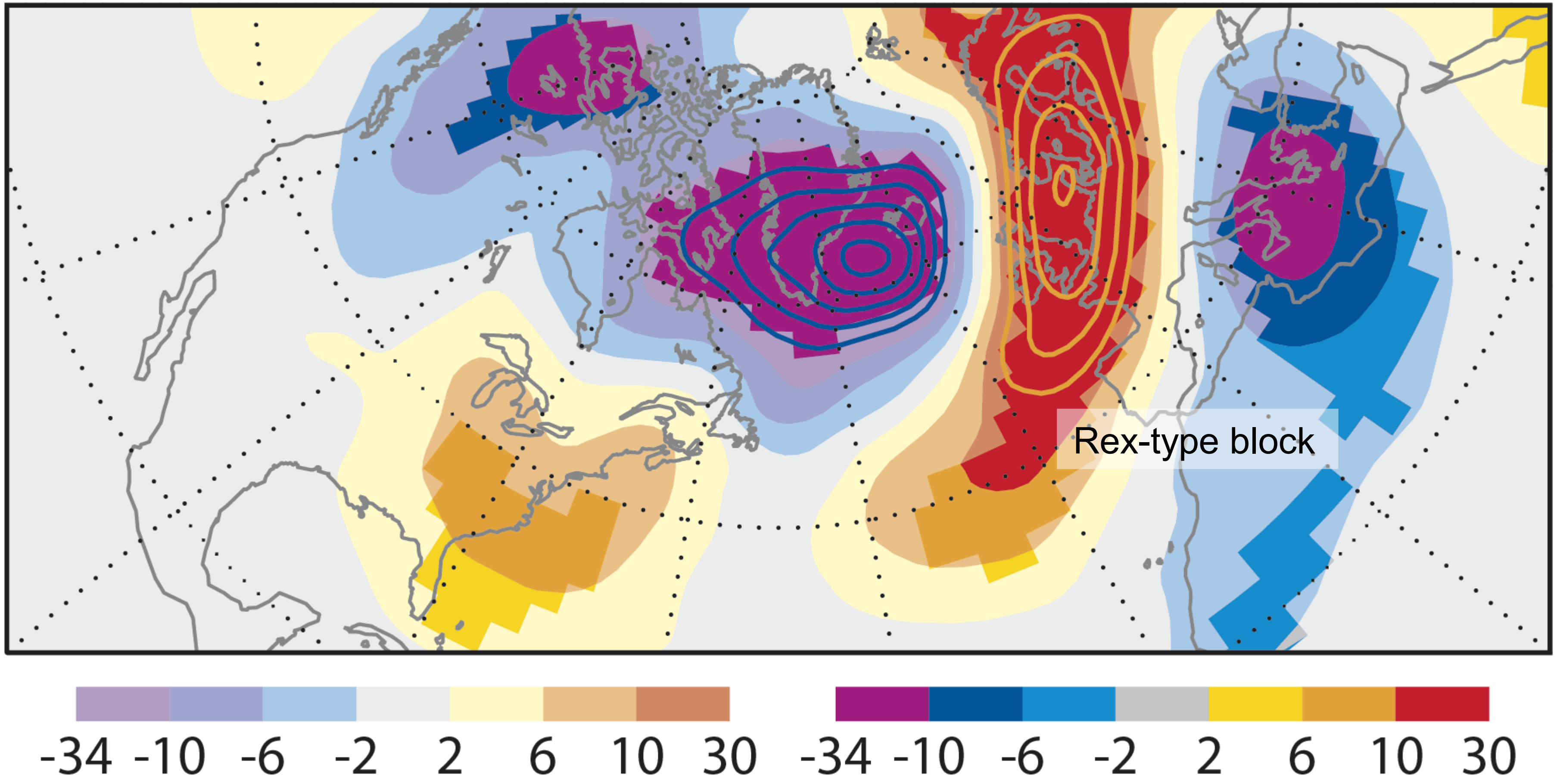
Chaos and Confidence in Weather Forecasting – further talks here



Verifying conditions composited over many bust forecasts

Rodwell et al, 2013, BAMS

500 hPa geopotential height (Z500) anomaly



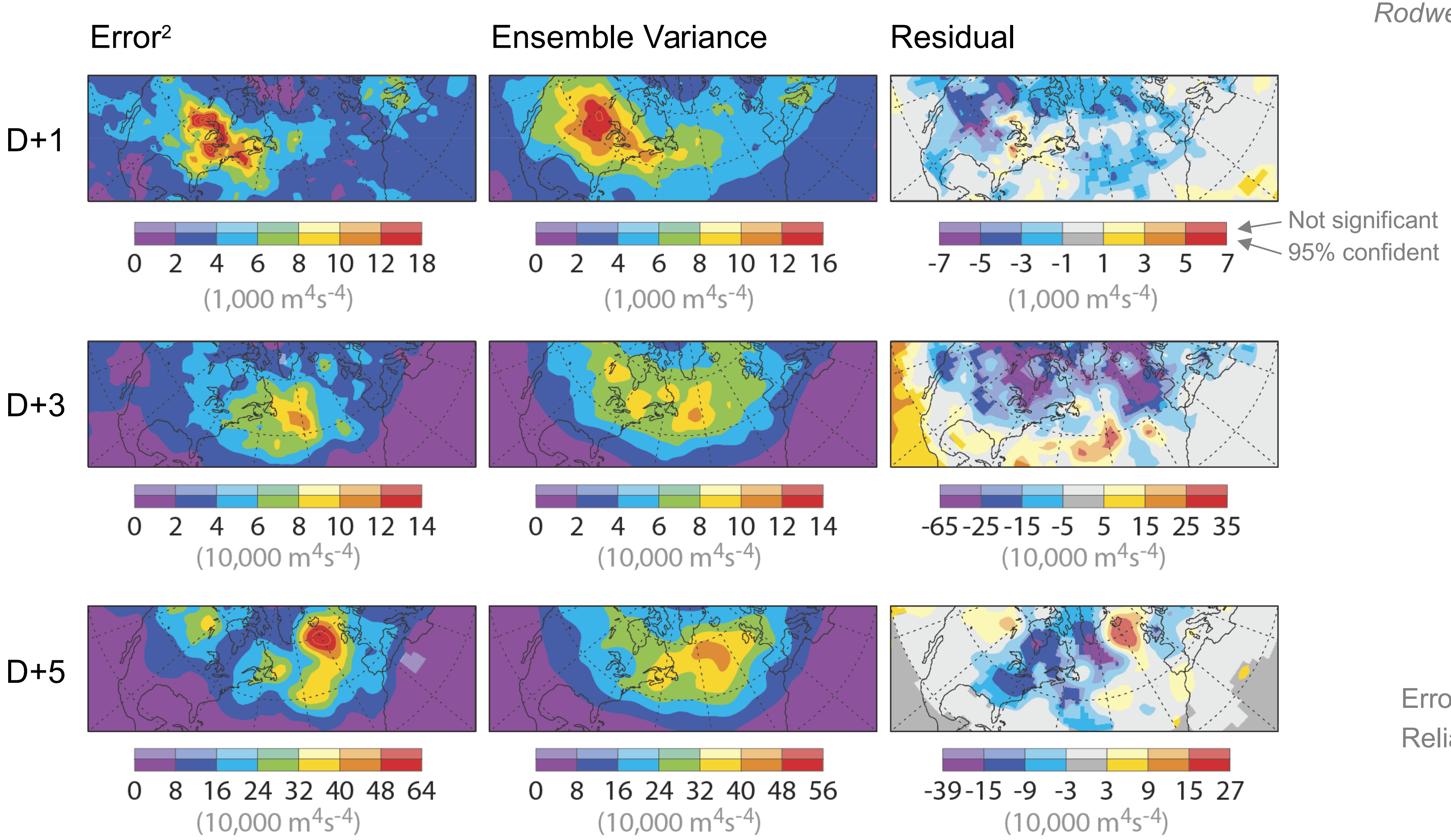
Unit = m   Bold colours = statistical significance at 5% level

Composite of 584 busts in ERA Interim forecast prior to 24 June 2010

Composite with North American trough & CAPE ($\Rightarrow$ Mesoscale convective systems)

Z200, 54 cases

Rodwell 2016, ECMWF Newsletter



$\text{Error}^2 = \text{EnsVar} + \text{Residual}$

Reliability $\Rightarrow \mathbb{E}[\text{Residual}]=0$

- *Following conditions conducive to MCS development, enhanced errors and spread propagate east towards Europe → 'Busts'*
- *Note: -ve residuals occur in non-trough/CAPE situation too.*
- *+ve residual at D+5 is not significant (Chaos? → use bigger sample or shorter leadtime? But analysis uncertainty at D+1?)*