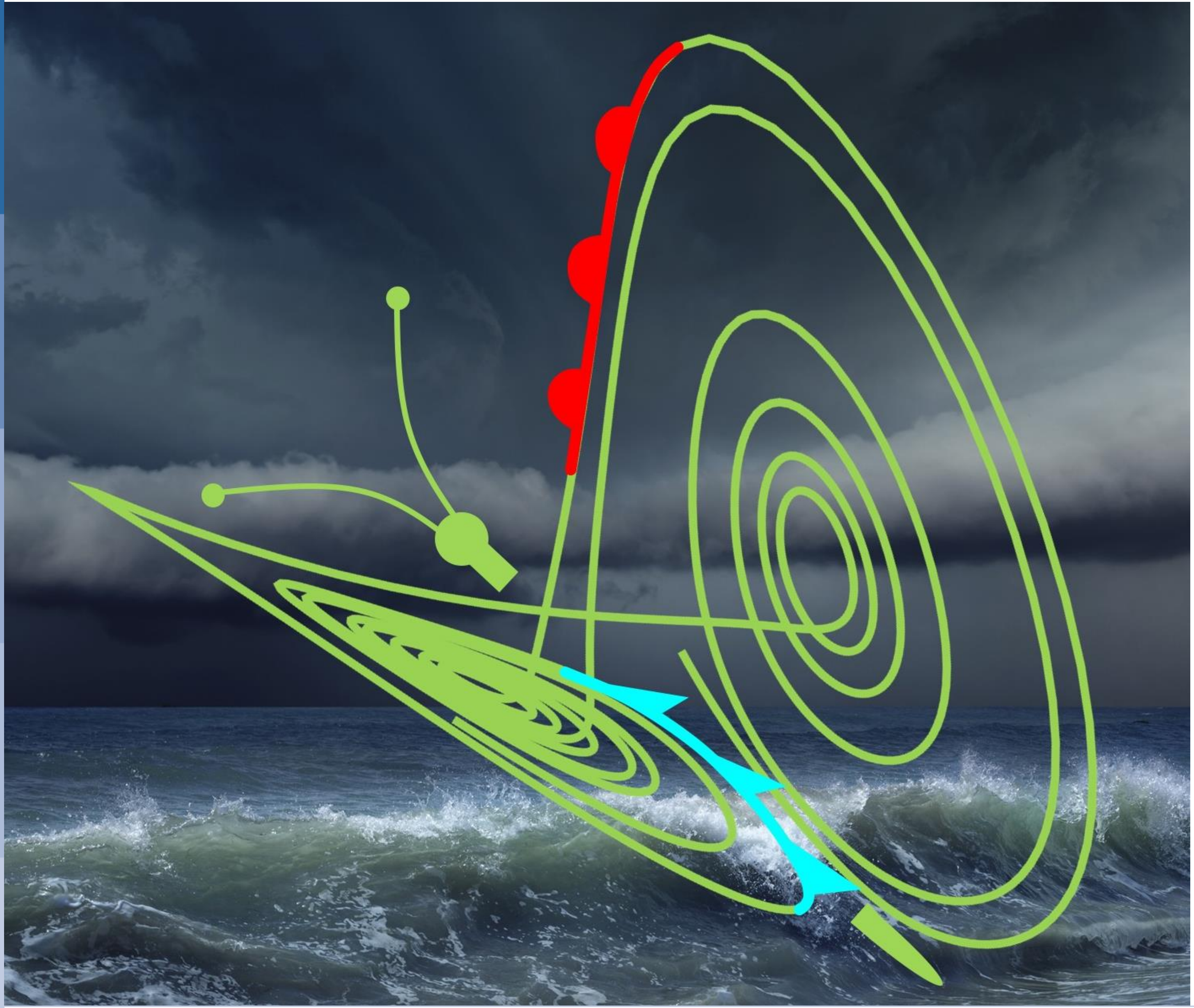


Diagnostics 2

Mark Rodwell

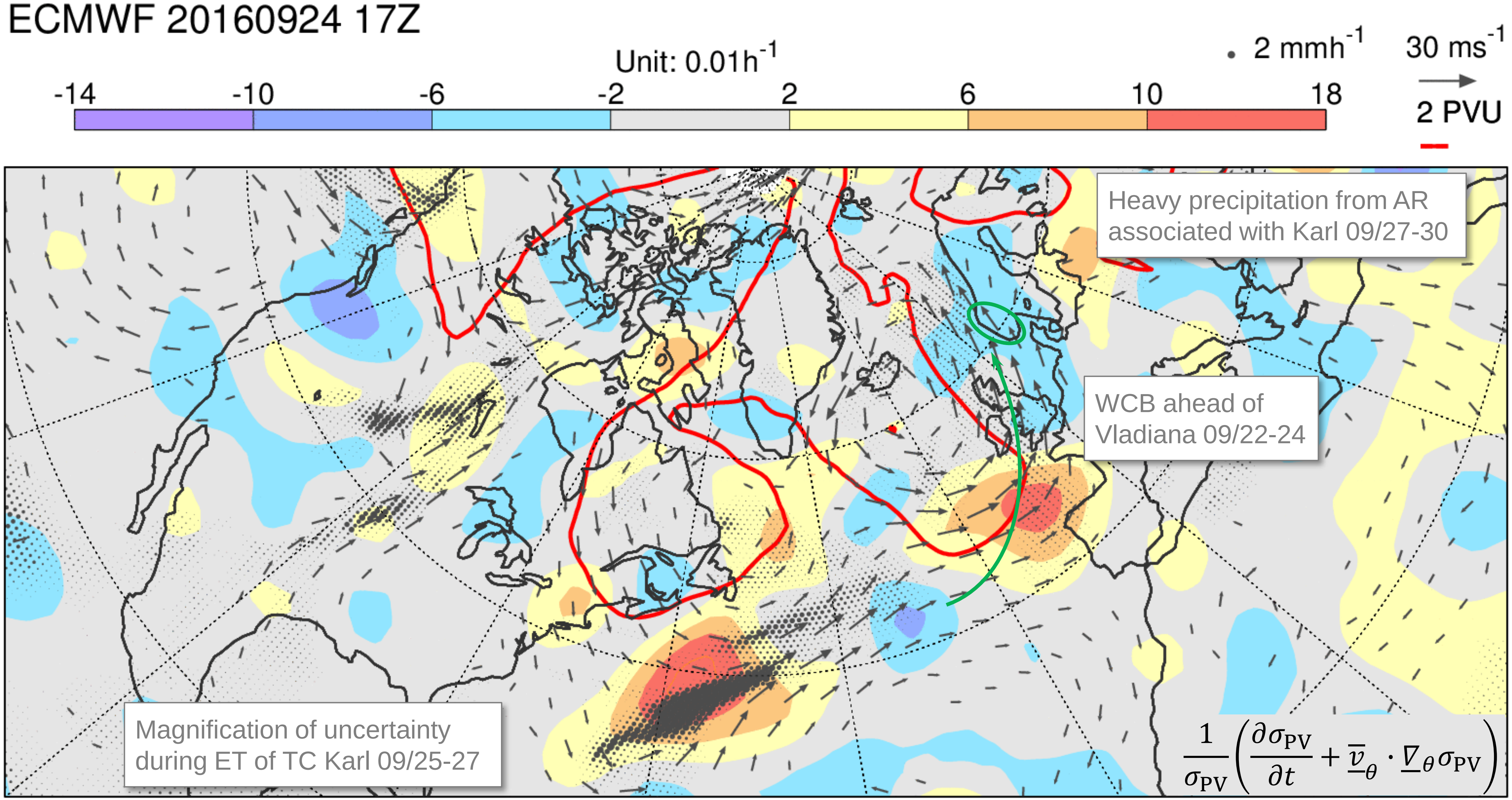
Collaborators
David Richardson, Dave Parsons,
Heini Wernli, Linus Magnusson, Elias Hólm

Training course: Predictability & ocean-atmosphere ensemble forecasting

1 March 2019, ECMWF

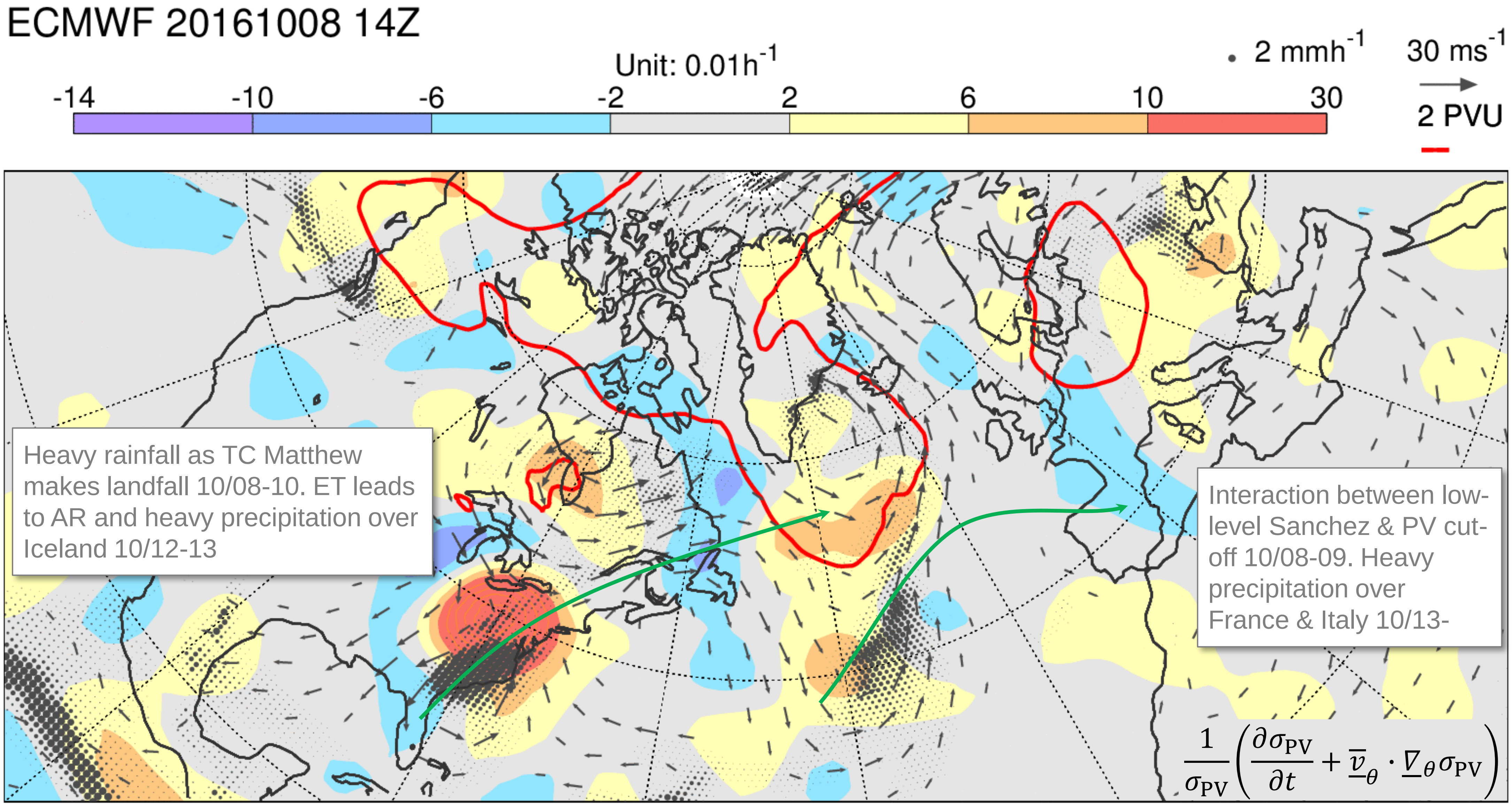
NWP continued

“Lagrangian” growth-rate for $\sigma_{PV_{315}}$: NAWDEX Vladiana & TC Karl



PV_{315}=2 & $\underline{v}_{850}$ from control forecast, precipitation is ensemble-mean. 1d running-mean gives 12h-integrated growth rate with any diurnal cycle removed. T21 smoothed

“Lagrangian” growth-rate for $\sigma_{PV_{315}}$: NAWDEX Sanchez



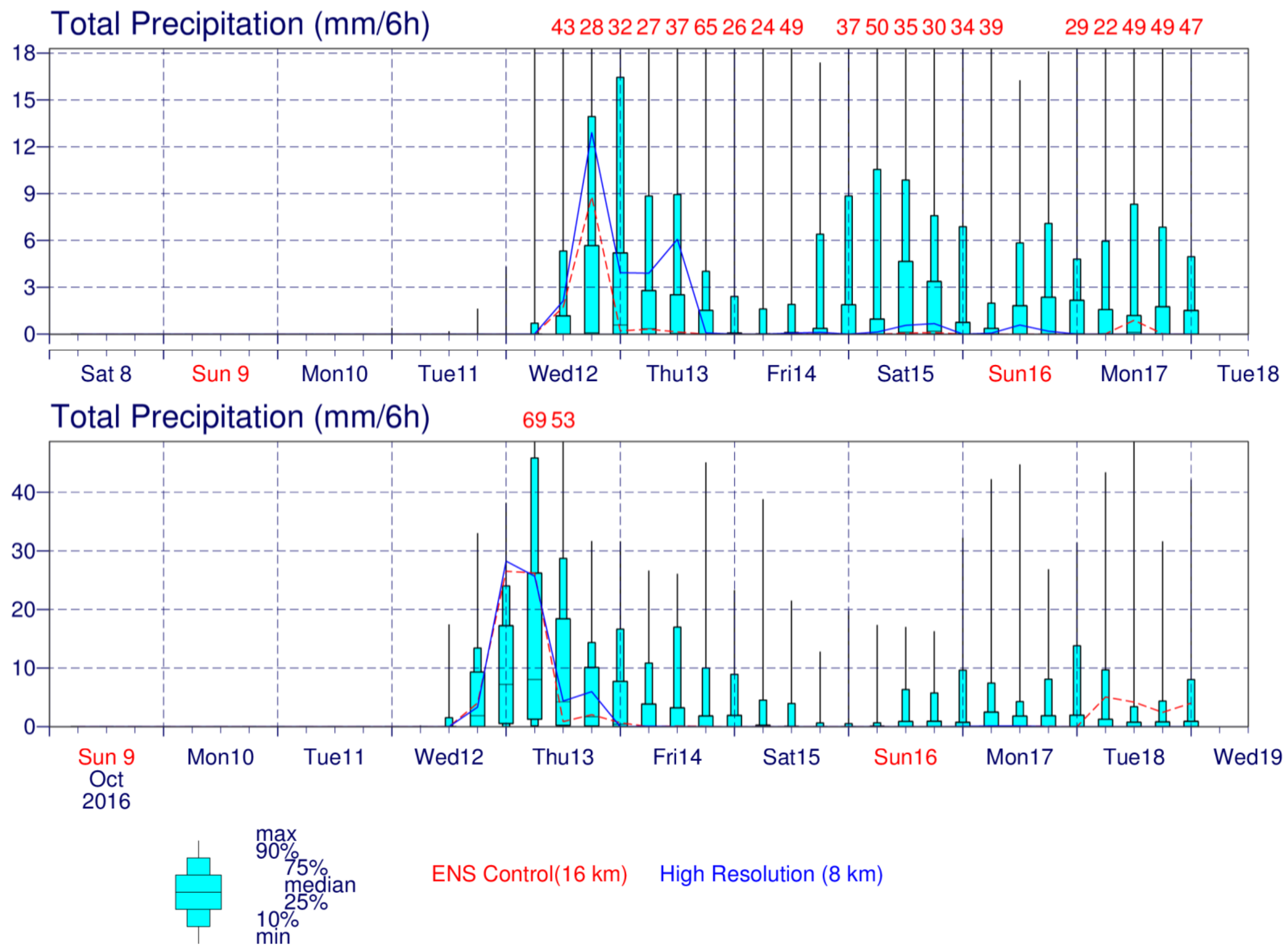
$$\frac{1}{\sigma_{\mathrm{PV}}}\left(\frac{\partial\sigma_{\mathrm{PV}}}{\partial t}+\underline{v}_\theta\cdot\underline{\nabla}_\theta\sigma_{\mathrm{PV}}\right)$$

PV_{315}=2 & $\underline{v}_{850}$ from control forecast, precipitation is ensemble-mean. 1d running-mean gives 12h-integrated growth rate with any diurnal cycle removed. T21 smoothed

Forecast for precipitation in Montpellier, southern France

FC start 10/08

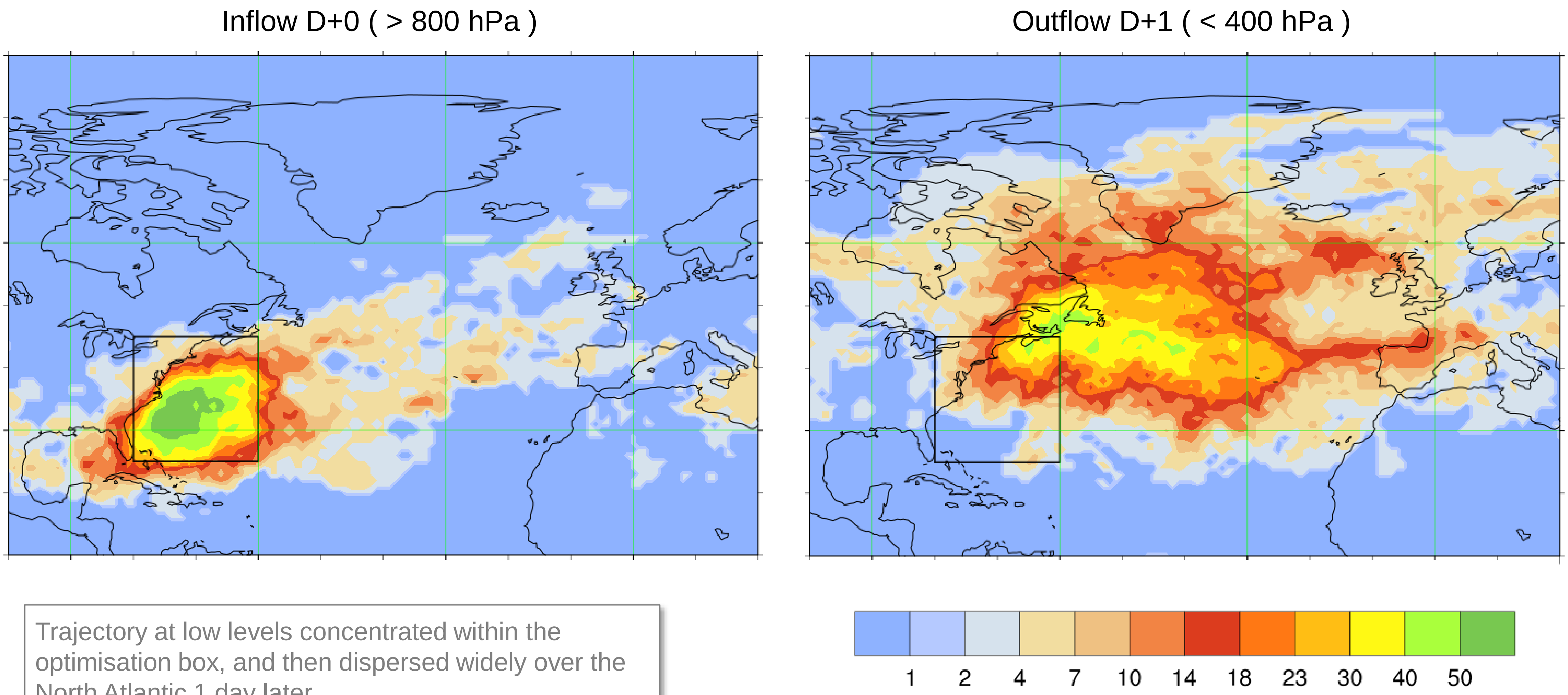
FC start 10/09



Once uncertainties associated with the interaction between Sanchez and the upper-level PV cut-off are resolved, the probability for strong precipitation over southern France firms-up. Note the different y-axes.

Top 50 Warm Conveyor Belt inflow events in box indicated from Nov 15 – Oct 16

From Heini Werni

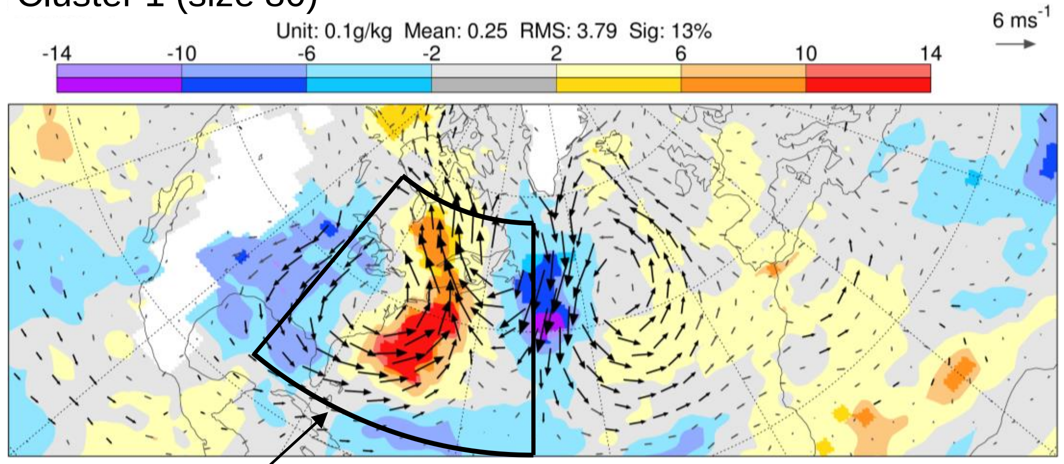


Trajectory at low levels concentrated within the optimisation box, and then dispersed widely over the North Atlantic 1 day later

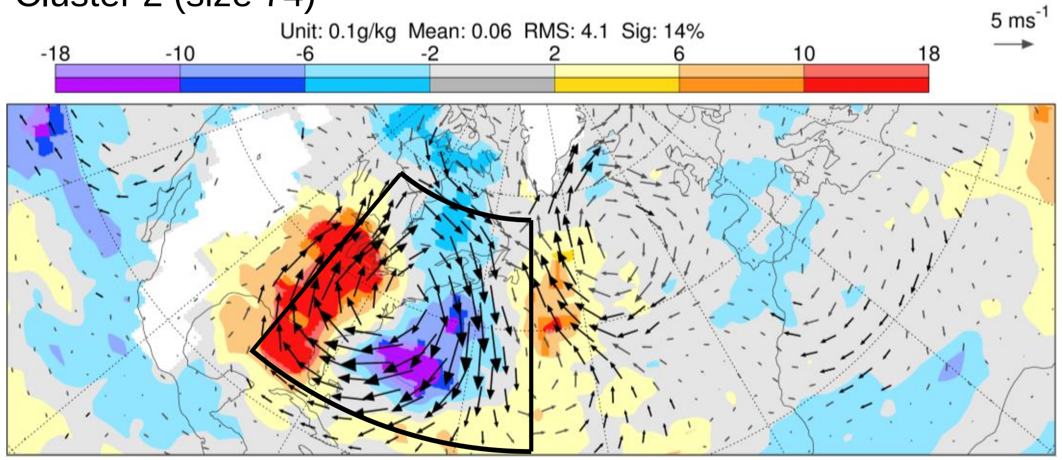
Based on trajectories ascending by more than 600 hPa in 2d

Clustering to identify 3 synoptic flow patterns. Mean anomalies in q850 and v850

Cluster 1 (size 80)

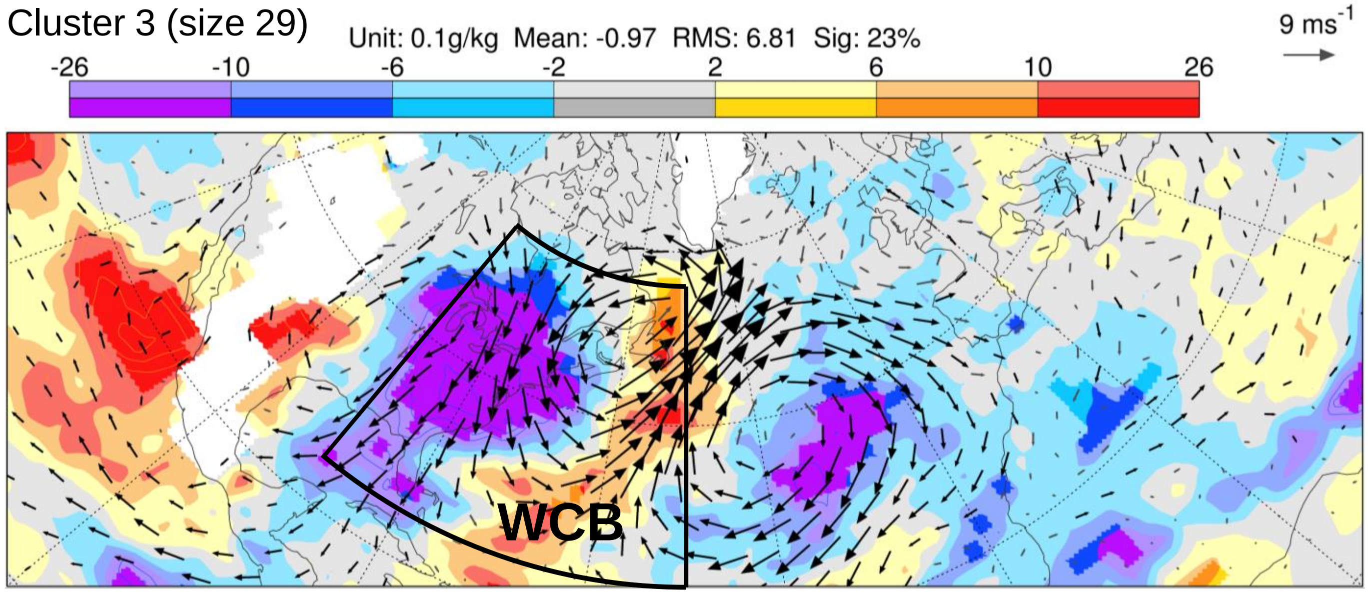


Cluster 2 (size 74)



Clustering region

Cluster 3 (size 29)



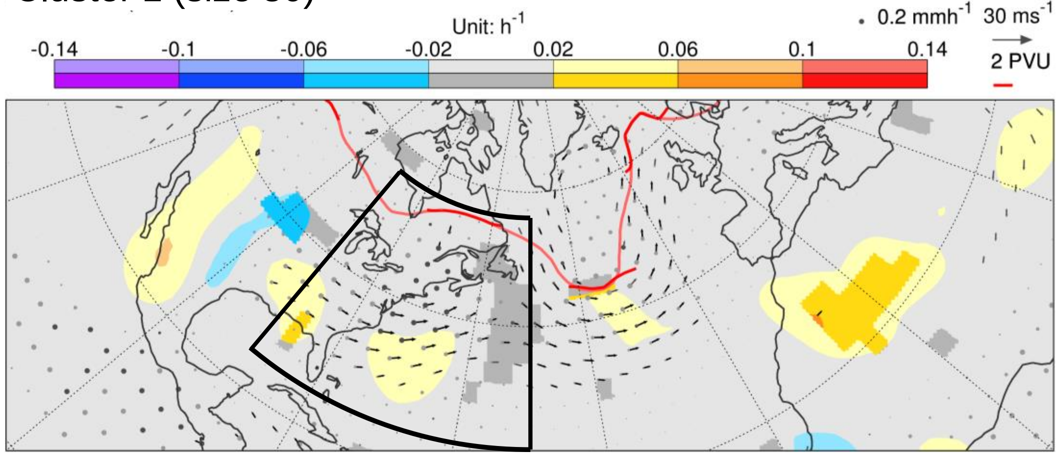
Here, the main focus is on the "Warm Conveyor Belt" cluster, with enhanced uncertainty growth-rates

Note everything is for the single season (MAM 2017) – important for operational development cycle

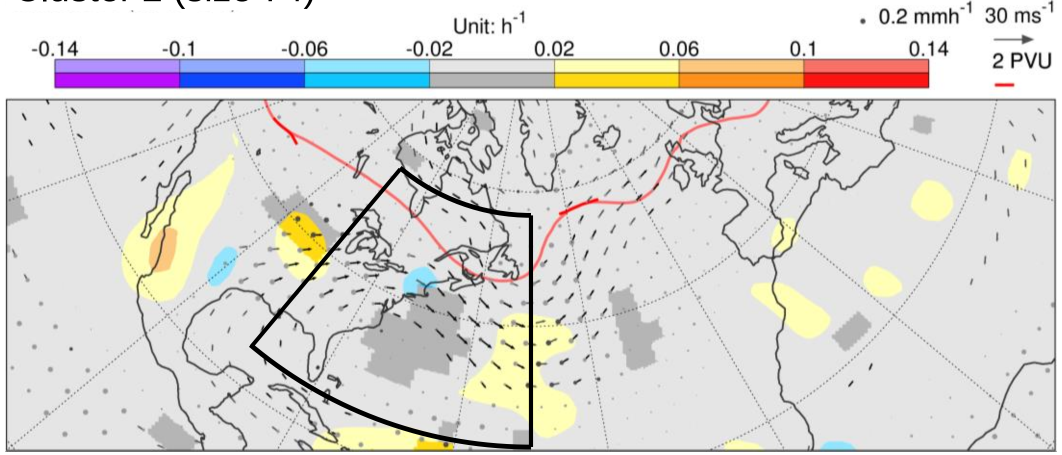
Based on K-means clustering of the synoptically-filtered and normalised fields from the PV animation in the region indicated, for MAM 2017. Bold colors = 5% significance

Cluster-means based on / showing PV_{315}, $\underline{v}_{850}$, precipitation & PV_{315} growth-rate

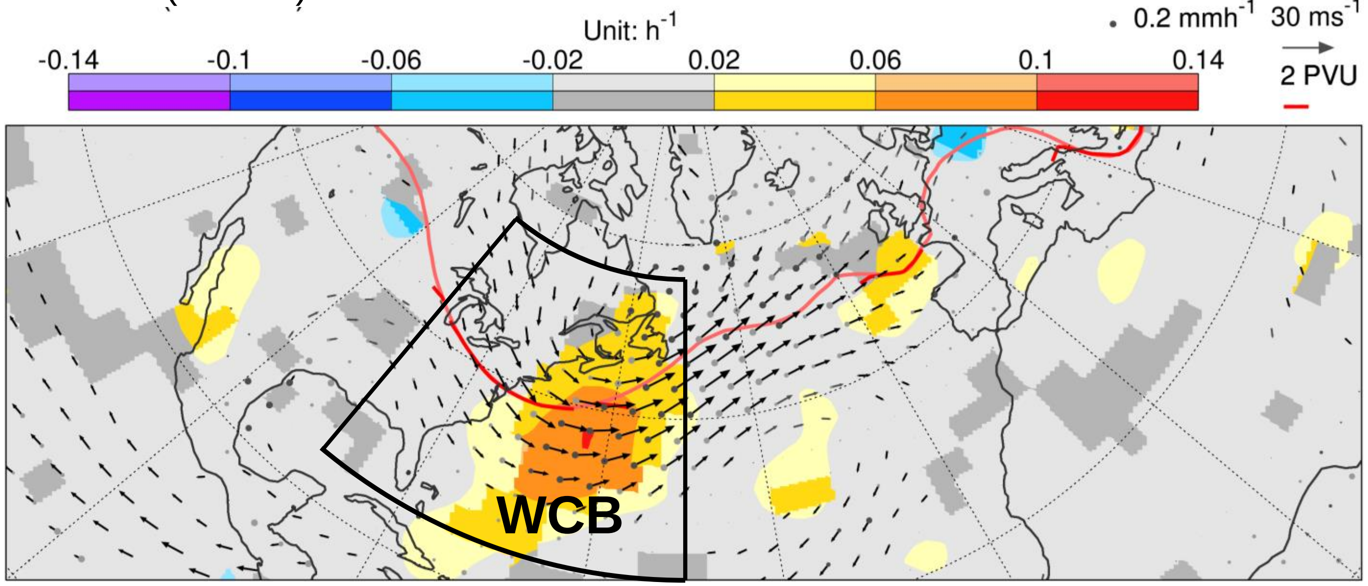
Cluster 1 (size 80)



Cluster 2 (size 74)



Cluster 3 (size 29)

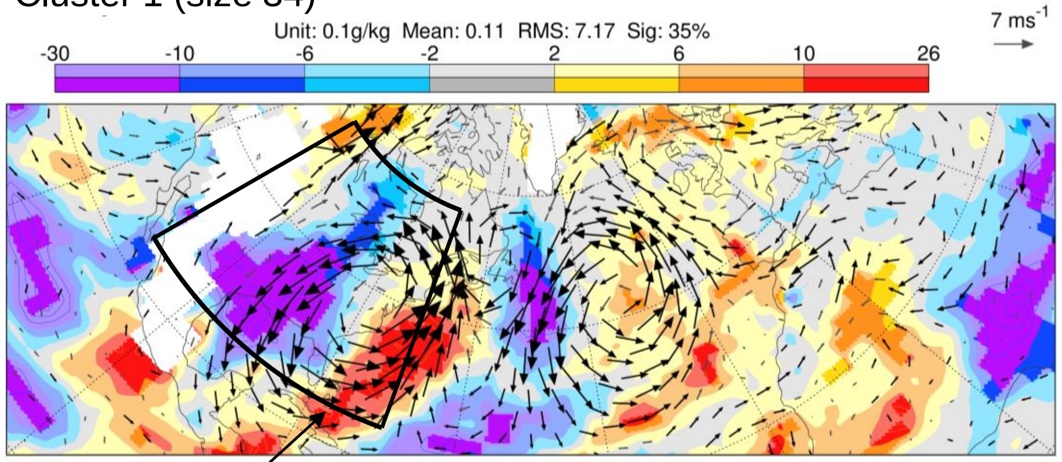


Here, the main focus is on the "Warm Conveyor Belt" cluster, with enhanced uncertainty growth-rates

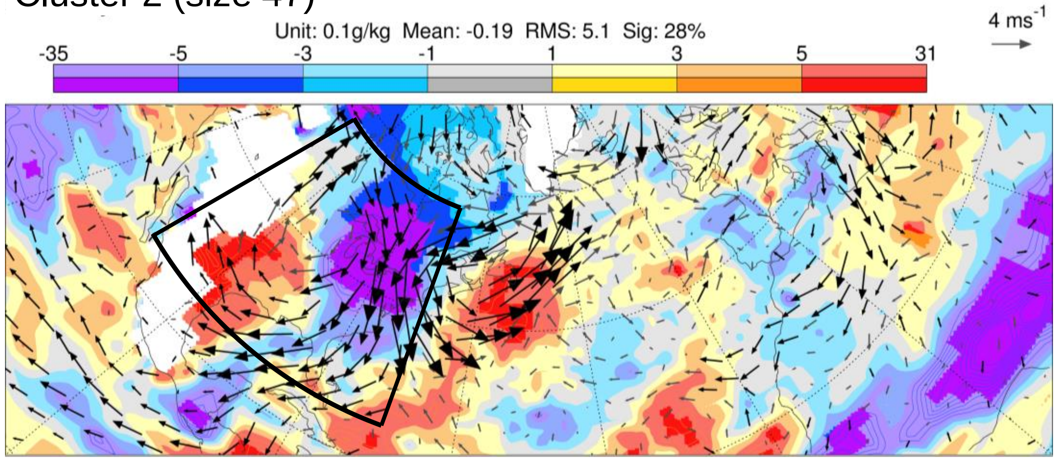
Based on K-means clustering of the synoptically-filtered and normalised fields from the PV animation in the region indicated, for MAM 2017. Bold colors = 5% significance

Clustering to identify 3 synoptic flow patterns. Mean anomalies in q850 and v850

Cluster 1 (size 34)

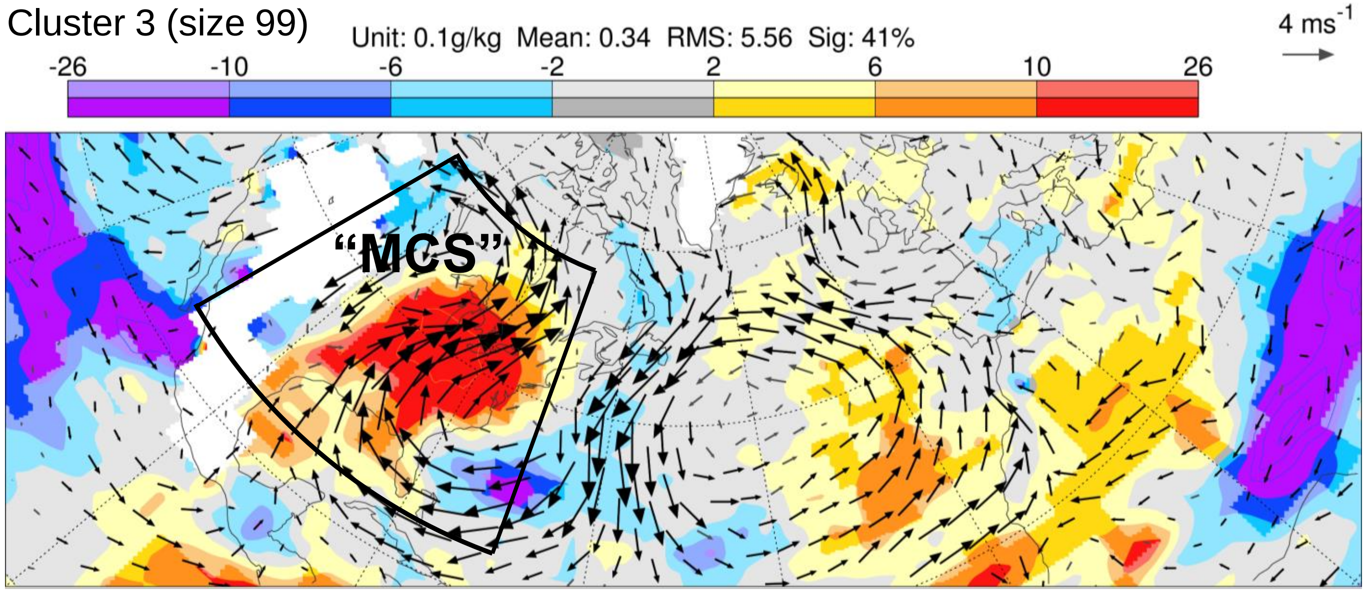


Cluster 2 (size 47)



Clustering region

Cluster 3 (size 99)

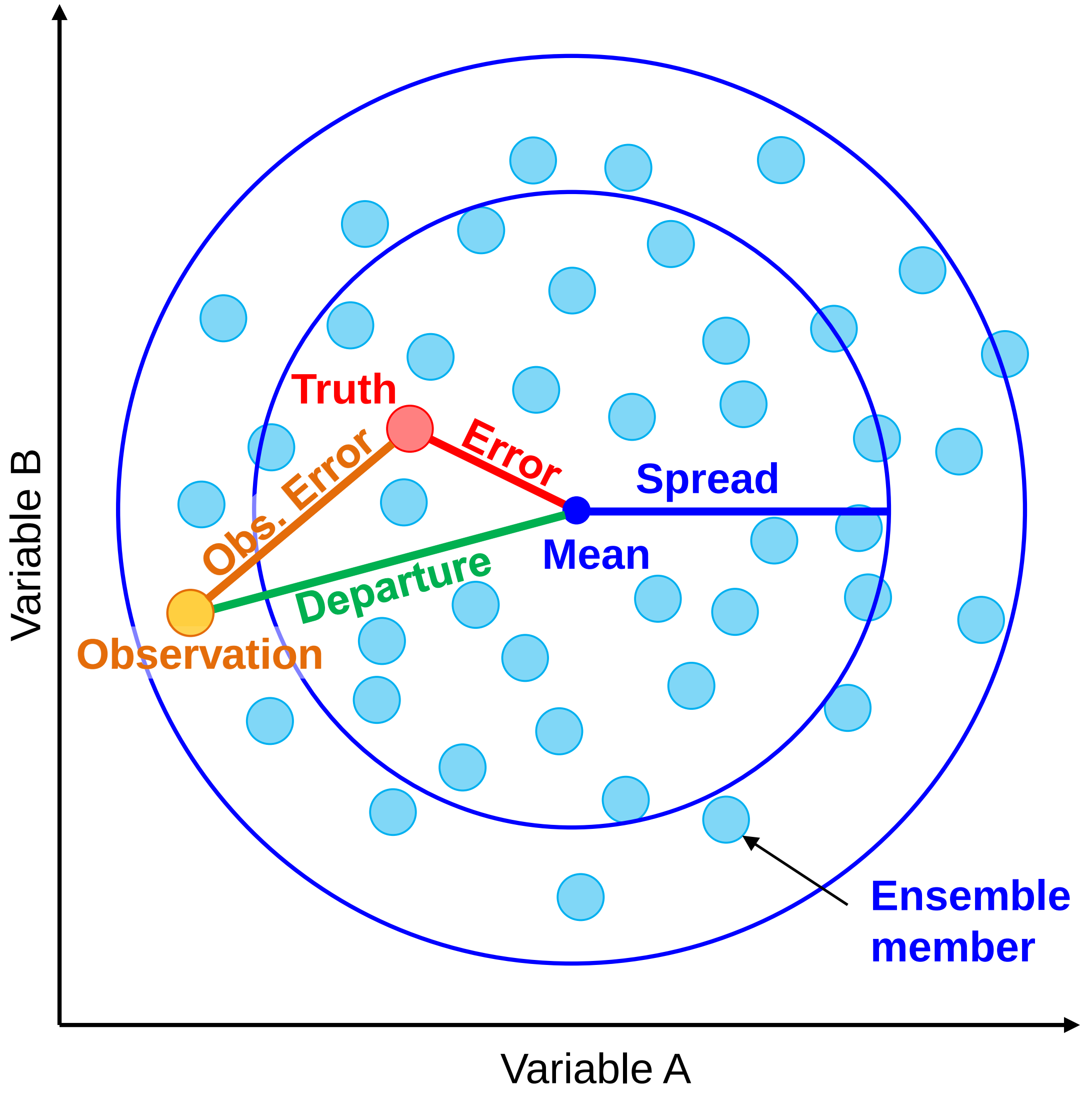


Here, the main focus is on "Mesoscale Convective System" cluster, which is the most highly populated

Based on K-means clustering of the synoptically-filtered and normalised fields from the PV animation in the region indicated, for MAM 2017. Bold colors = 5% significance

Reliability in ensemble data assimilation

After Rodwell et al, 2015, QJRMS



Reliability ⇒

$$\overline{\mathbf{Error^2}} = \overline{\mathbf{Spread^2}} \quad (\equiv \overline{\mathbf{EnsVar}})$$

(averaged over many forecast start dates)

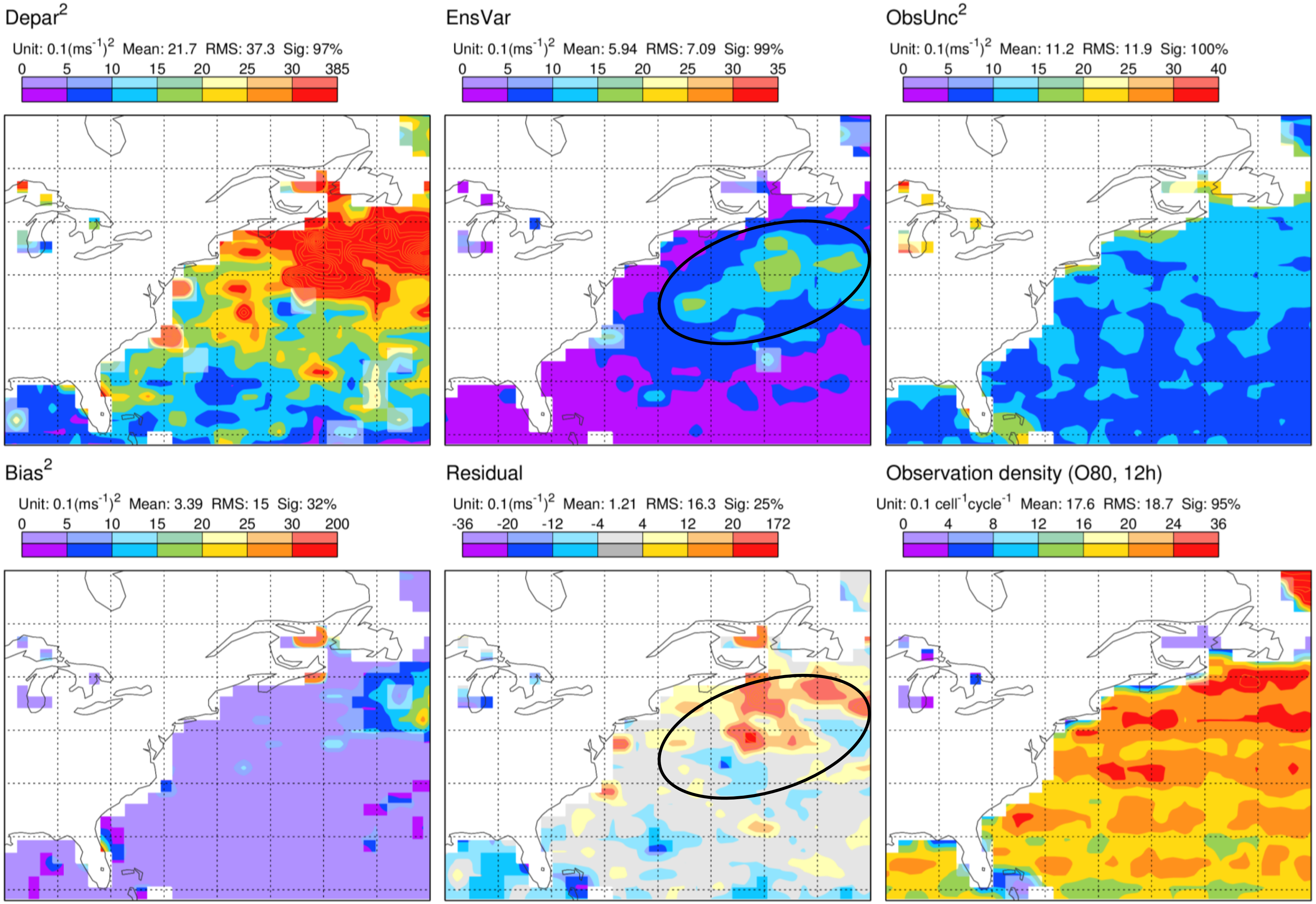
If we do not know the truth well-enough to calculate the error, use‡

$$\overline{\mathbf{Departure^2}} = \overline{\mathbf{EnsVar}} + \overline{\mathbf{Obs.Unc^2}}$$

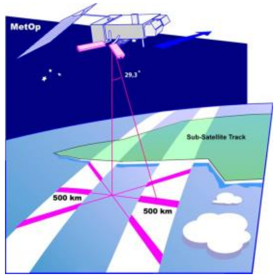
Any imbalance in this equation indicates that the (initialization of) the ensemble forecast is unreliable

‡Assuming the observation error is uncorrelated with the error of the ensemble-mean

EDA error growth evaluation for WCB cluster using scatterometer surface winds



MAM 2017



Large departures for this WCB case

Observation error variance is over-estimated for this data (not shown)

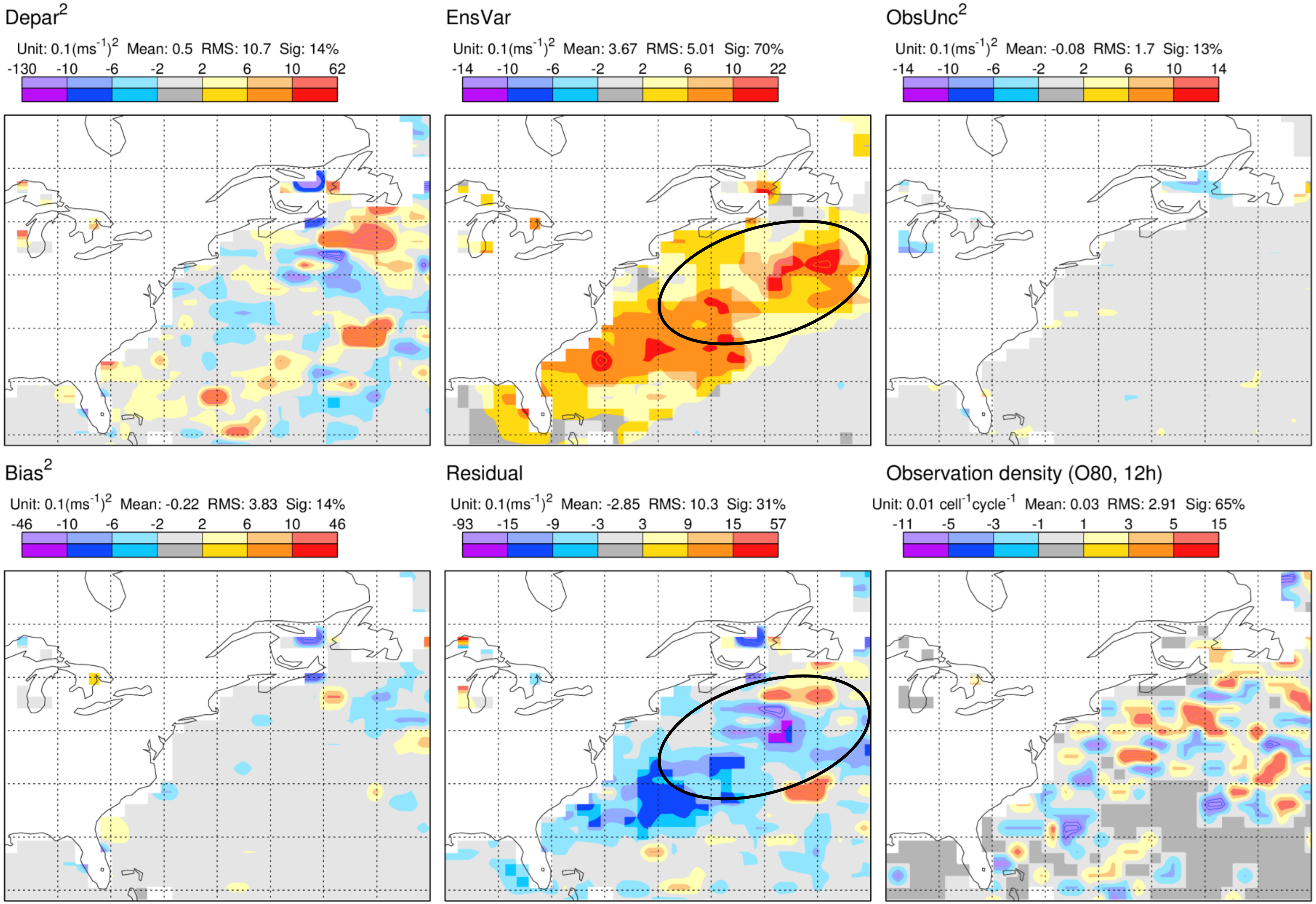
Hence EnsVar is under-represented

What happens if we increase stochastic physics?

Cluster	Size	Residual $(ms^{-1})^2$	(Un)reliability (%)
1	80	0.65	35
2	74	0.67	33
3	29	1.63	32

RMS of residual over the clustering region shows that WCB cluster contributes 1/3 to overall unreliability

Change to EDA variance budget when stochastic physics included in boundary-layer



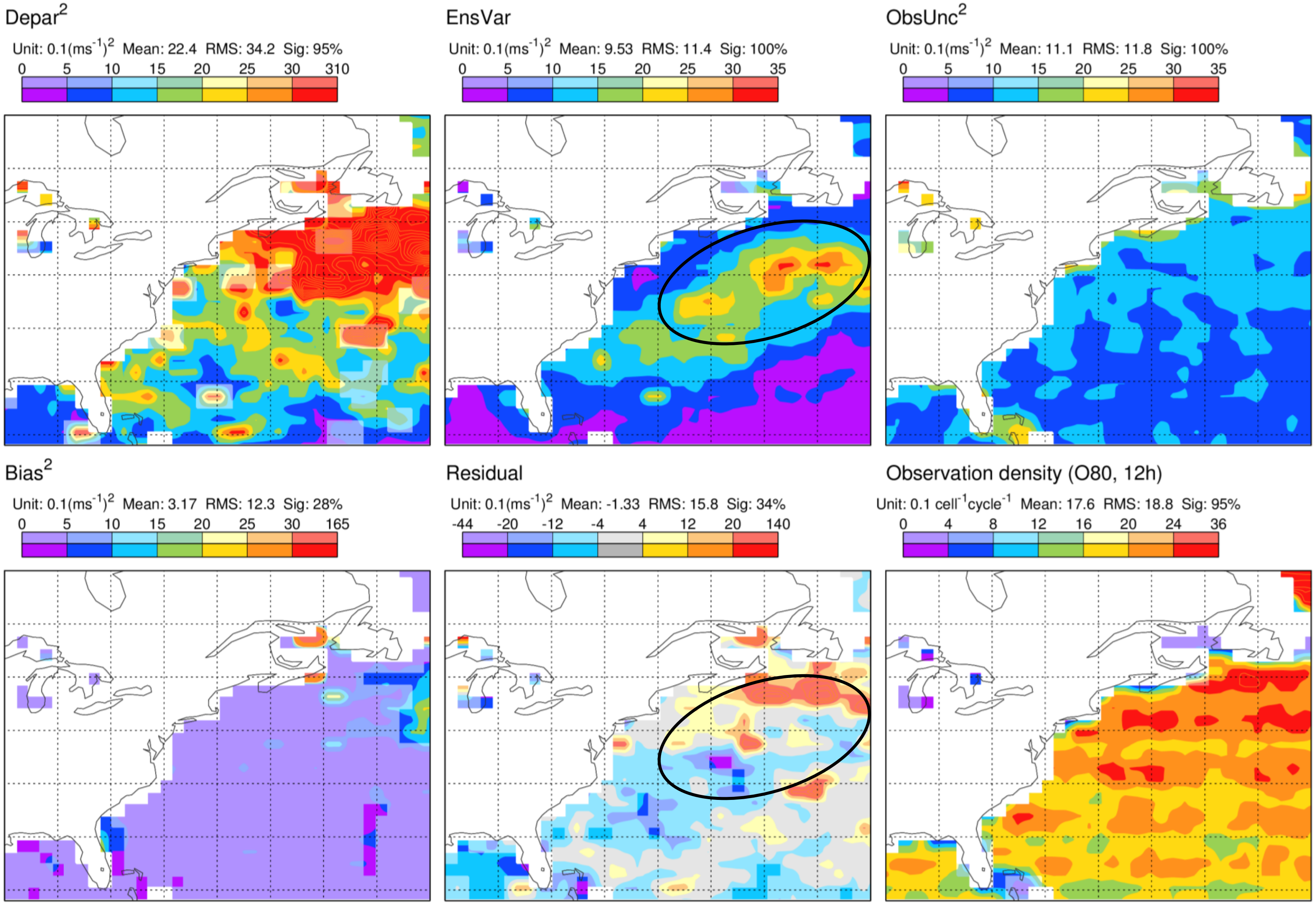
MAM 2017

Model cycle 45r1 – cycle 43r1

Removing the boundary-layer tapering of stochastic physics increases background variance (EnsVar) by 60%

Residual is made more negative (an improvement – see next slide)

Ensemble Data Assimilation variance budget based on scatterometer winds (45R1)



MAM 2017

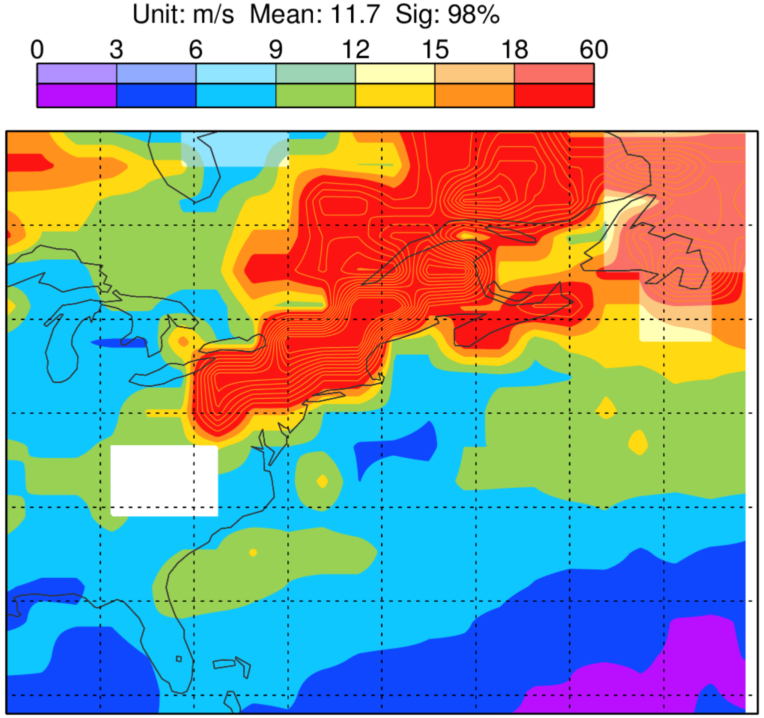
5% improved residual. Mixed signals for squared departures (mean 3% increase, RMS 8% decrease)

Future tuning (down) of observation error variance (ObsUnc^2) should act to remove the blue from the Residual

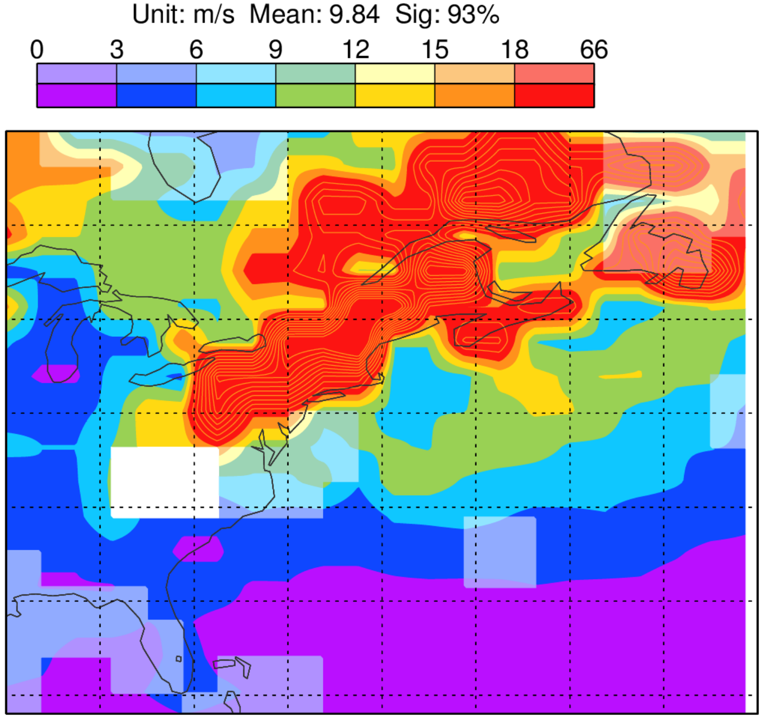
Potential to reduce departures and ensemble variance (in subsequent cycle) …

Impact of processes on u_{925} for Cluster 2. (RMS of 12h accumulated tendencies)

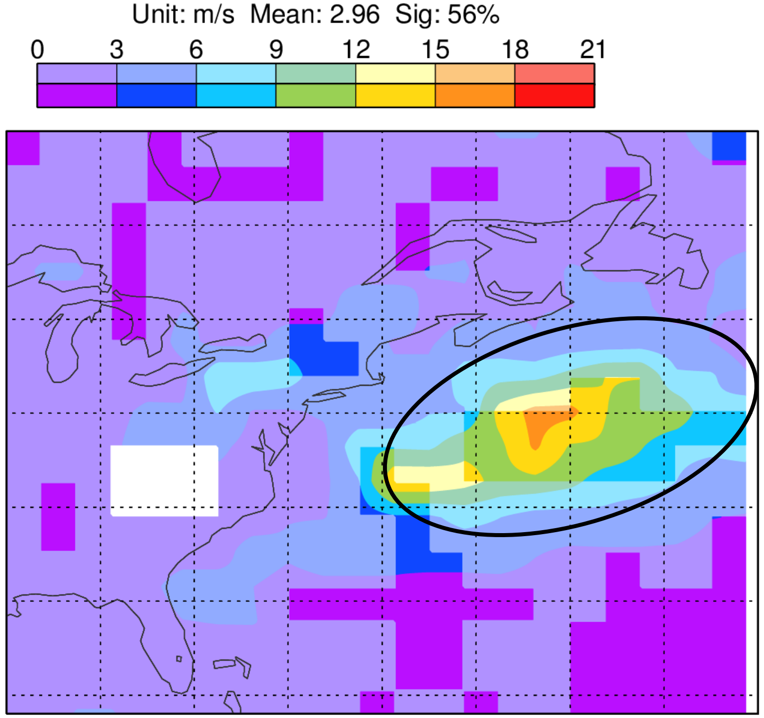
Dynamic tendency



Vert. diffusion (linked to surface drag)



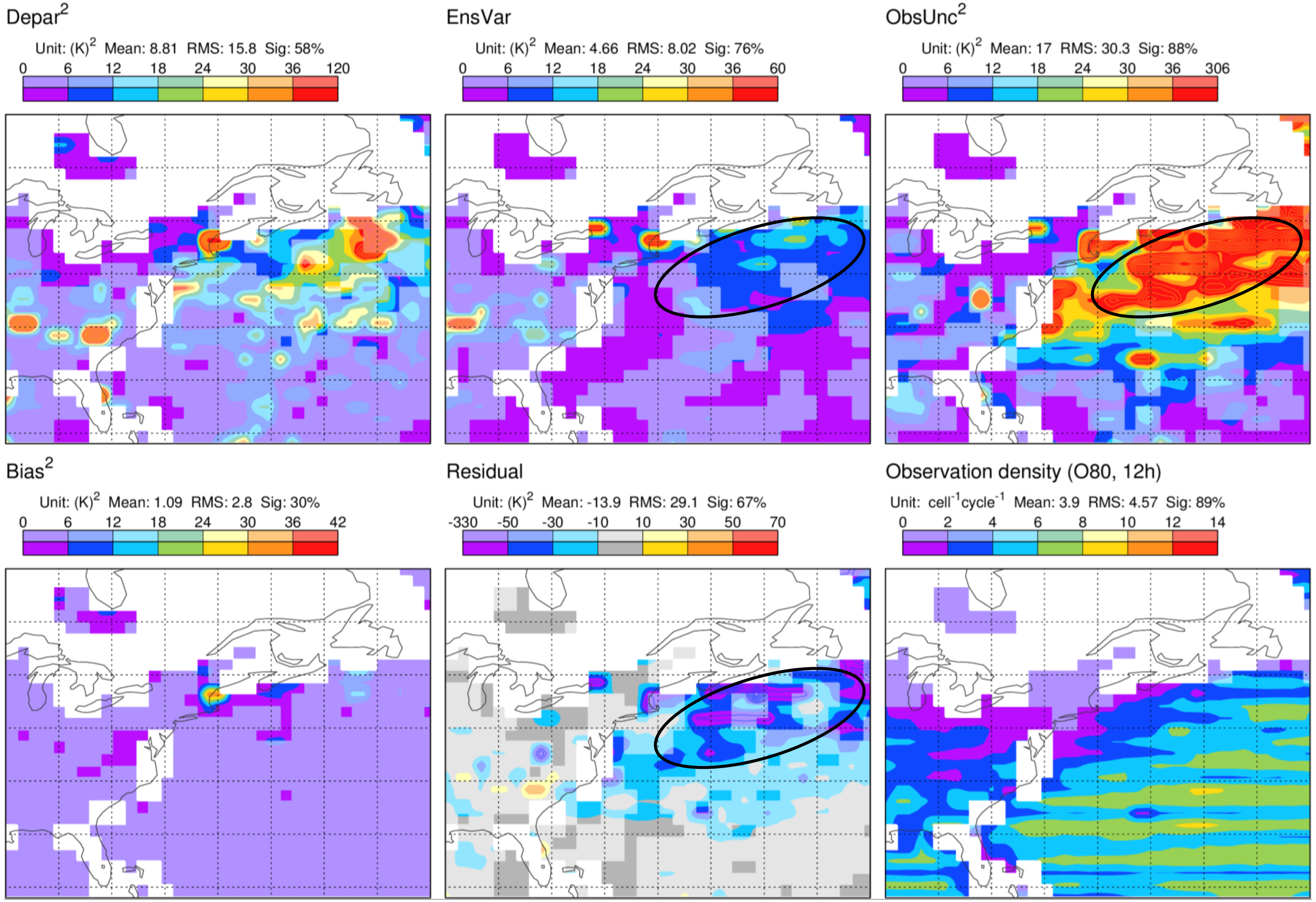
Convective momentum transport



Comparison shows that convective activity is collocated with the EDA budget residual. Suggests increased emphasis of stochastic physics on convection would also be beneficial

MAM 2017

Flow-dependent EDA error growth-rate evaluation using MHS All-Sky channel 5



Warm Conveyor Belt cluster
Sensitive to H_2O 750-400 hPa

MAM 2017

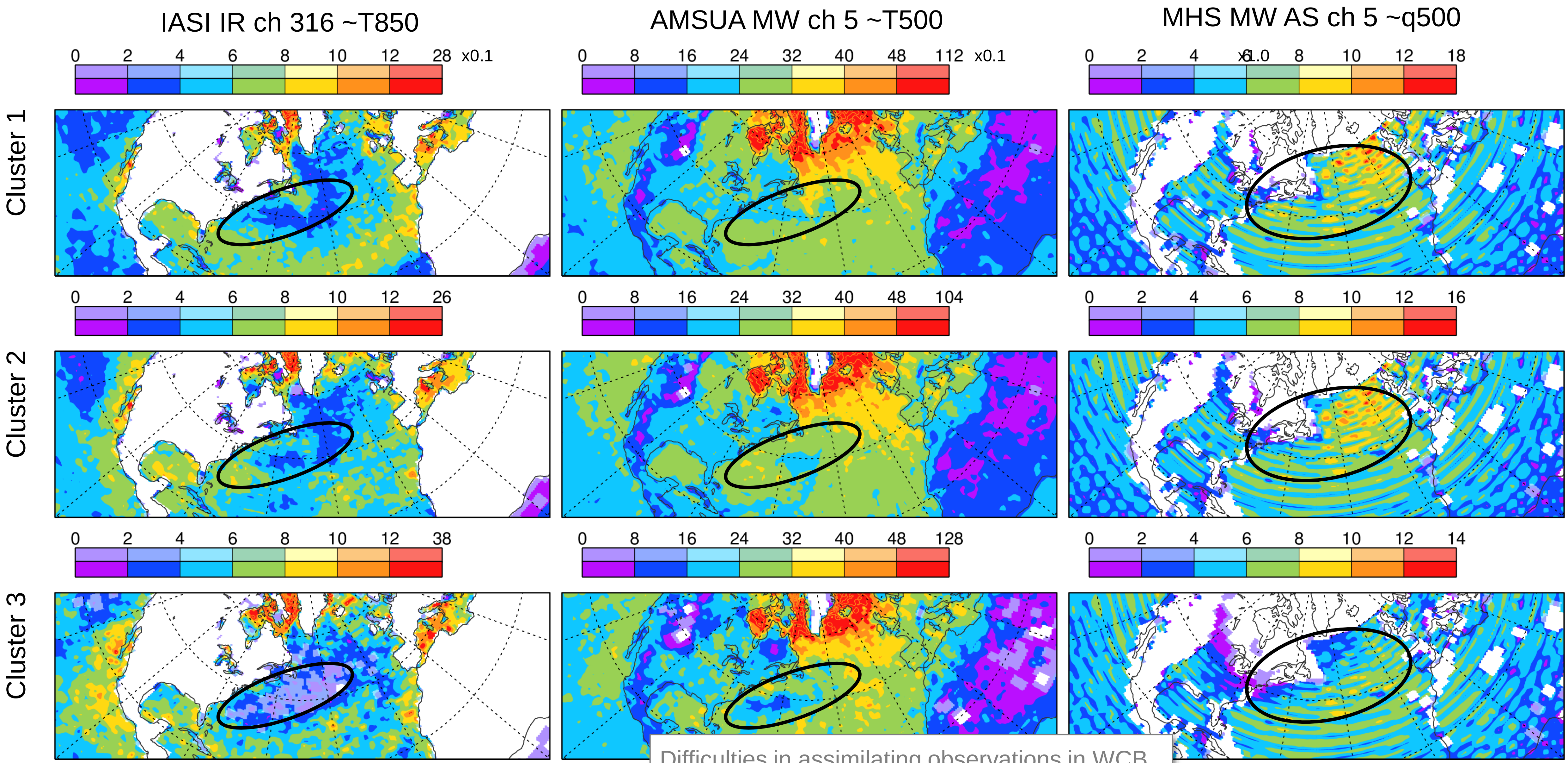
Budget suggests these observations are currently down-weighted — no doubt for good reason*. While their impact on analyses could be beneficial relative to an OSE, there is a potential to extract more information from them — thereby sharpening the EDA's distribution of analyses.

*Discussion of these reasons could be useful. What role does the model play here? Can observations alter boundary layer height, existence of cloud?

Bias is relatively small. Little physics and observational bias or a consequence of Var-BC?

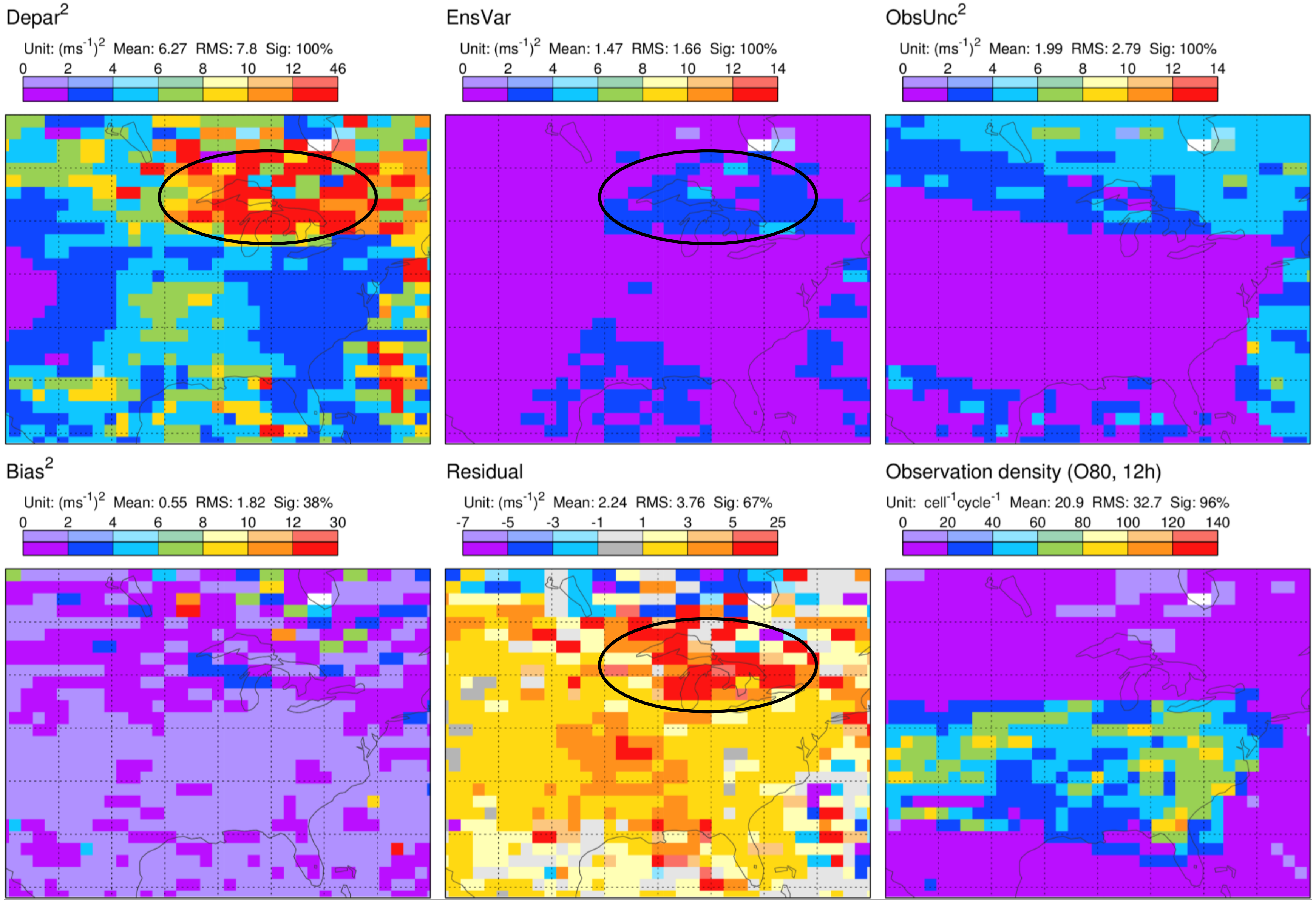
Density of observations assimilated for each cluster MAM 2017

unit=cell^{-1}cycle^{-1}≈(125km)$^{-2}$(12h)$^{-1}$



Difficulties in assimilating observations in WCB cluster just when we need them most!

EDA error growth evaluation for MCS cluster using aircraft observations of U_{200}



K-means clustering
Meso-Scale Convective cluster

MAM 2017

Large departures for this MCS case

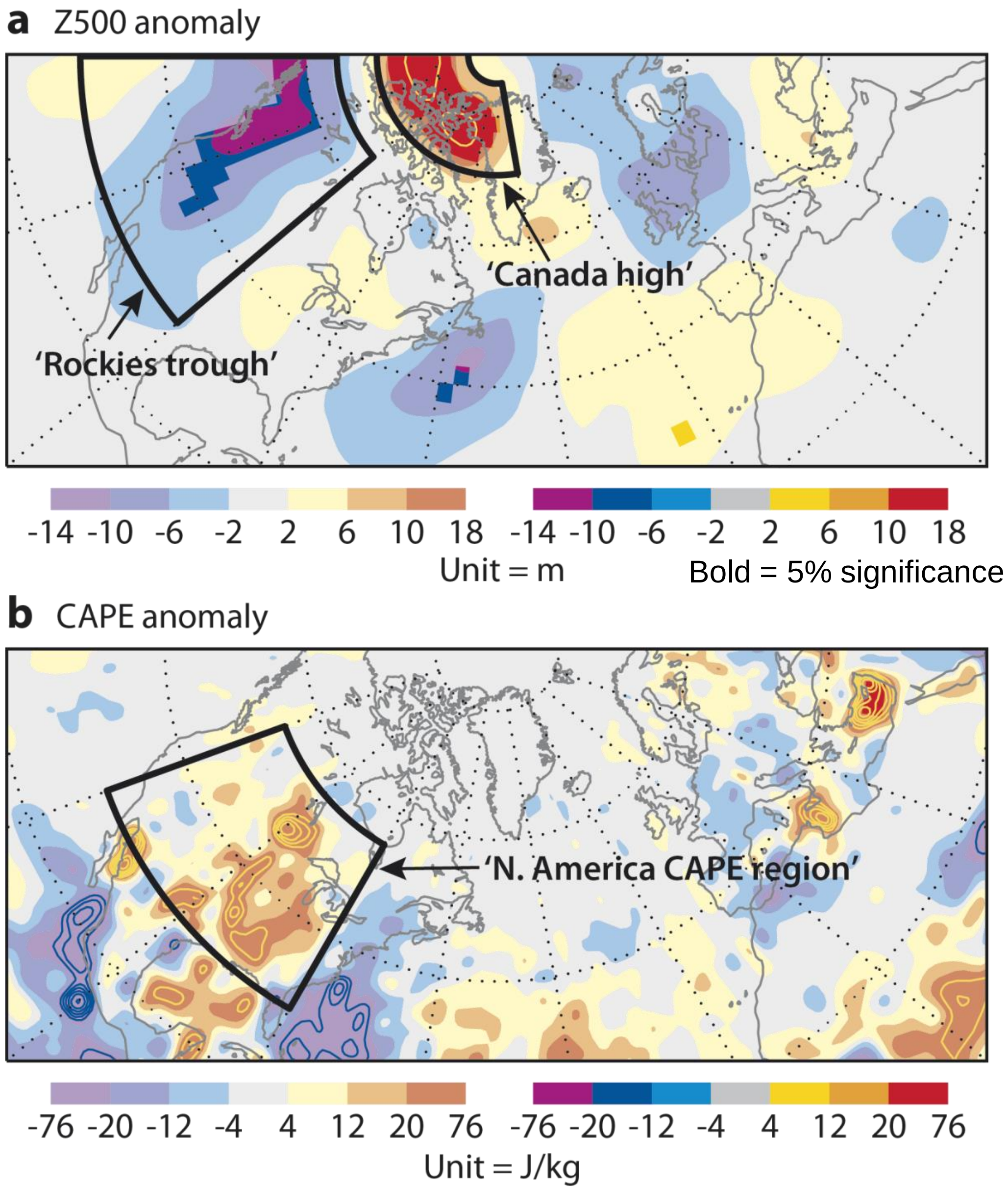
Likely that EnsVar is under-represented

Cluster	Size	Residual $(ms^{-1})^2$	(Un)reliability (%)
1	34	3.35	19
2	50	2.60	21
3	99	3.76	60

RMS of residual over the clustering region shows that MCS cluster contributes 60% to overall unreliability. Clearer argument to focus efforts.

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

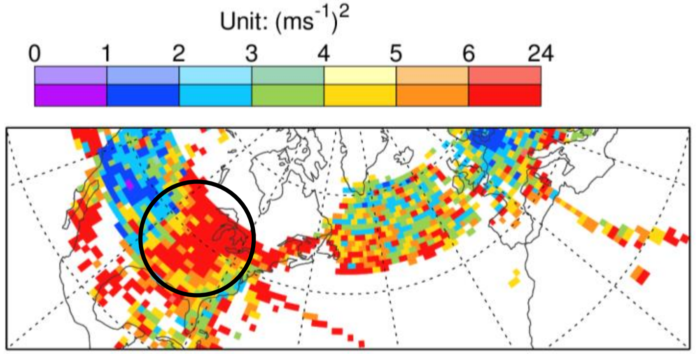
EDA reliability in u200 against aircraft observations in “Rocky trough/CAPE” situations

Departure^2 = Background variance + Estimated observation error variance (+ Bias^2 + Residual)

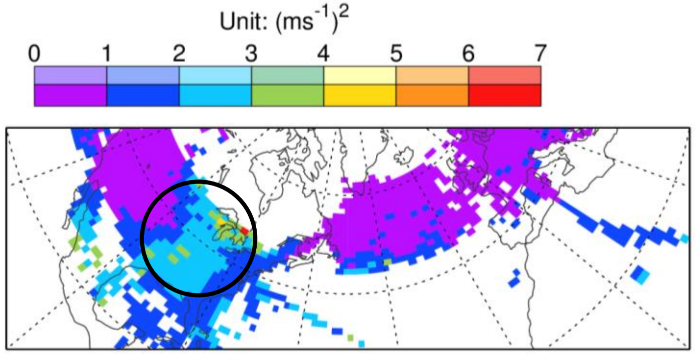
(Basically the error-spread relationship taking into account uncertainty in our knowledge of the truth)

Rodwell, Richardson, Parsons & Wernli. 2018, BAMS

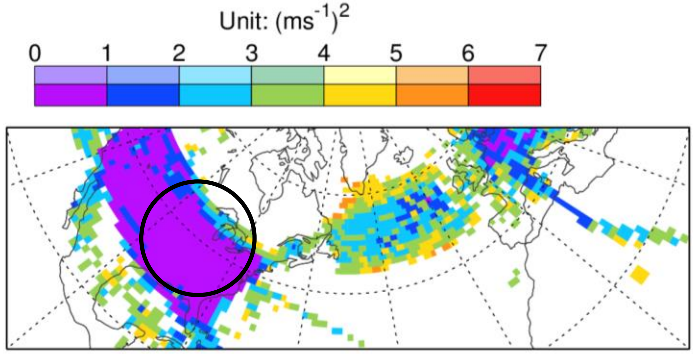
Departure^2



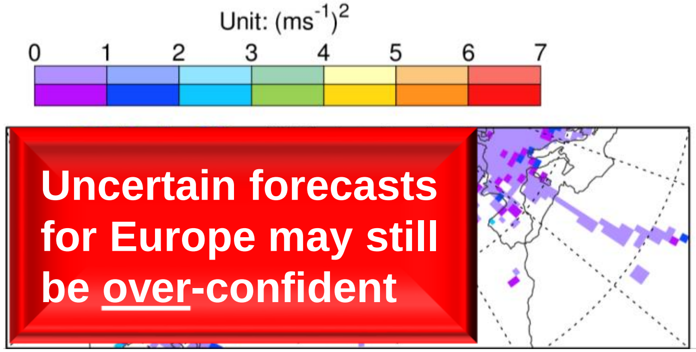
Background variance



Estimated observation error variance

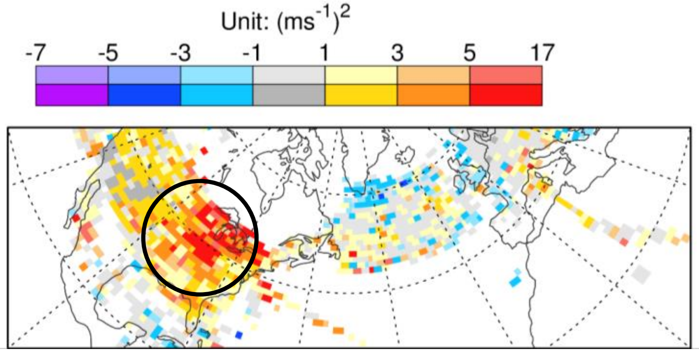


Bias^2

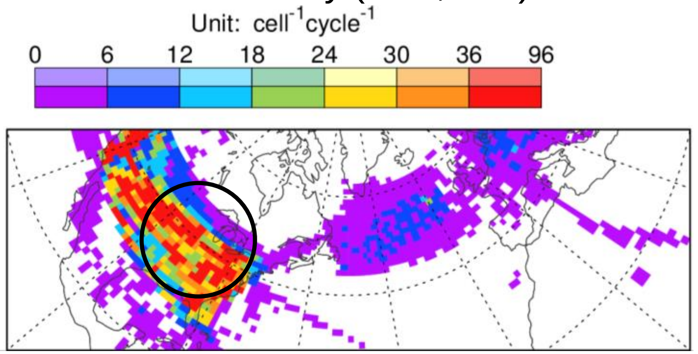


Uncertain forecasts for Europe may still be <u>over</u>-confident

Residual



Observation density (O80, 12h)



54 cases, 12h window

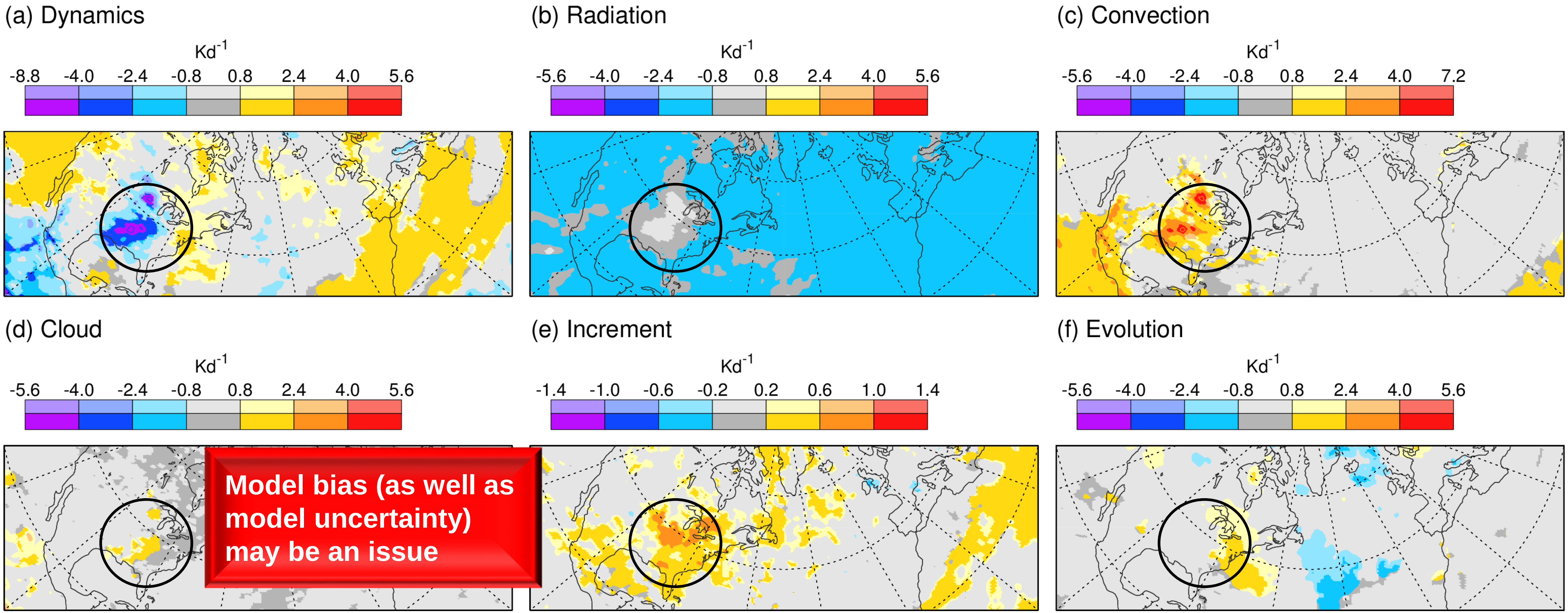
Enhanced uncertainty (Background variance) in Great Lakes / Mississippi River region

Even larger Departures^2 of the ensemble-mean from the observations ensue

Bias^2≈0 (important for reliability), but Residual ≫ 0 indicates insufficient Background variance (since Estimated observation error variance and Observation density are similar over Northwestern North America where Residual is smaller)

EDA unperturbed initial tendency budget for T300 in “Rocky trough/CAPE” situations

Rodwell, Richardson, Parsons & Wernli. 2018, BAMS



Budget: **Evolution = Dynamics + Radiation + Convection + Cloud micro-physics + analysis Increment**

54 cases, 12h window

Shows how the model represents dynamics and physics of MCS

Positive (and statistically significant) increment suggests that the background forecast is too cold near the top of the convection

The Jetstream and mesoscale convection: “The piano string and hammer”

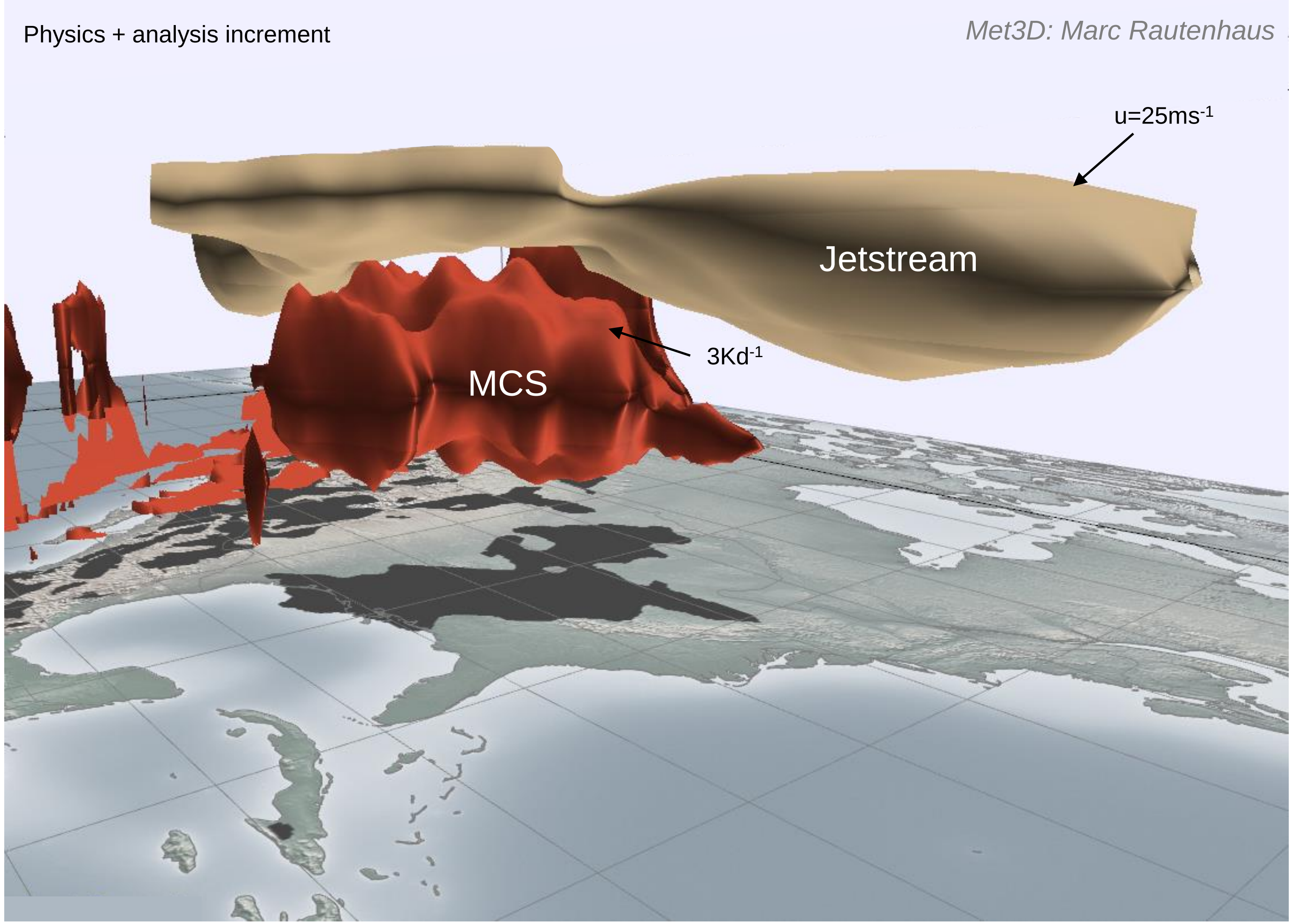
54 cases

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

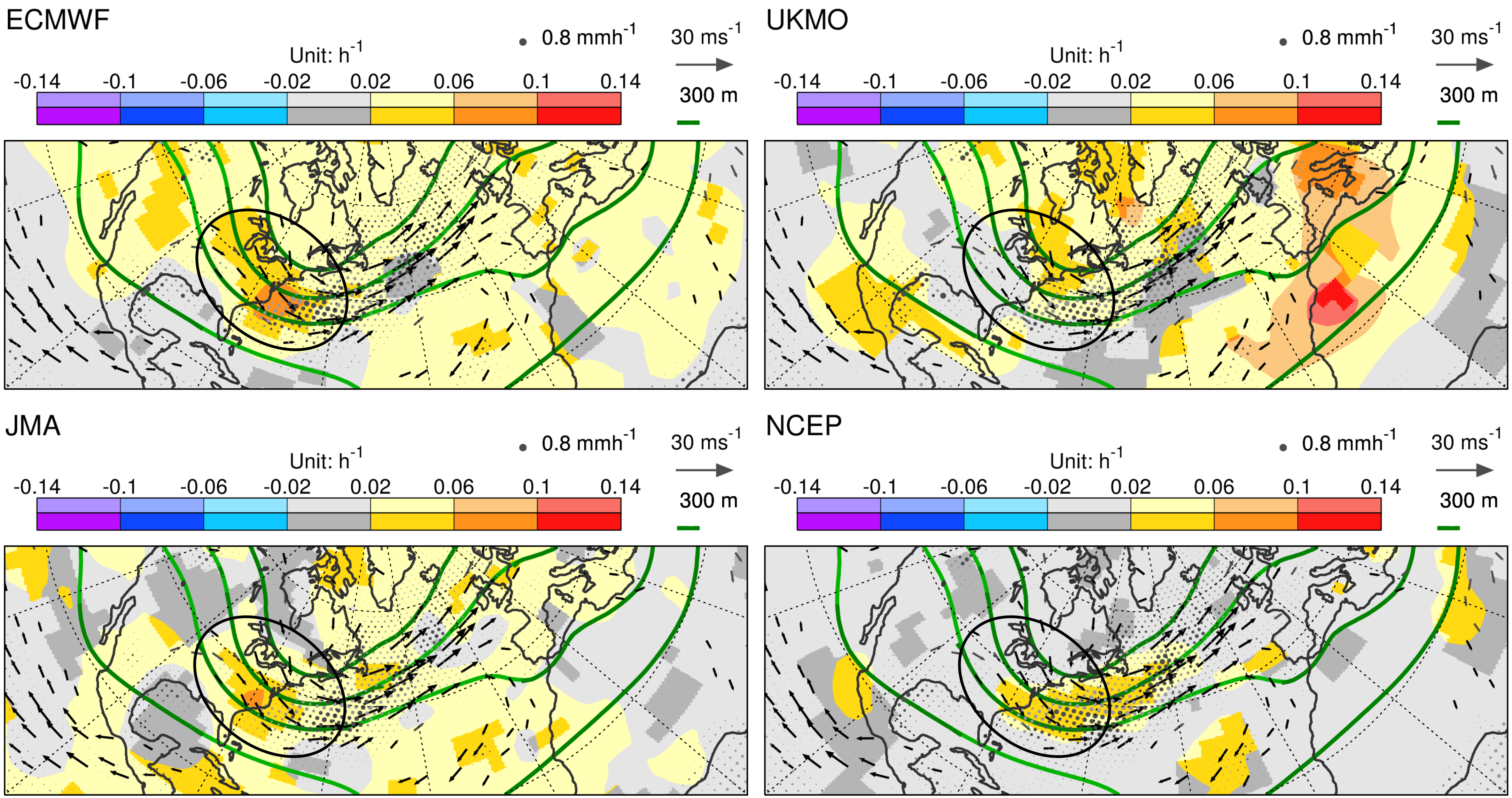
If we hit the string at the wrong time, the wave will arrive over Europe at the wrong time

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



TIGGE model Z250 growth rates for WCB cluster



ECMWF: Substantial growth-rate in trough region.

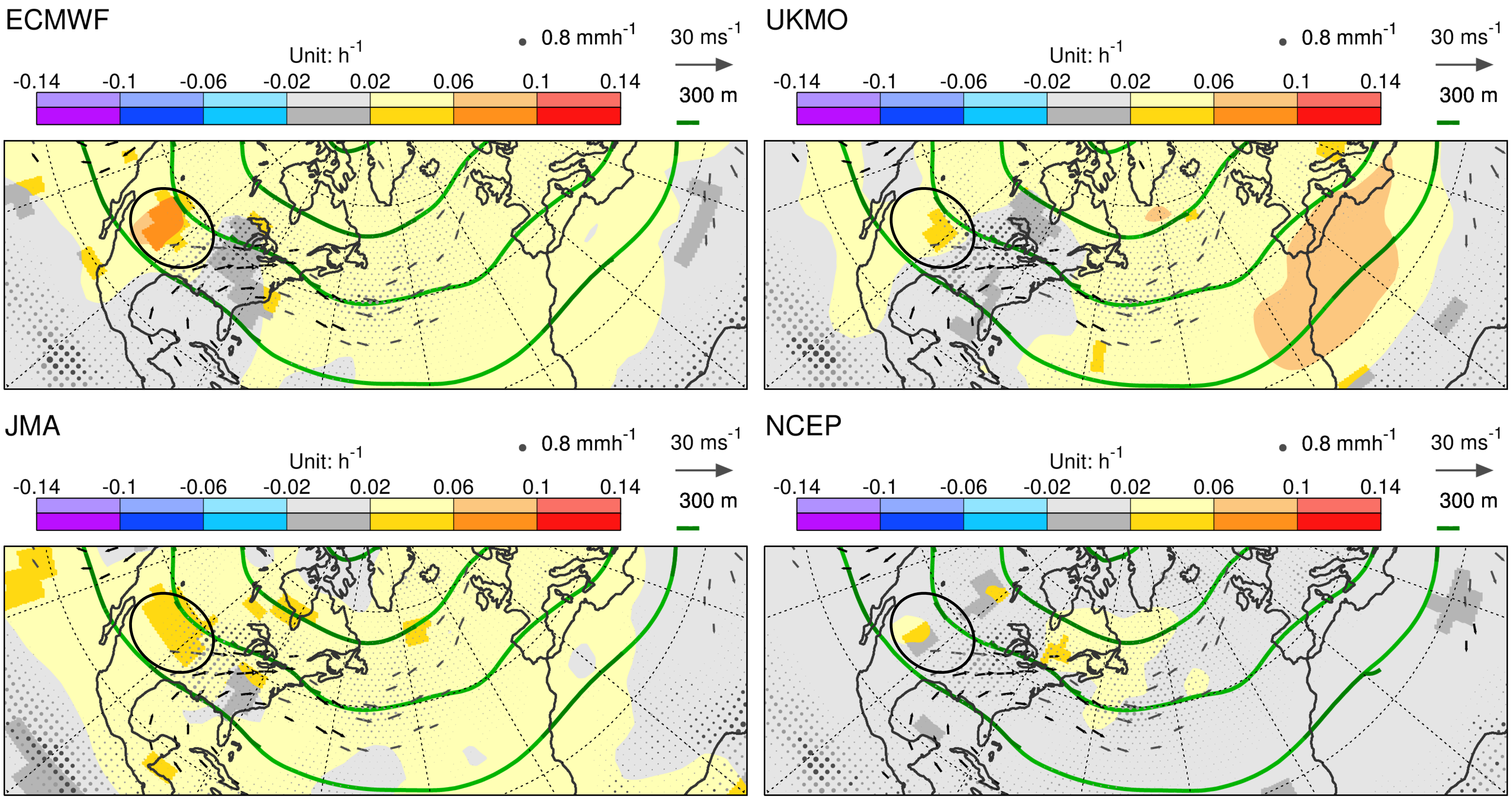
JMA: ≈ECMWF

UKMO: stronger growth-rate over Europe/Africa

NCEP: generally weaker growth-rate

MAM 2017. Clustered on ψ200 and T850 in same region for clean comparison. Some dates missing from TIGGE archive so WCB cluster includes 25 EDA cycles

TIGGE model Z250 growth rates for MCS cluster



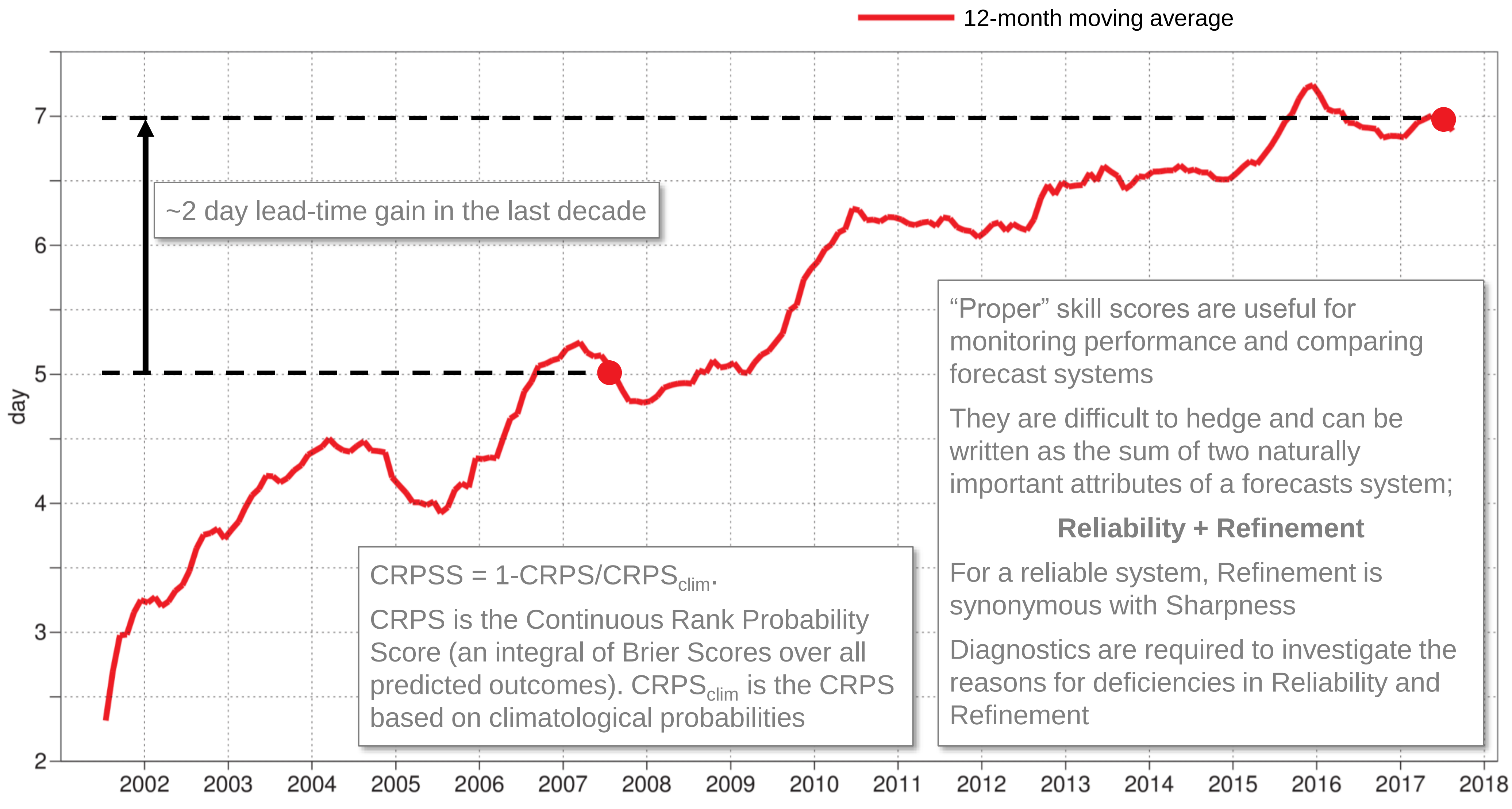
Weaker growth-rate for MCS cluster (larger cluster ⇒ more smeared-out?)

ECMWF: strongest MCS growth-rate(but still in deficit over Great Lakes region)

MAM 2017. Clustered on PV315 growth, CF PV315 & v850, EM precipitation. Some dates missing from TIGGE archive so MCS cluster includes 87 EDA cycles

Trend in probabilistic forecast performance. Leadtime at which CRPSS drops to 0.1

Continuous Rank Probability Skill Score (CRPSS) for extratropical precipitation verified against 24h observed accumulations



Climate prediction

In climate prediction, or assessment of sensitivity to a doubling of CO_2, and important issue is the distribution of outcomes purely associated with model biases.

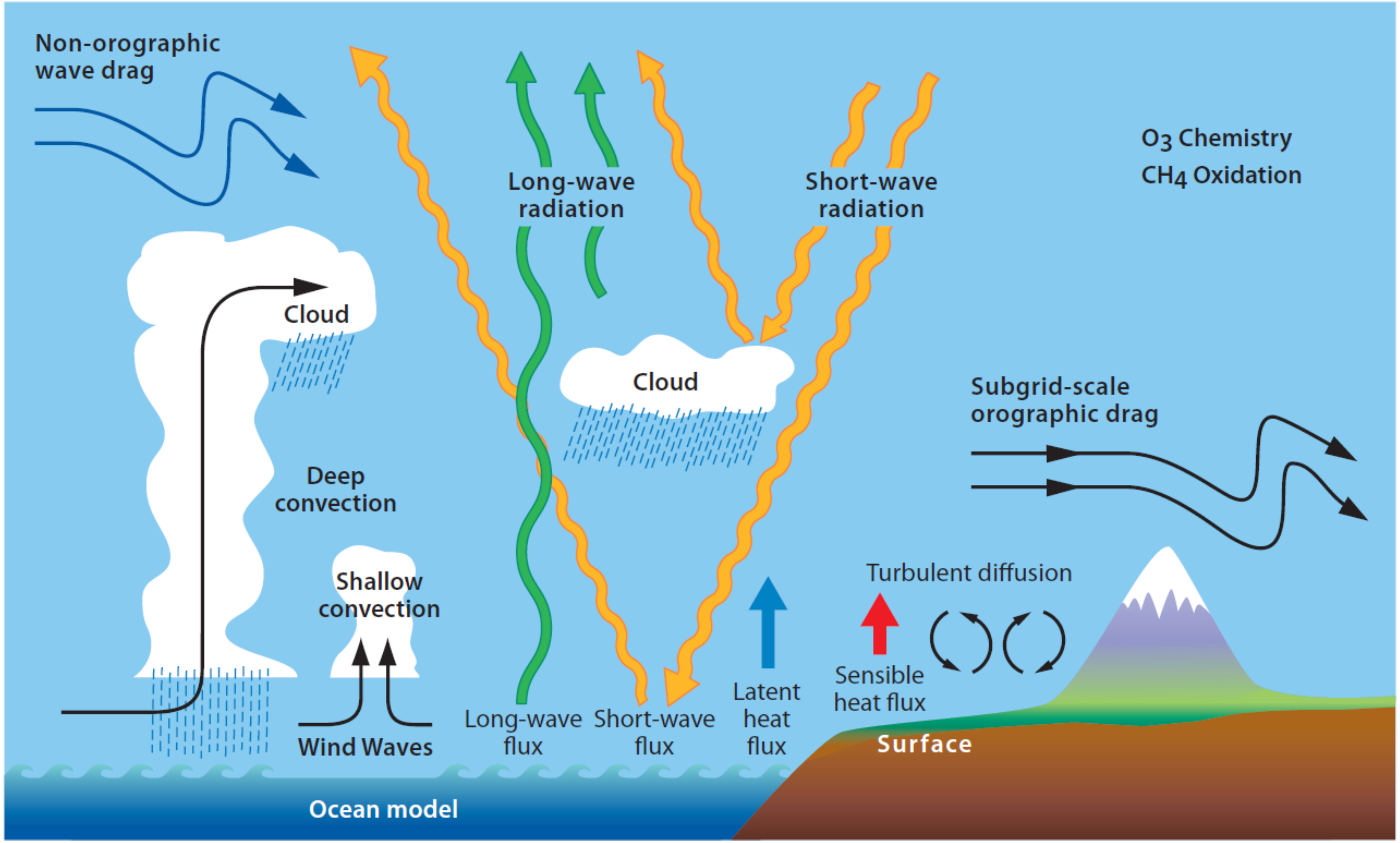
Can we use NWP diagnostics to constrain this aspect of climate uncertainty?

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 feedbacks with other processes or the evolution of the resolved flow.



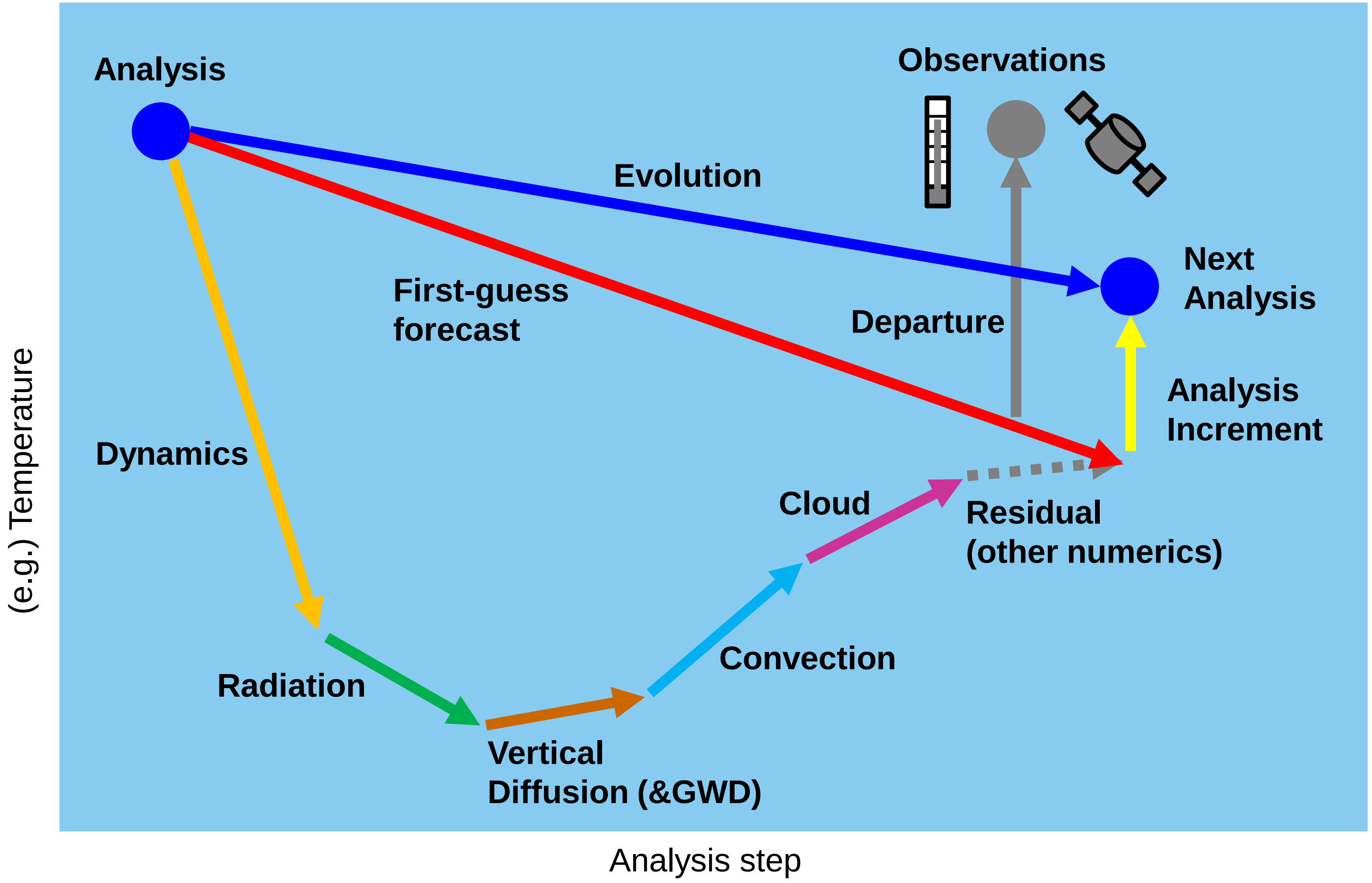
The Initial Tendency approach to diagnosing model error

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.

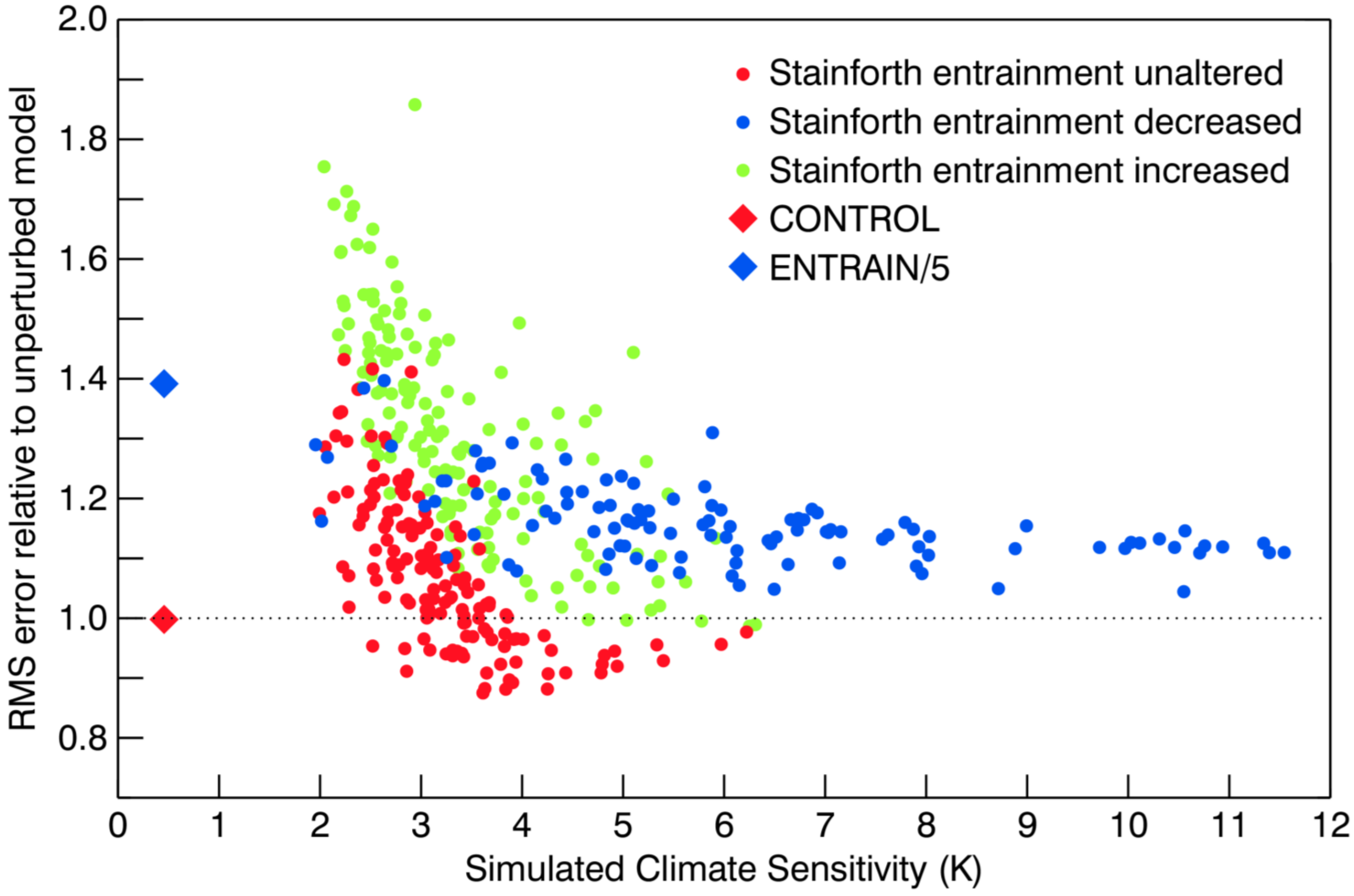
Schematic of the data assimilation process – a diagnostic perspective



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

Climate sensitivity of perturbed climate models

Rodwell and Palmer (2007) with data from Stainforth et al (2005)

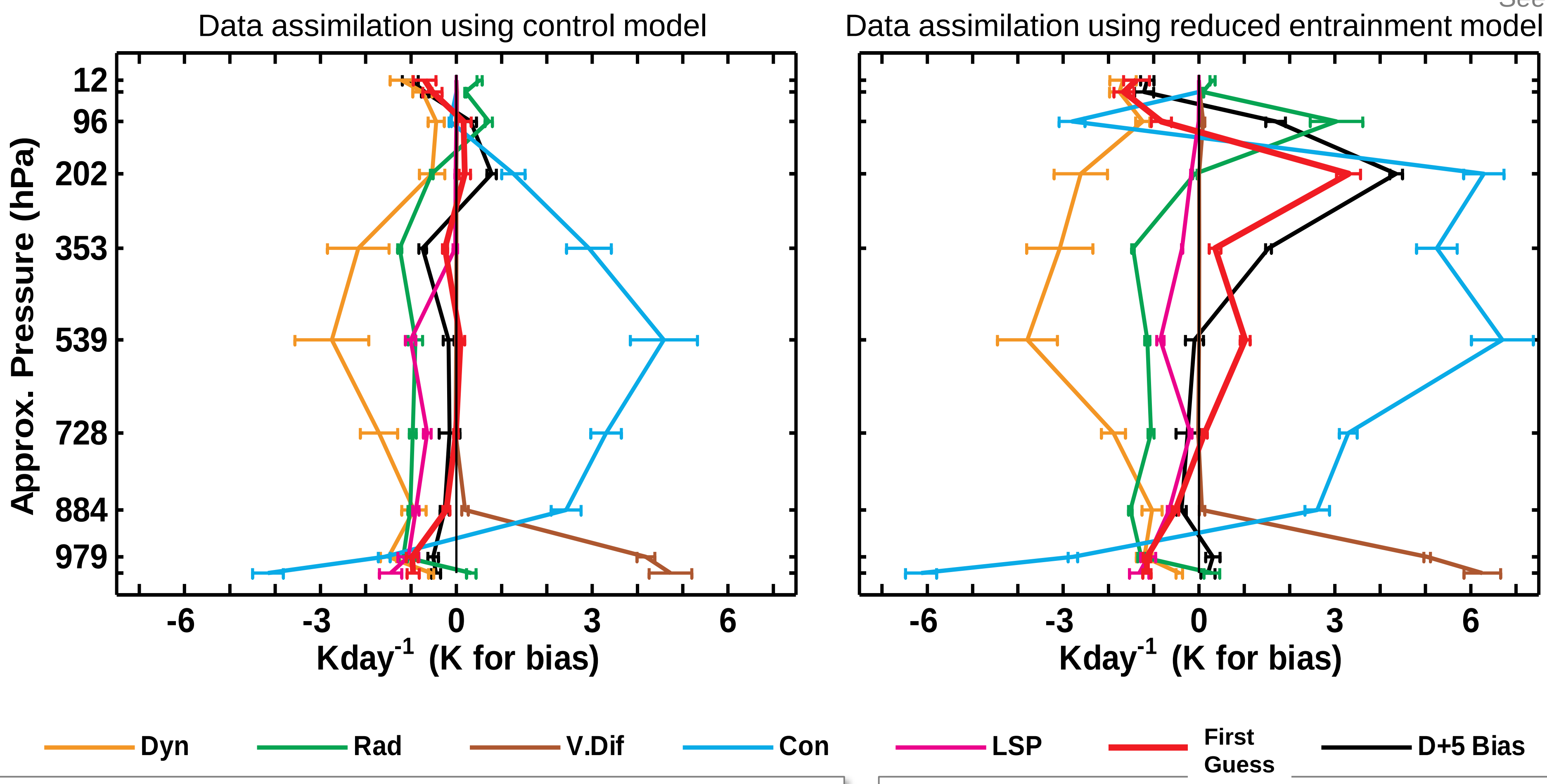


Using Initial Tendencies to investigate 12K warming possibility in climate ensemble

Temperature tendency profiles over the Amazon (300-320^{o}E, 20^{o}S-0^{o}N)

Rodwell and Palmer (2007)

See also Sexton et al. (2019)



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

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

Summary & Discussion

- The task of forecast system development can be summarised as *improving sharpness subject to maintaining or improving reliability*.
- For seasonal-mean, northern Hemispheric 500hPa geopotential heights, we see that mean spread ≈ mean error at lead-times 1-10 days. To this extent the ECMWF ensemble is reliable.
- Some flow-types associated with very low predictability have a large impact on the seasonal-mean scores. They may not be modelled very reliably or assimilated very sharply.
- Flow-dependent diagnostics can help guide research to the most important issues.
- Initial tendency diagnostics can also help us better quantify our current uncertainty in climate change associated with model bias.

Thank you

Summary: A practical path to more skillful ensemble forecasts

A theoretical path

Proper scoring rules reward reliability and refinement. *e.g.*:

$$\text{Brier Score} = \frac{1}{N}\sum_{t=1}^{N}(p_t - o_t)^2$$

p_t = forecast prob.
o_t = 0 or 1 outcome

$$\approx \underbrace{\frac{1}{N}\sum_{k=1}^{K} n_k (p_k - o_k)^2}_{\textbf{Reliability}} + \underbrace{\frac{1}{N}\sum_{k=1}^{K} n_k o_k (1 - o_k)}_{\textbf{Refinement}}$$

p_k = forecast prob.
o_k = outcome freq.

Instead of directly binning on forecast probabilities, think of this as a partition over initial flow-types. The probabilities that arise for a given flow-type should be similar enough to be represented by a single probability if the flow-types are defined tightly enough, if the events being forecast are local to the flow-type, and if short-enough leadtimes are considered. Modelling developments which improve short-range flow-dependent reliability (even if this involves increasing uncertainty growth-rates) should lead to improvements in the Brier Score*, and other proper scores

*The refinement term is only directly dependent on the initial flow-types and the verifying observations, and should be less affected

A practical path to improving flow-dependent **reliability**

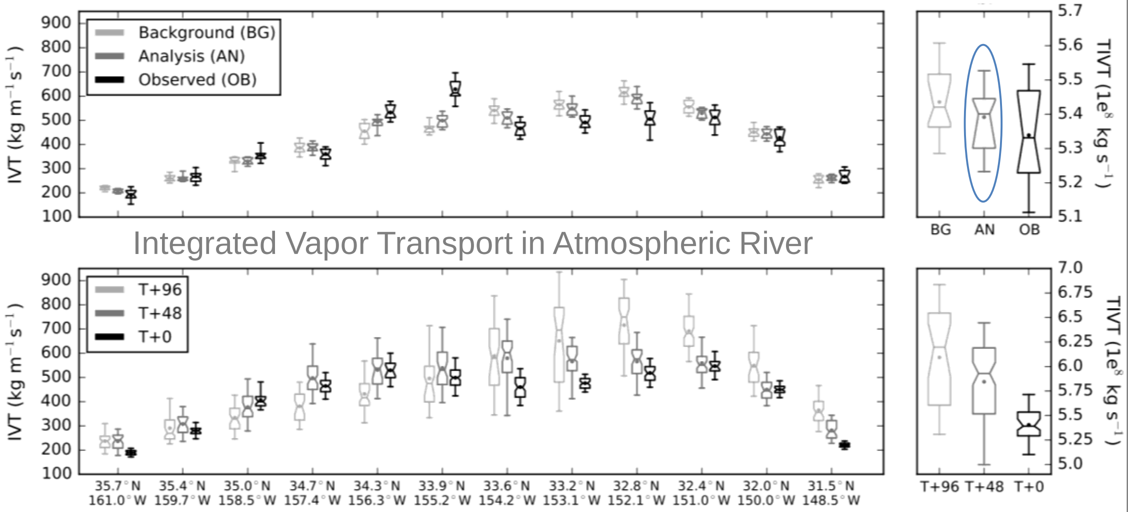
By analogy, for a small set of initial flow-types (clusters), reduce local short leadtime Bias^2 and Residual in:

$$\overline{\mathrm{Departure}_k^2} = \overline{\mathrm{EnsVar}_k} + \overline{\mathrm{ObsUnc}_k^2} + \overline{\mathrm{Bias}_k^2} + \overline{\mathrm{Residual}_k}$$

This should improve reliability of ensemble initialization, and uncertainty growth-rates applicable at all leadtimes

Note that the theoretical assumptions are not required here

Assuming reliability, **refinement** benefits from better Observations



Better Obs $\Rightarrow$ Reduced analysis uncertainty $\Rightarrow o_k \approx p_k \rightarrow \in \{0,1\}$