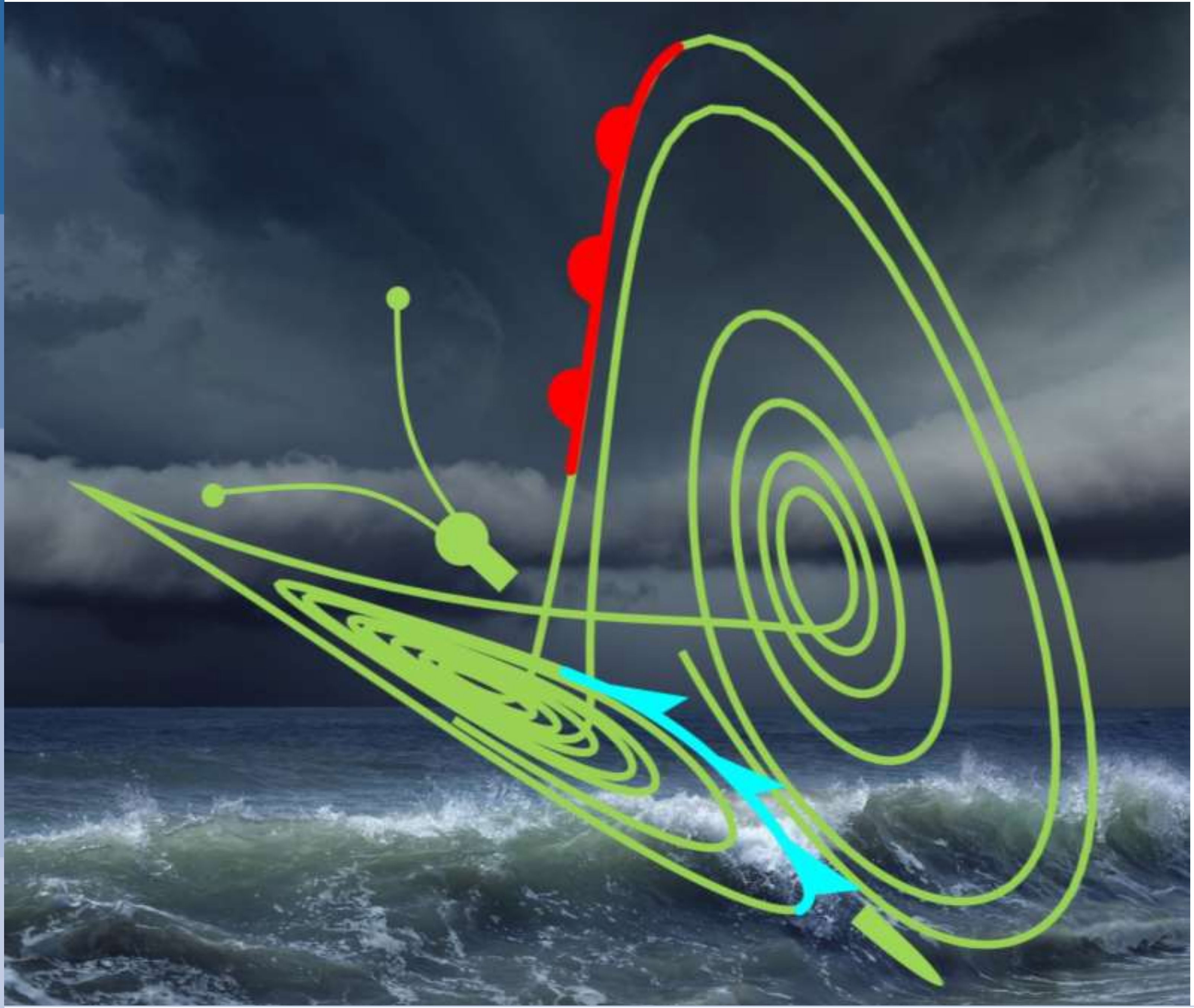


Flow-Dependent Error Growth in TIGGE Models

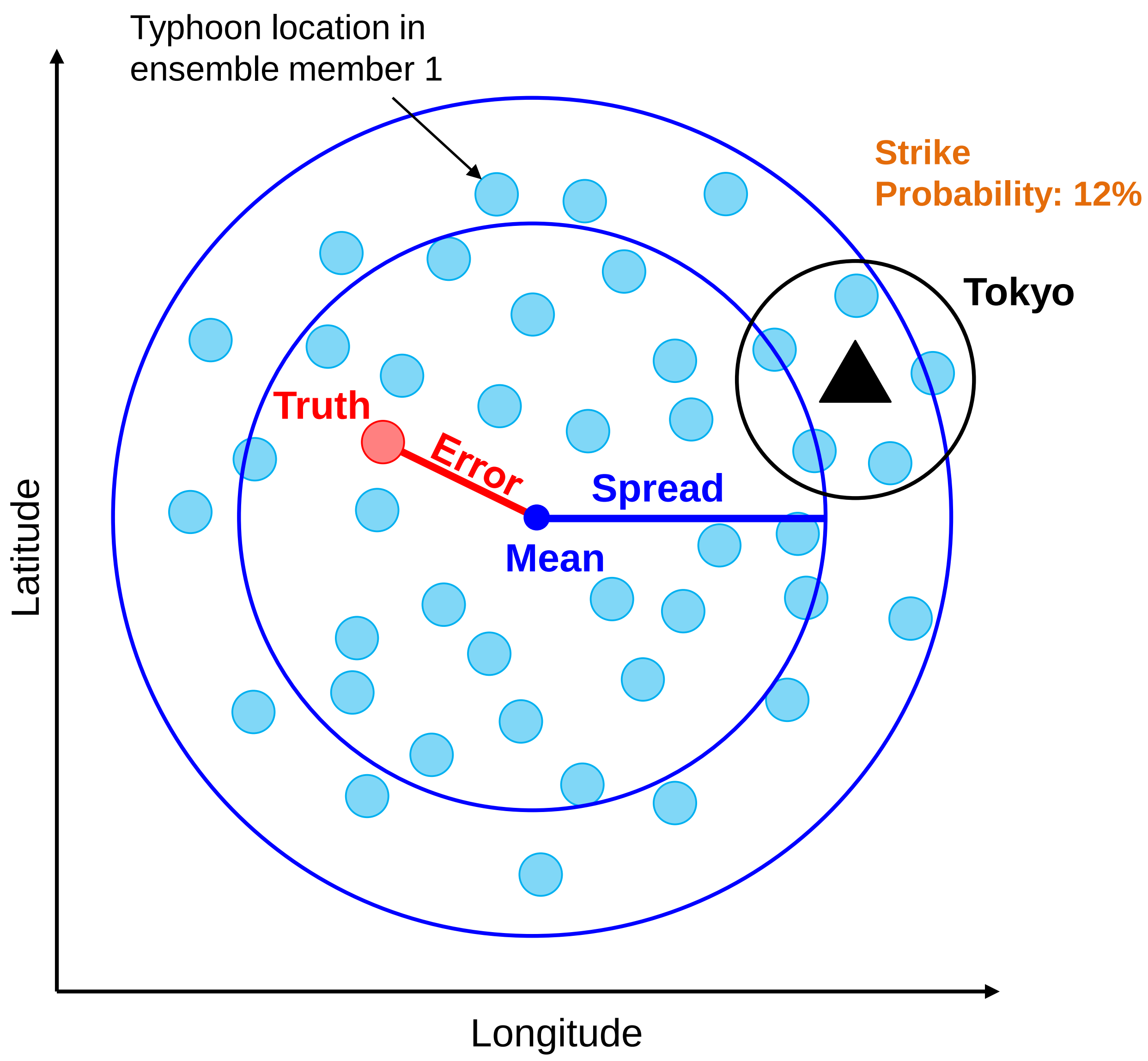
Mark Rodwell

Following recent paper with David Richardson, Dave Parsons, Heini Wernli, *BAMS* (2018)

WGNE/PDEF Joint Meeting

10 October 2018, Tokyo

Flow-dependent Reliability and Sharpness: Typhoon strike example



In a **reliable** forecast system, the truth should be statistically indistinguishable from the individual ensemble members

Reliability is very useful: an event predicted to occur with probability 12% will happen with frequency 12%

An easily testable consequence of reliability is that

$$\overline{\mathbf{Error}^2} = \overline{\mathbf{Spread}^2}$$

(averaged over many forecast start dates)

“The task of NWP research is to **maintain reliability** while decreasing spread (improving refinement)”

Ensemble spread and error

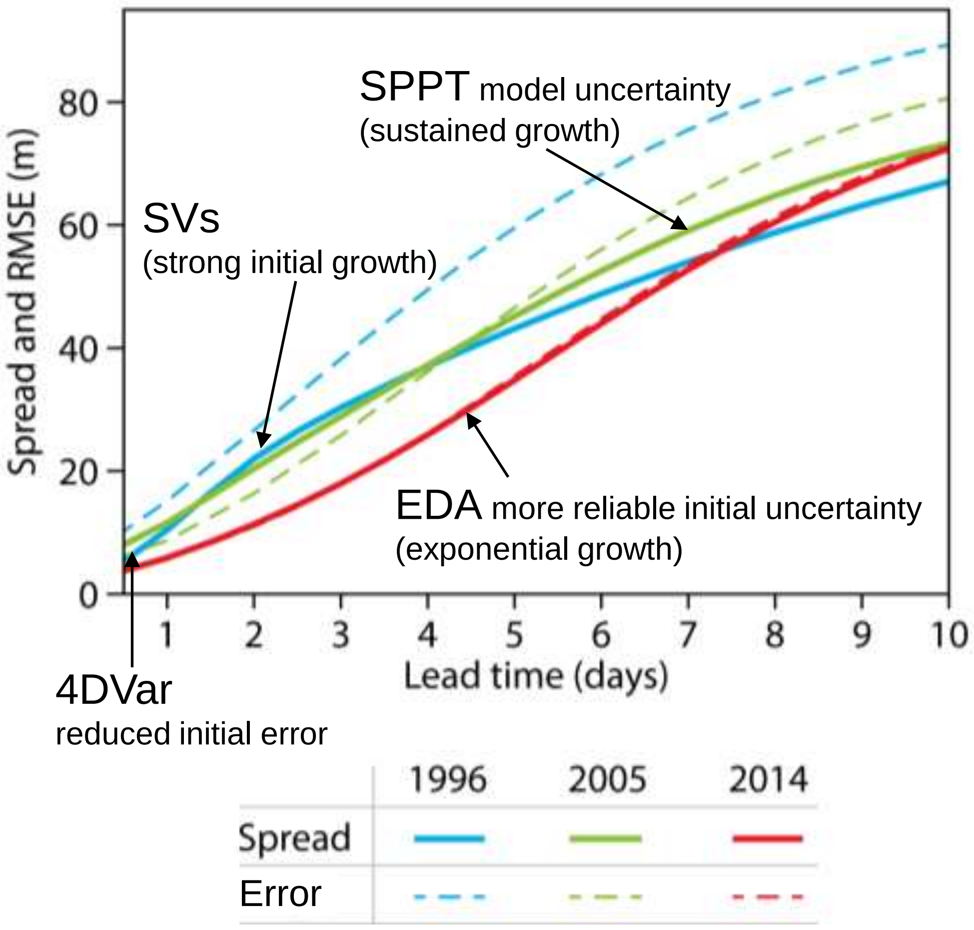
Z500

Rodwell et al. 2018, BAMS

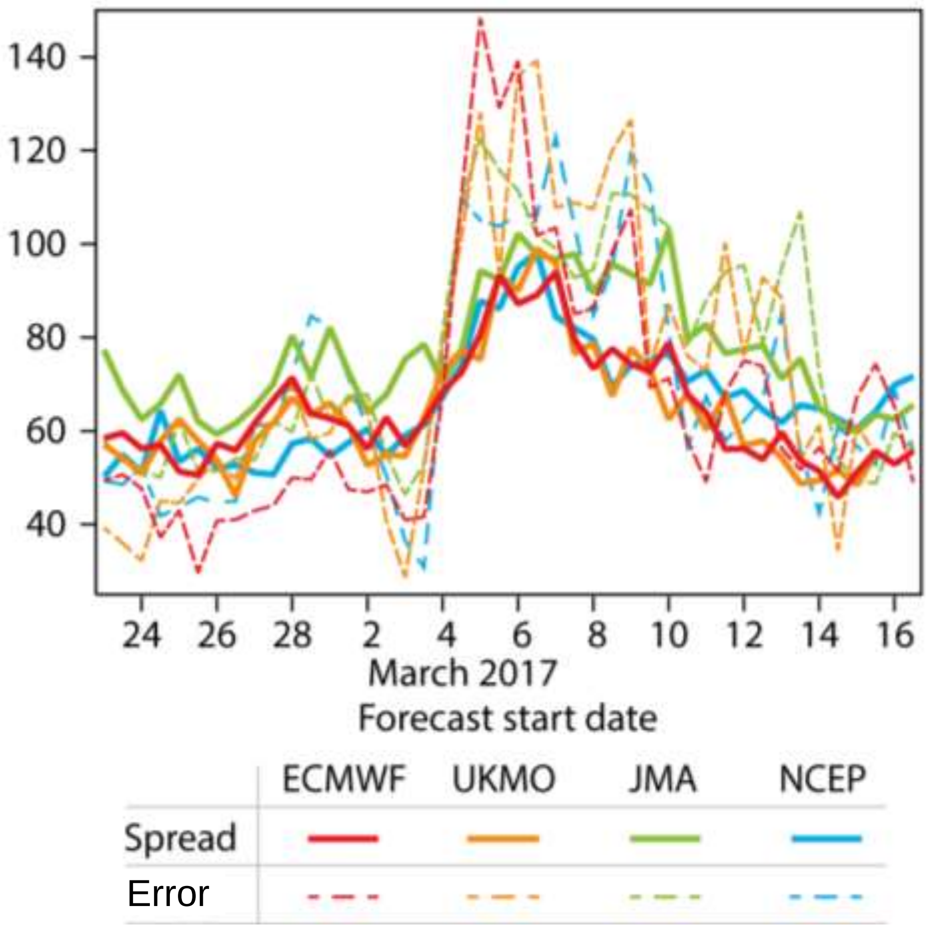
Overall Error and Spread have reduced and come into alignment; due to better observations, initial conditions, forecast model and better representation of uncertainty

…but we make ensemble forecasts to represent the day-to-day variations in predictability and uncertainty. Can we evaluate it in our forecasts?

Annual means N.Hem. (ECMWF)



Timeseries for Europe at D+6 (TIGGE)



500 hPa geopotential height (Z500). "Error" is RMS of ensemble-mean error

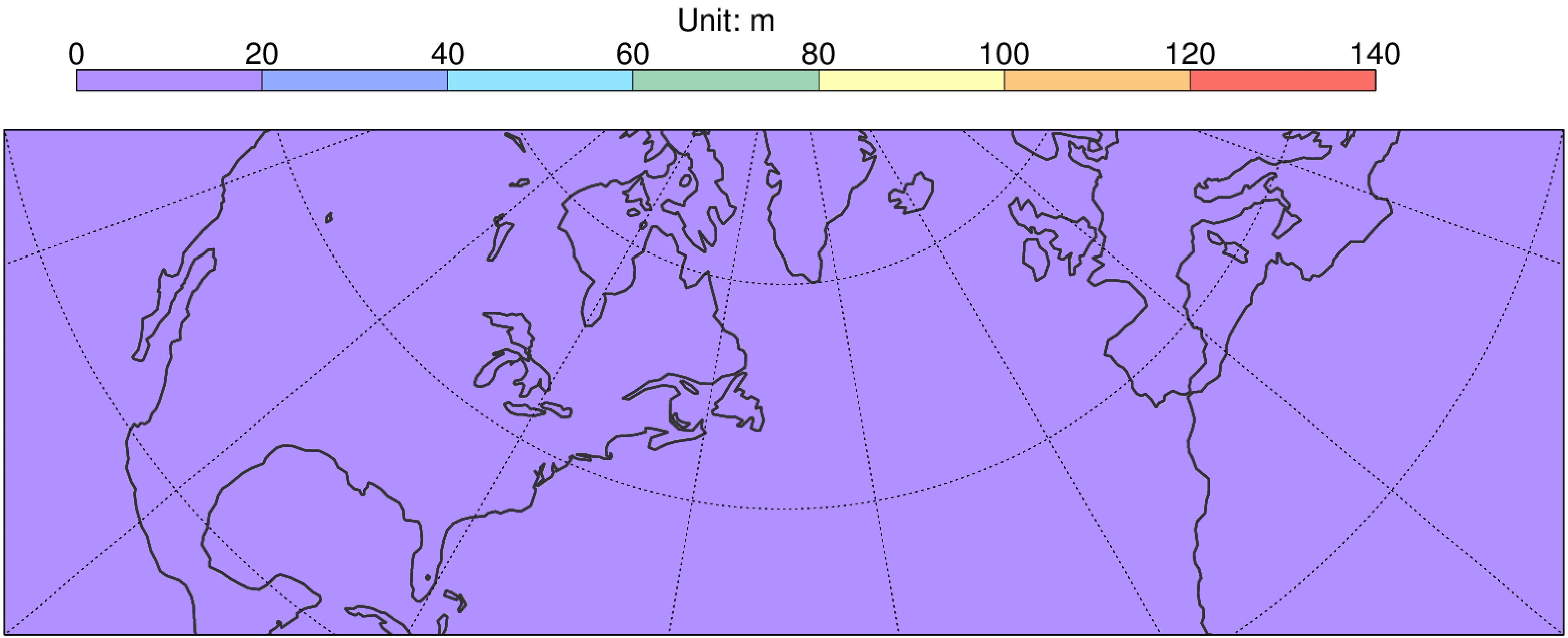
Spread = ensemble standard deviation (scaled to take account of finite ensemble size)

Animation of ECMWF ensemble forecast spread 20170305 12Z D+0 to 6: σ_{Z500}

ECMWF ENS stdev Z_{500hPa} (shaded).

20170305 12Z

Uncertainty growing from various sources, is itself advected, and becomes large over Europe by D+6



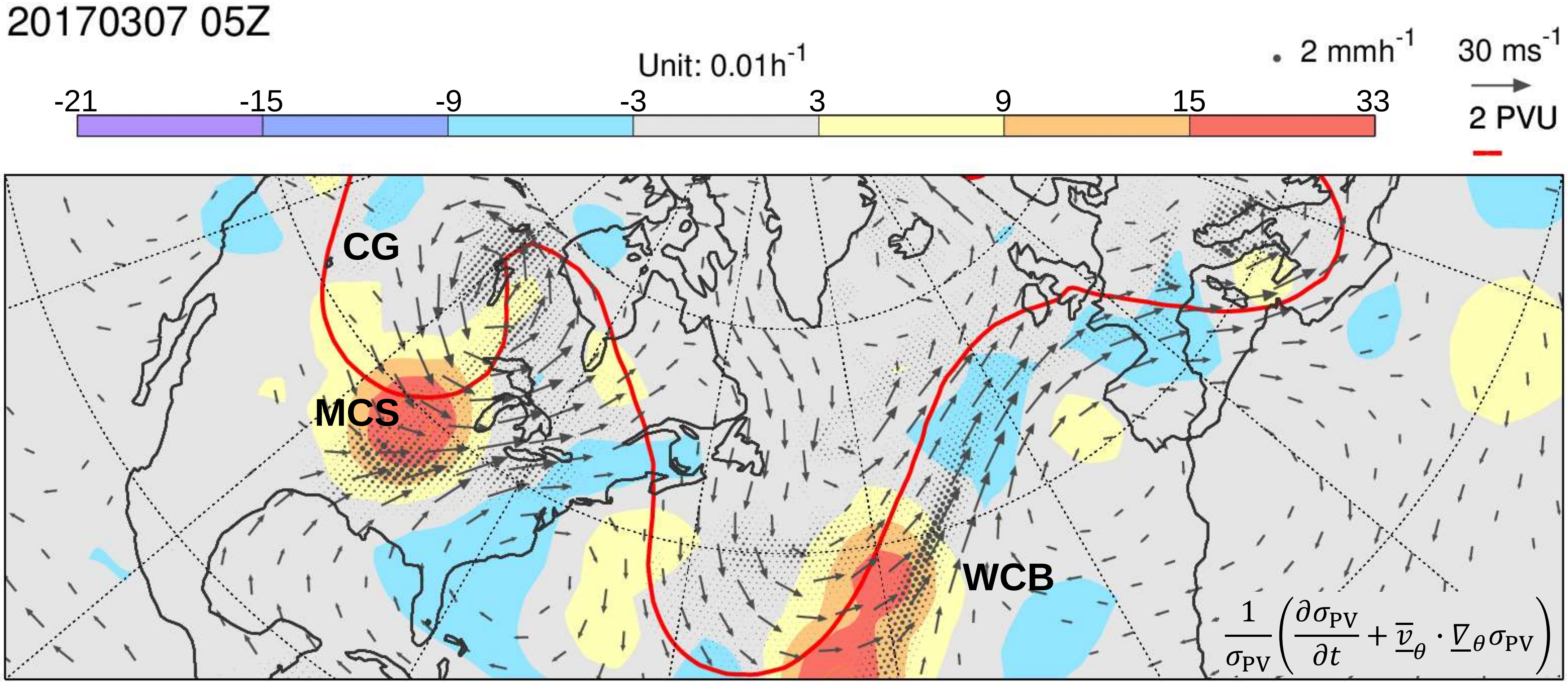
“Lagrangian” growth-rate (following EnsMn horizontal flow) for EDA background $\sigma_{PV_{315}}$

Much uncertainty growth associated with moist processes: **W**arm **C**onveyor-**B**elts, and **M**eso-**S**cale **C**onvection

Interaction of uncertain features, large ENS spread & poor prediction of Euro blocking at D+6

Aim: Evaluate short-range synoptic flow-dependent representation of uncertainty

Q: Is sensitivity to moist processes real or due to deficiencies in model uncertainty representation? TIGGE?



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

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

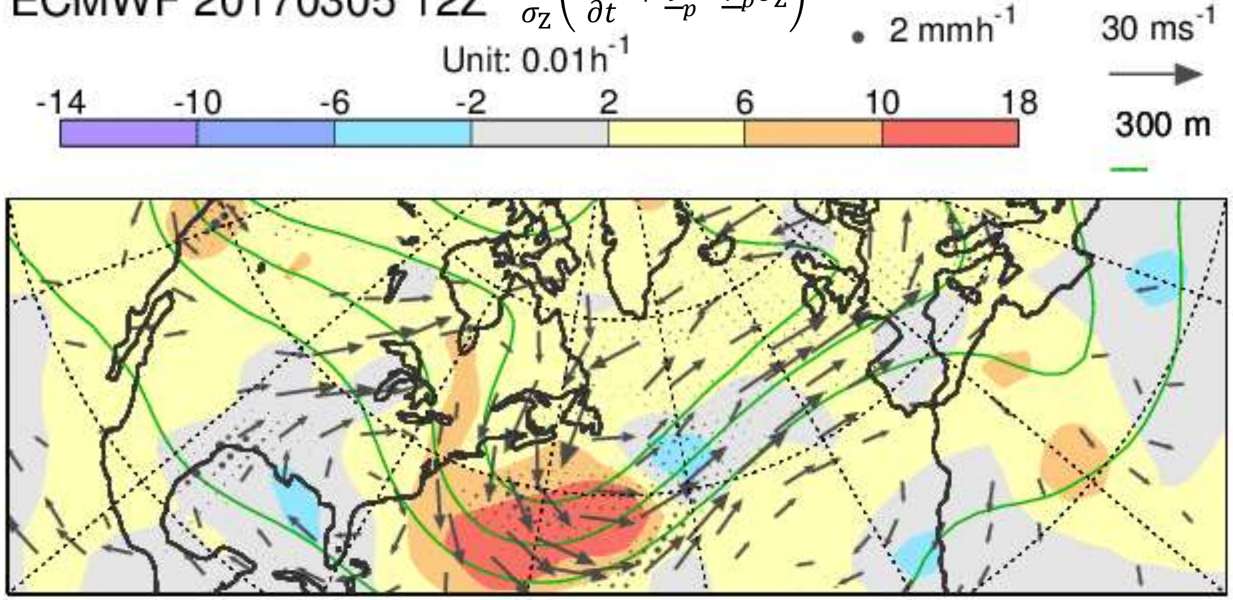
TIGGE 12h “Lagrangian” growth rate (following EnsMn horizontal flow) for Z_{250hPa} (shaded)

CF 850hPa winds (vectors), CF Z_{250} (green contour) and EM precip (dots). 1d running-mean applied.

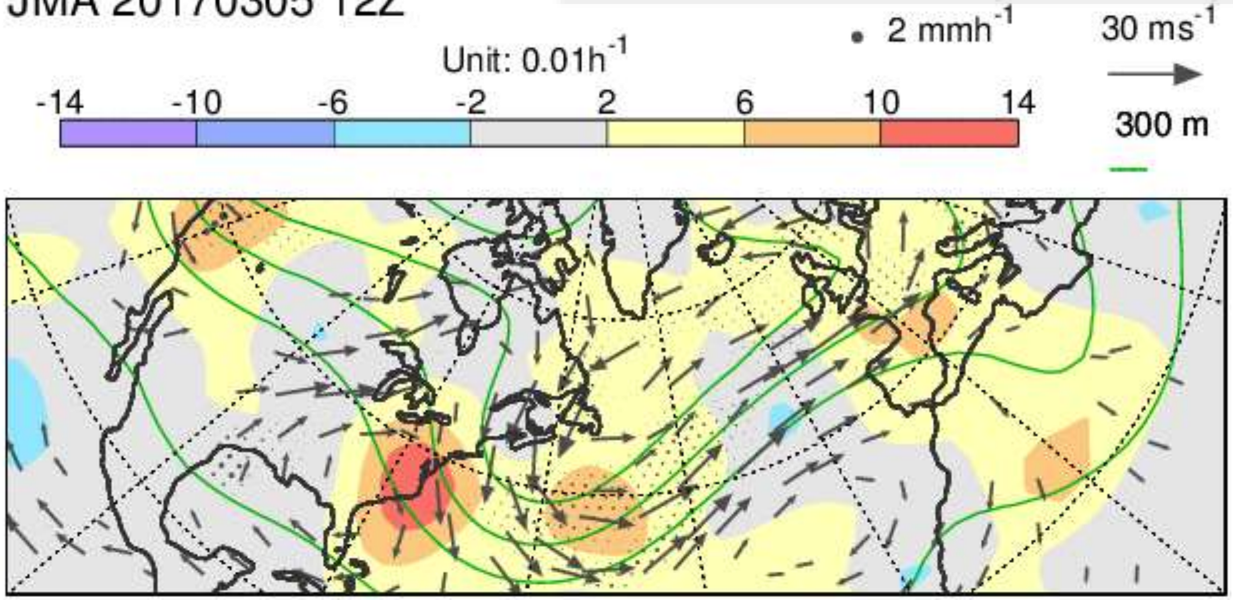
EDA(PV_{315K}) ≈ ENS(Z_{250hPa}) ≈ JMA. UKMO: MCS less, Europe more. KMA less everywhere(?) Which is best?

ECMWF 20170305 12Z

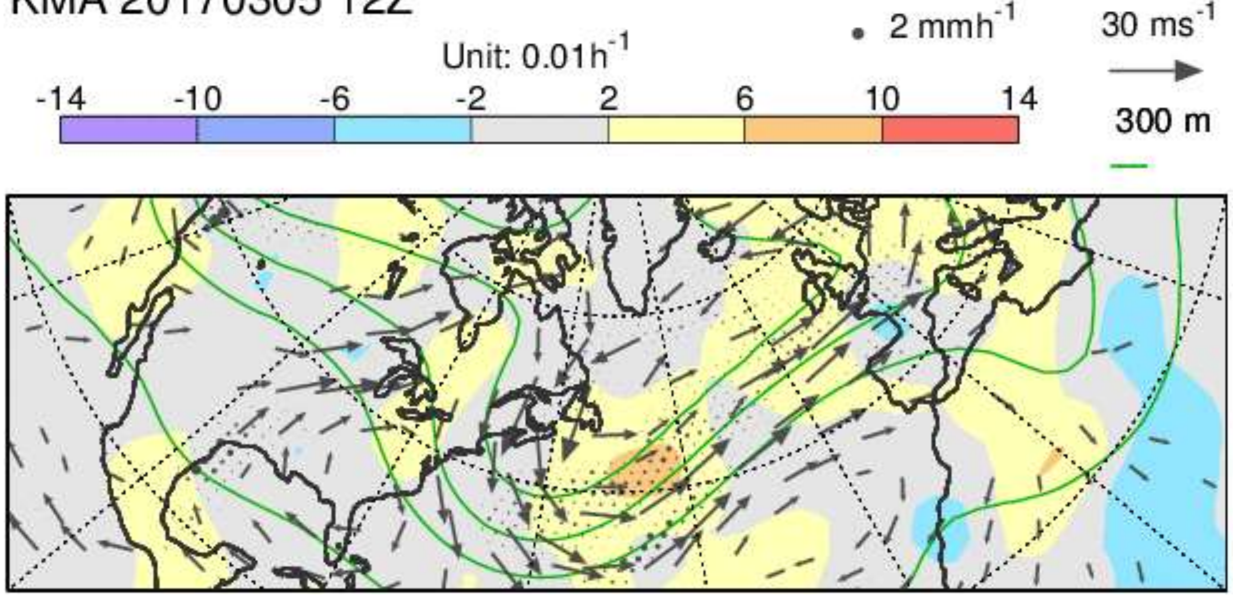
$$\frac{1}{\sigma_Z}\left(\frac{\partial \sigma_Z}{\partial t} + \overline{\underline{v}}_p \cdot \underline{\nabla}_p \sigma_Z\right)$$



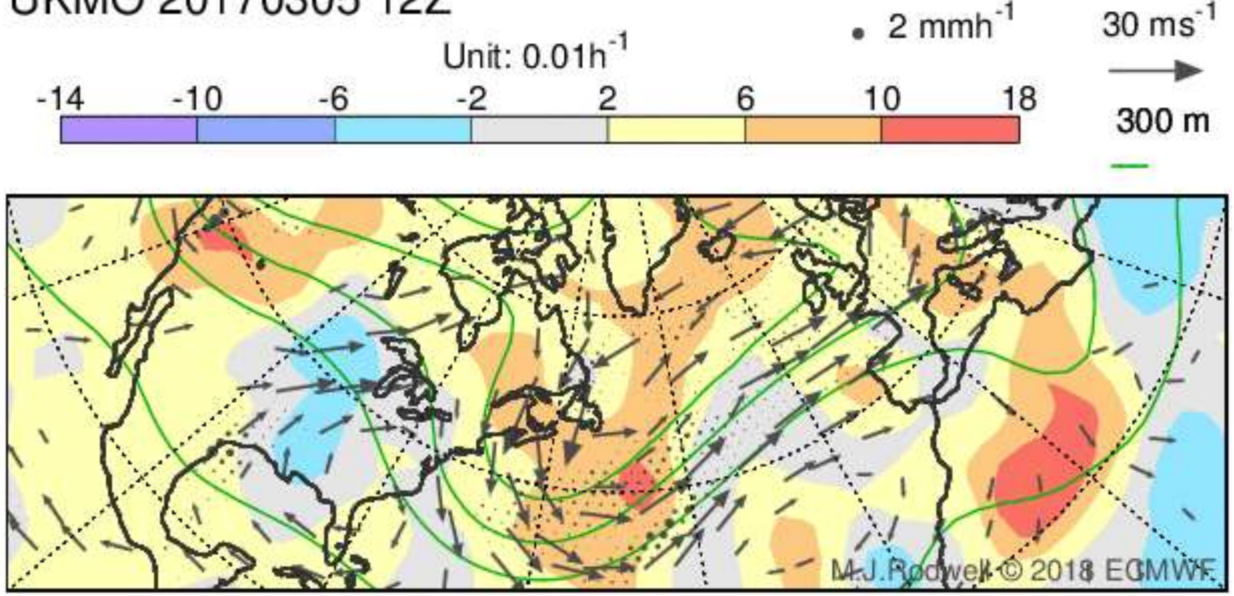
JMA 20170305 12Z



KMA 20170305 12Z



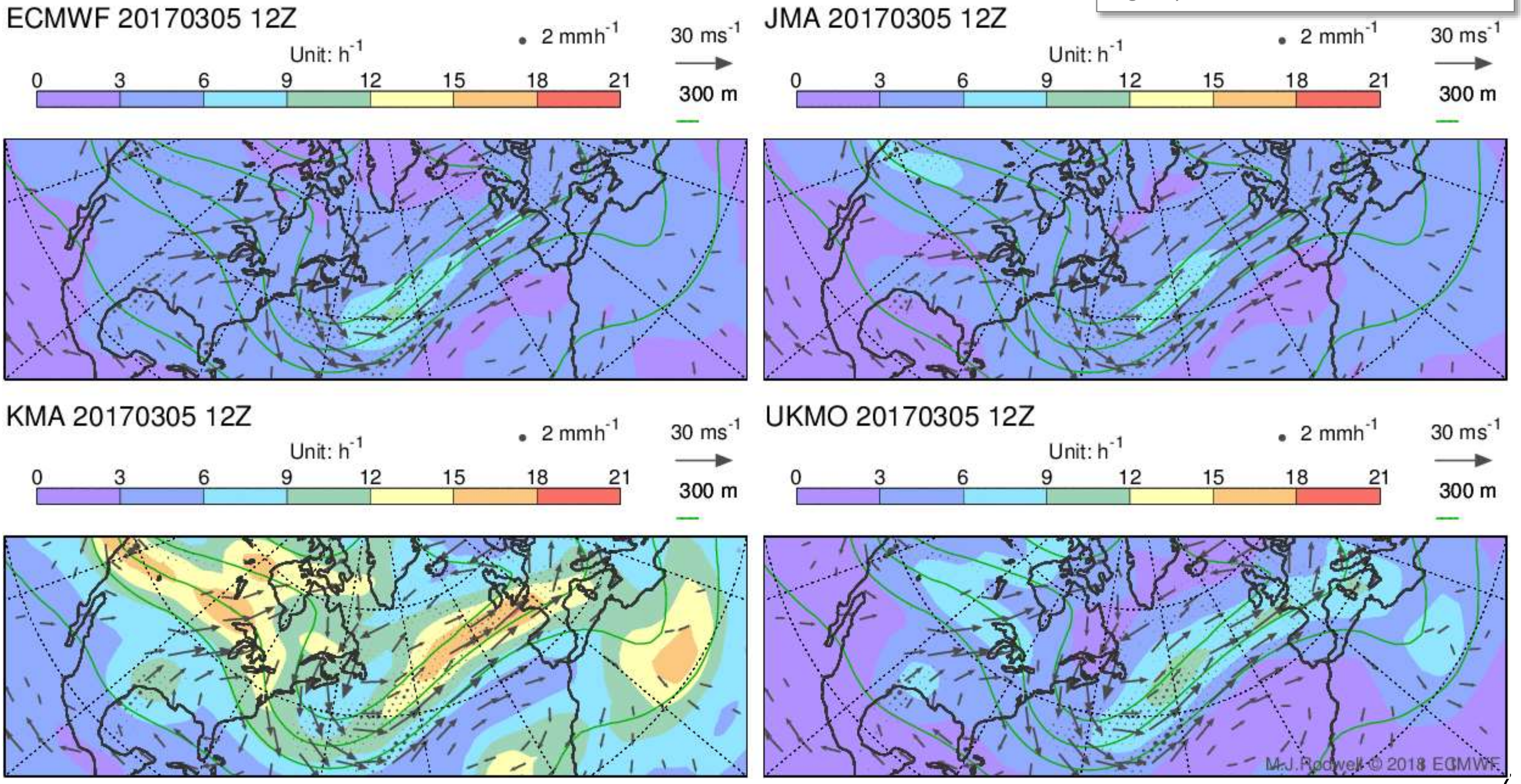
UKMO 20170305 12Z



TIGGE 12h Standard deviation in Z_{250hPa} (shaded)

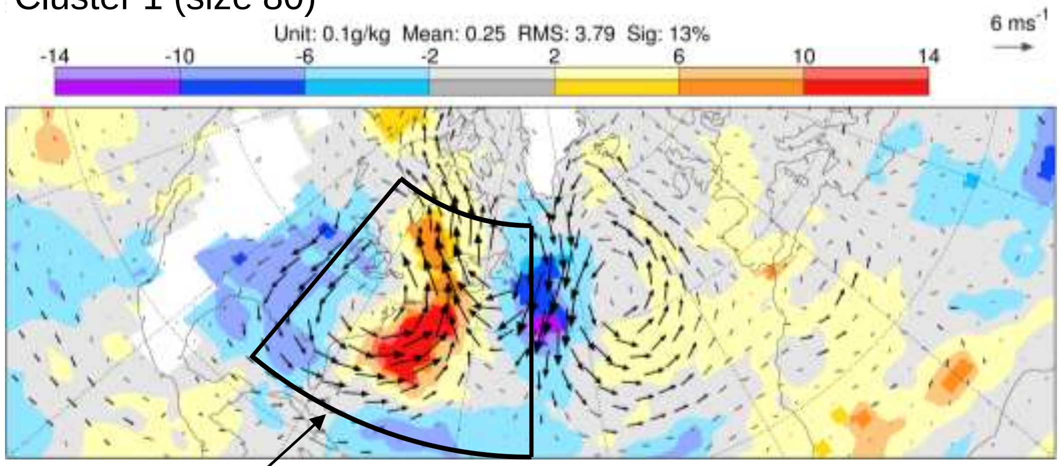
CF 850hPa winds (vectors), CF Z_{250} (green contour) and EM precip (dots). 1d running-mean applied.

Ensembles start from a different initial uncertainties, but growth-rate (in linear regime) should be unaffected

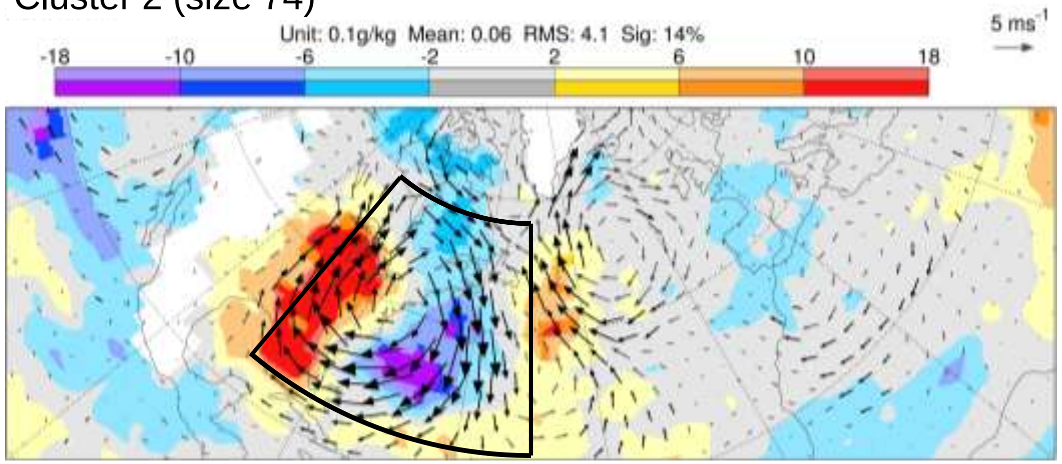


Mean anomalies in q850 and v850 for 3 clusters, identifying ‘synoptic’ flow patterns

Cluster 1 (size 80)

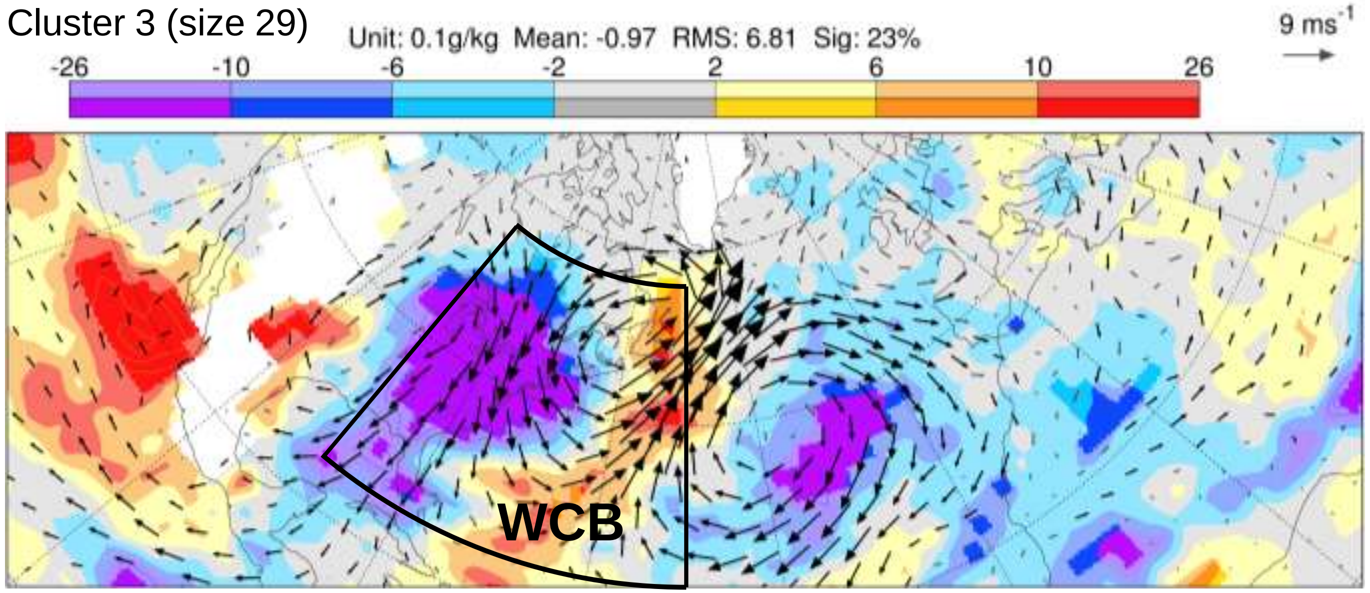


Cluster 2 (size 74)



Clustering region

Cluster 3 (size 29)



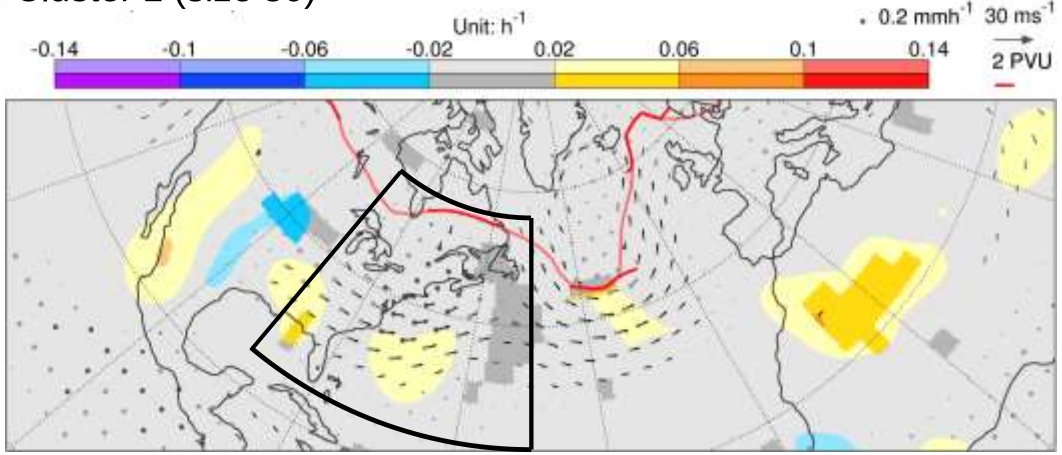
Here, the main focus is on Cluster 3. This includes a strong Warm Conveyor Belt and enhanced uncertainty growth-rates.

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

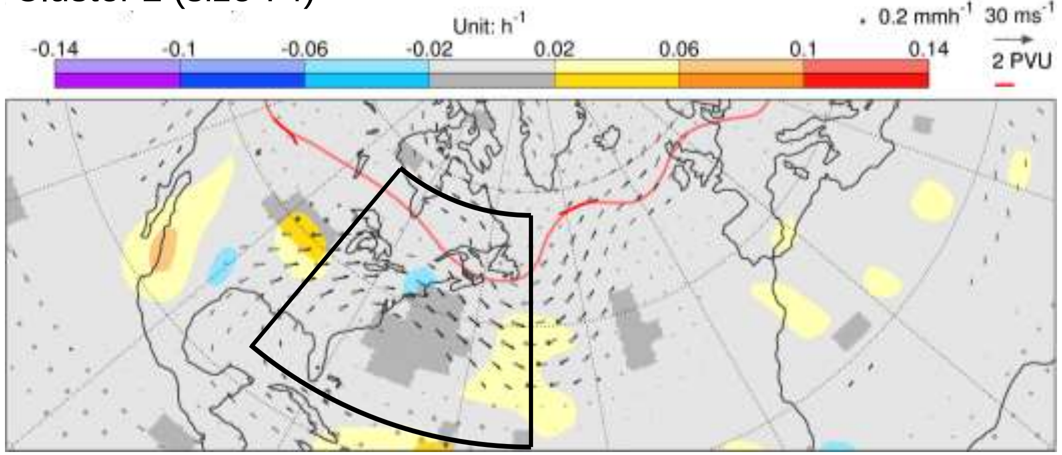
Based on K-means clustering of the fields from the animation in the region indicated, for MAM 2017. Bold colors = 5% significance

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

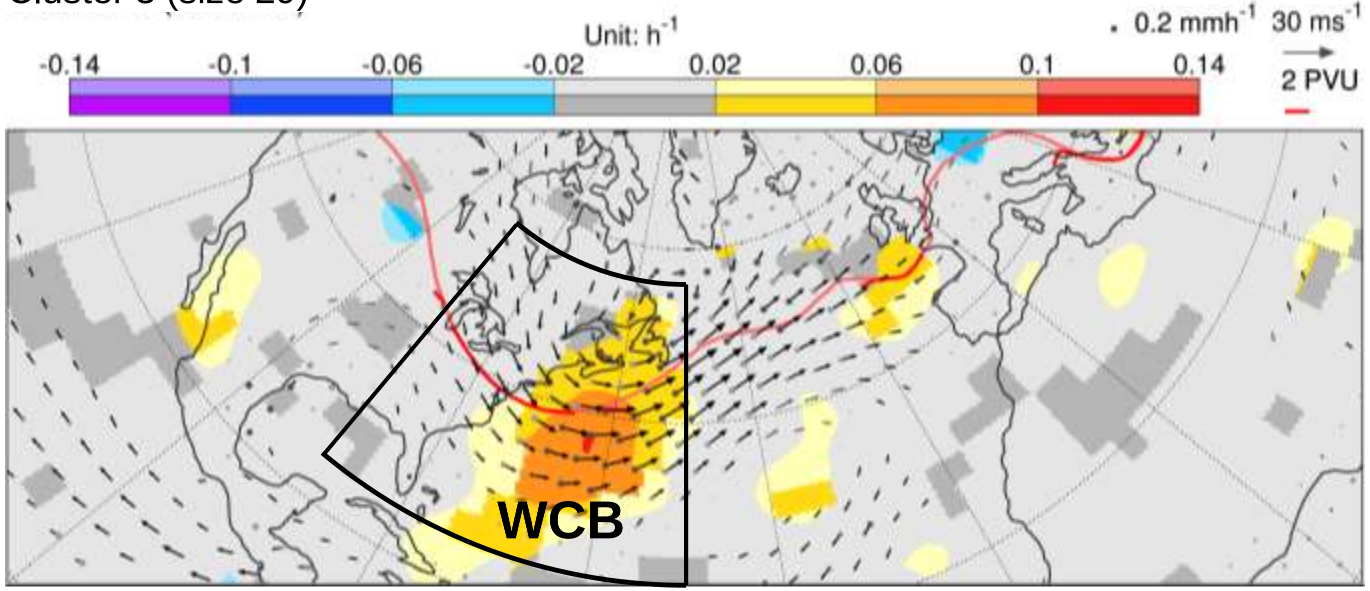
Cluster 1 (size 80)



Cluster 2 (size 74)



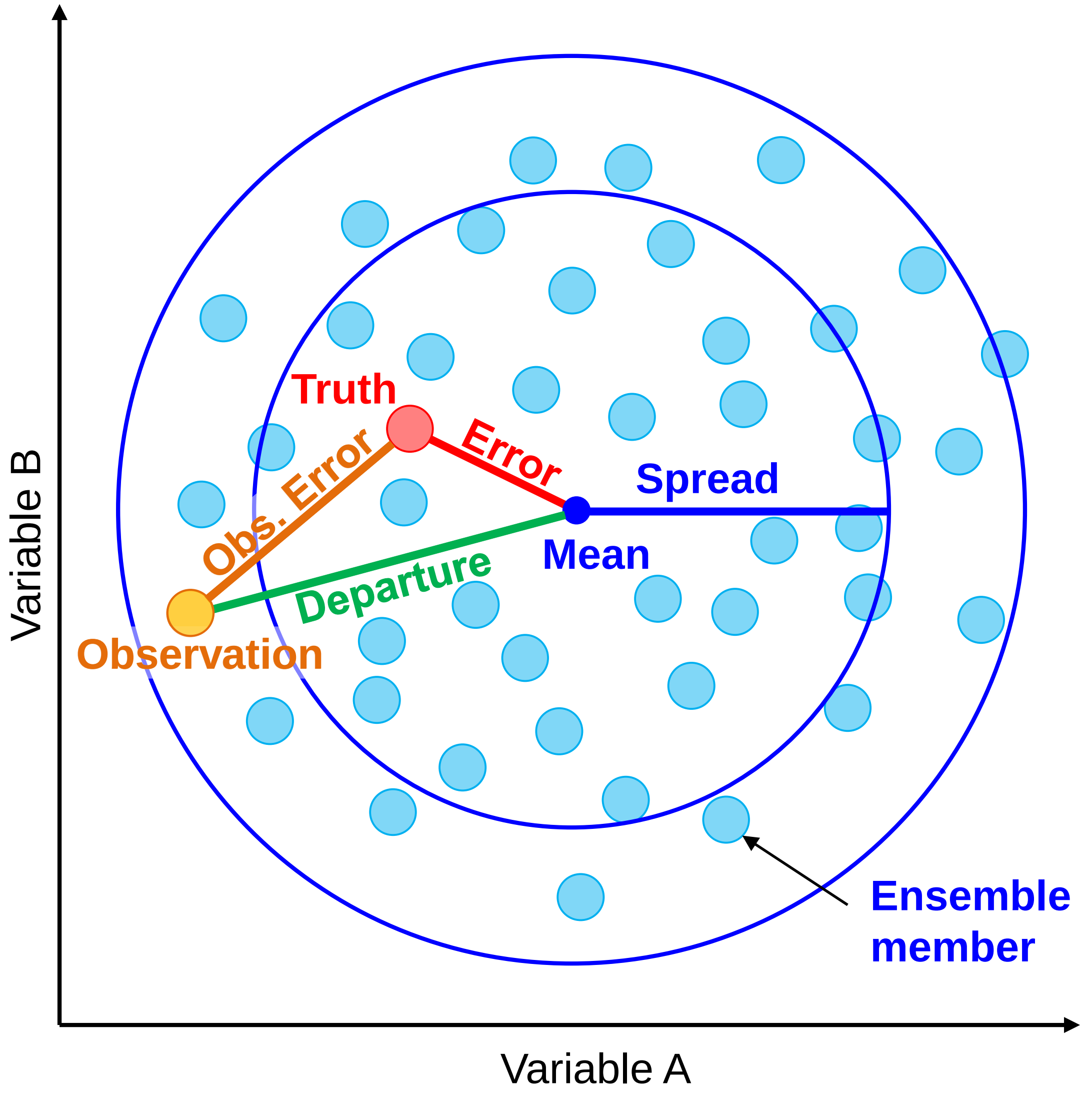
Cluster 3 (size 29)



Here, the main focus is on Cluster 3. This includes a strong Warm Conveyor Belt and enhanced uncertainty growth-rates

Based on K-means clustering of the synoptically-filtered and normalised fields (previously shown in the animation) in the region indicated for a single season (MAM 2017)

Reliability in ensemble data assimilation



Reliability $\Rightarrow$

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

(averaged over many forecast start dates)

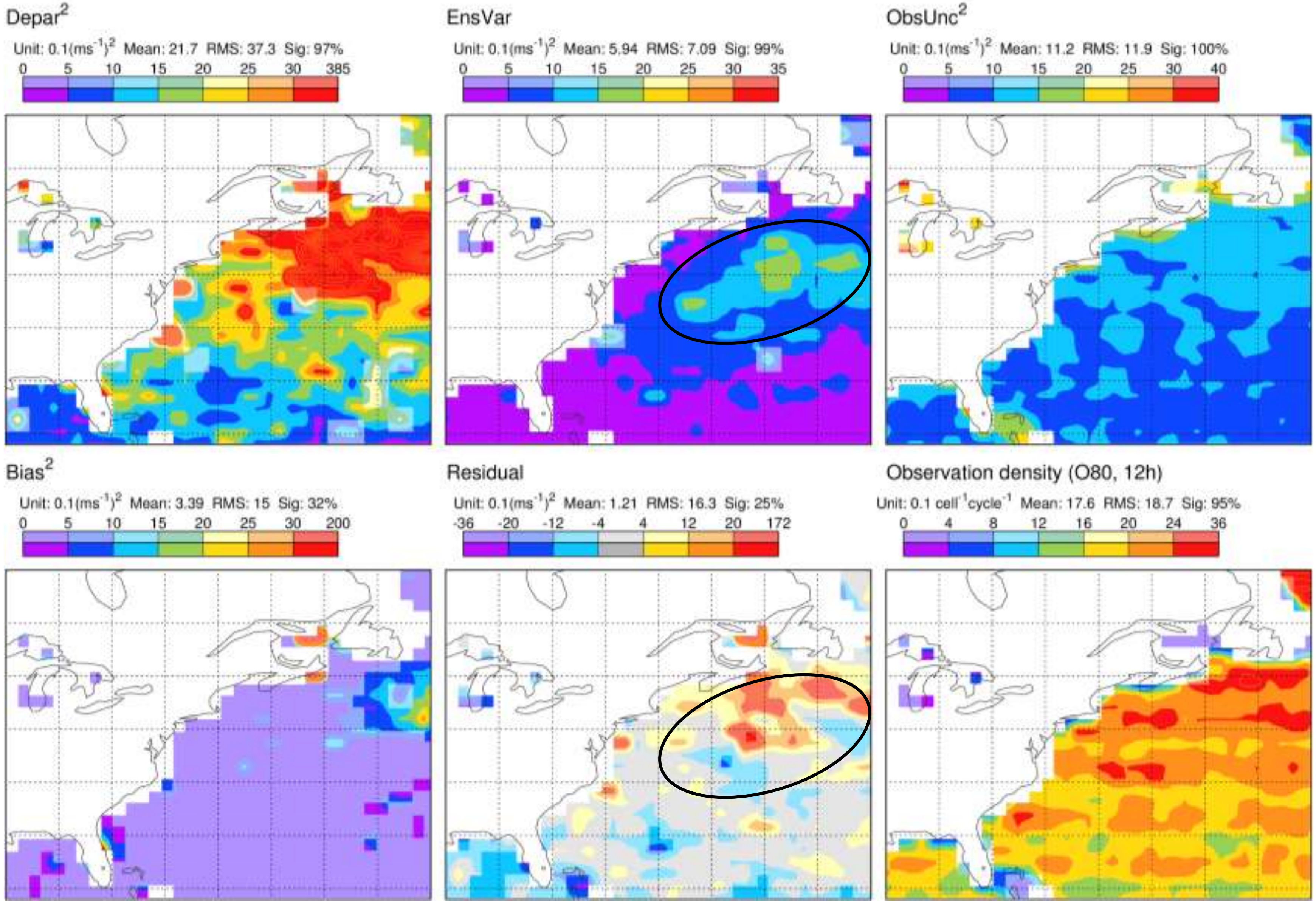
If we do not know the truth well-enough to calculate the error, use‡

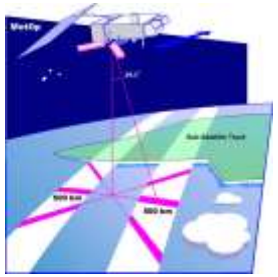
$$\overline{\mathbf{Departure}^2} = \overline{\mathbf{EnsVar}} + \overline{\mathbf{Obs.Unc}^2}$$

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





Large departures for this WCB case

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

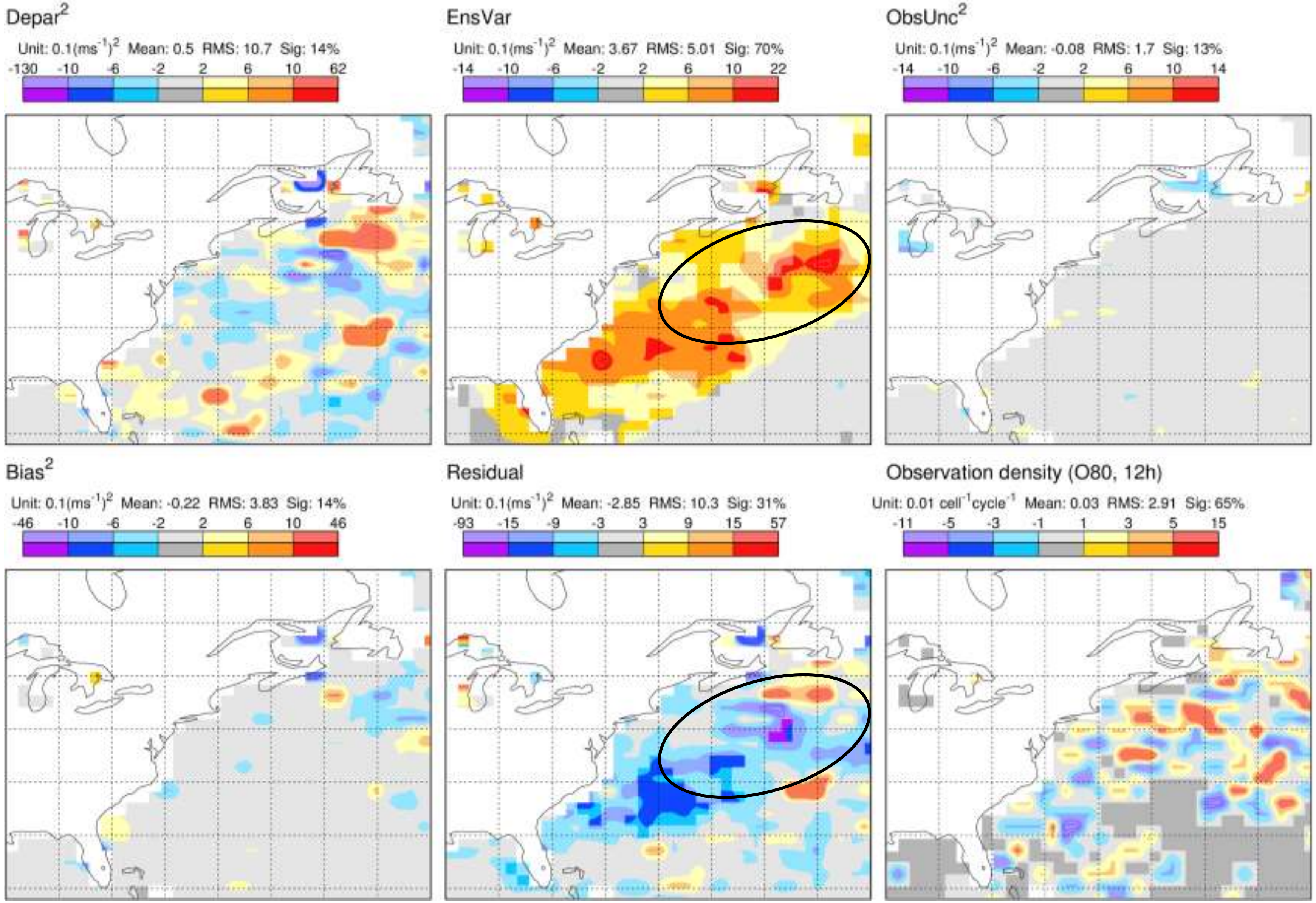
Hence EnsVar is under-represented

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 cluster 3 contributes 1/3 to overall unreliability

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

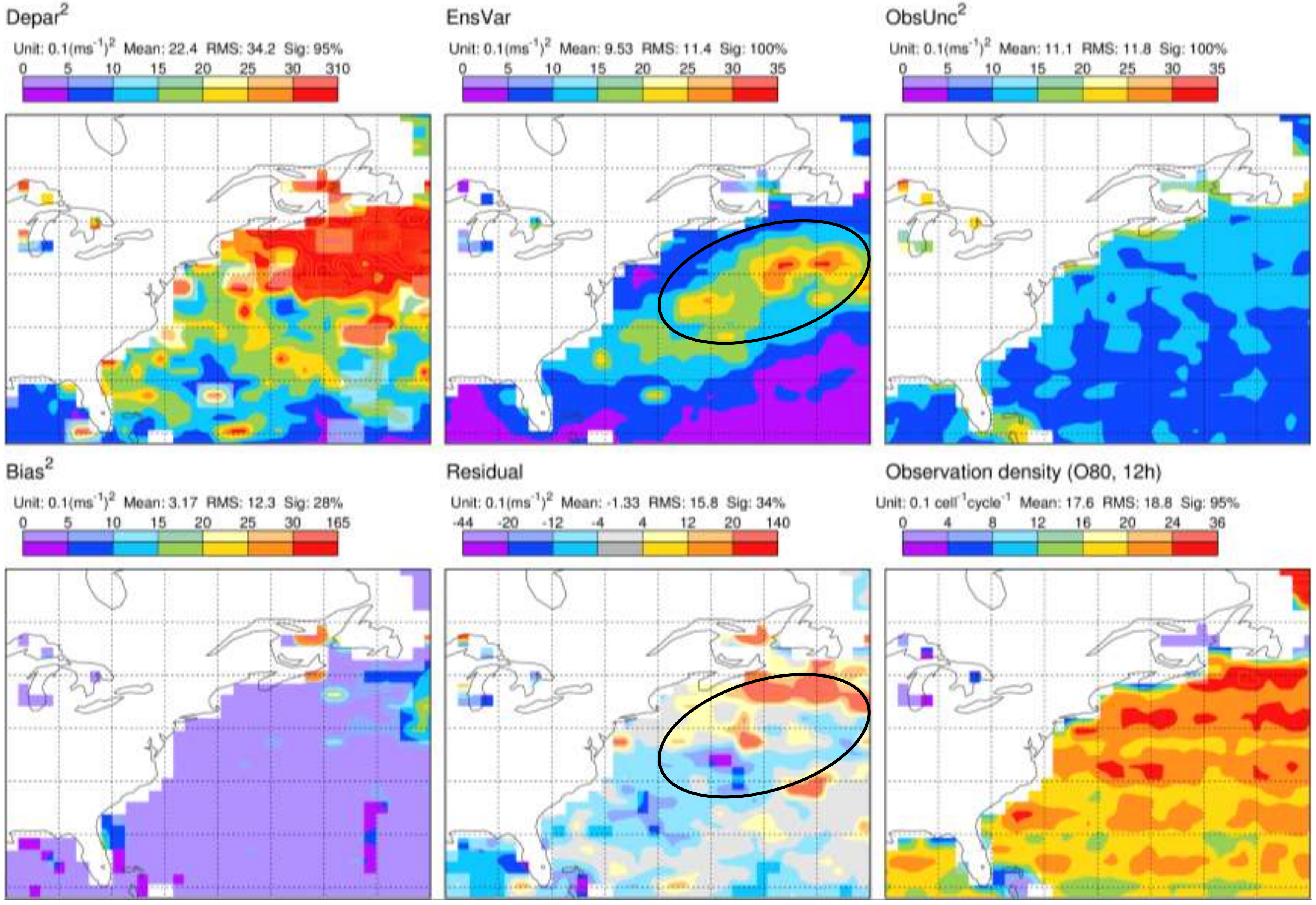


Model cycle 45r1 – cycle 43r1

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)

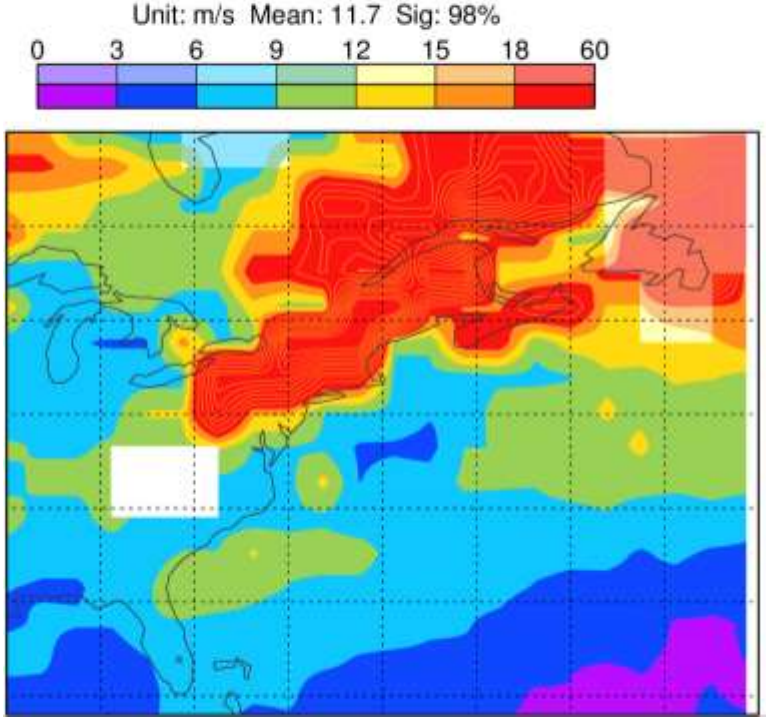


5% improved residual. Mixed signals for squared departures (mean 3% increase, RMS 8% decrease)

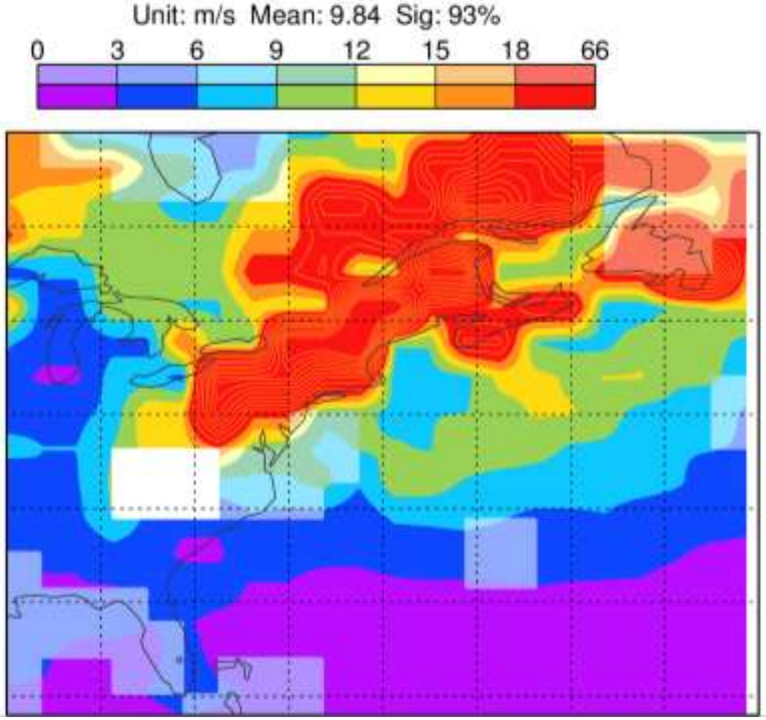
Future tuning (down) of observation error variance (ObsUnc^2) should act to remove the blue from the Residual, and could improve squared departures

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

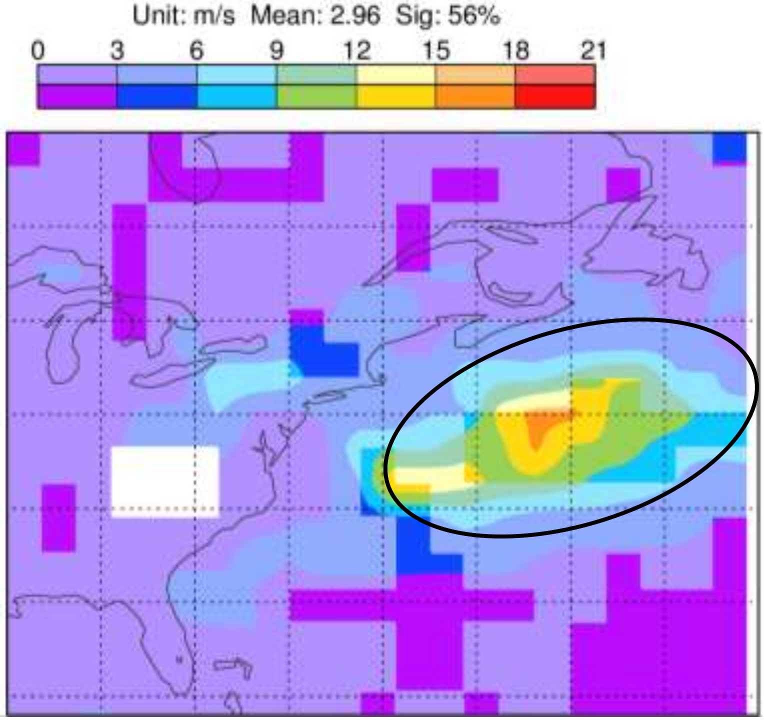
Dynamic tendency



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

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**

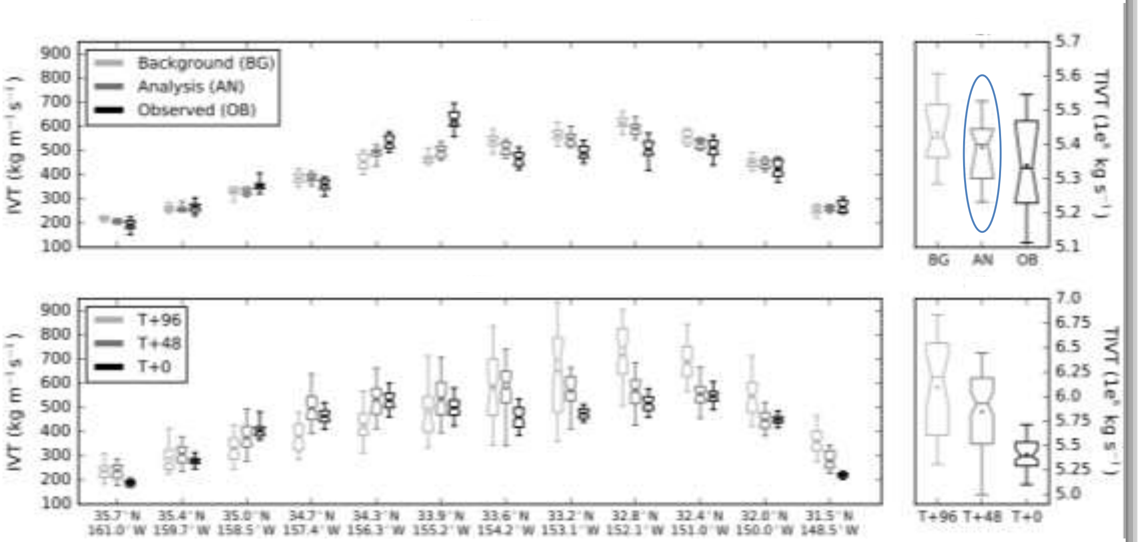
By analogy, for a small set of initial flow-types (clusters), reduce local short leadtime Bias^2 and Residual in:

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

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

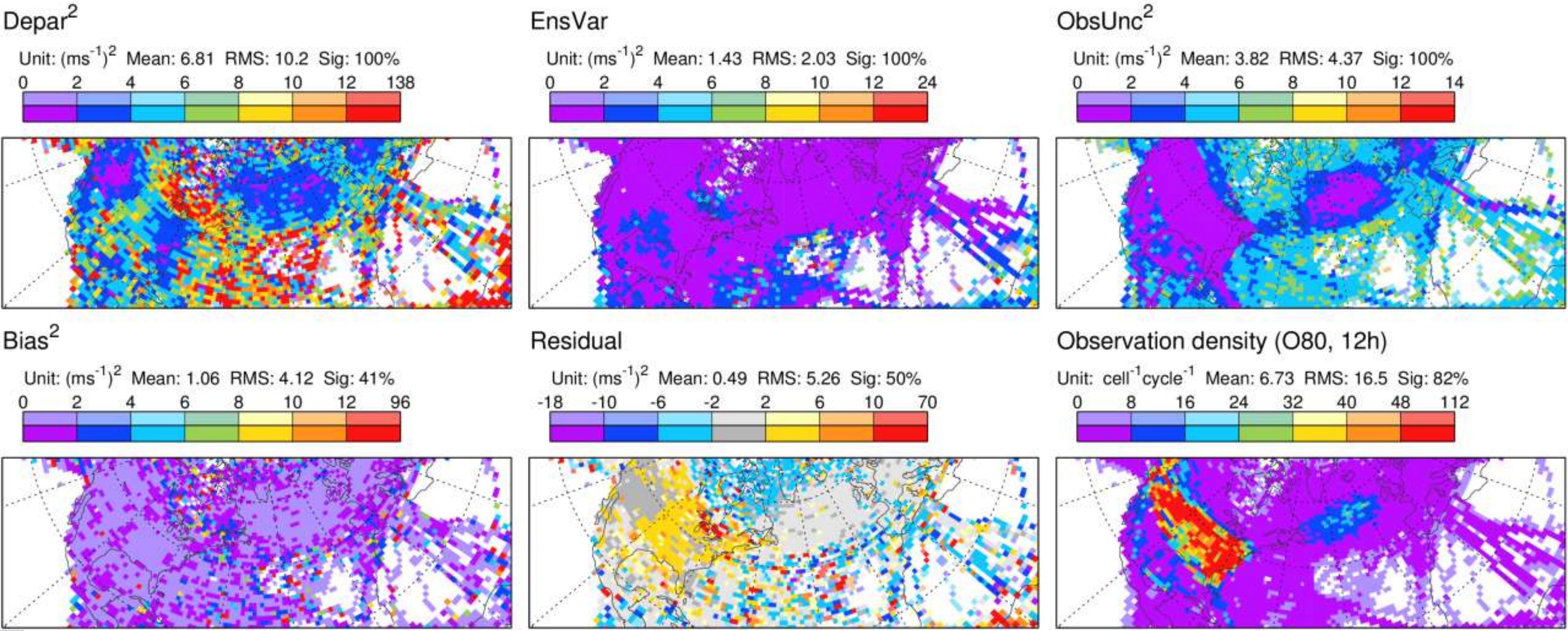


Better Obs $\Rightarrow$ Reduced analysis uncertainty $\Rightarrow o_k \approx p_k \rightarrow \in \{0,1\}$

Thank you

Extra slides

Eda Observations. AIREP ~u200 (185-215hPa) for CL5b_2_2017030512-2017052200. Deep colours = 5% sig.
Aircraft observations (BRPA)



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

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

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

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

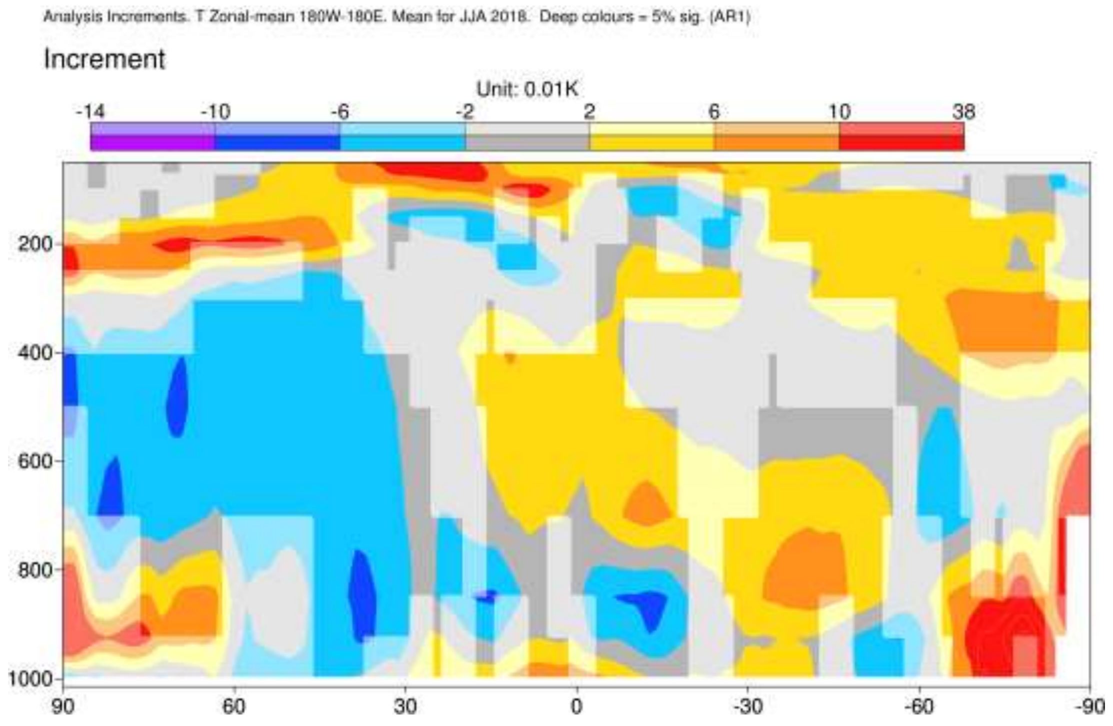
*Improve refinement through use of more observations, or extraction of more information from the observations

Implications for system development

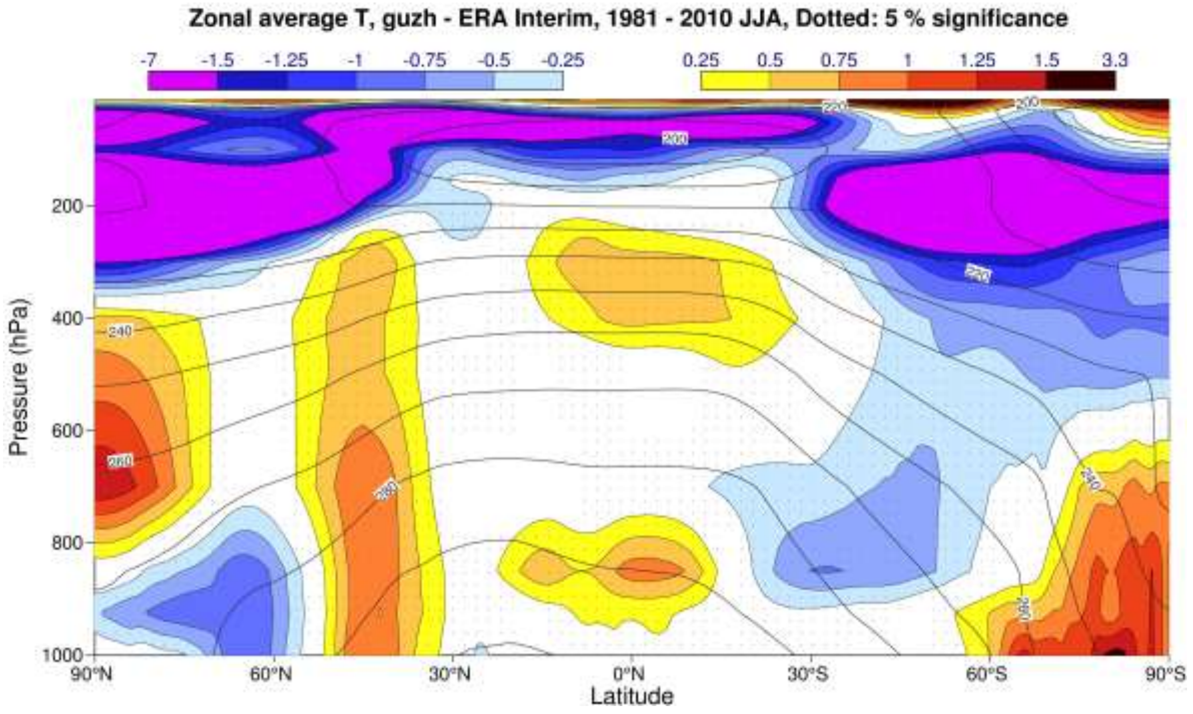
Development of physics/stochastic physics at least partially within the context of DA/EDA

A season of forecasts is just about enough

For ECMWF: Potential to decrease scatterometer wind error estimates and more emphasis of stochastic physics on convection



45r1 implemented 5 June 2018



45r1 – ERA Interim

Ensemble spread and error

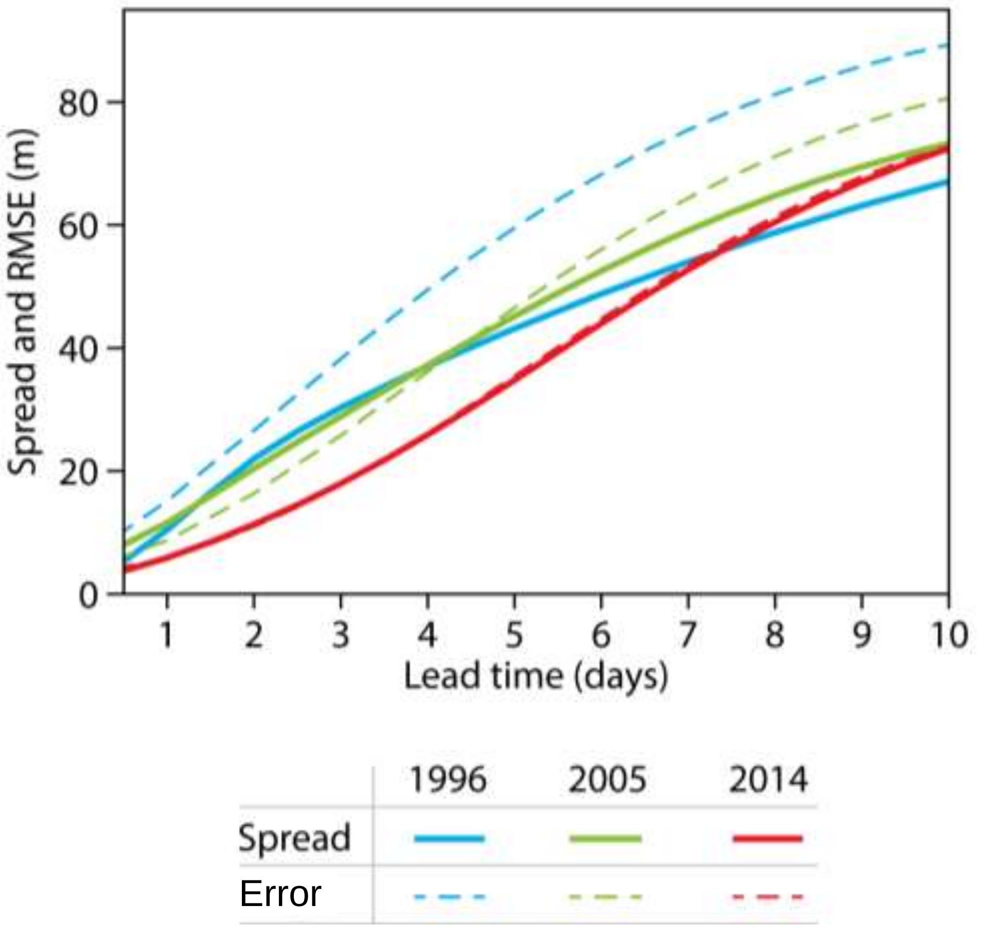
Z500

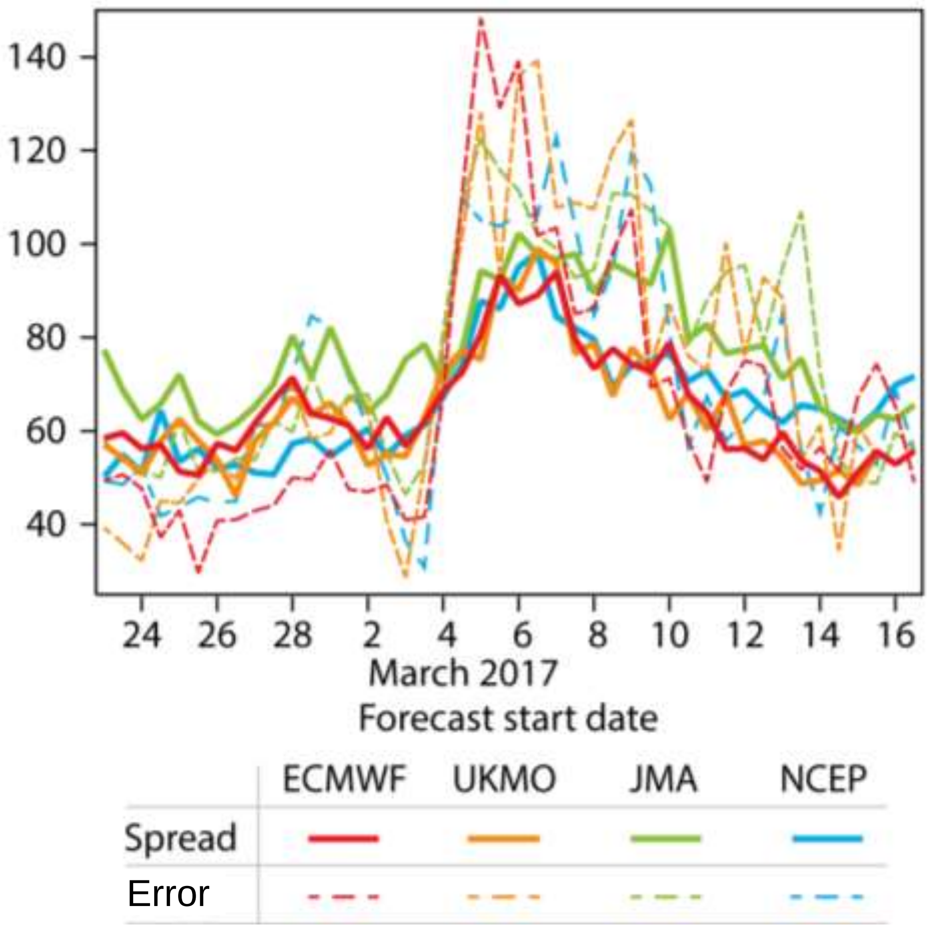
Rodwell et al. 2018, BAMS

Annual means N.Hem. (ECMWF)

Timeseries for Europe at D+6 (TIGGE)

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



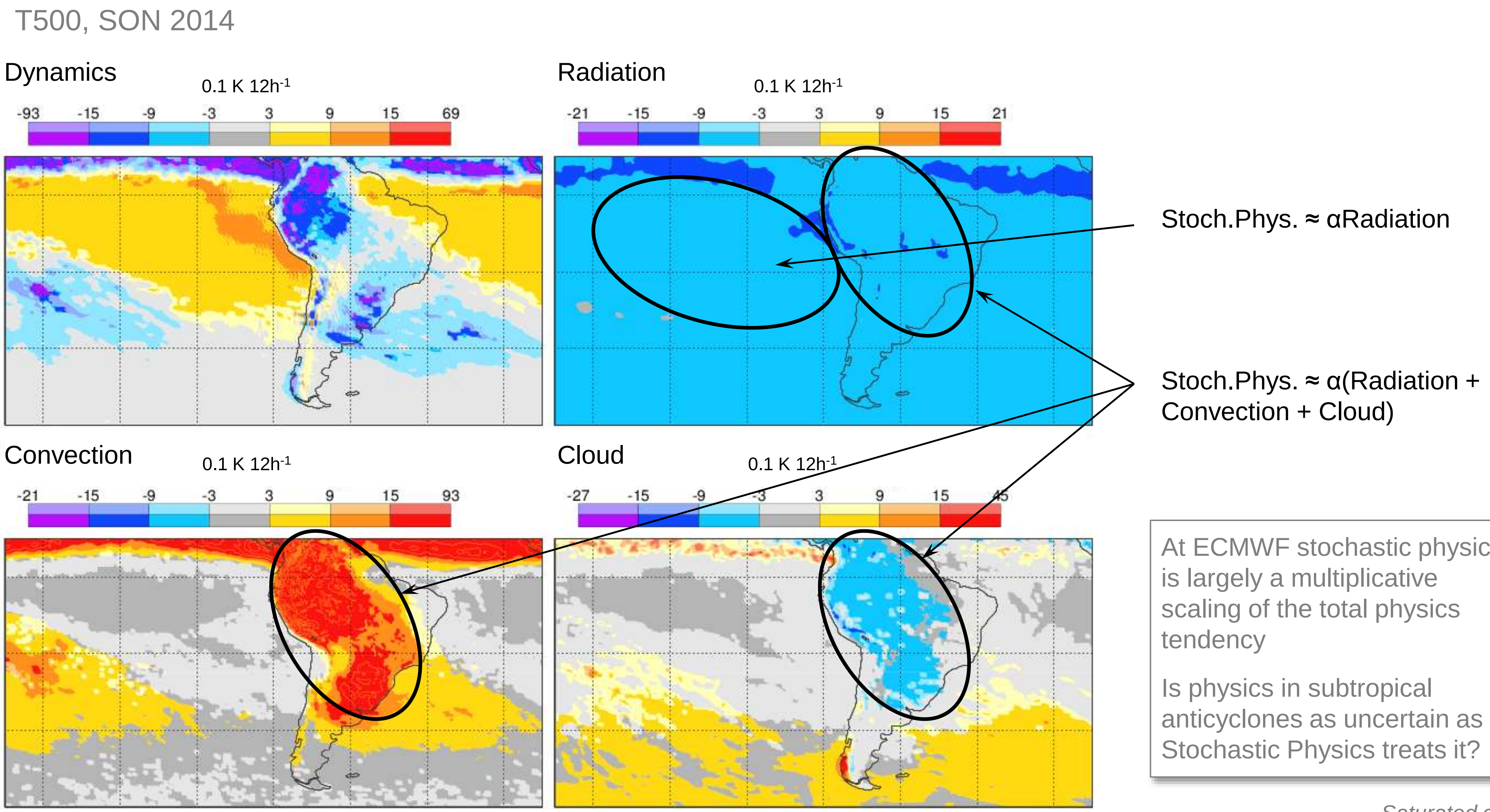


…but we make ensemble forecasts to represent the day-to-day variations in predictability and uncertainty. What causes this uncertainty? How can we evaluate it in our forecasts?

500 hPa geopotential height (Z500). "Error" is RMS of ensemble-mean error

Spread = ensemble standard deviation (scaled to take account of finite ensemble size)

Initial tendencies from control forecast: SON 2014



At ECMWF stochastic physics is largely a multiplicative scaling of the total physics tendency

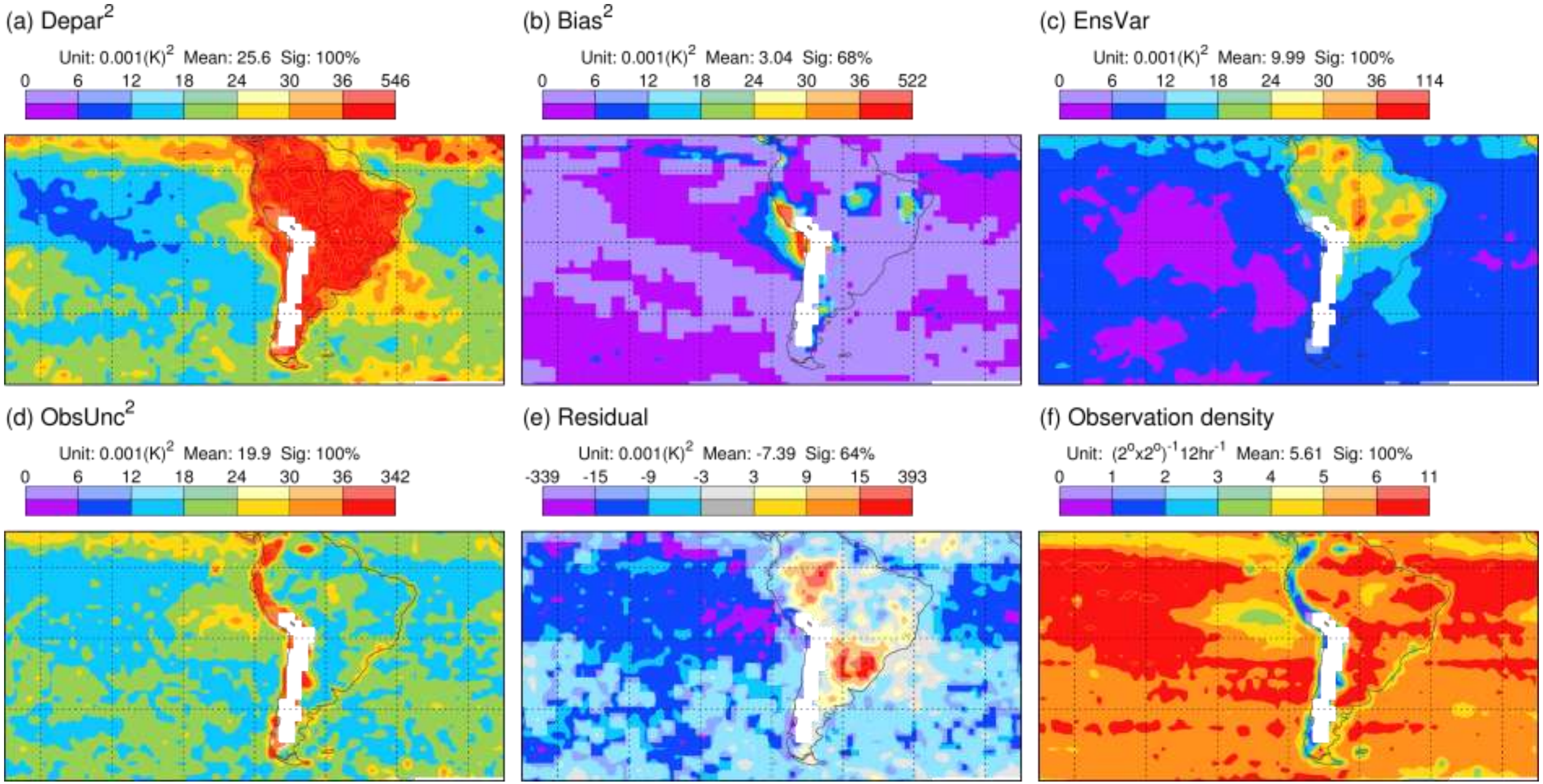
Is physics in subtropical anticyclones as uncertain as Stochastic Physics treats it?

Saturated colours: 5% sig.

EDA reliability for AMSUA satellite observations of mid-tropospheric temperature

Rodwell et al (2015). This experiment by Simon Lang

One explanation for the Residual was that the ensemble variance is too large in the clear-sky subtropical anticyclone regions and too small in the convective regions over South America



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

AMSUA microwave channel 5.12 August – 16 November 2011

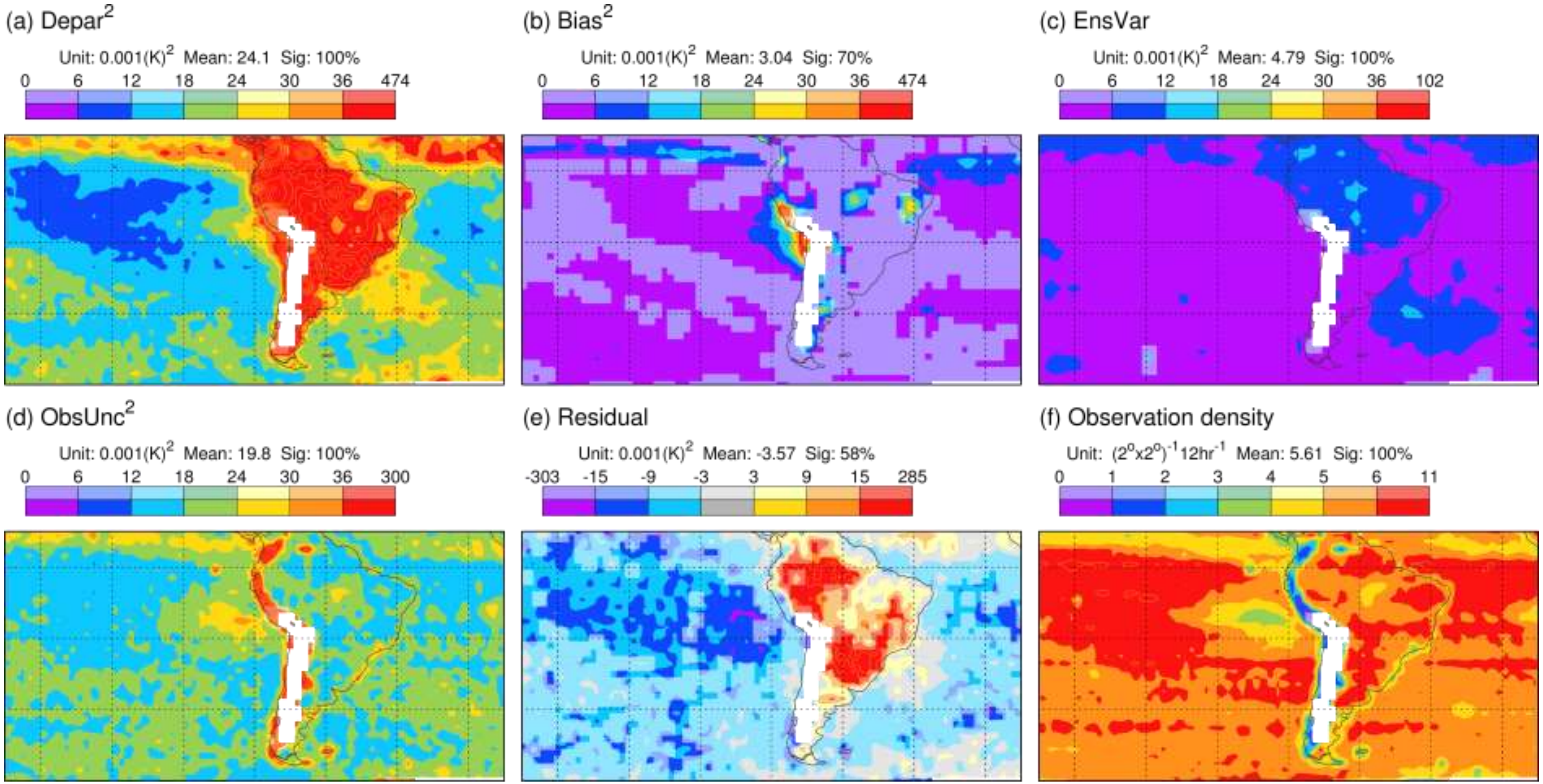
EDA reliability for AMSUA observations of mid-tropospheric T (no stochastic physics)

Rodwell et al (2015). This experiment by Simon Lang

Turning off stochastic physics (SPPT) reduces the residual in the clear-sky regions and increases it in the convection regions

Operational model requires less stochastic physics in clear-sky situations and more in convective situations?

These results were possibly a motivation for current system developments



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

AMSUA microwave channel 5.12August – 16 November 2011

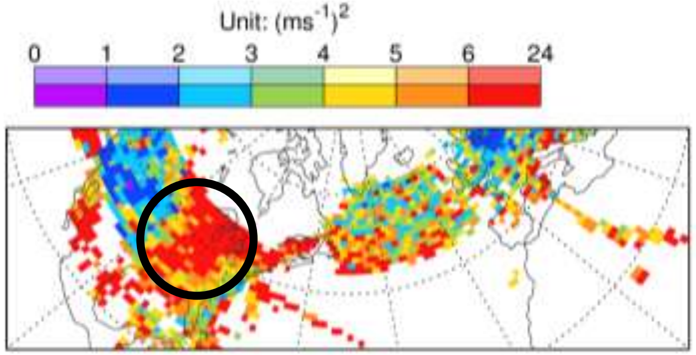
EDA reliability in u200 against aircraft observations in “Rocky trough/CAPE” situations

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

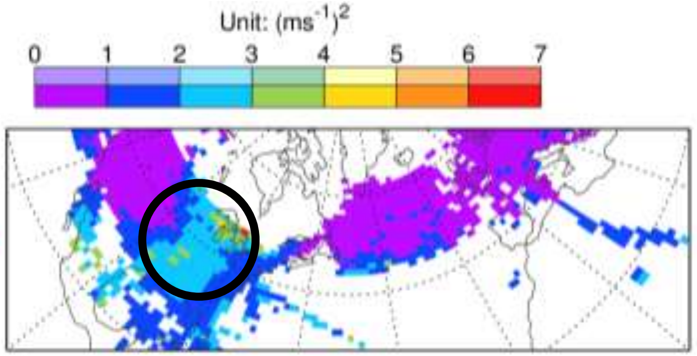
(Basically the error-spread relationship taking into account uncertainty in our knowledge of the truth)

Rodwell et al. 2018, BAMS

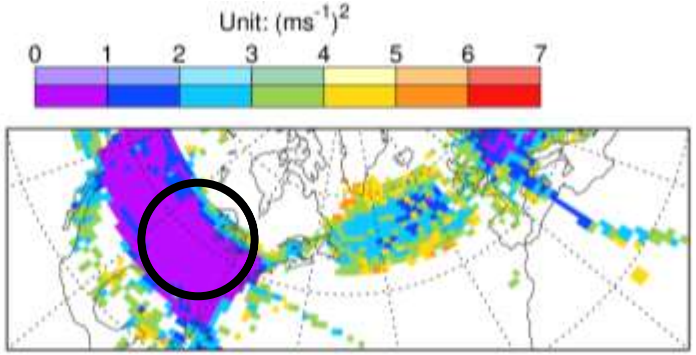
Departure^2



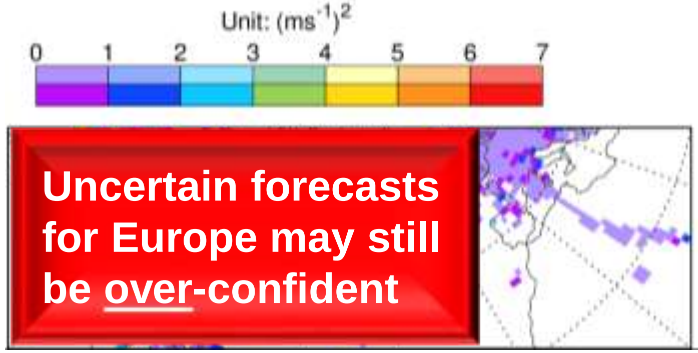
Background variance



Estimated observation error variance

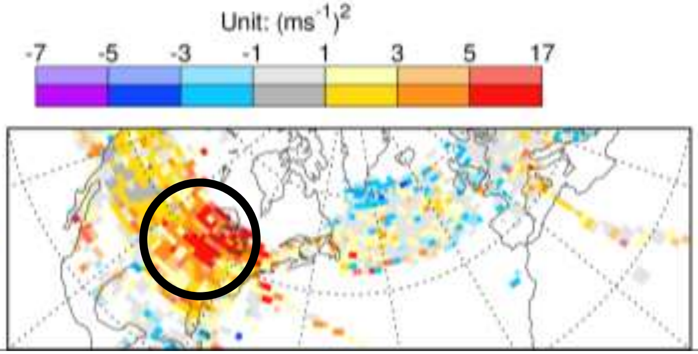


Bias^2

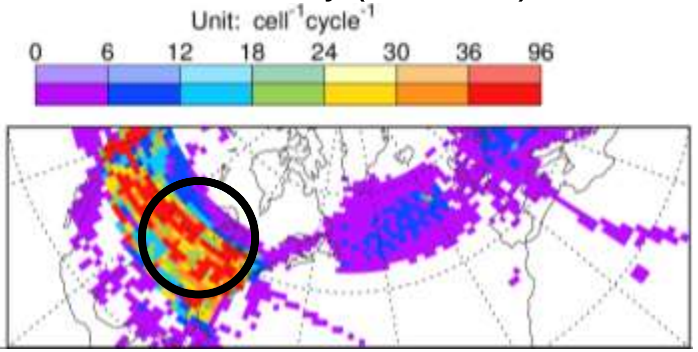


Uncertain forecasts for Europe may still be <u>over</u>-confident

Residual



Observation density (O80, 12h)



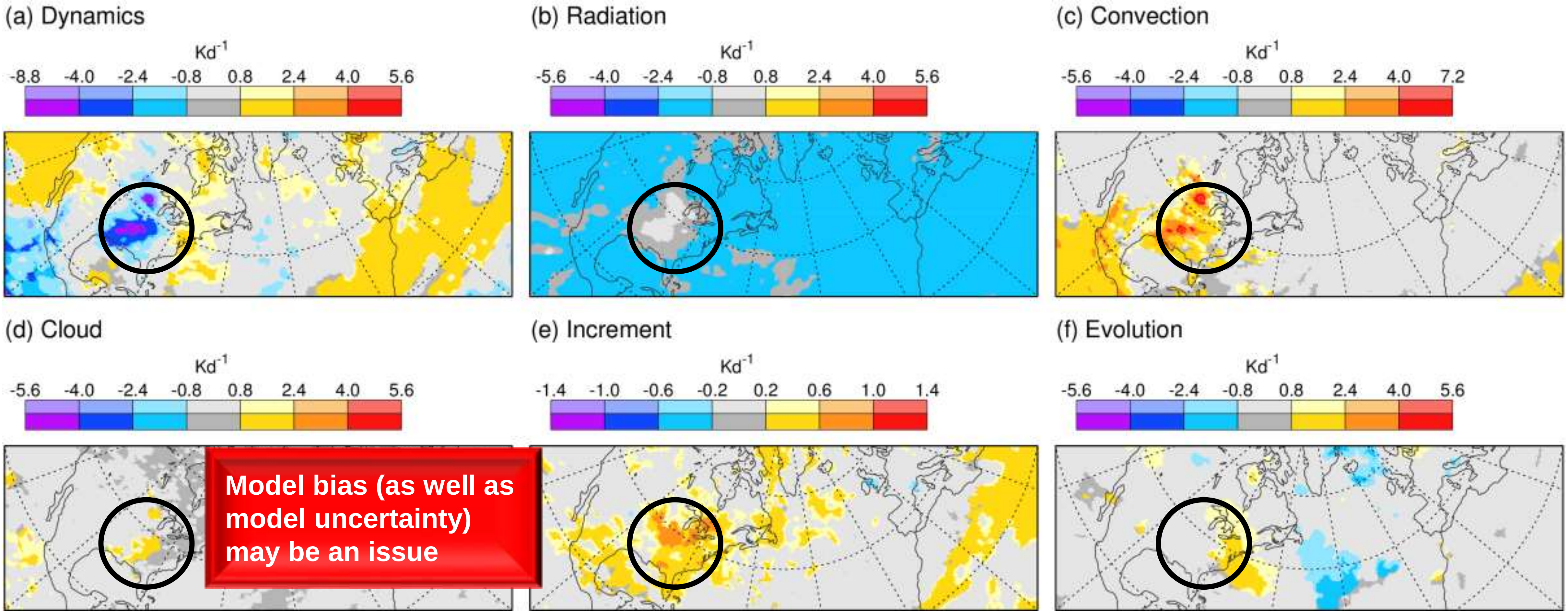
54 cases, 12h window

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



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”

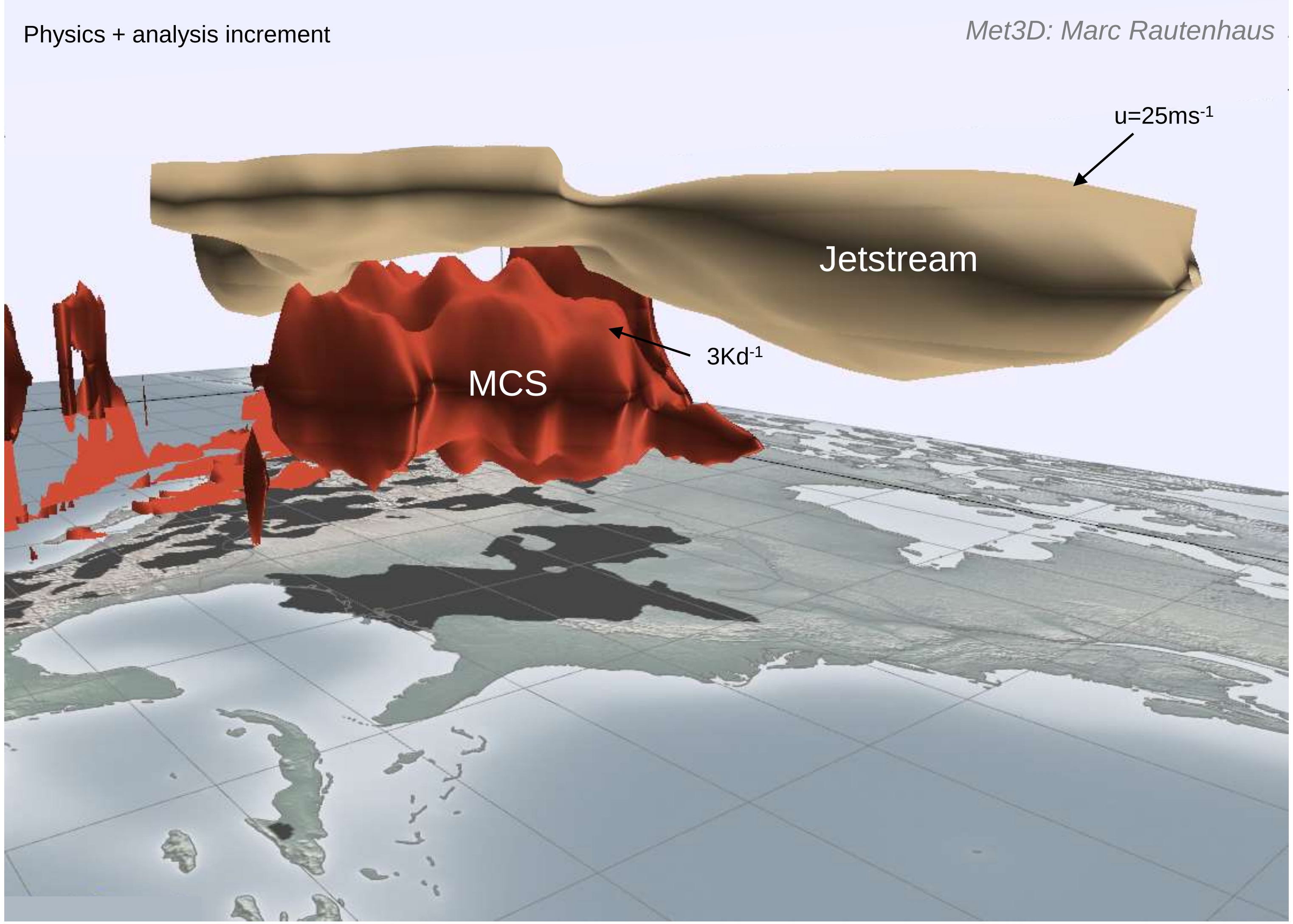
54 cases

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

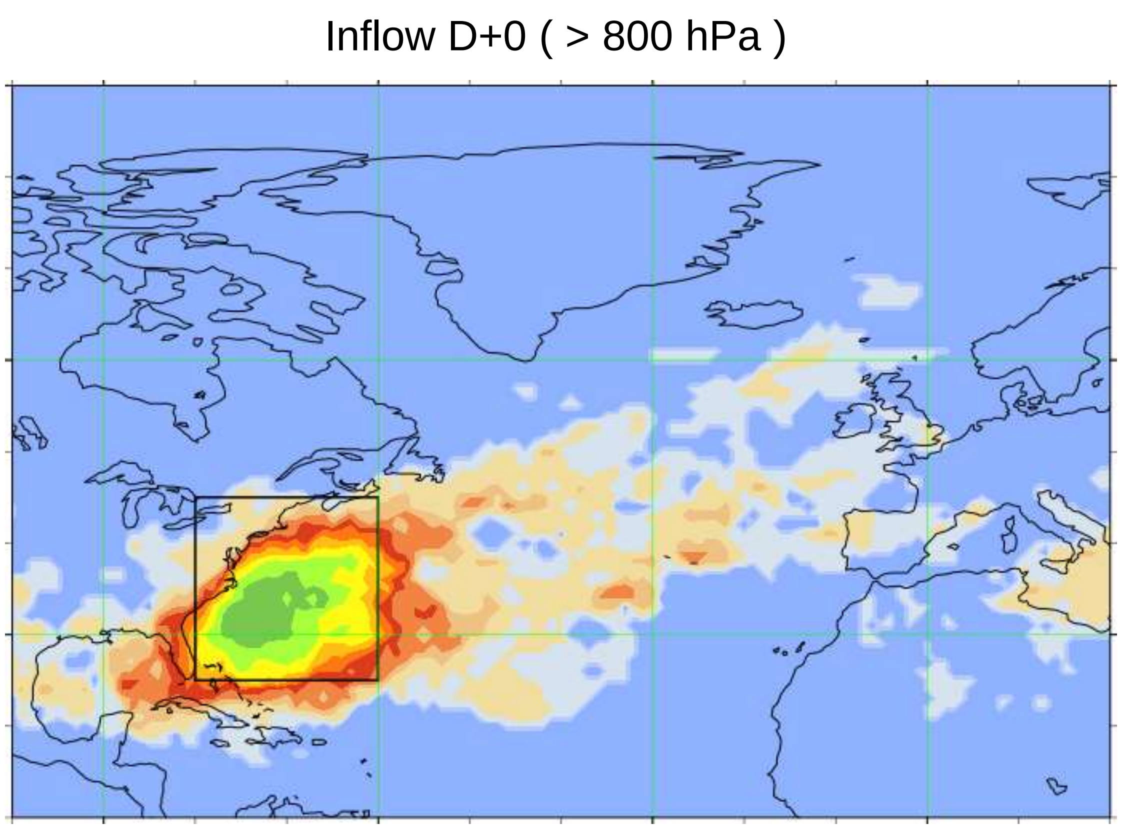
If we hit the string at the wrong time, the wave will arrive over Europe at the wrong time

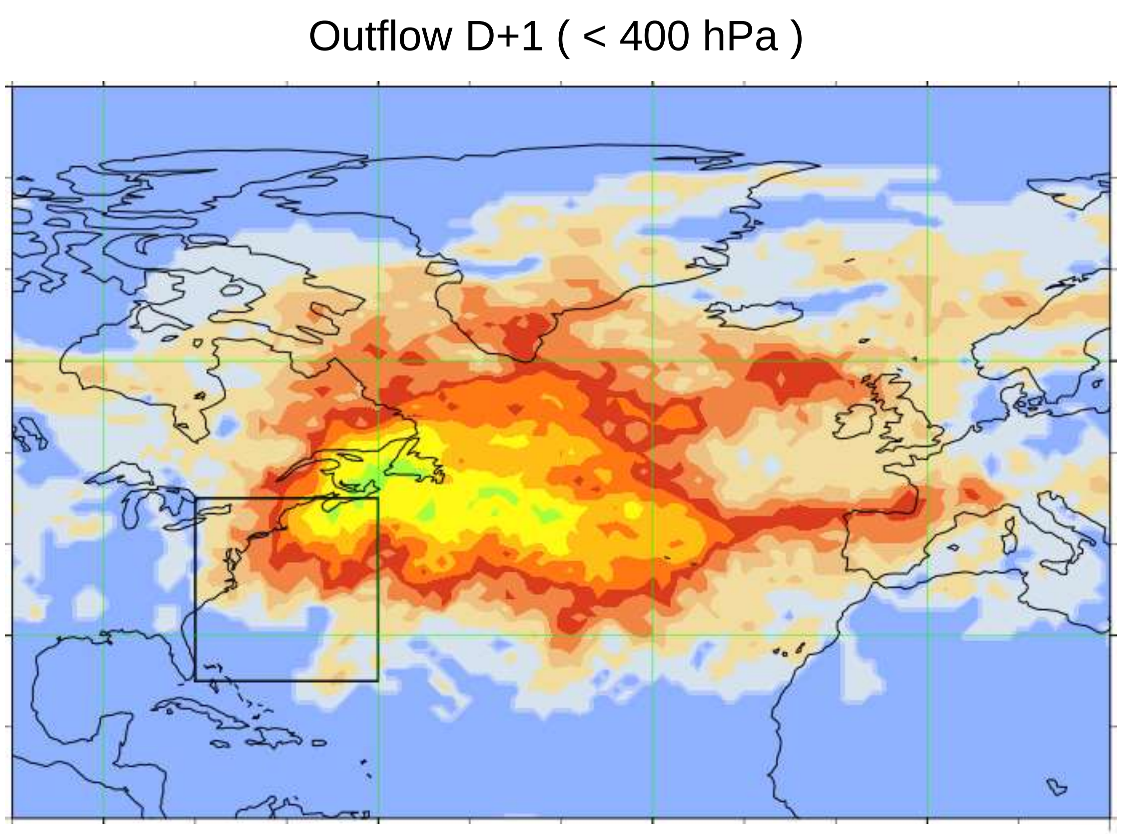
We do not know when to press the key (mesoscale convection itself involves chaotic uncertainty)

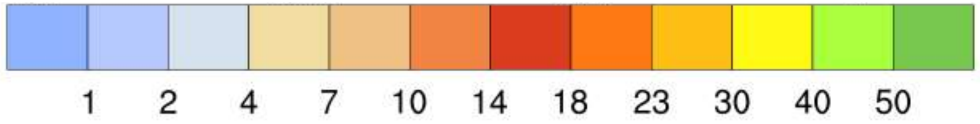
What we want is that the ensemble members generate such convection with the “right” uncertainty



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



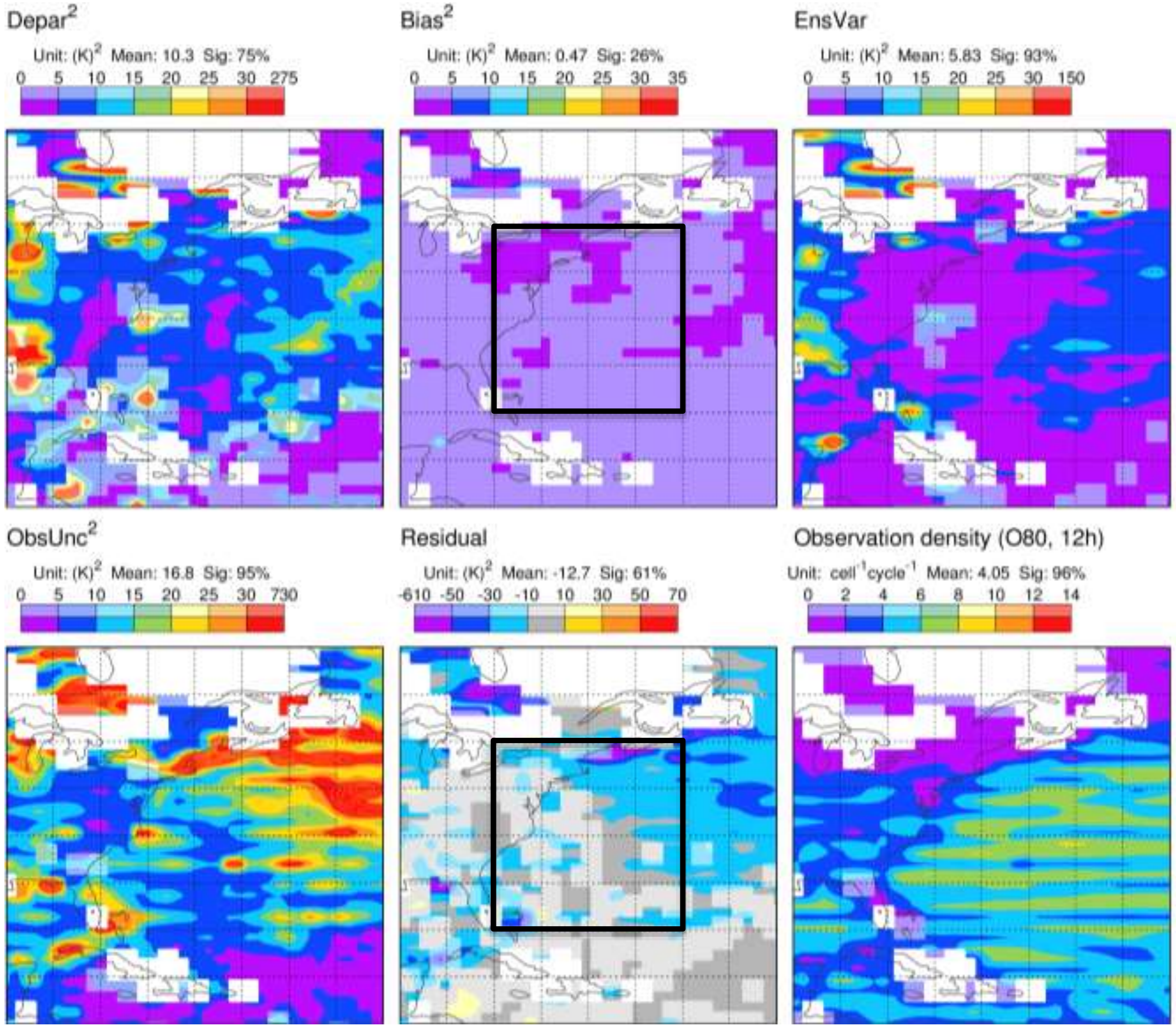




From Heini Werni. Based on trajectories ascending by more than 600 hPa in 2d

EDA variance assessment with MHS “all sky” mid-tropospheric humidity: Non-WCB

87 cases



Bias and residual are not significant in absence of WCBs ✓

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

Microwave channel 5

EDA variance assessment with MHS "all sky" mid-tropospheric humidity: WCB events

50 cases

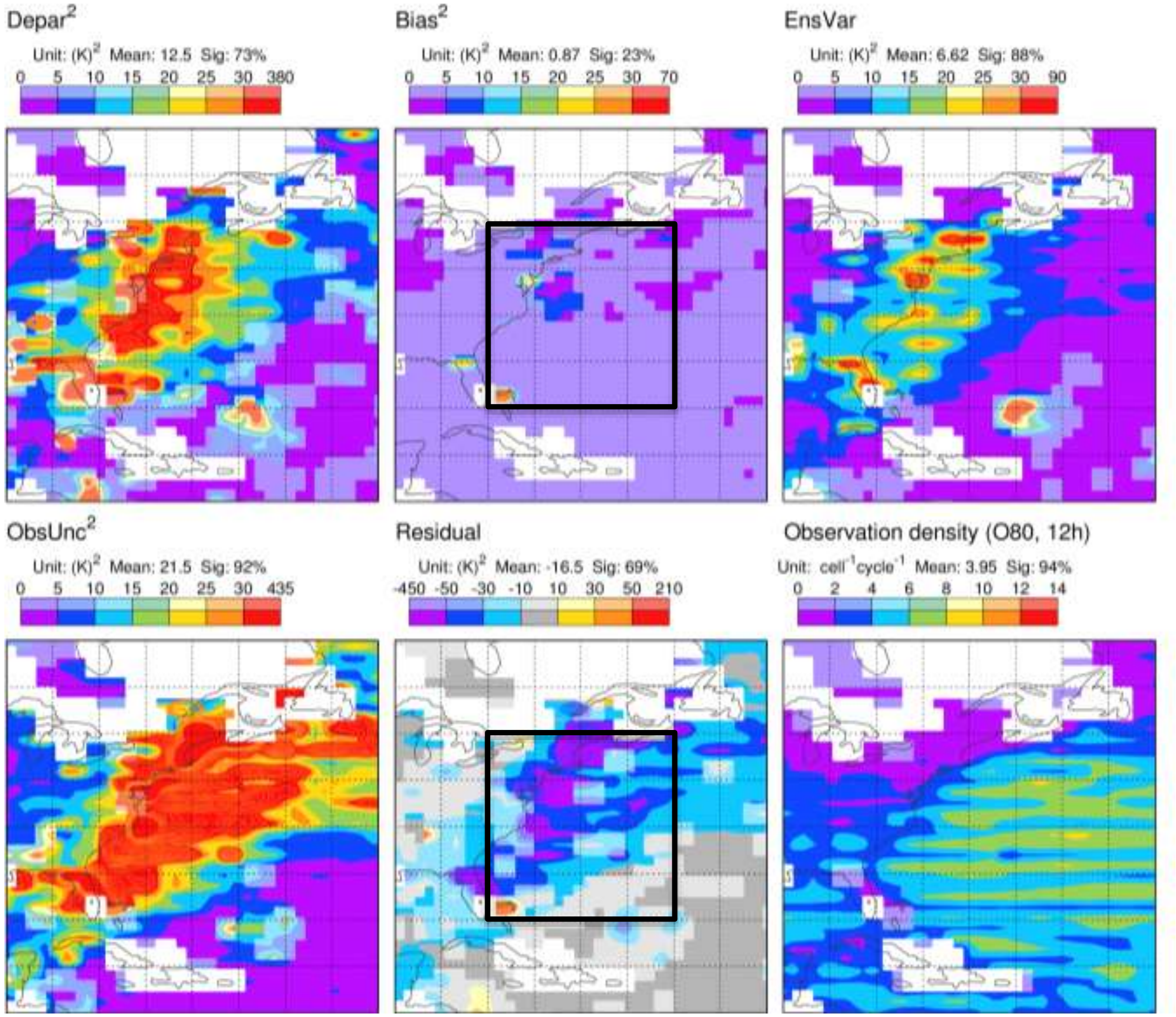
Increased Depar^2 and EnsVar in WCB situations

Negative residual largely due to large ObsUnc^2 (larger than the departures) in cloudy regions

No simple fix here:

- Sometimes ObsUnc^2 inflated as surrogate for spatial and inter-channel observation error correlations
- Good model representation of (*e.g.*) planetary boundary layer depth important for assimilation of observations with deep weighting functions

Diagnostic highlights potential and areas where work focus could help



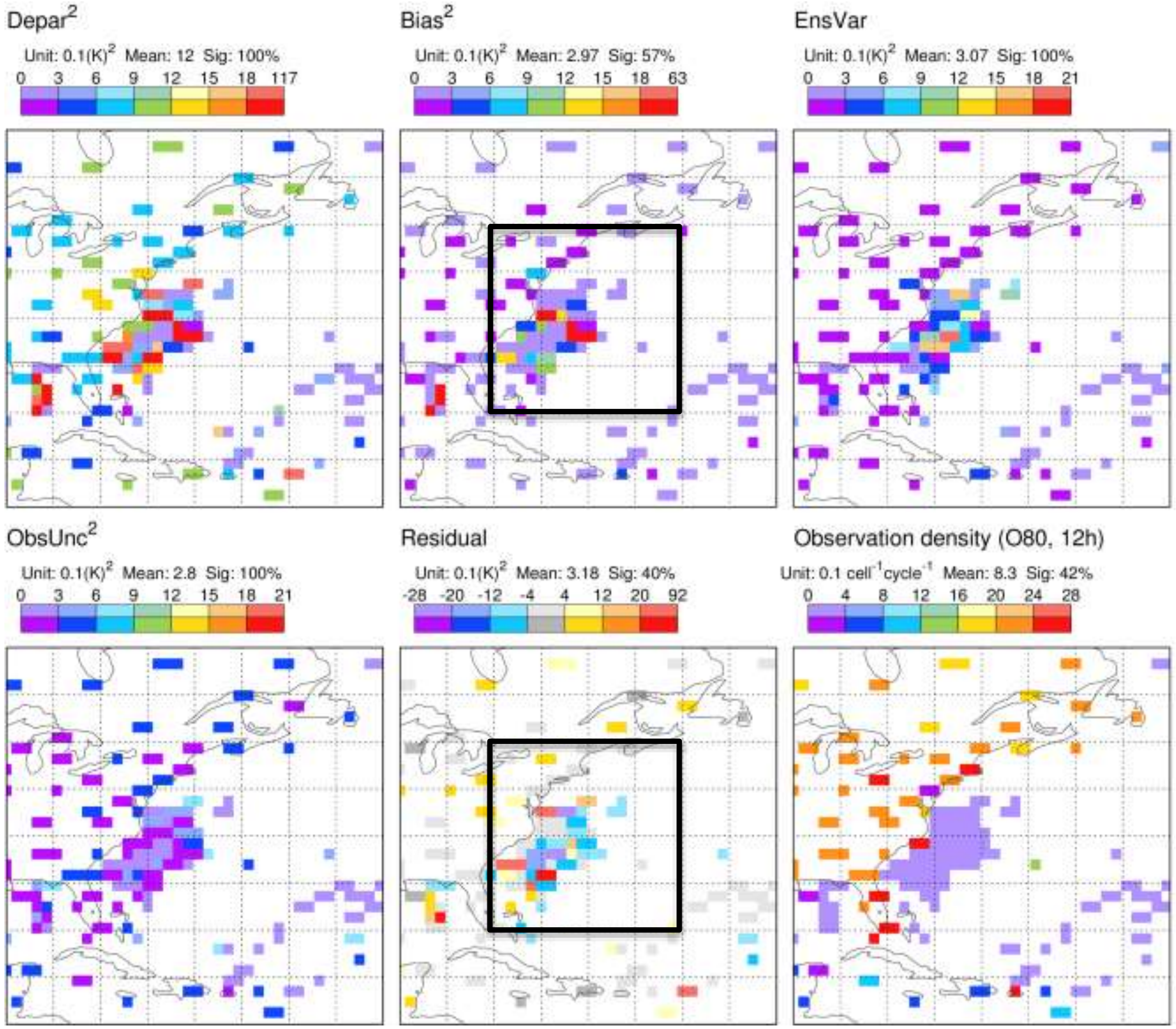
$$\text{Depar}^2 = \text{Bias}^2 + \text{EnsVar} + \text{ObsUnc}^2 + \text{Residual}$$

Microwave channel 5

EDA variance assessment with TEMP T700: WCB events

50 cases

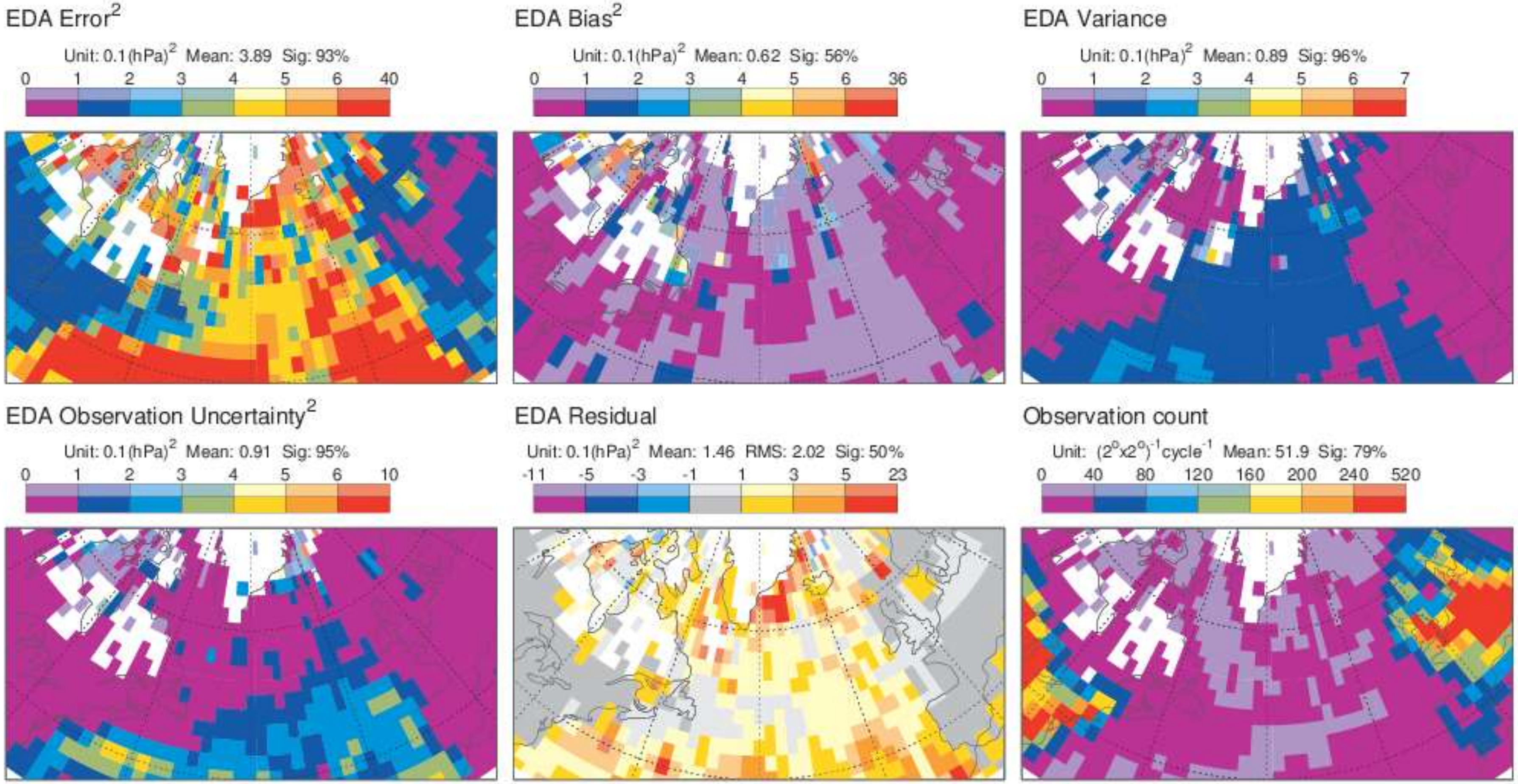
More tricky to see systematic issues as data is more sparse, etc



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

EDA reliability budget: p_* (land, ship, buoy)

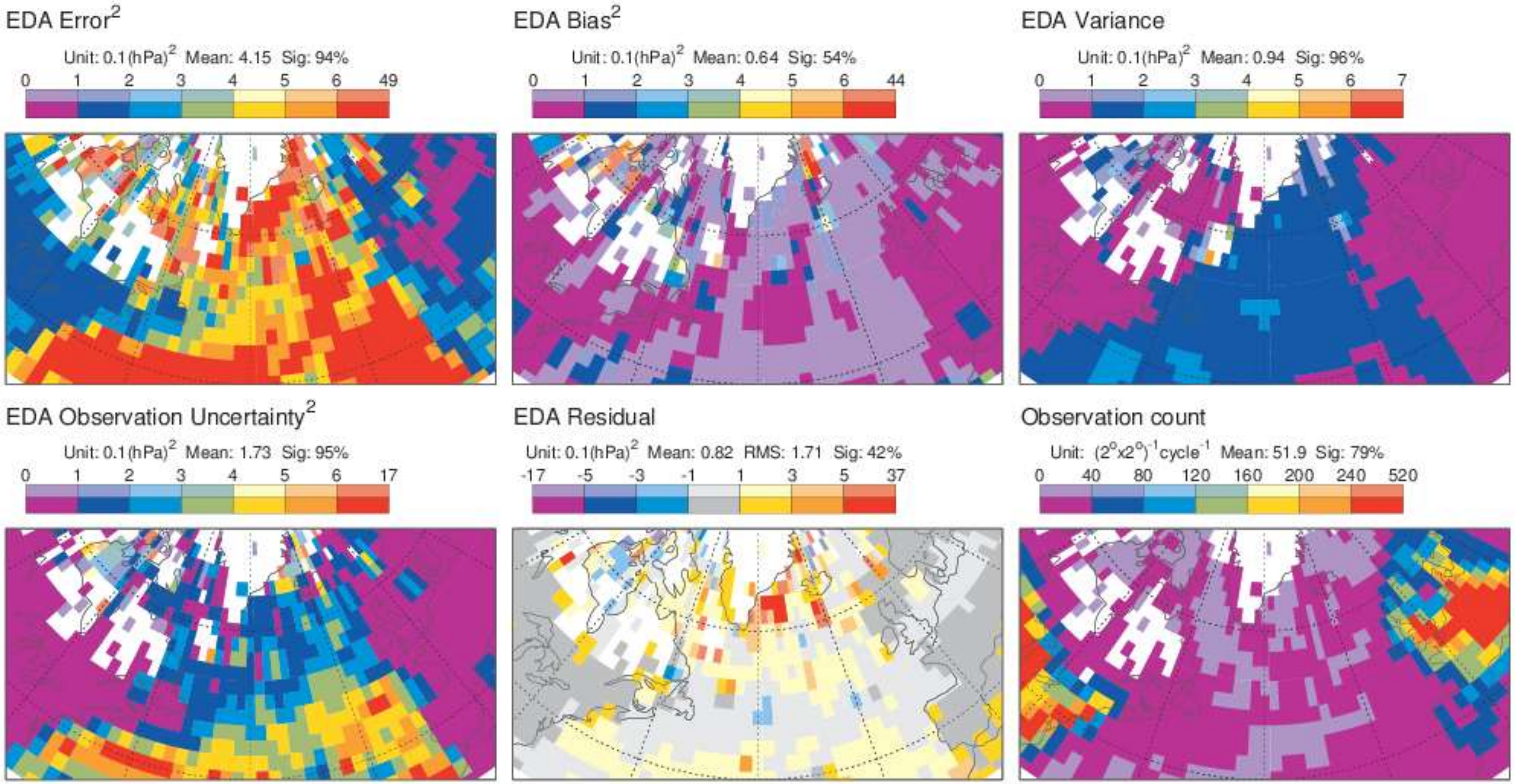
Relative to conventional observations of surface pressure



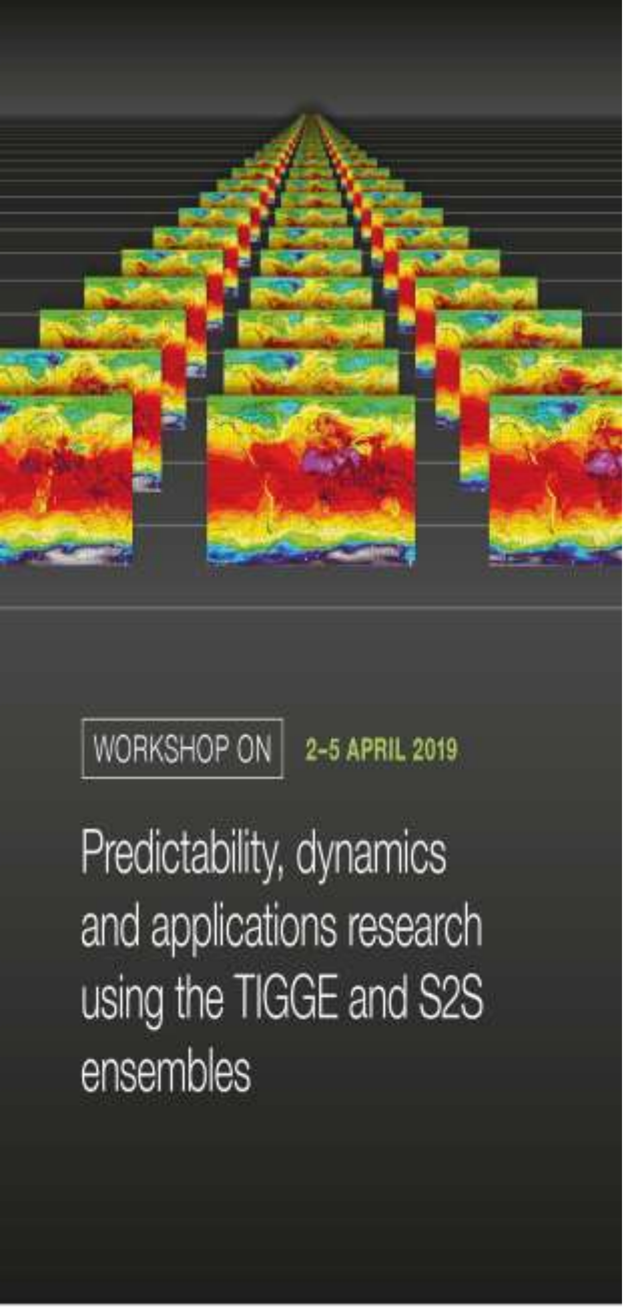
- Reference experiment (3 members, 20140901-20141130)
- Marine observation errors not large enough?

EDA reliability budget: Larger marine Obs Errors

Relative to conventional observations of surface pressure



- Larger marine observation errors consistent with better reliability
- Ultimately reliable ENS initialisation requires joint improvements in Obs Errors and Stochastic Physics



Workshop on Predictability, dynamics and applications research using the TIGGE and S2S ensembles

ECMWF, Reading, 2-5 April 2019

To bring together the users and data providers of the TIGGE and S2S databases around the following themes:

Thank you

- Database technical development
- Predictability and Dynamics
- Prediction and Verification
- Multi Model approaches
- Application Studies

Please take a flyer

Registration and abstract submission opens on 4 June and closes on 30 November 2018

https://www.ecmwf.int/en/learning/workshops/workshop-predictability-dynamics-and-applications-research-using-tigge-and-s2s-ensembles