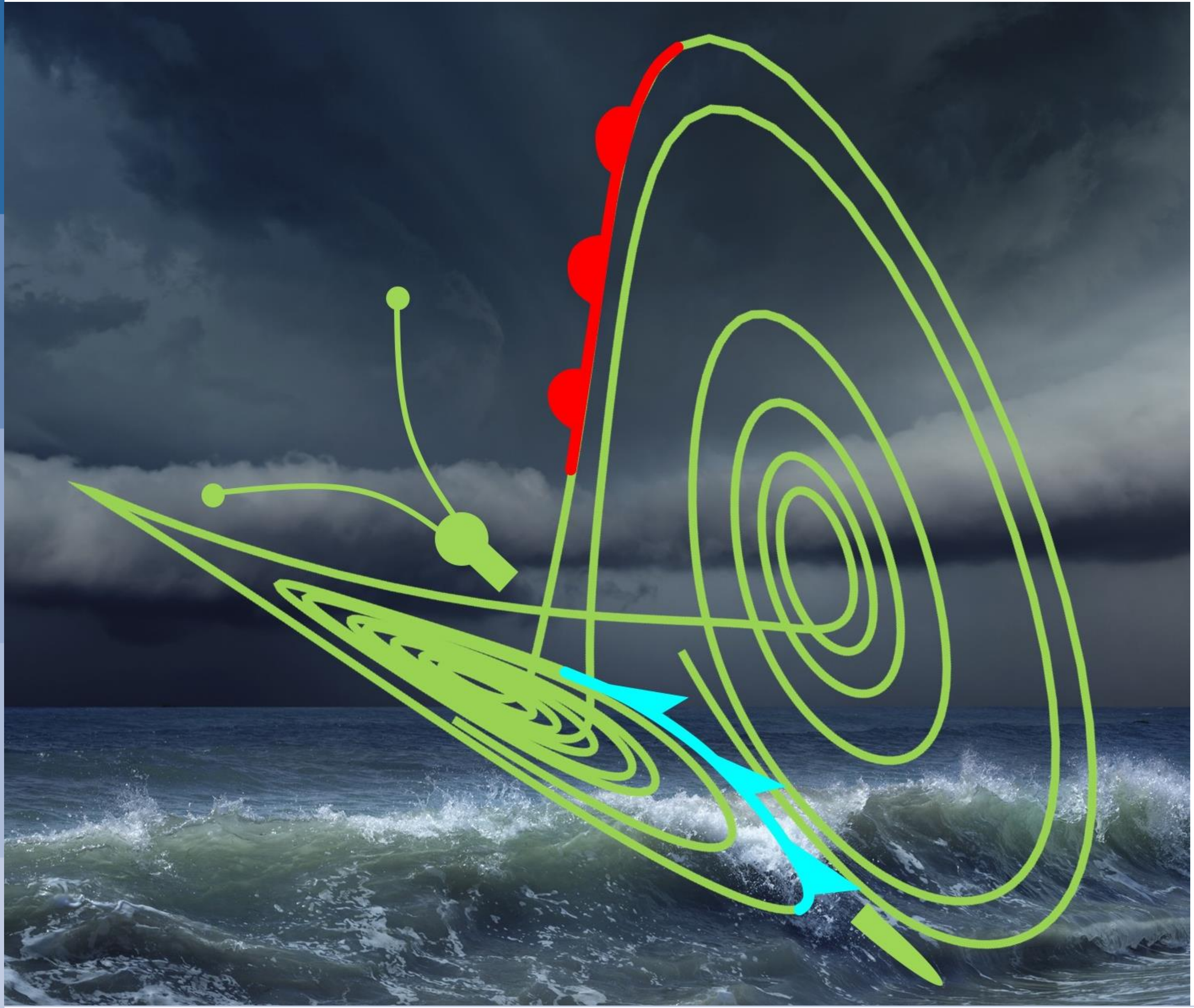


Diagnostics of ensemble data assimilation and ensemble forecasts

Mark Rodwell

Acknowledgements: David Richardson, Dave Parsons, Heini Wernli, Simon Lang, Linus Magnusson, Elias Hólm, Laura Ferranti

Ensemble prediction: past, present and future

14 September 2017, ECMWF

  mark.rodwell@ecmwf.int

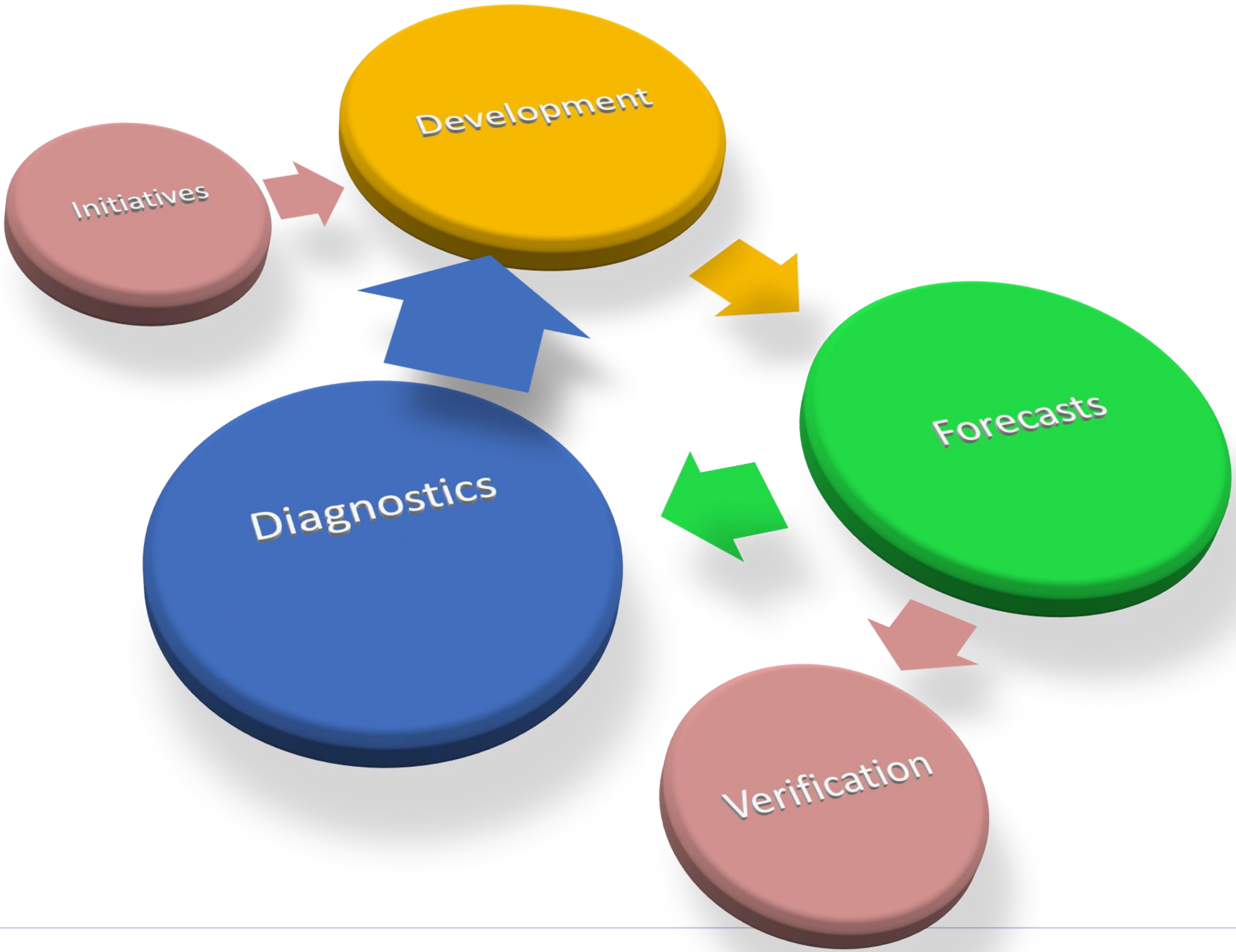
Outline

- The aim of operational diagnostics
- The issues
- Diagnosis of flow-dependent reliability
- Possible useful framework for diagnosis of ensemble forecasting systems

Outline

- The aim of operational diagnostics
- The issues
- Diagnosis of flow-dependent reliability
- Possible useful framework for diagnosis of ensemble forecasting systems

The role of Diagnostics in the development process

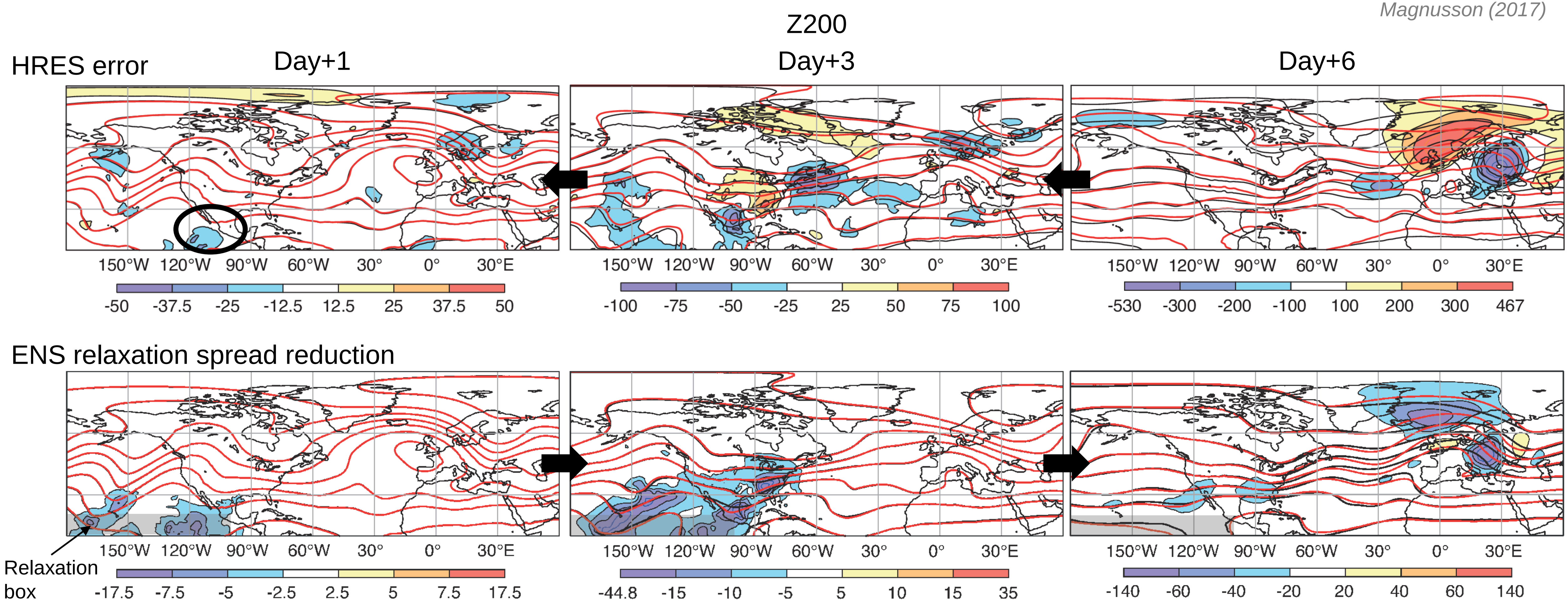


Outline

- The aim of operational diagnostics
- The issues
- Diagnosis of flow-dependent reliability
- Possible useful framework for diagnosis of ensemble forecasting systems

Forecast busts: Error tracing/correlation and confirmation through relaxation expts

Magnusson (2017)



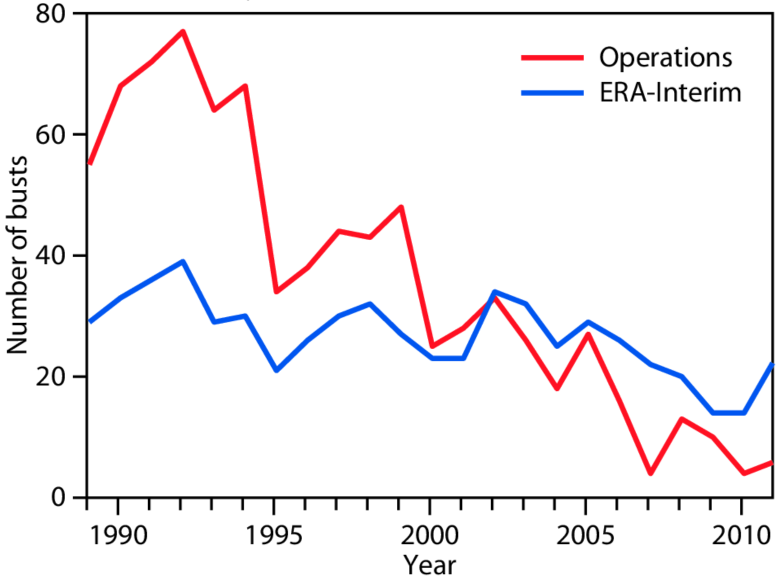
Such work improves knowledge of initial errors (equatorial Rossby waves) and amplifying factors (convection over North America, cyclogenesis over the North Atlantic,…) that can be associated with European weather uncertainty

Forecasts started 0UTC 15 March 2014

Forecast busts: Systematic identification and confirmation of key initial uncertainties

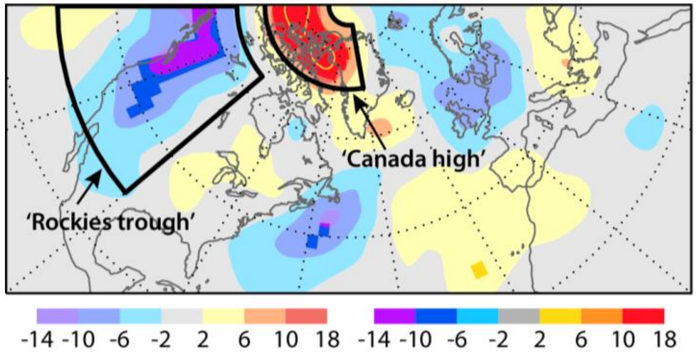
Rodwell et al (2013), Rodwell (2016)

Frequency of deterministic forecast 'busts'

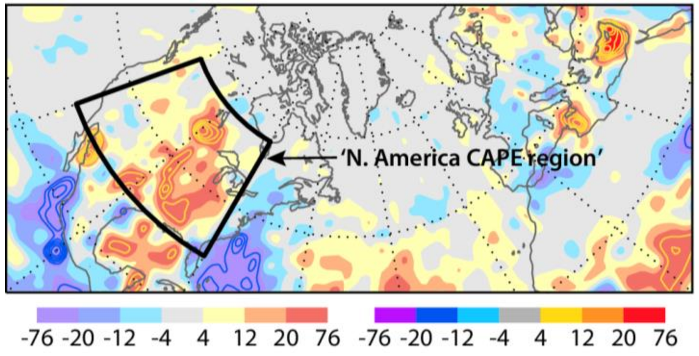


Identification of key initial flow-regime

Z500 anomaly Unit=m



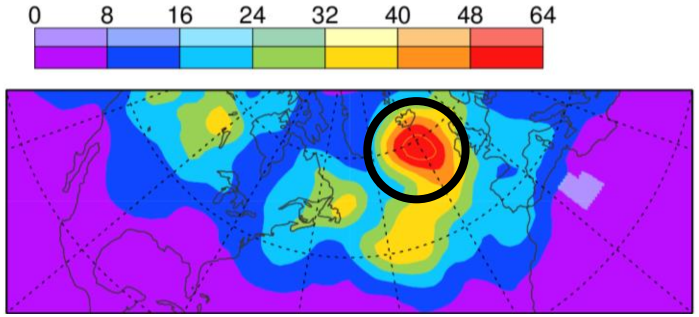
CAPE anomaly Unit=Jkg-1



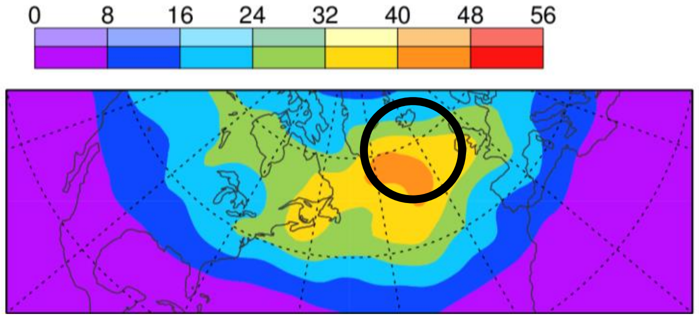
Ensemble forecast composited on "Trough/CAPE" initial regime (independent data)

Z200
Day+5
Unit≈100m²

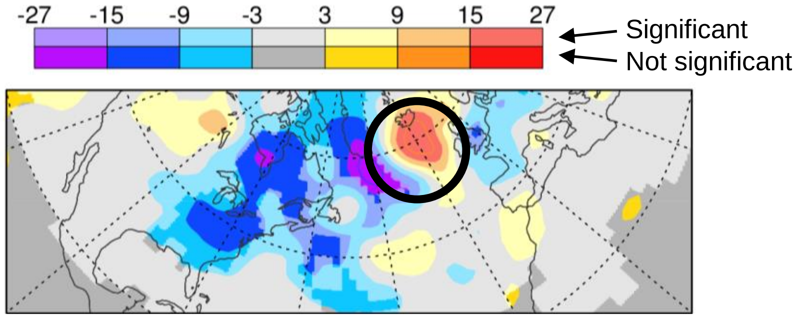
ENS mean error²



ENS variance



ENS error²-variance



Identification of initial flow-regime partly fortuitous/already known. Increased error²/variance, but large +ve residual not significant for feasible data-sets

Both Magnusson (2017) and Rodwell et al (2013) conditioned on deterministic busts. What about 'ensemble busts'?

TIGGE models’ spread and error: Z500 D+6 Europe

Centres’ spread agreement ⇒

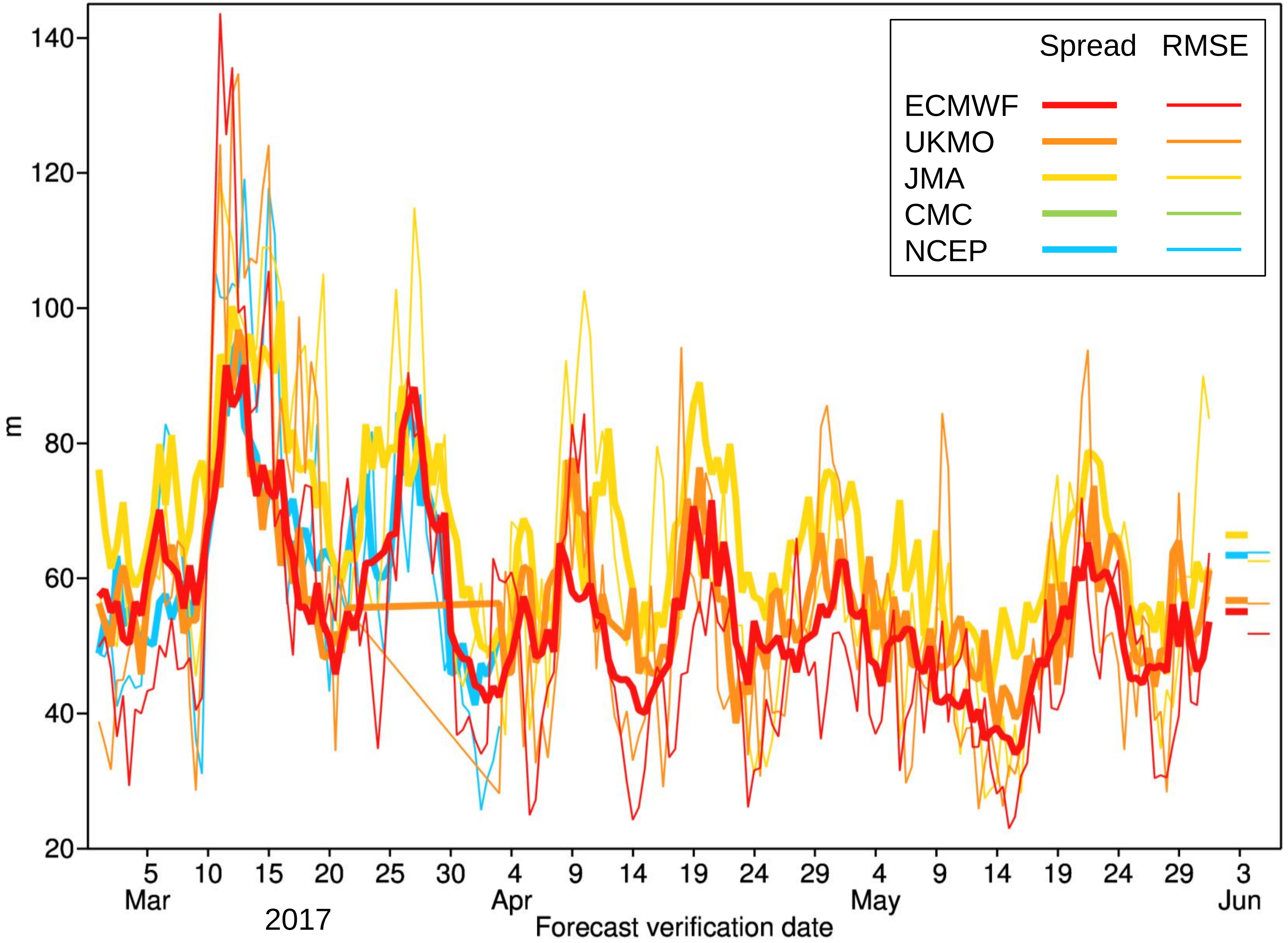
Real variations in predictability

Spread-error agreement ⇒

Degree of flow-dependent reliability

11 March RMSE>>Spread. ‘Ensemble bust’**?**

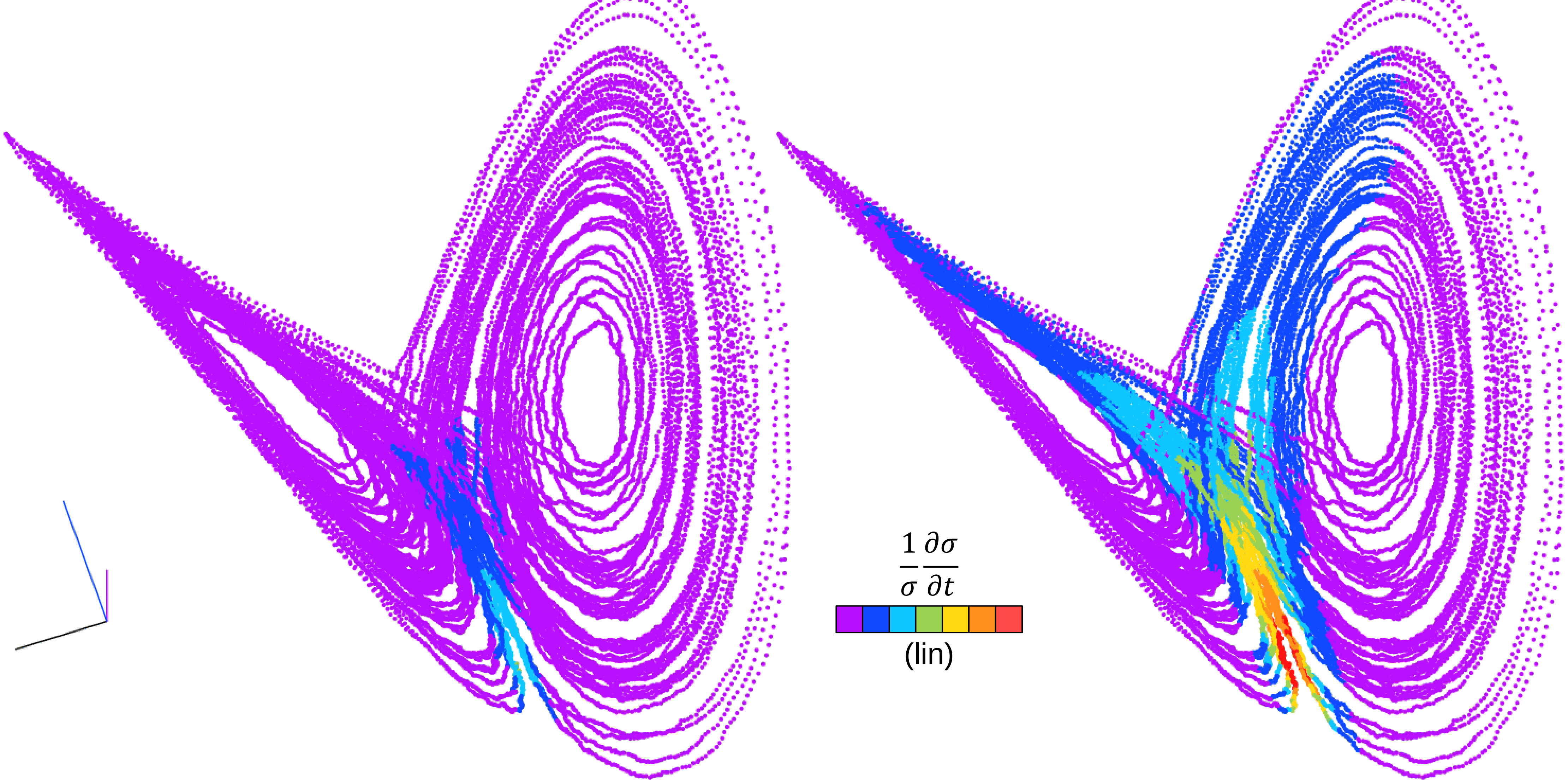
TIGGE valuable, but with data drop-outs



Attractor of Lorenz ’63 model with stochastic noise. Shading = uncertainty growth-rate

Perfect model

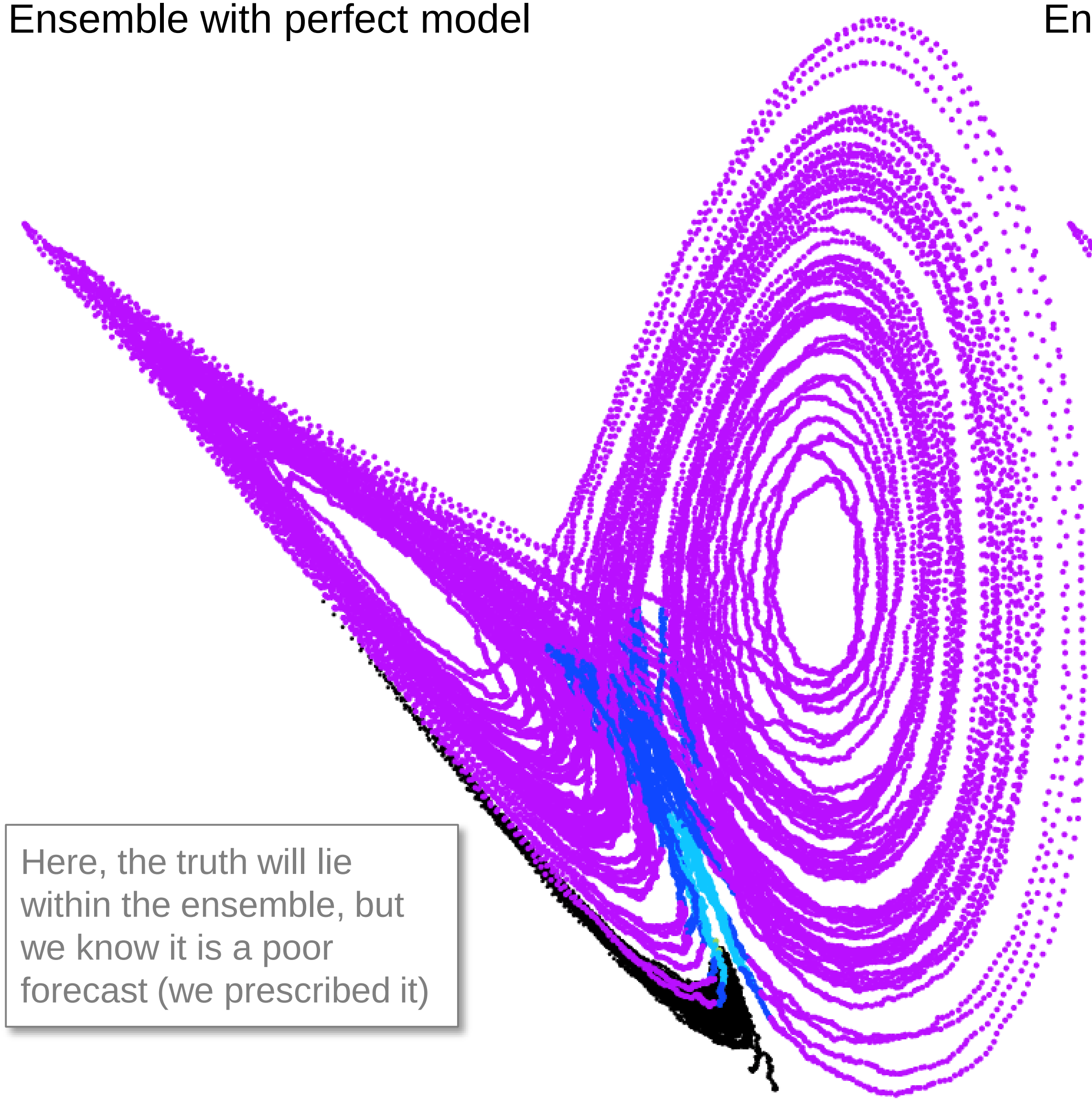
Growth-rate with 2.5x model uncertainty



Lorenz ‘63 model uses original parameter settings. Ensembles initial perturbations (to the truth run) σ_0, and model uncertainty σ_{X_t}, with $\sigma_0 \sim \sigma_{X_t}\delta t$ where δt is timestep

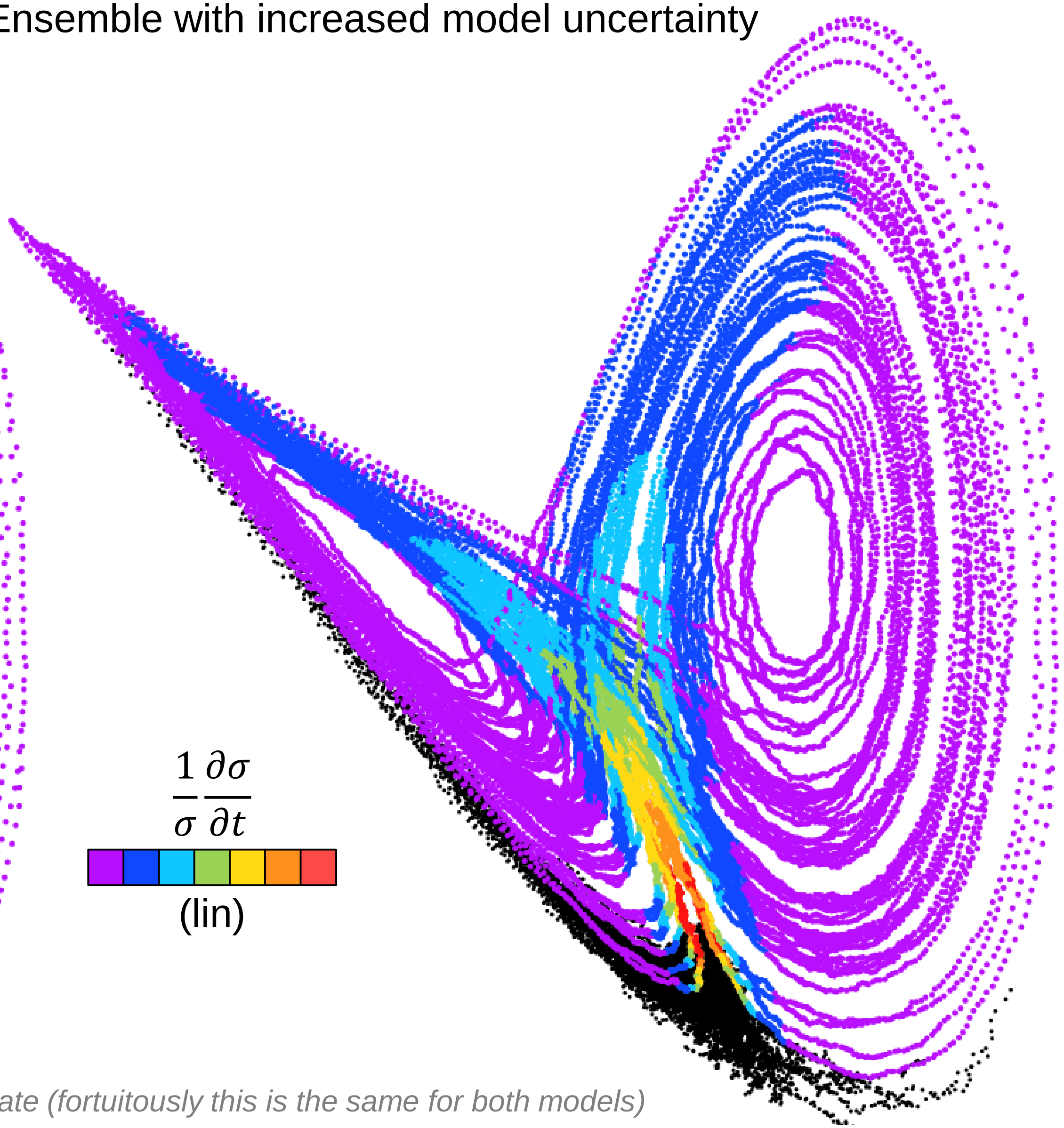
“van Lorenz” attractor: Forecast with fastest uncertainty growth-rate (black)

Ensemble with perfect model



Here, the truth will lie within the ensemble, but we know it is a poor forecast (we prescribed it)

Ensemble with increased model uncertainty



The highlighted ensemble forecast is the one with largest uncertainty growth-rate (fortuitously this is the same for both models)

Outline

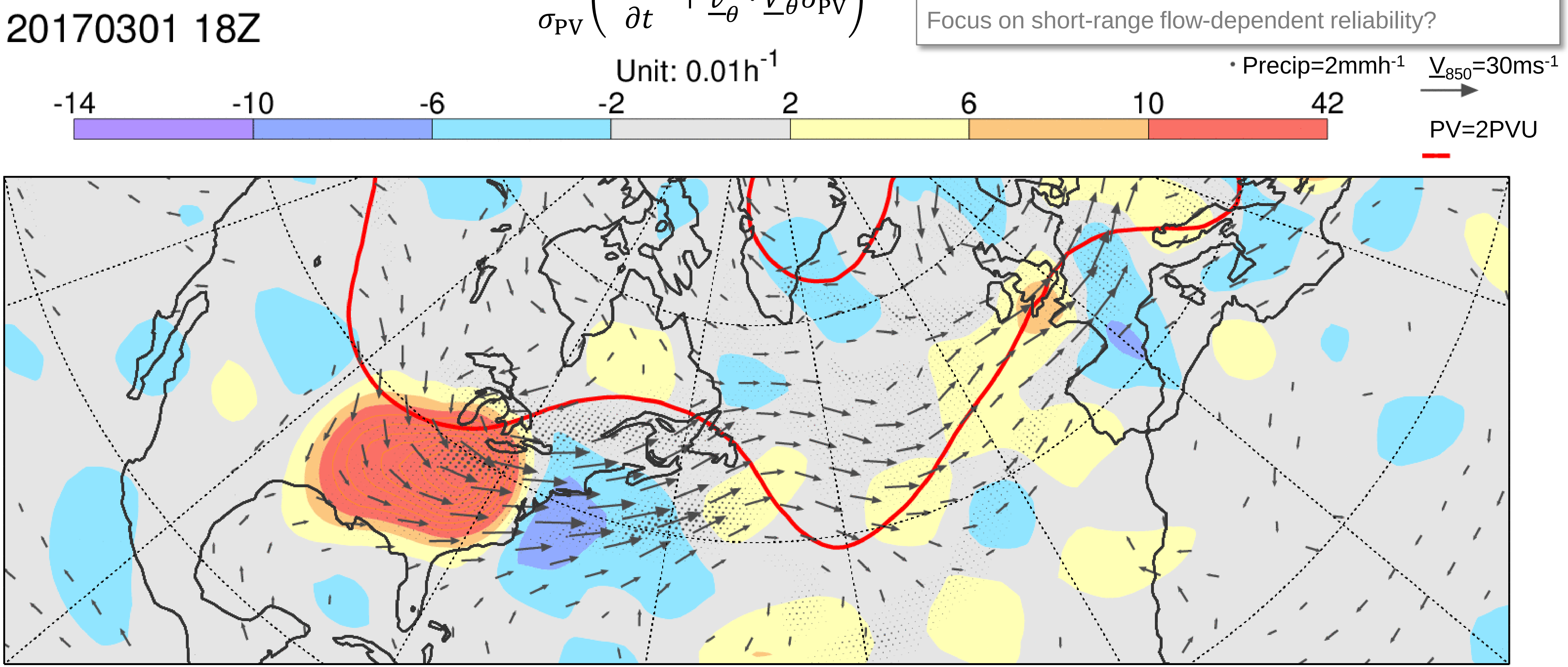
- The aim of operational diagnostics
- The issues
- Diagnosis of flow-dependent reliability
- Possible useful framework for diagnosis of ensemble forecasting systems

“Instantaneous” (0-12h) “Lagrangian” uncertainty growth-rates for $PV_{\theta=315K}$

Complicated interactions hinder direct diagnosis of medium-range ensemble deficiencies

Focus on short-range flow-dependent reliability?

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



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

The EDA reliability budget

Working at short lead-times, need to extend “spread-error” relationship to include observation error variances (and bias)

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

$$\textbf{Depar}^2 = \textbf{Bias}^2 + \textbf{EnsVar} + \textbf{ObsUnc}^2 + \textbf{Residual} \quad ✓$$

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

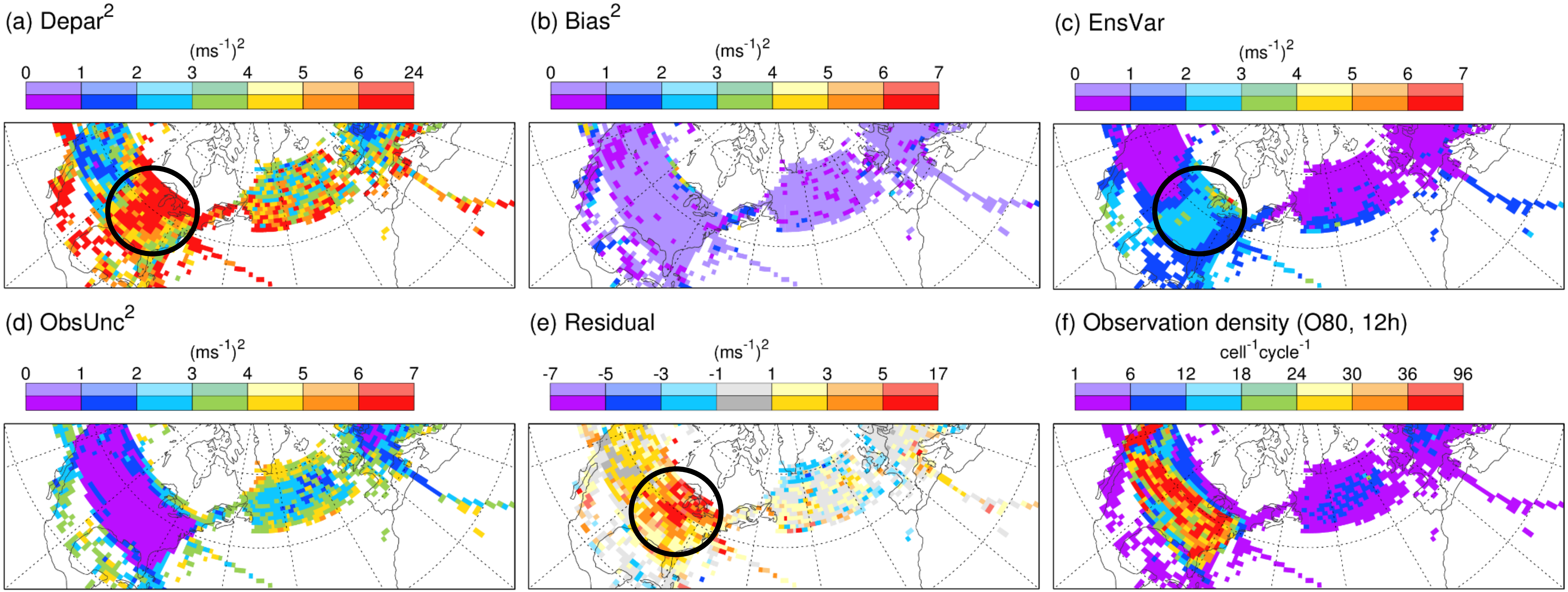
(similar to equations / aspirations of data assimilation)

Short-range variance assessment for u200 in “trough/CAPE” situations using EDA

54 cases

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

Rodwell 2016, ECMWF Newsletter



Enhanced uncertainty (EnsVar) around Great Lakes / Mississippi Region, large ‘errors’ ($Depar^2$)

Observation uncertainty ($ObsUnc^2$) quite small so a statistically significant positive Residual $\Rightarrow$

ENS does not inject enough uncertainty into global circulation. Forecasts will be too confident

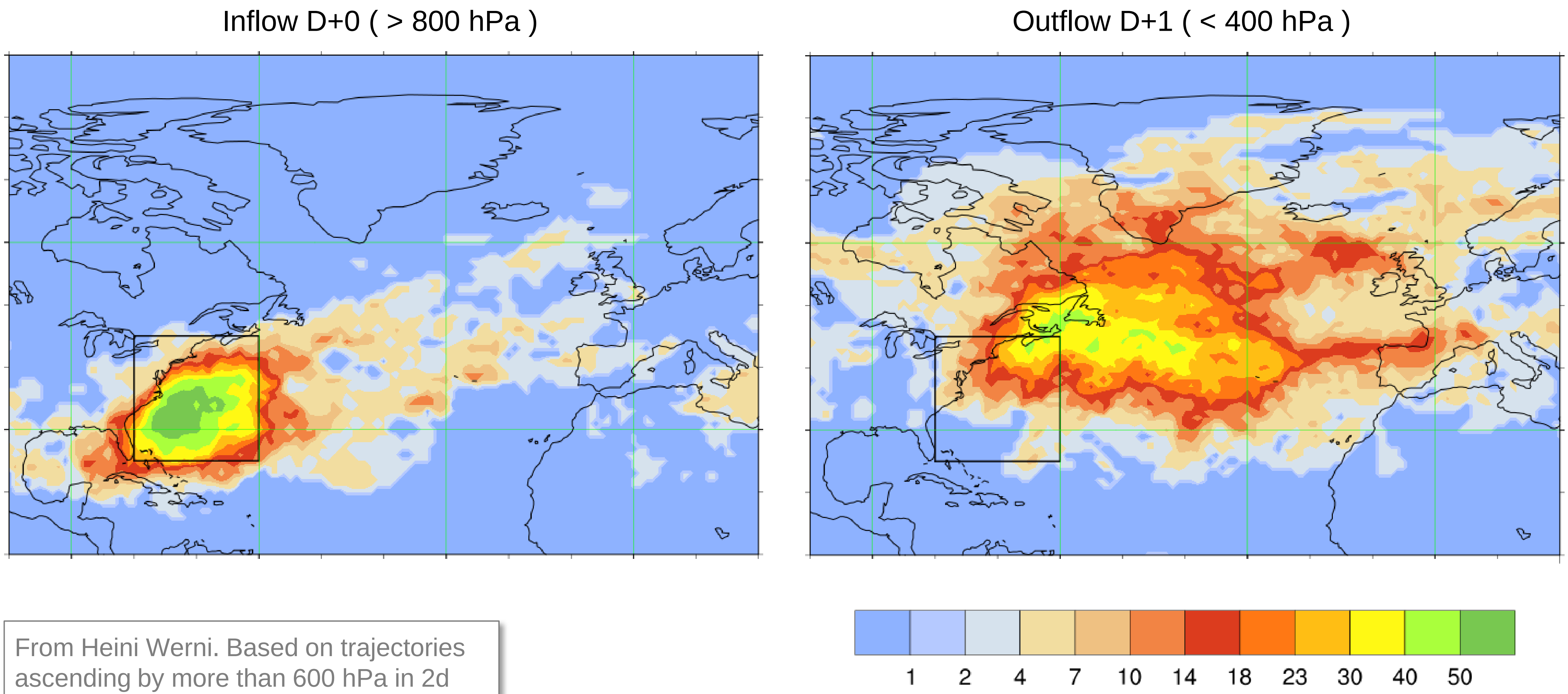
Extend “spread-error” relation to include obs error variances and bias (similar to data assimilation)

~~$Error^2$ = EnsVar + Residual~~

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

Reliability $\Rightarrow$ E[Residual]=0

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



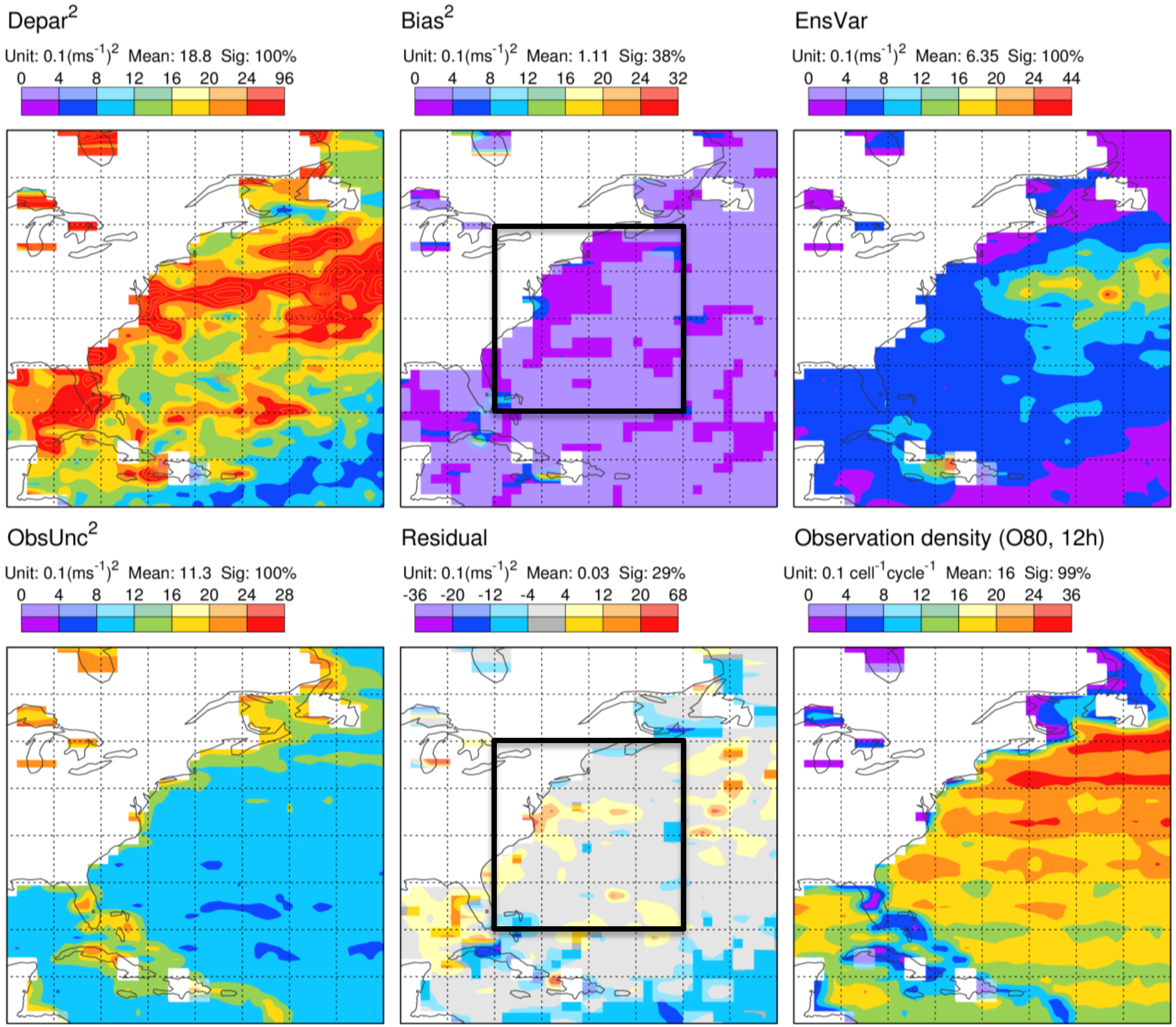
From Heini Werni. Based on trajectories ascending by more than 600 hPa in 2d

EDA variance assessment with ASCAT surface v wind: Non-WCB composite

87 cases

ObsUnc^2 a large component of budget

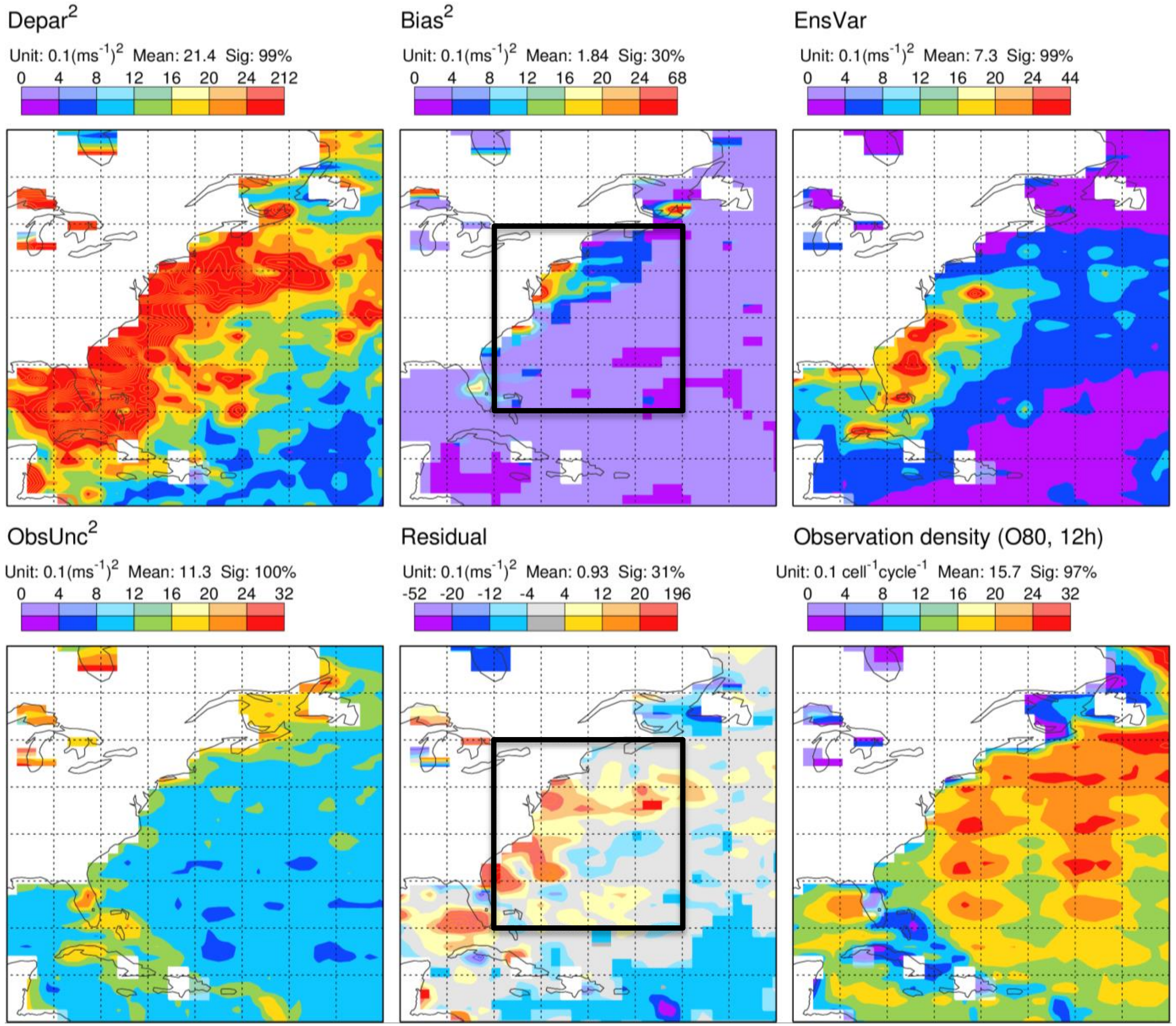
Bias^2 and Residual are not significant in absence of WCBs ✓



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

EDA variance assessment with ASCAT surface v wind: WCB (inflow) composite

50 cases



Larger EnsVar and Depar^2 so a more uncertainty situation

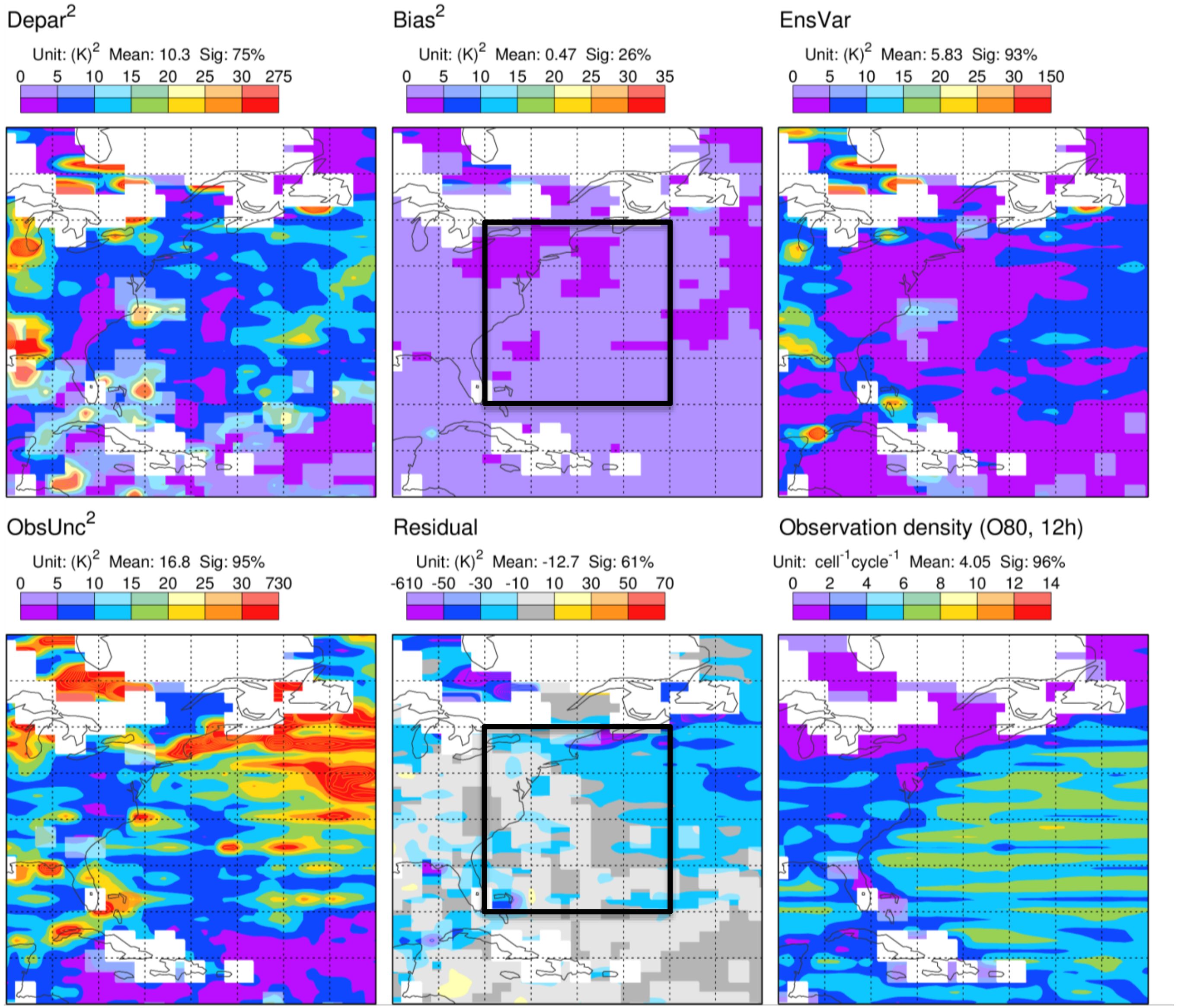
Increased Bias^2 (first moment error) along coast

Positive Residual (not significant): Insufficient spread associated with cyclogenesis, or underestimation of observation uncertainty within the WCB region?

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

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

87 cases



Bias and residual are not significant in absence of WCBs ✓

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

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

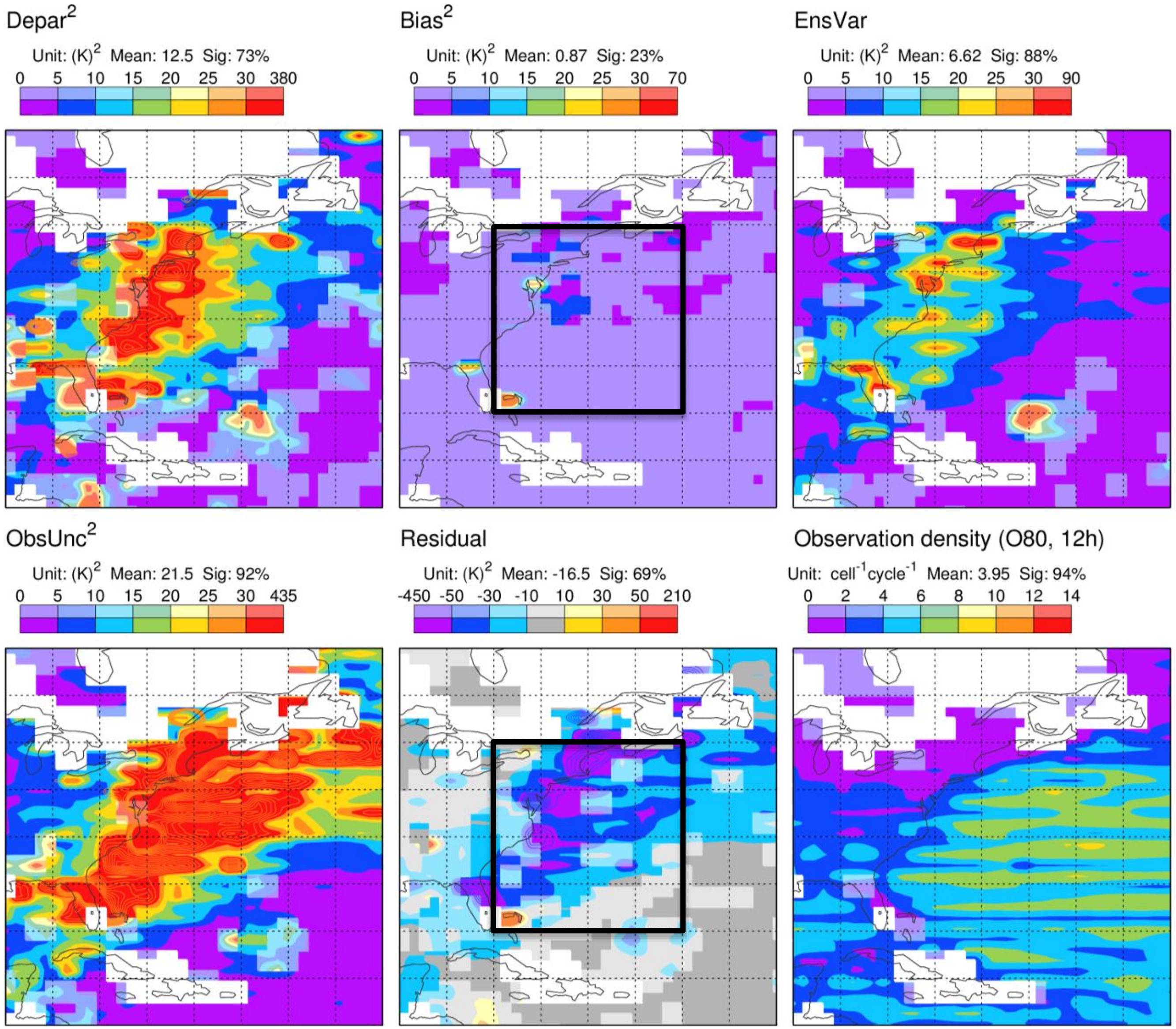
Negative residual largely due to large ObsUnc^2 (larger than the departures) in cloudy regions

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

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

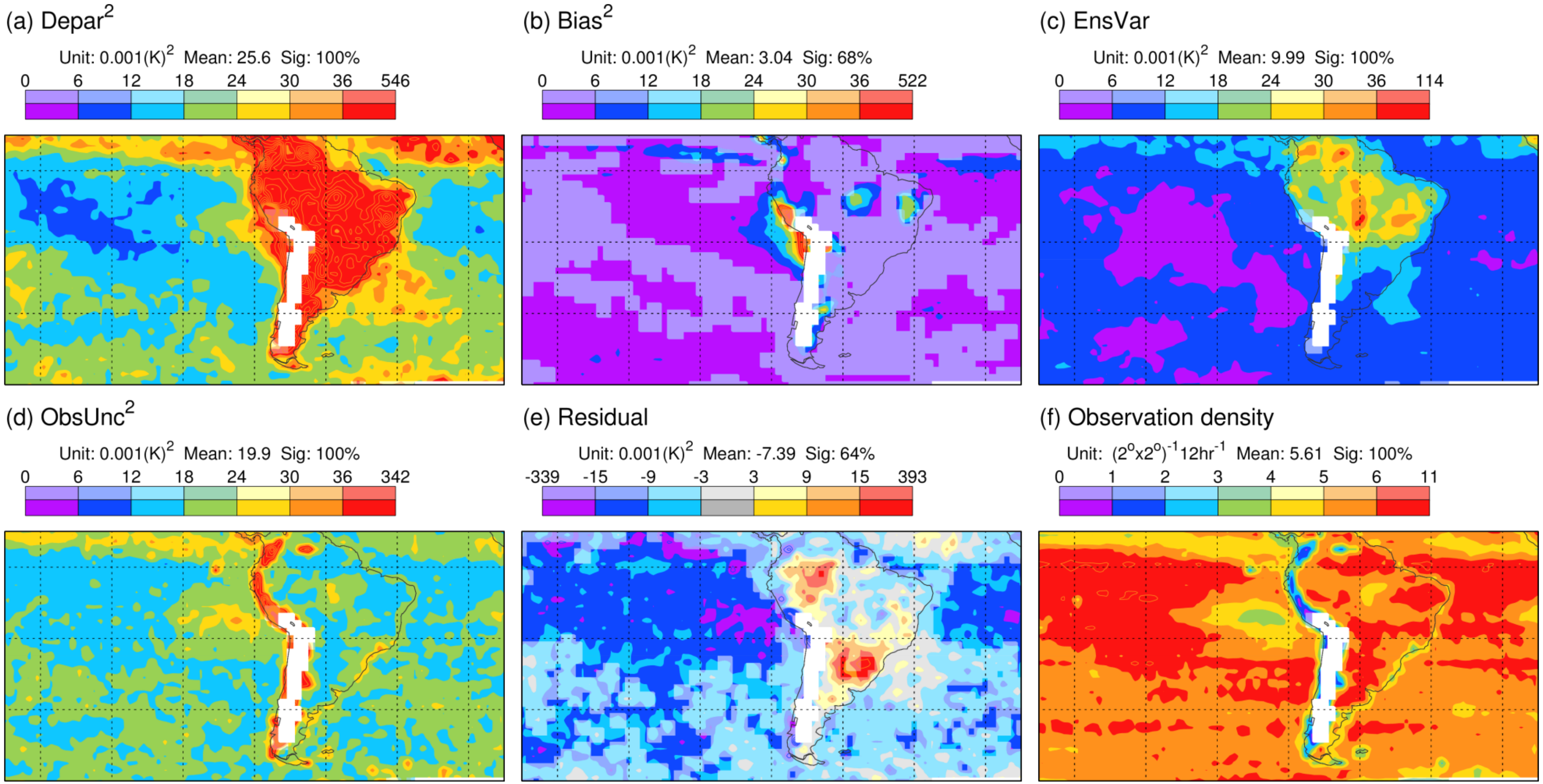


Microwave channel 5

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



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

AMSUA microwave channel 5.12 August – 16 November 2011

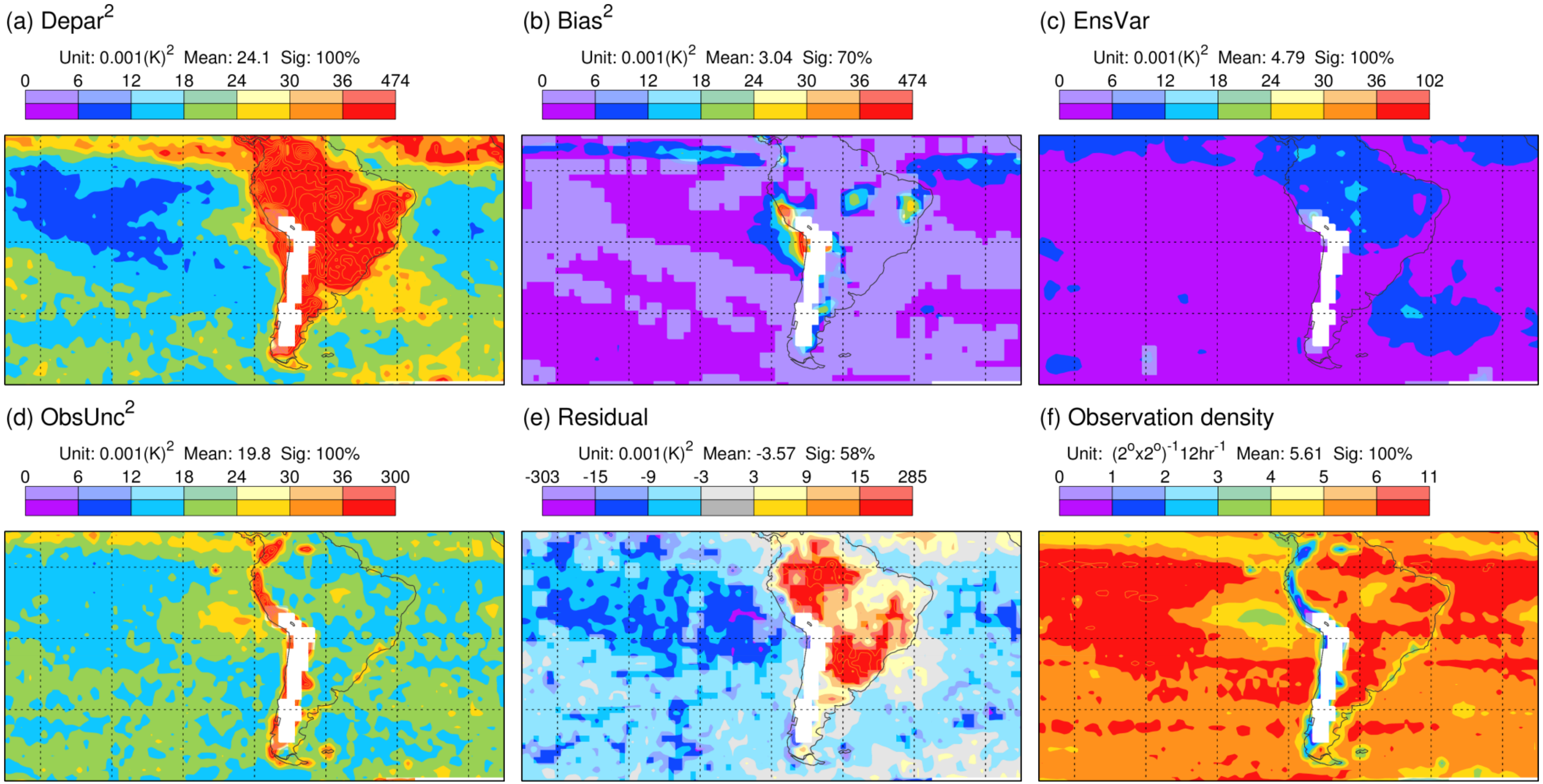
EDA reliability for AMSUA observations of mid-tropospheric T (no stochastic physics)

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 (see Sarah-Jane's talk)



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

AMSUA microwave channel 5.12August – 16 November 2011

Outline

- The aim of operational diagnostics
- The issues
- Diagnosis of flow-dependent reliability
- Possible useful framework for diagnosis of ensemble forecasting systems

Possible useful framework for diagnosis of ensemble forecasting systems

Local state space partitioned into a set of synoptic flow types

Brier Score (*e.g.*)

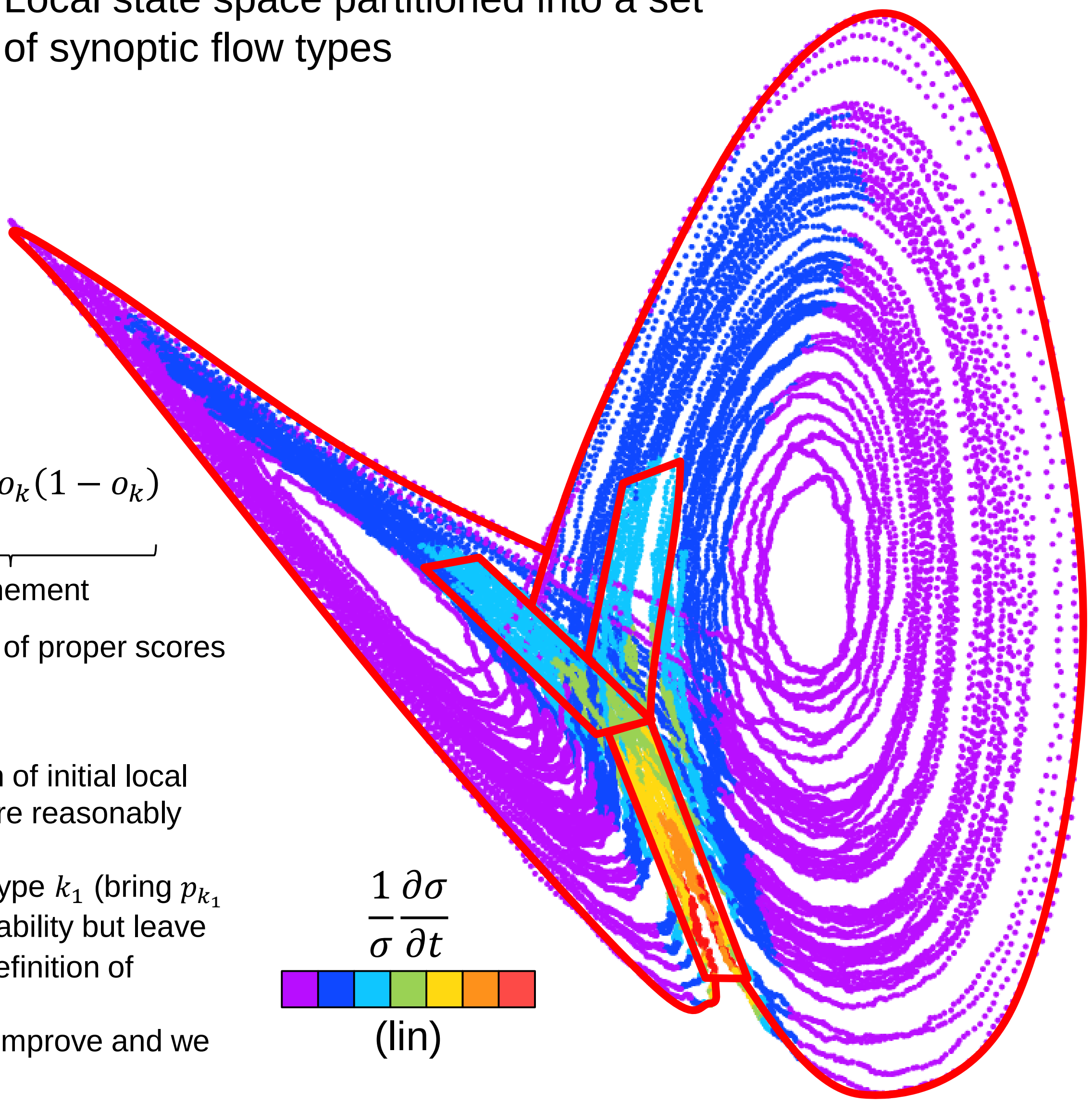
$$\mathrm{BS} = \frac{1}{N}\sum_{t=1}^{N}(p_t - o_t)^2$$

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

Reliability – Refinement decomposition of proper scores

Thought experiment:

- Think of the $k = 1, \ldots, K$ as a partition of initial local flow type and assume probabilities are reasonably constant for a given flow type.
- Improving reliability for a given flow type k_1 (bring p_{k_1} closer to o_{k_1}) will improve overall reliability but leave refinement essentially unchanged (definition of synoptic flow-type does not change).
- Hence local short-range skill should improve and we have a better model.



Focusing on short-range local flow-dependent reliability, should obtain:

- Better short-range skill
- Better model and representation of uncertainty at all lead-times

Prioritise efforts and monitor progress on flow-types that contribute most to reliability aspect of a proper score (refinement essentially unaffected)

Refinement contributions might highlight particularly desirable observational information to constrain initial uncertainty

“Diagnostics Toolbox” at ECMWF, which draws on the work of many at the Centre, allows users to compute EDA and Initial tendency diagnostics, *etc*. Should provide a common framework with which to discuss issues (as we already have in verification)

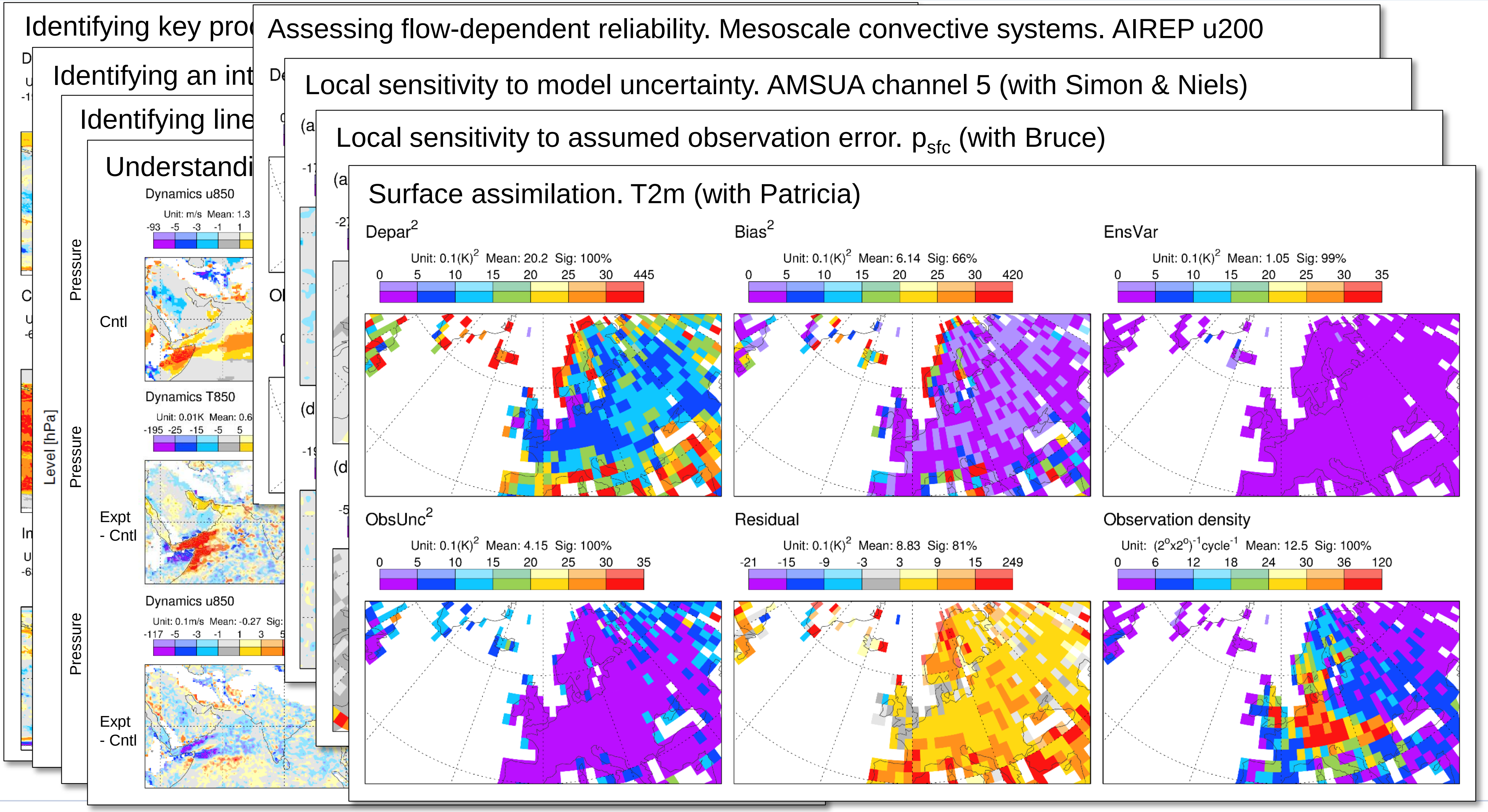
Thank you

Summary

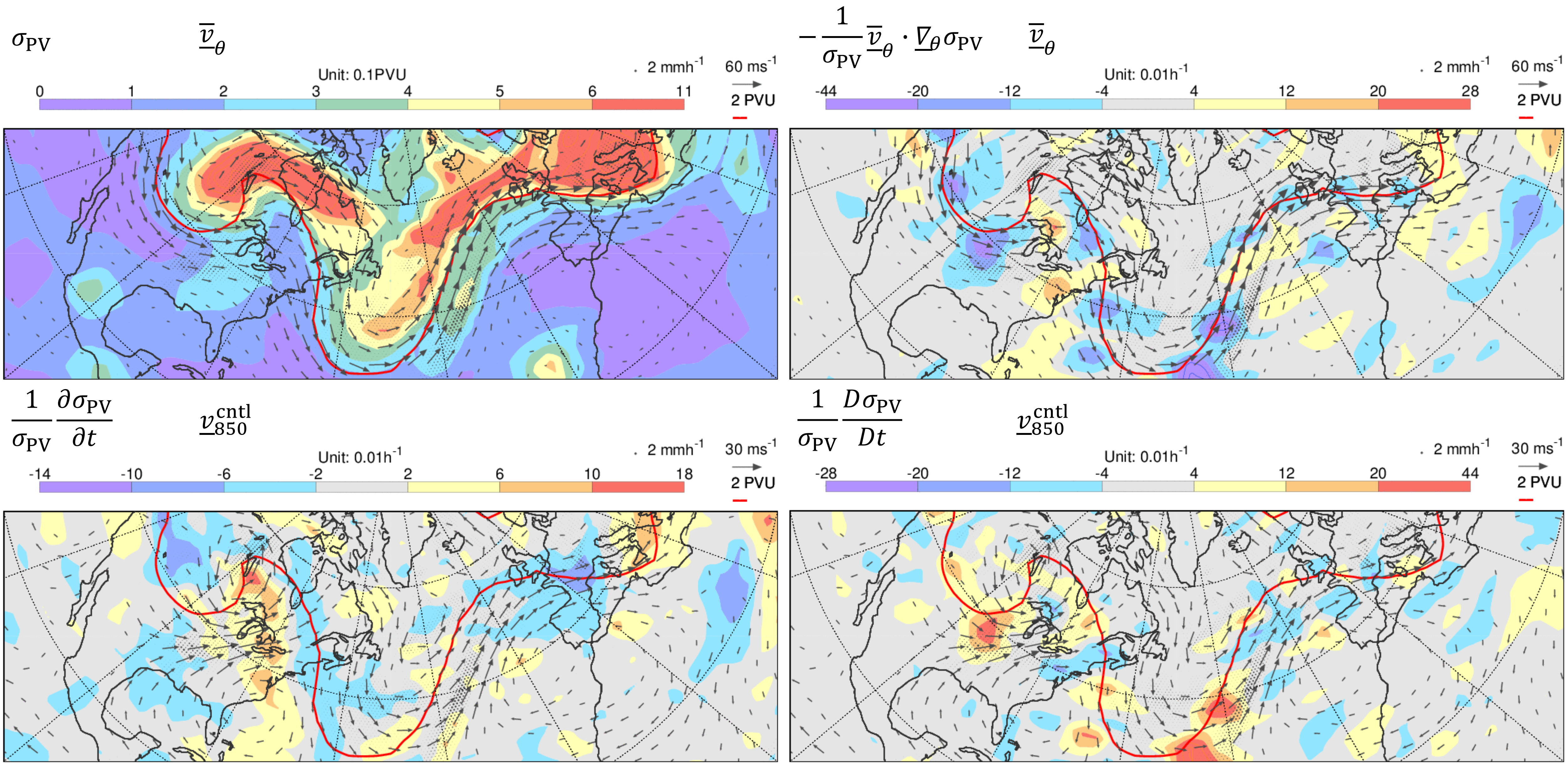
- Improving **short-range local reliability for a given synoptic flow situation**‡ should improve overall reliability, leaving refinement essentially unaltered, and thus increasing short-range proper skill (even if spread is increased). Requires improvements to the model(s) and/or representation of uncertainties (not trivial), and these will be used throughout the forecast range.
- "**Refinement is an attribute of the observations** rather than of the forecasts" (Murphy 1972). Can be improved (subject to predictability limits) by reducing initial uncertainty through the assimilation of increased observational information
- **Hence the diagnosis and improvement of flow-dependent reliability (and observational information content) may represent a useful framework for future forecast system development**
- Integrated nature of forecasting systems means that there is a need for a common framework in which to discuss issues, so some centralisation of diagnostics is desirable. **Diagnostics Toolbox**, which draws on the work of many at the Centre, includes tools like the "EDA reliability budget" and the "Initial tendencies / analysis increments budget" aims to be as efficient and easy-to-use as possible.

‡ while leaving other flow-types unaffected

Initial tendency / analysis increment and EDA reliability budgets



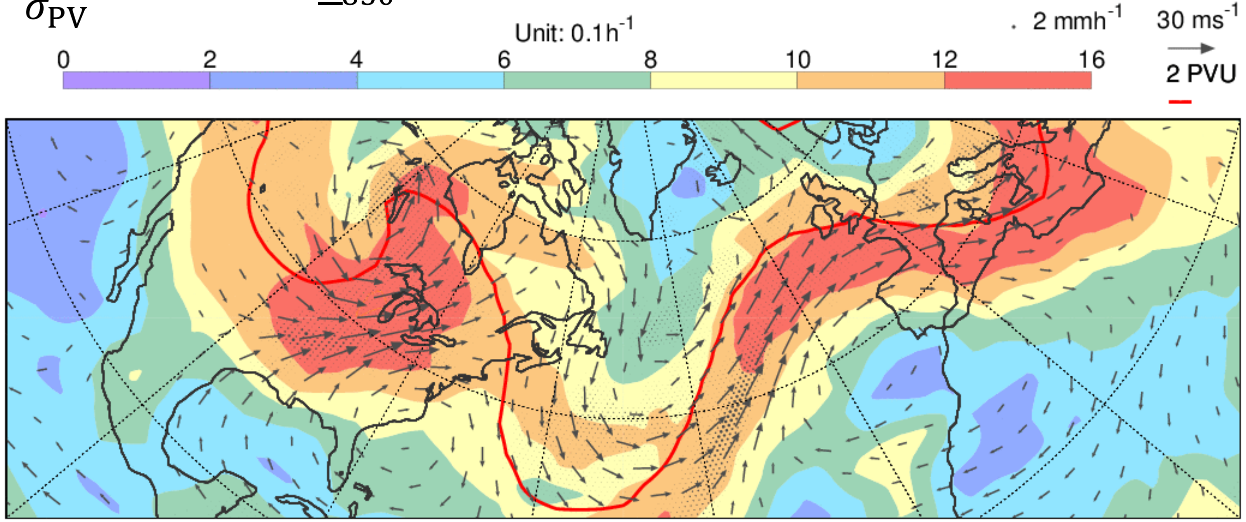
Uncertainties & 12h growth-rates on 20170307: PV on θ=315K, PV_{315}=2, $\underline{v}_{850}$, Precip.



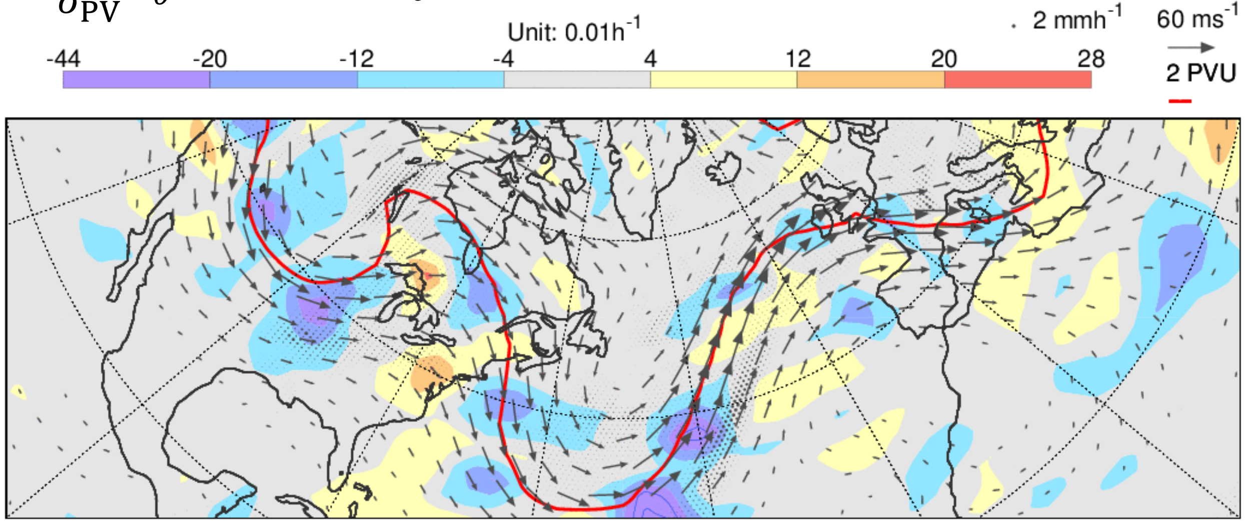
12h growth-rate sensitivity to correlation ρ_{X,X_t} for X=PV$_{315K}$

$$\frac{1}{\sigma_X}\frac{\partial\sigma_X}{\partial t} = \rho_{XX_t}\frac{\sigma_{X_t}}{\sigma_X}$$

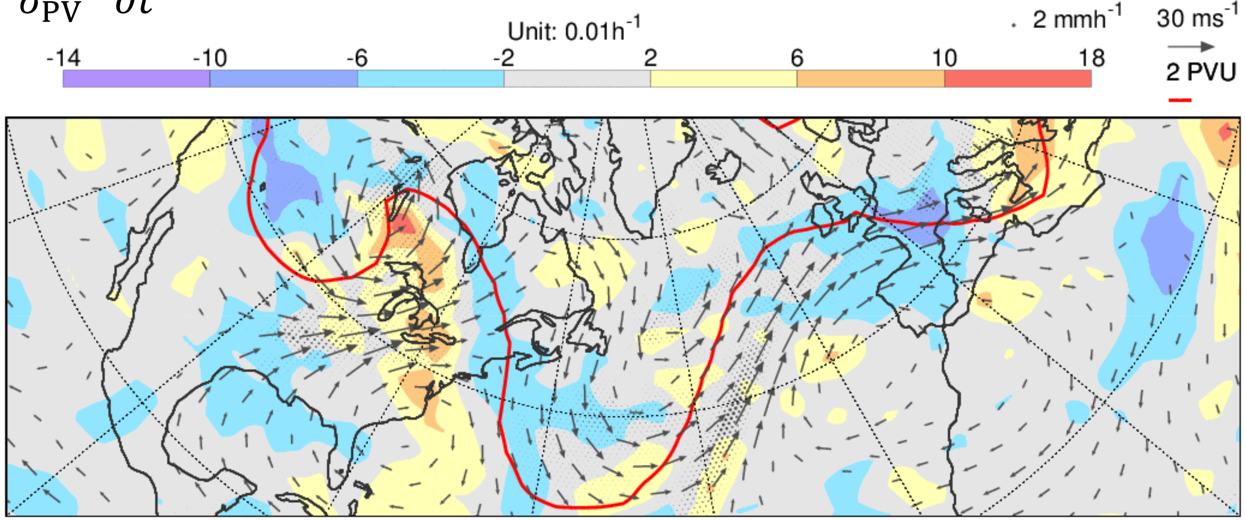
$\dfrac{\sigma_{PV_t}}{\sigma_{PV}}$ $\underline{v}_{850}^{cntl}$



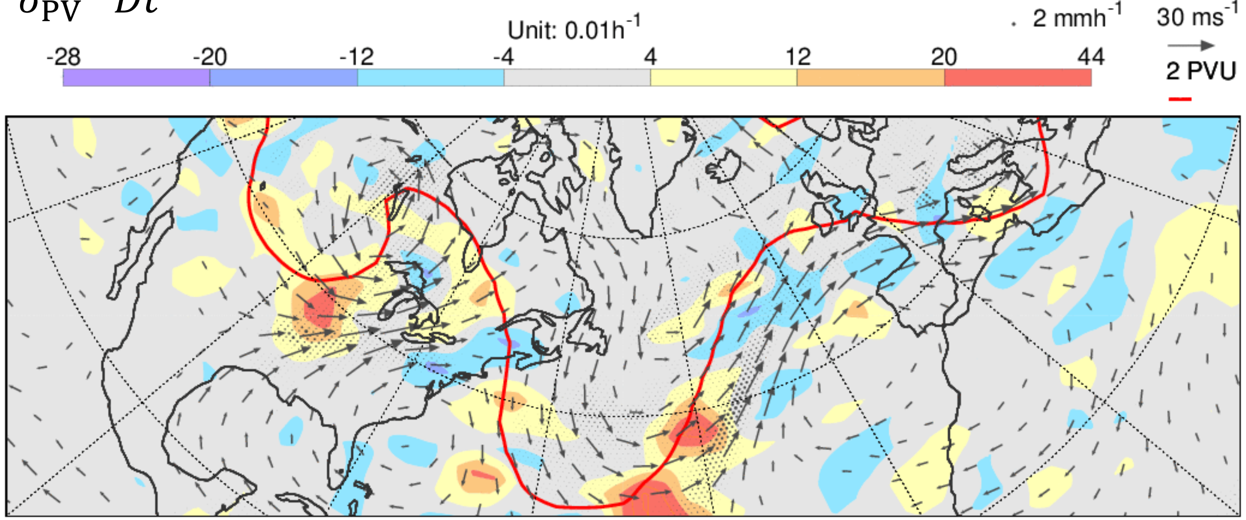
$-\dfrac{1}{\sigma_{PV}}\underline{\bar{v}}_\theta\cdot\underline{\nabla}_\theta\sigma_{PV}$ $\underline{\bar{v}}_\theta$



$\dfrac{1}{\sigma_{PV}}\dfrac{\partial\sigma_{PV}}{\partial t}$ $\underline{v}_{850}^{cntl}$



$\dfrac{1}{\sigma_{PV}}\dfrac{D\sigma_{PV}}{Dt}$ $\underline{v}_{850}^{cntl}$



Abstract

The task of numerical weather prediction research is to improve ensemble reliability and resolution. Clearly, diagnostics play a role in this process, but a key question is how can they best inform system development? A series of more specific questions then follow from this. For example, where should we target diagnostics to optimally identify the root-causes of forecast deficiency? What is the role for centralised diagnostic tools, and how do we facilitate their use by the broader research community? For the first of these specific questions, I will argue that all of the issues point to the same solution - diagnostics targeted at the shortest lead-times possible (i.e. within the assimilation process) permit an evaluation which is localised in space, can be flow-specific, and increases signal-to-noise ratios. For the second question, I will argue that the integrated nature of forecasting systems means that there is a need for a common framework in which to discuss issues. This, and the investment of time required to develop and optimise diagnostic tools means that some centralisation is desirable. However, for several reasons, it is often better for us to "find our own bugs" (as one eminent scientist at ECMWF once said!) and so these tools do need to be usable by the broader community. Throughout this talk, I will draw on results from ECMWF's "Diagnostics Toolbox" which, itself, draws on the work of many at the Centre.