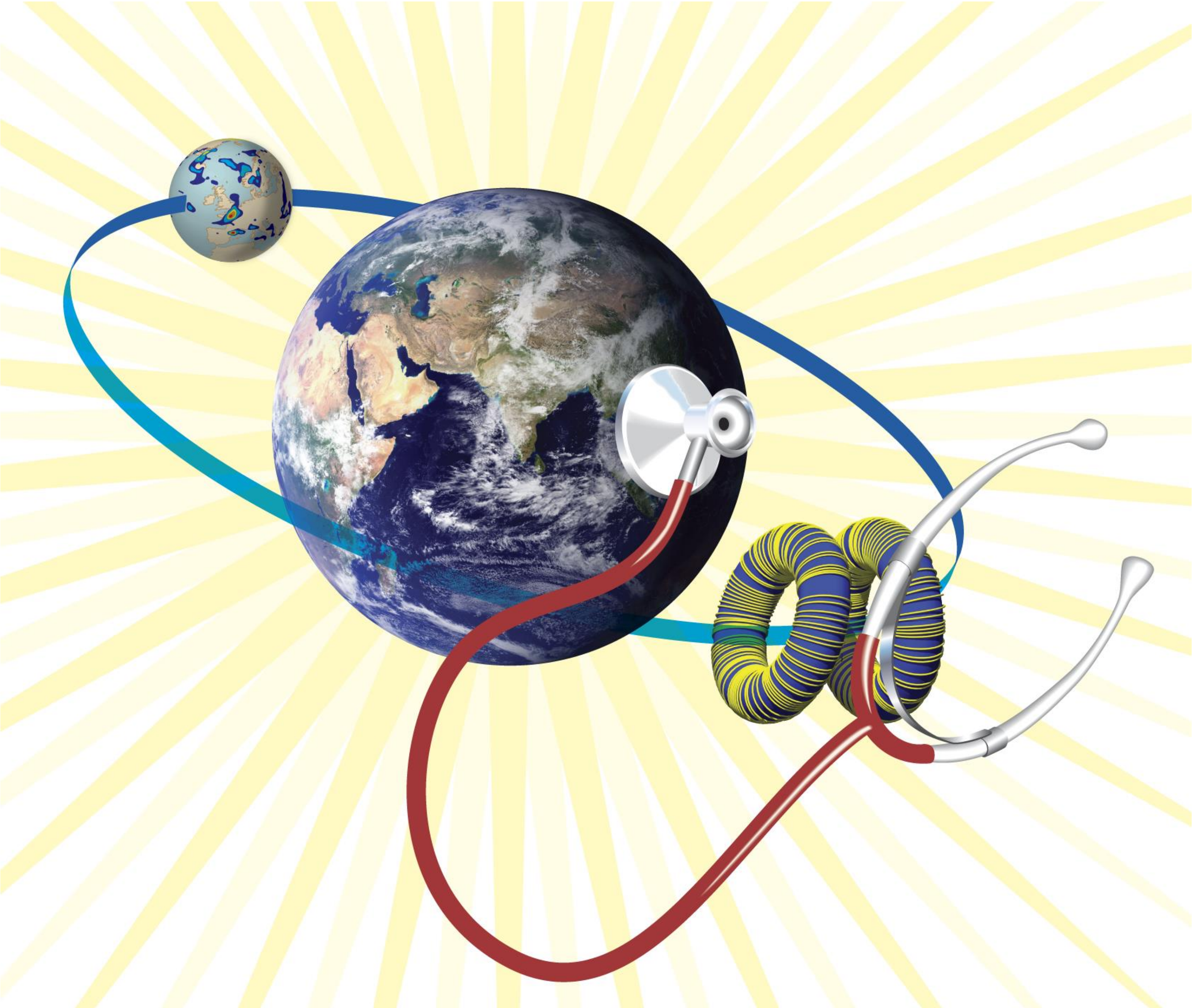


Model errors and diagnostic tools

Mark Rodwell

Forecasters' Training Course

29 January 2018, ECMWF

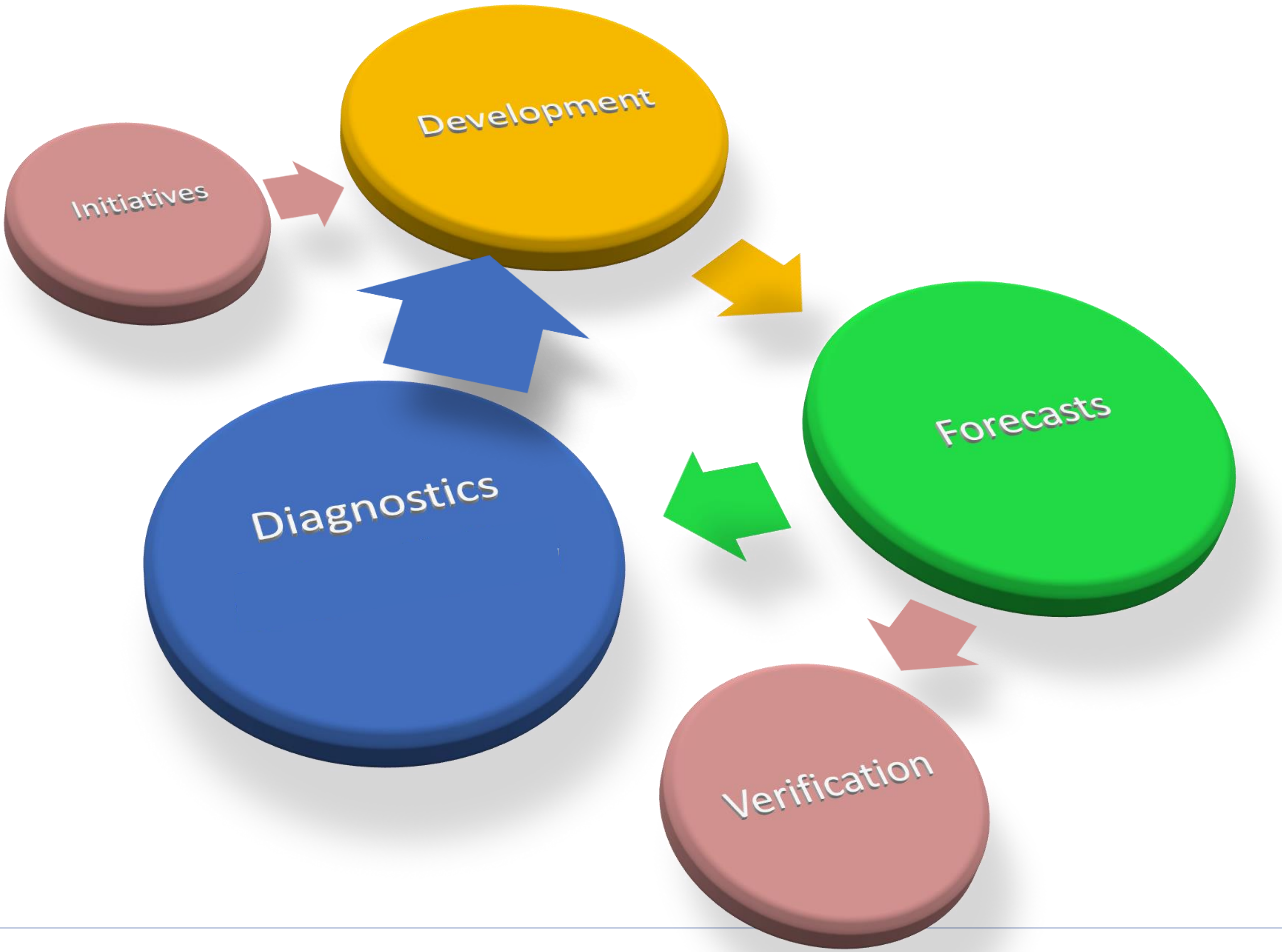
Outline

- The role of operational diagnostics
- Mean error
- Variance error (& predictability)
- A diagnostic framework for forecast system development

Outline

- The role of operational diagnostics
- Mean error
- Variance error (& predictability)
- A diagnostic framework for forecast system development

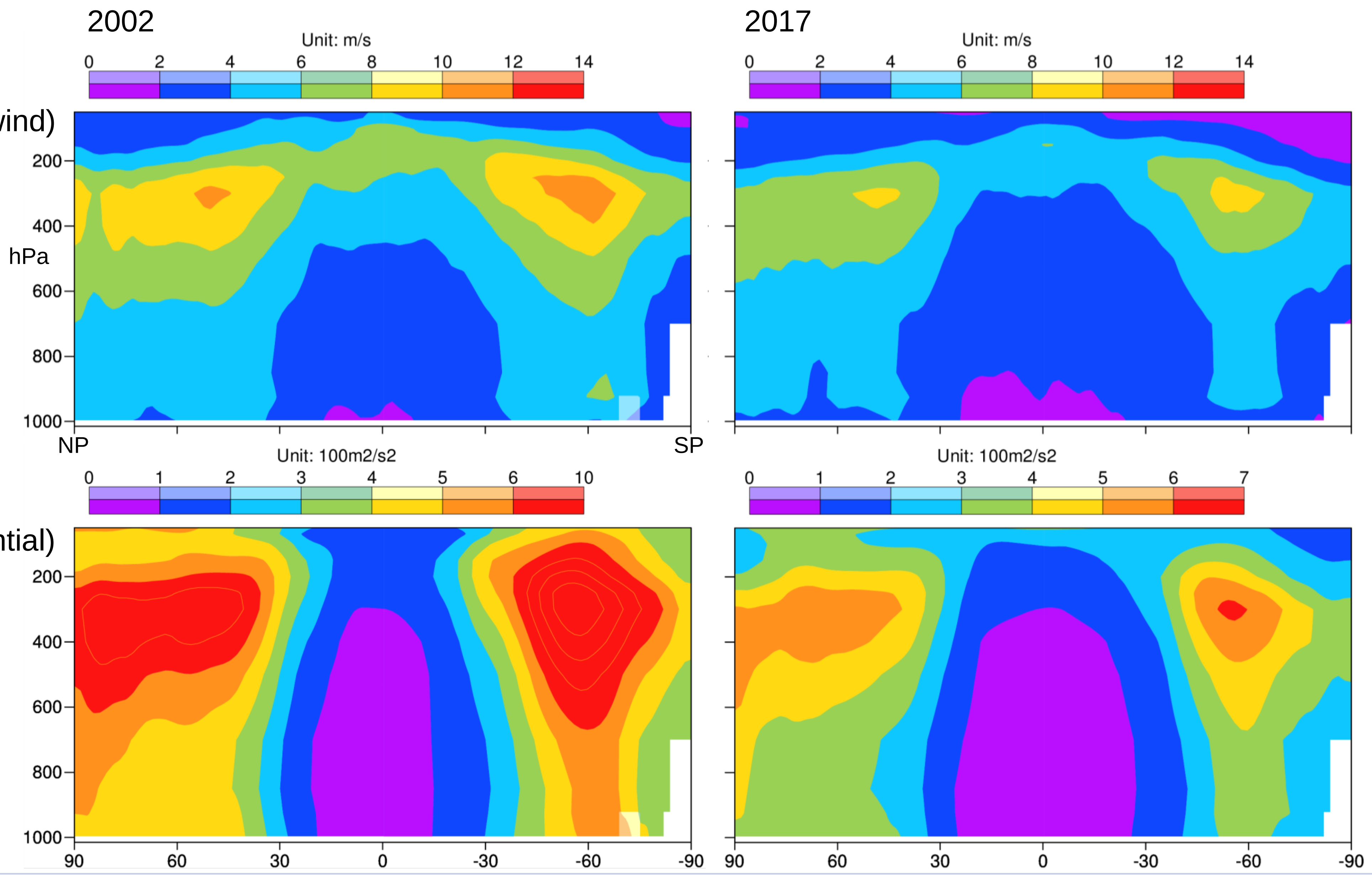
The role of Diagnostics in the development process



Progress over the last 15 years - [RMSE] latitude-pressure cross-sections DJF D+5

U (zonal wind)

Z (geopotential)



Diagnostics research needs to find ways of identifying diminishing errors in an increasing complex model

Also need to consider ensemble aspects

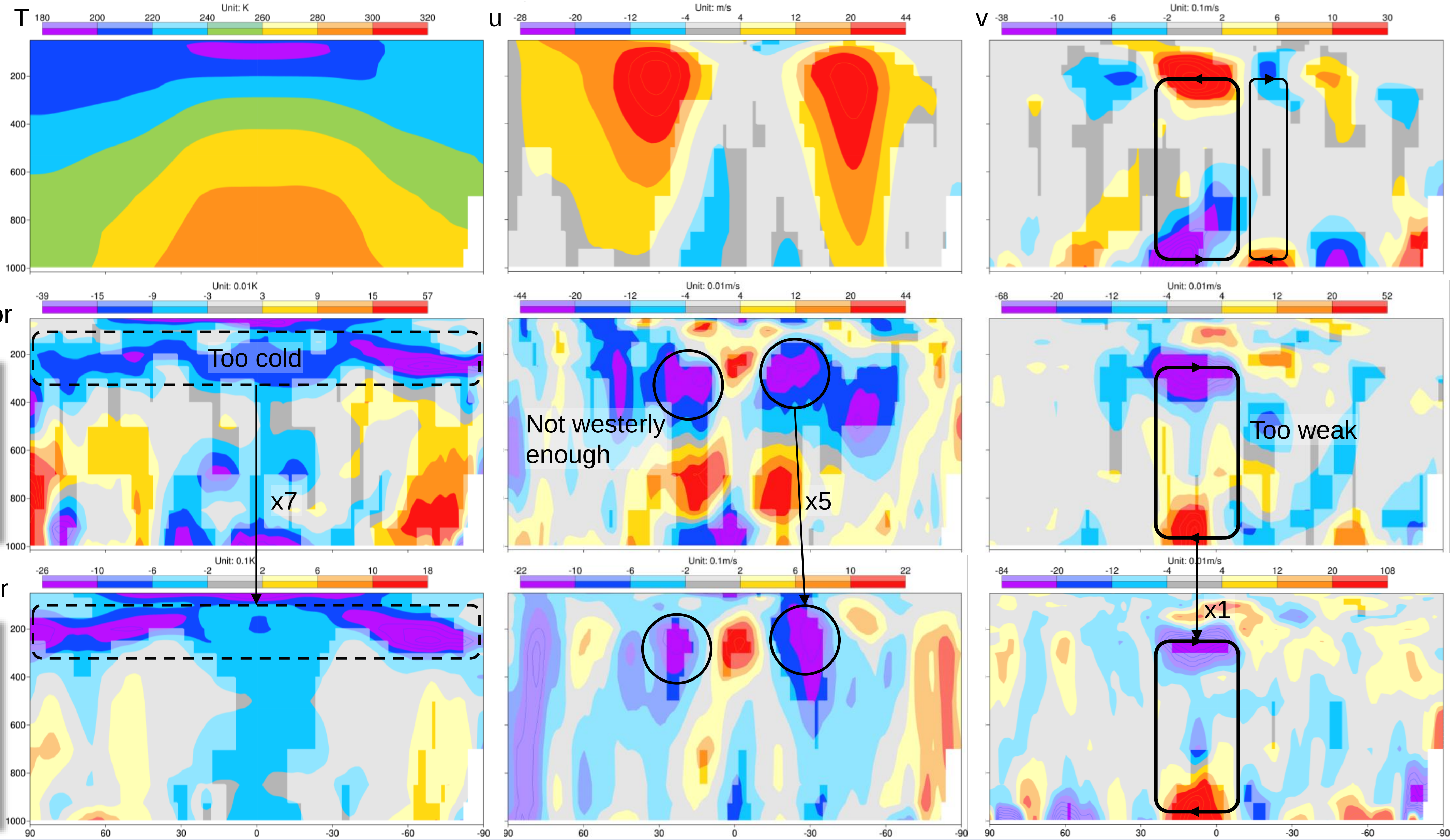
Contours are used to extend the shading to the maximum and minimum values indicated at the ends of the colour bars. Values which are significantly different from 0 at the 5% level are shaded in deep colours (almost everything in this case)

Outline

- The role of operational diagnostics
- Mean error
- Variance error (& predictability)
- A diagnostic framework for forecast system development

Key forecast biases - DJF 2016/17

Note the different units and numbers on colour bars!



Cold tropopause: Excessive (and poorly observed) moisture leads to radiative cooling and growing T error?

Weak Hadley Circulation: Constant v error leads to weak pumping of upper-level westerly momentum and growing u error?

Geographical view of mean wind errors – DJF 2016/17

Day 1: Errors are localised at their sources and statistically significant

Day 10: 'Errors' have strengthened but become less local and less statistically significant. They start to reflect the lack of predictability of the seasonal-mean anomaly rather than model error

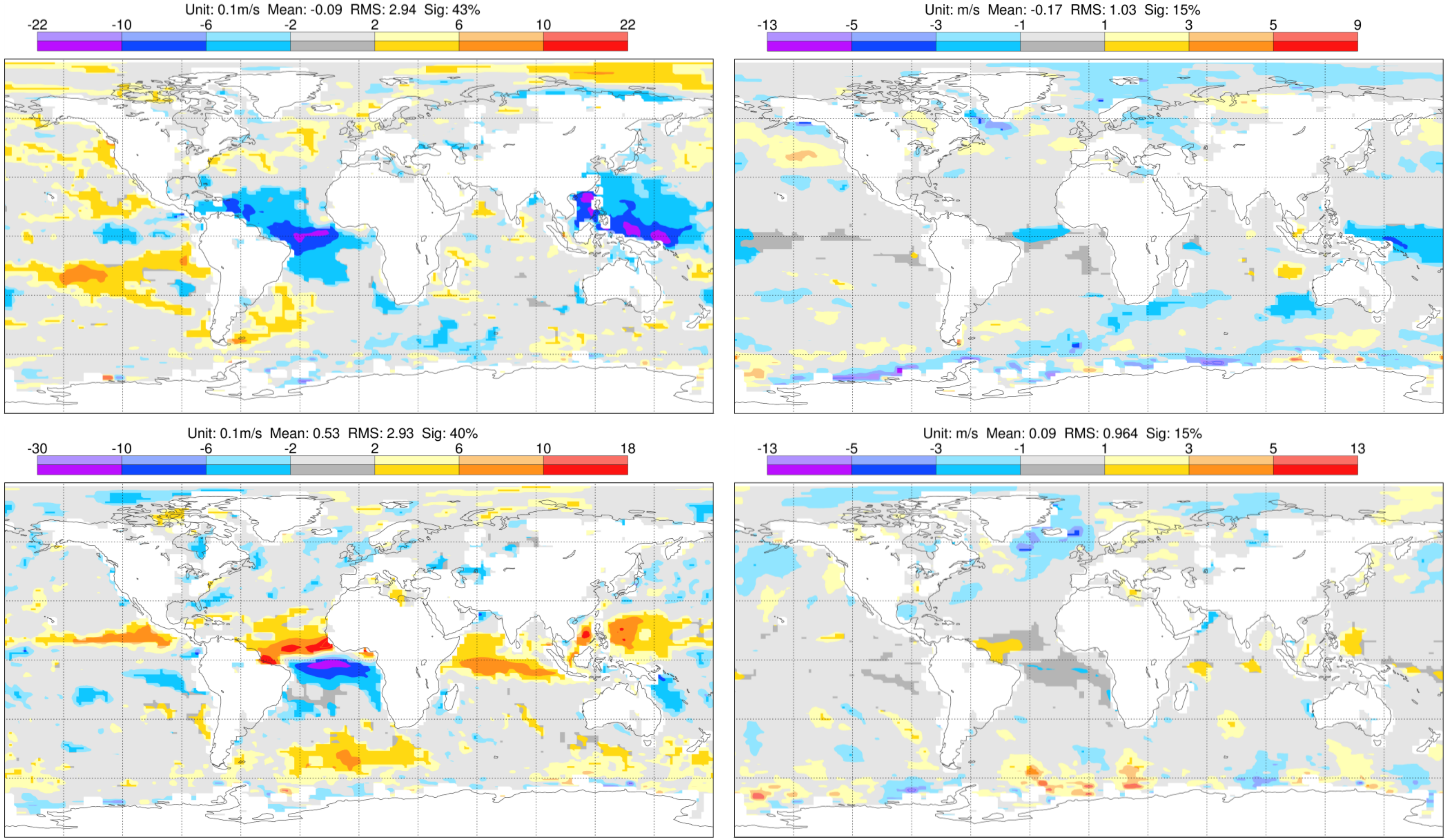
Argument for diagnosis of model bias at short lead-times – e.g. within the data assimilation system

Day 1

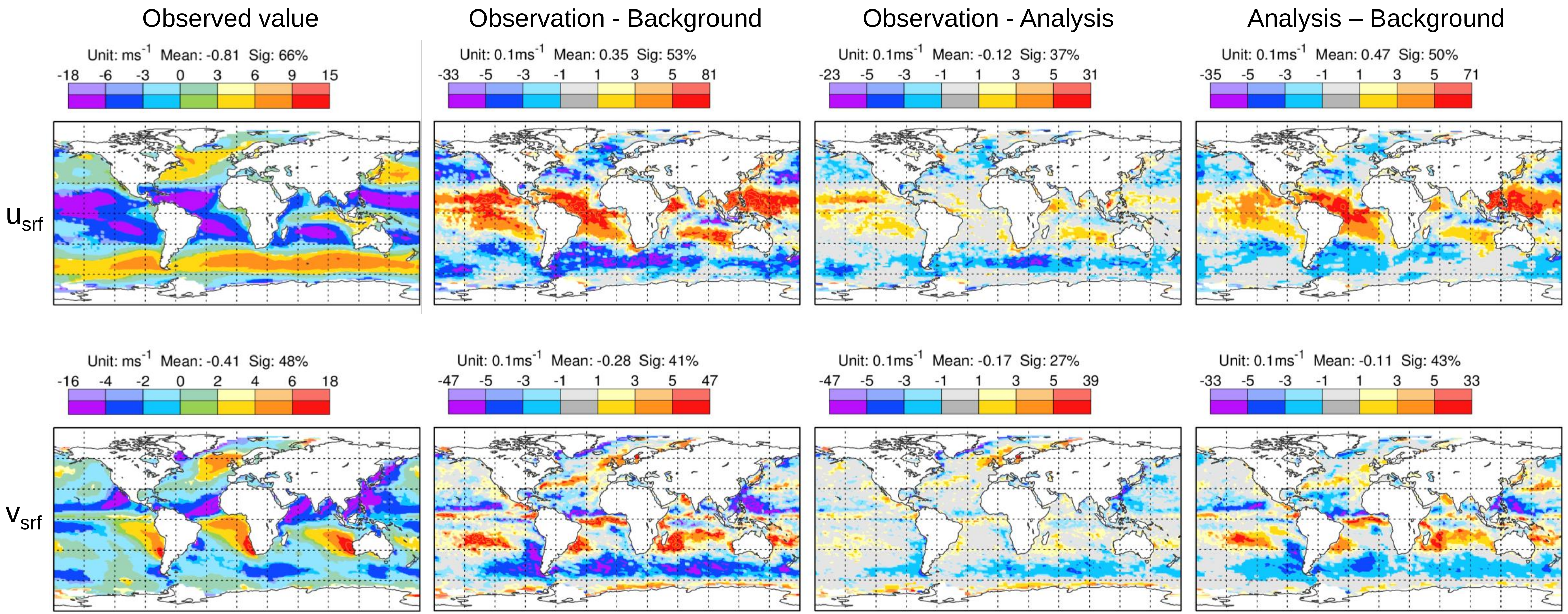
Day 10

u1000

v1000



Mean assimilation diagnostics for “ASCAT” surface winds



Data assimilation effective at drawing analyses away from the background (i.e. “first guess”) and towards the observations

Also known as the “analysis increment”

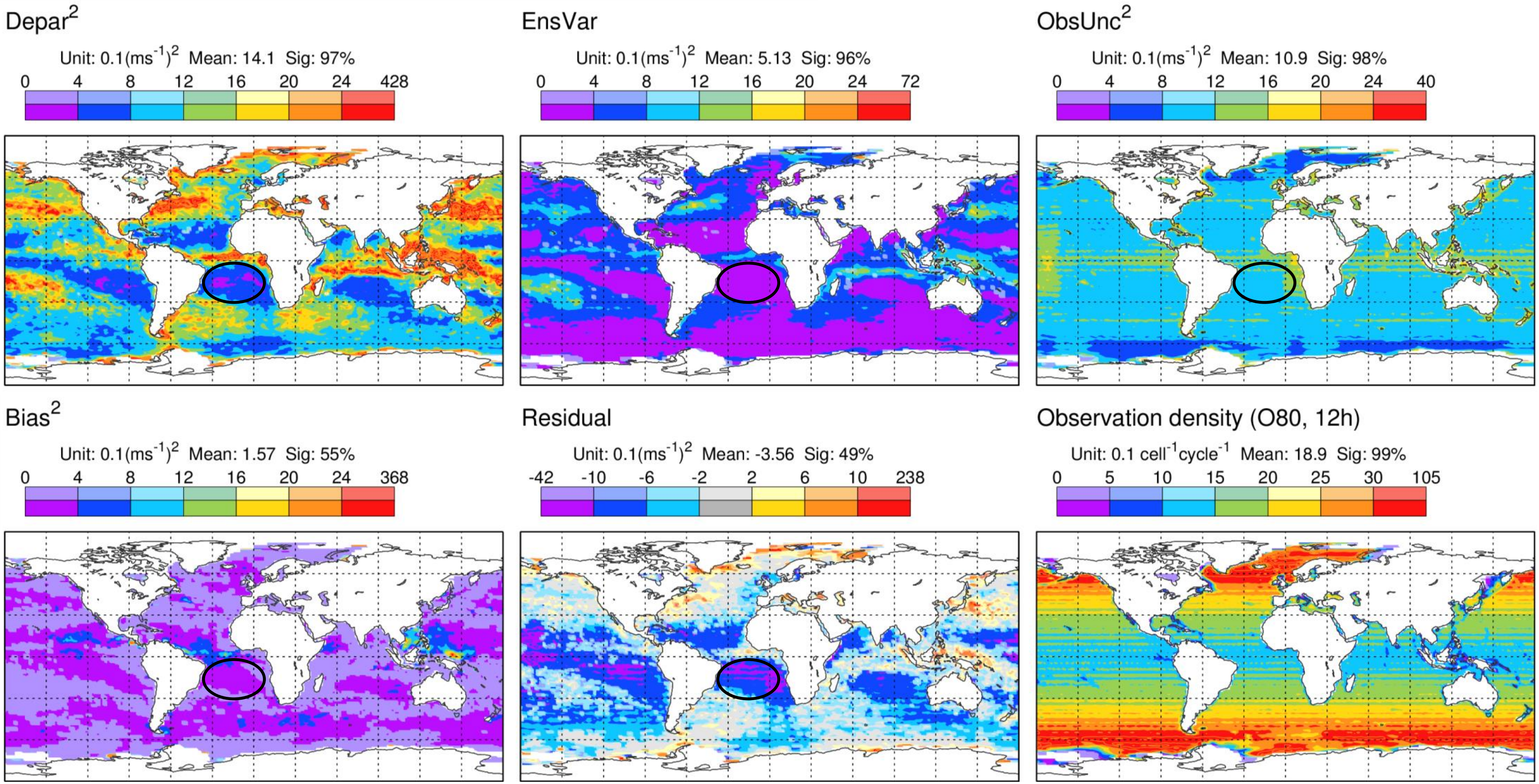
DJF 2016/17 0 & 12Z Operational HRES analyses

Do we under-estimate ASCAT observation uncertainty? – EDA says “No”!

EDA variance budget: **$Depar^2$ = EnsVar + $ObsUnc^2$ (+ $Bias^2$ + Residual)**

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

Rodwell et al. (2015)



Look in region of very small EnsVar (subtropical anticyclones)

Still have a large Residual ($<-1m^2s^{-1}$)

Hence $ObsUnc^2$ is most likely to be **over-**estimated

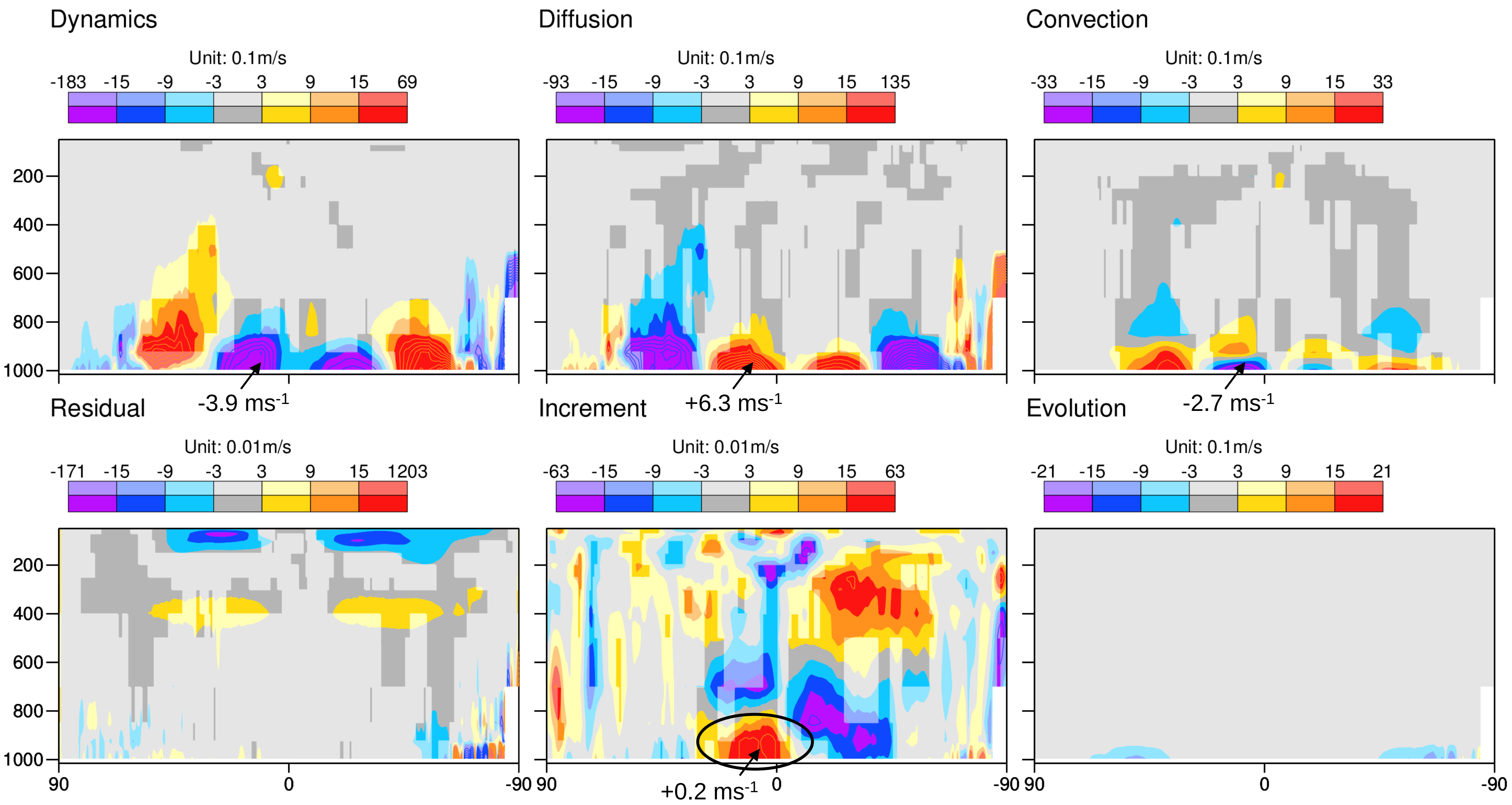
EDA DJF 2016/17

Budget of mean background process tendencies and analysis increments for [u]

To weak surface drag is the likely issue

Vertical structure of tropical low-tropospheric increment projects better onto the vertical diffusion term than onto the convective (momentum transport) term

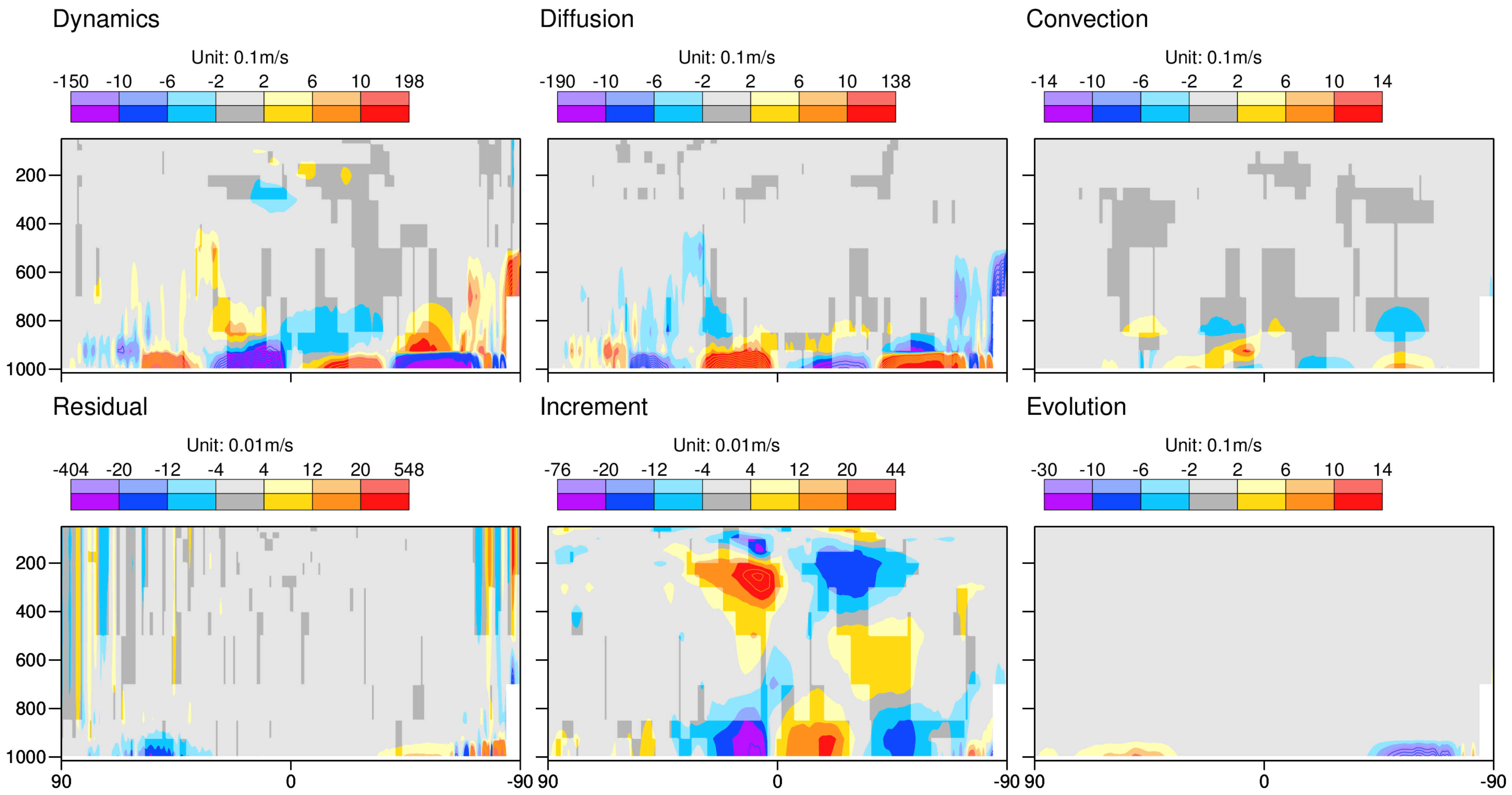
Motivation for drag experiments



Data based on background forecast of the EDA control for DJF 2015/16. Tendencies are integrated over the data assimilation window (12 hours)

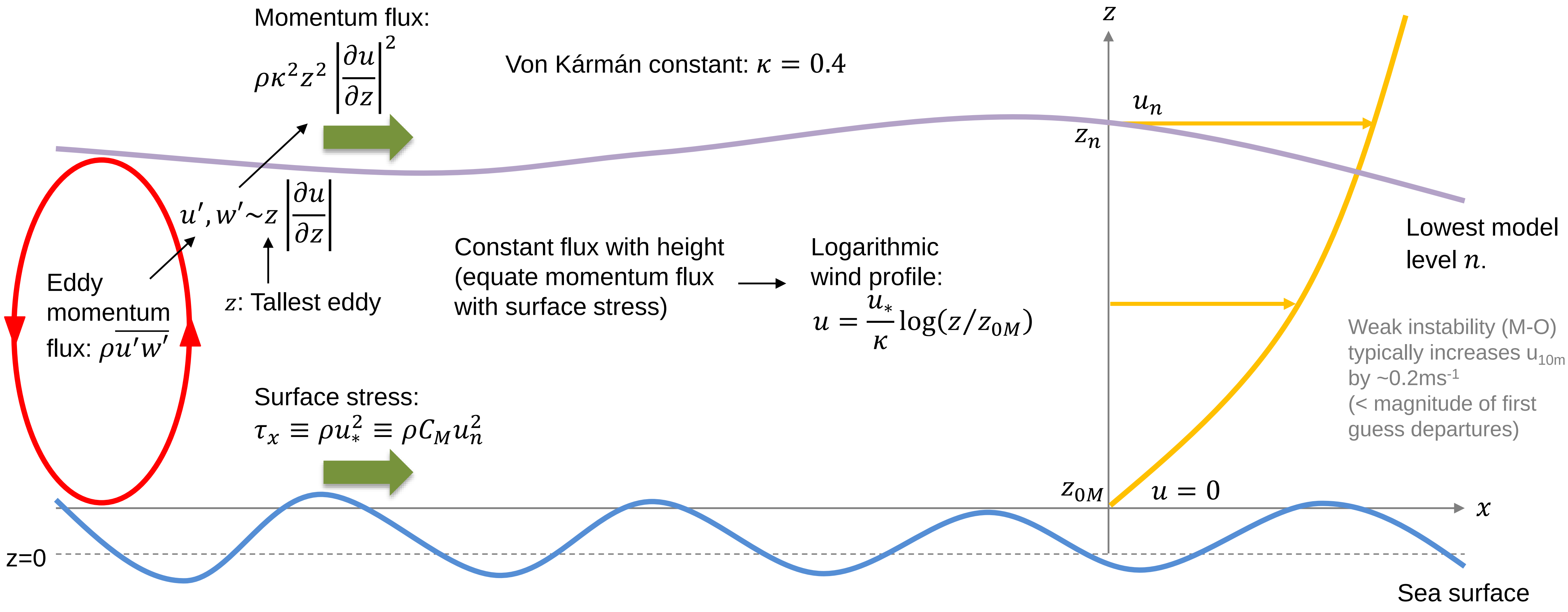
Budget of mean background process tendencies and analysis increments for [v]

Increments act to strengthen convergence in the ITCZ. In doing so, they are acting against the diffusive tendency associated with surface drag. Will suggest this is an indirect effect



Data based on background forecast of the EDA control for DJF 2015/16. Tendencies are integrated over the data assimilation window (12 hours)

Momentum fluxes in the Surface Boundary Layer (neutrally stable, zonal flow)



Friction velocity: u_*

Transfer coefficient:

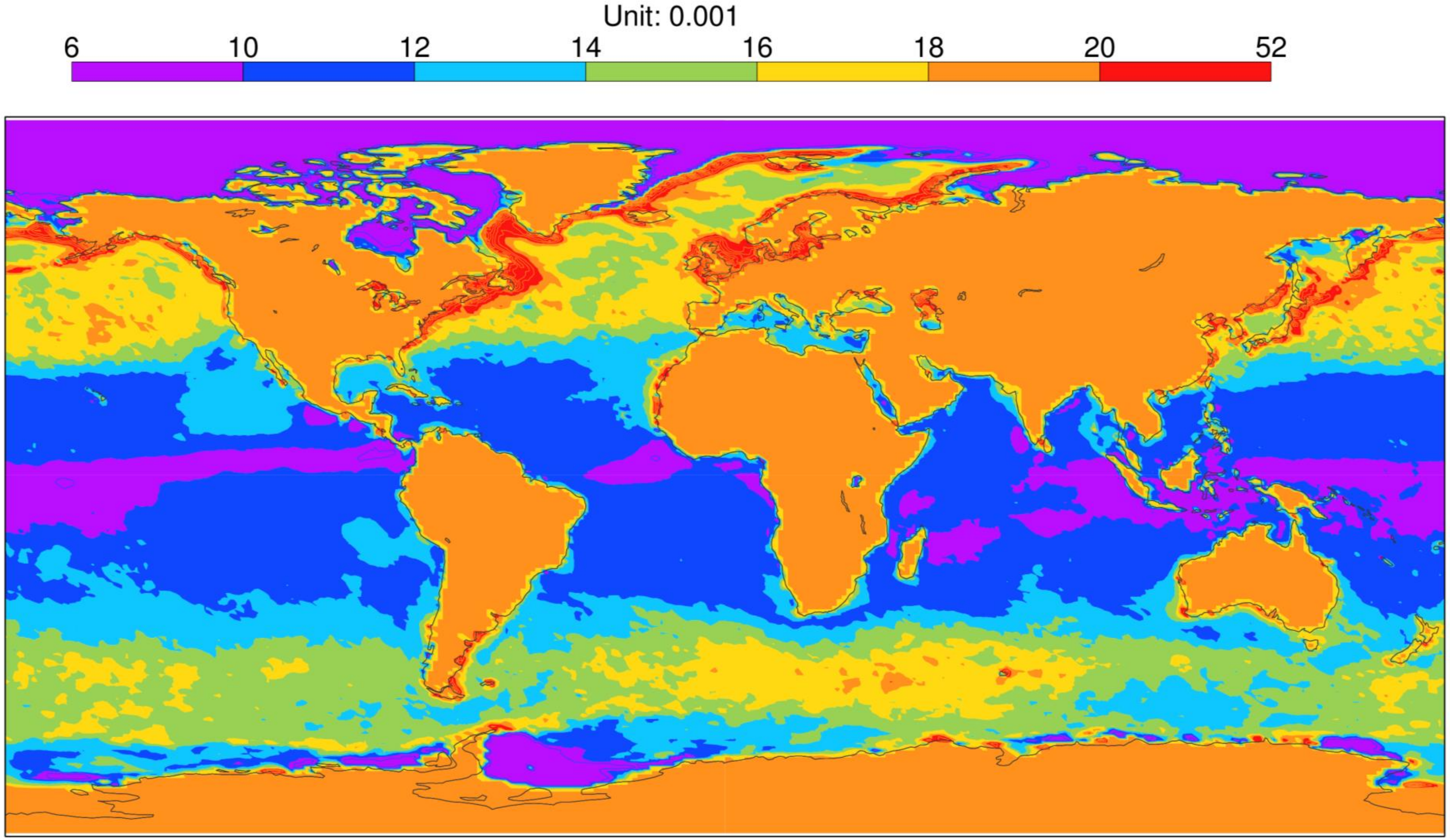
$$C_M = \frac{u_*^2}{u_n^2} = \frac{\kappa^2}{\left(\log(z_n/z_{0M})\right)^2}$$

Roughness length: $z_{0M} = \alpha_{Ch}\frac{u_*^2}{g}$

Charnock Parameter: α_{Ch} (from Wave model)

Turning off wave model fixes $\alpha_{Ch} = 0.018$ and affects coefs for momentum, heat and moisture C_M, C_H, C_Q. In drag expts, C_M alone is scaled.

Charnock parameter from wave model (DJF 2016)



The wave model produces values in the tropics below 0.010

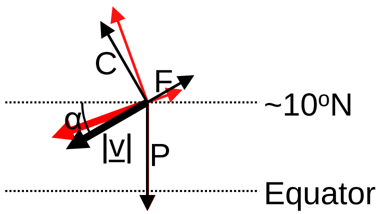
The historical uniform value is 0.018

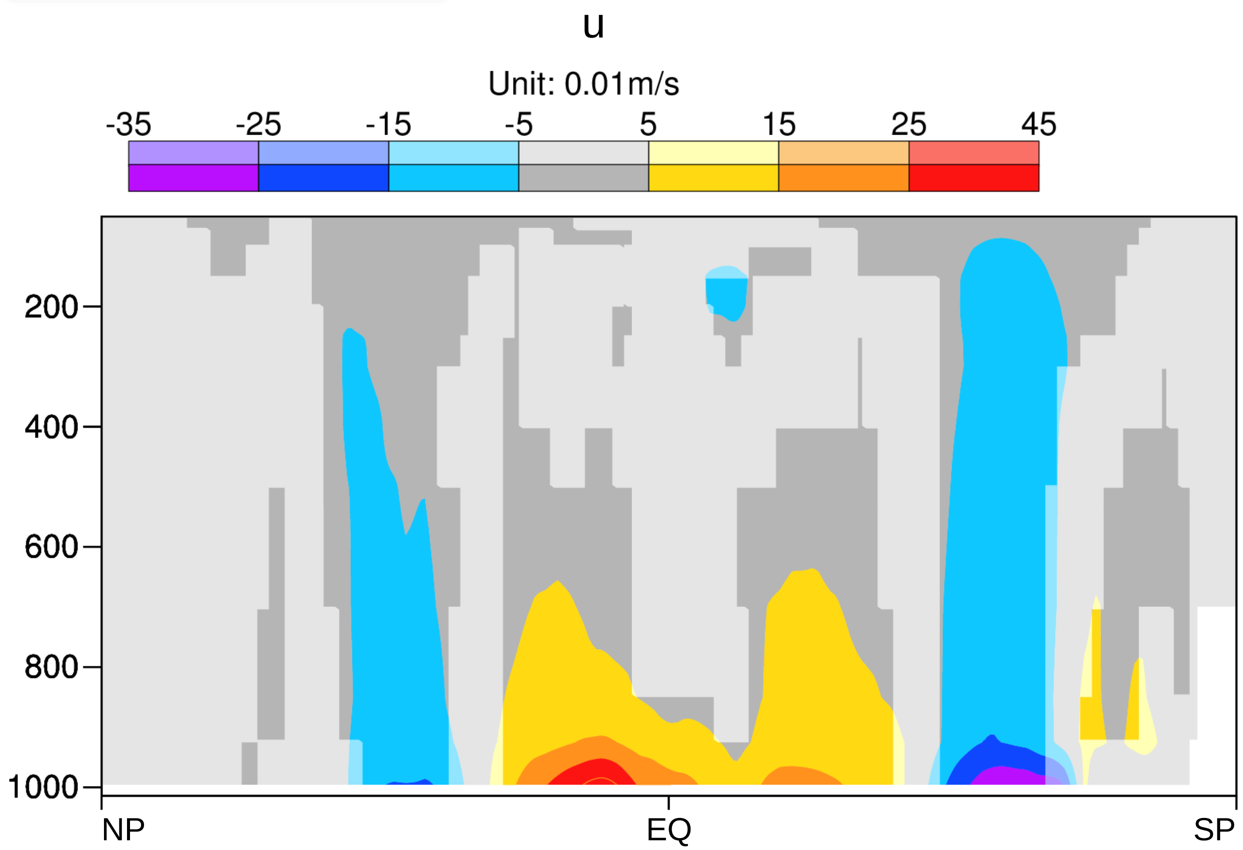
Based on HRES analyses for 0 and 12Z 20151201-20160228

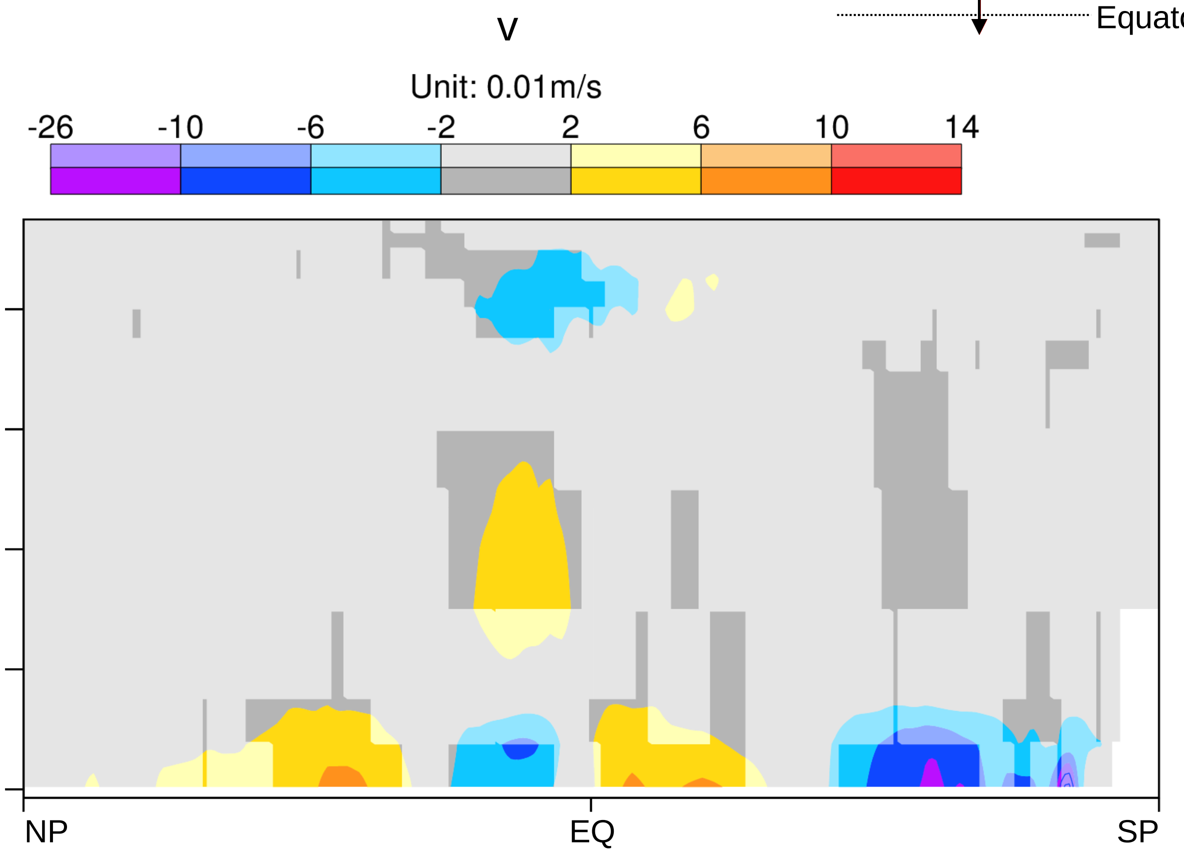
Zonal-mean change at day 1 for 110% C_M – 90% C_M (20% increase in transfer coef.)

Change would reduce mean boundary-layer wind errors almost everywhere

$F \cos(\alpha) - C \sin(\alpha) = 0$
$F \sin(\alpha) + C \cos(\alpha) = P$
$C = f\,|\underline{v}|$







- Tropical zonal wind error reduced by up to 100%
- Extra-tropical zonal wind errors greatly reduced

- Tropical meridional wind errors reduced by ~20%
- Antarctic convergence error reduced by ~50%

Control and drag experiments for 28 forecasts started at 0Z 20140201-20140228

Outline

- The role of operational diagnostics
- Mean error
- Variance error (& predictability)
- A diagnostic framework for forecast system development

Ensemble spread and error

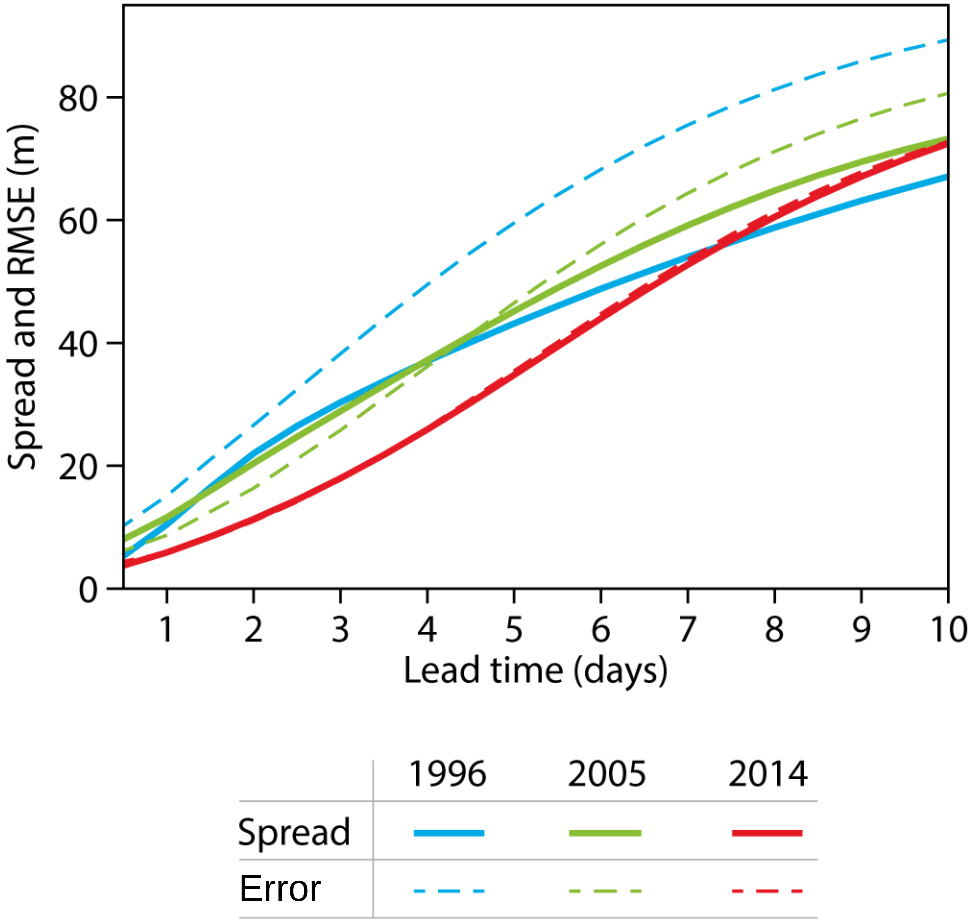
Z500

Rodwell et al (2018)

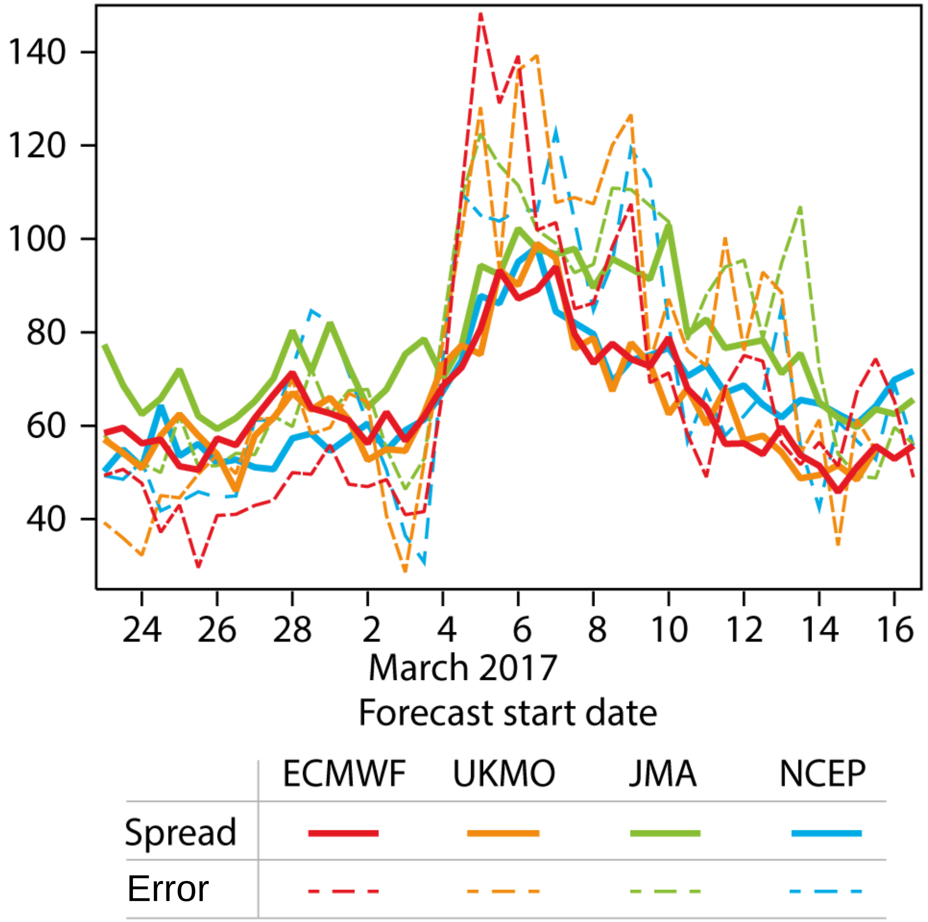
> *Overall* Error and Spread have reduced and come into alignment; due to better observations, initial conditions, forecast model and better representation of uncertainty

> *…but we make ensemble forecasts to represent the day-to-day variations in predictability and uncertainty…(so the job is not complete)*

Annual means (ECMWF)



Timeseries for 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)

"Instantaneous" (0-12h) uncertainty growth-rates for $PV_{\theta=315K}$ following the flow

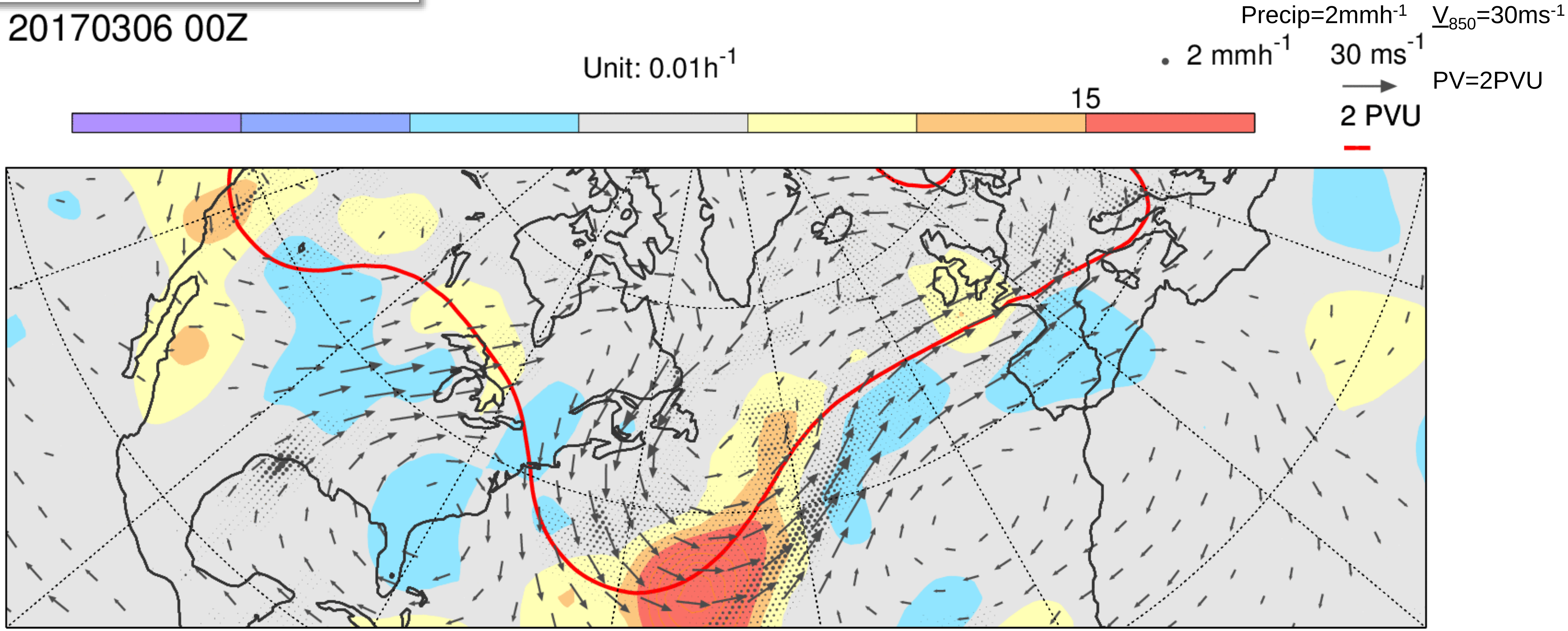
Large growth rates associated with:

- Mesoscale convection over North America associated with Trough and CAPE
- Warm Conveyor Belts over the North Atlantic

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

Complicated interactions hinder direct diagnosis of medium-range ensemble deficiencies

Focus on short-range flow-dependent reliability?



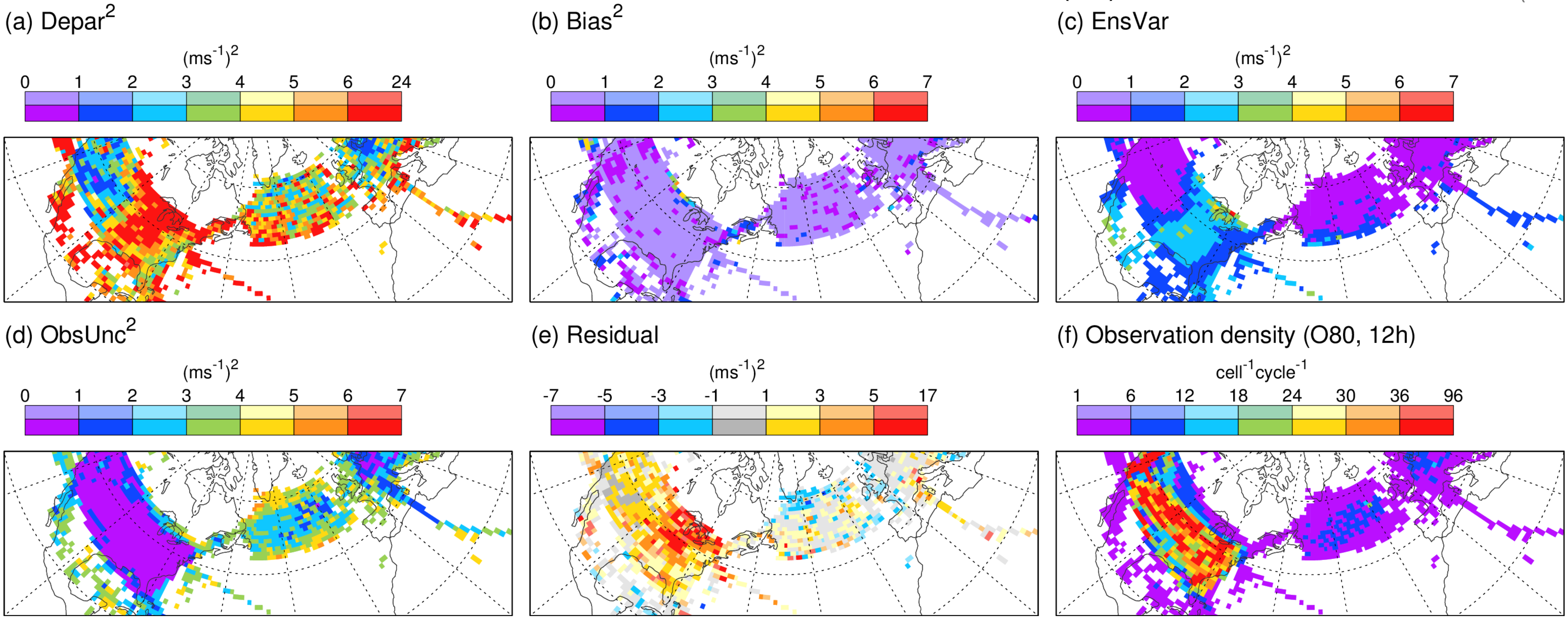
PV_{315}=2 & $\underline{v}_{850}$ from control forecast, precipitation is ensemble-mean. 1d running-mean gives 12h-integrated growth rate with any diurnal cycle removed. T21 smoothed

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 et al (2018)



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

Observation uncertainty (ObsUnc^2) quite small so a statistically significant positive Residual ⇒

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

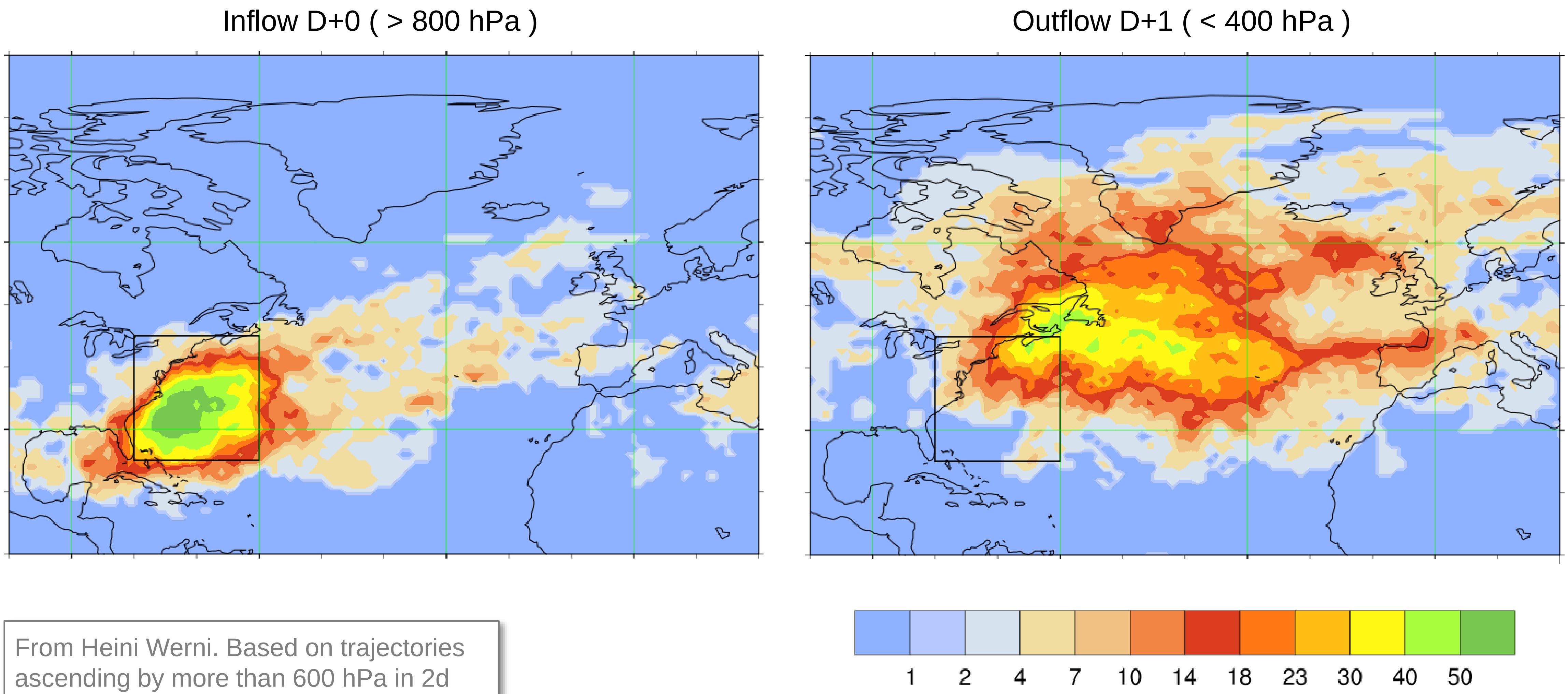
Extend "spread-error" relation to include obs error variances and bias (similar to data assimilation)

~~Error^2 = EnsVar + Residual~~

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

Reliability ⇒ E[Residual]=0

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

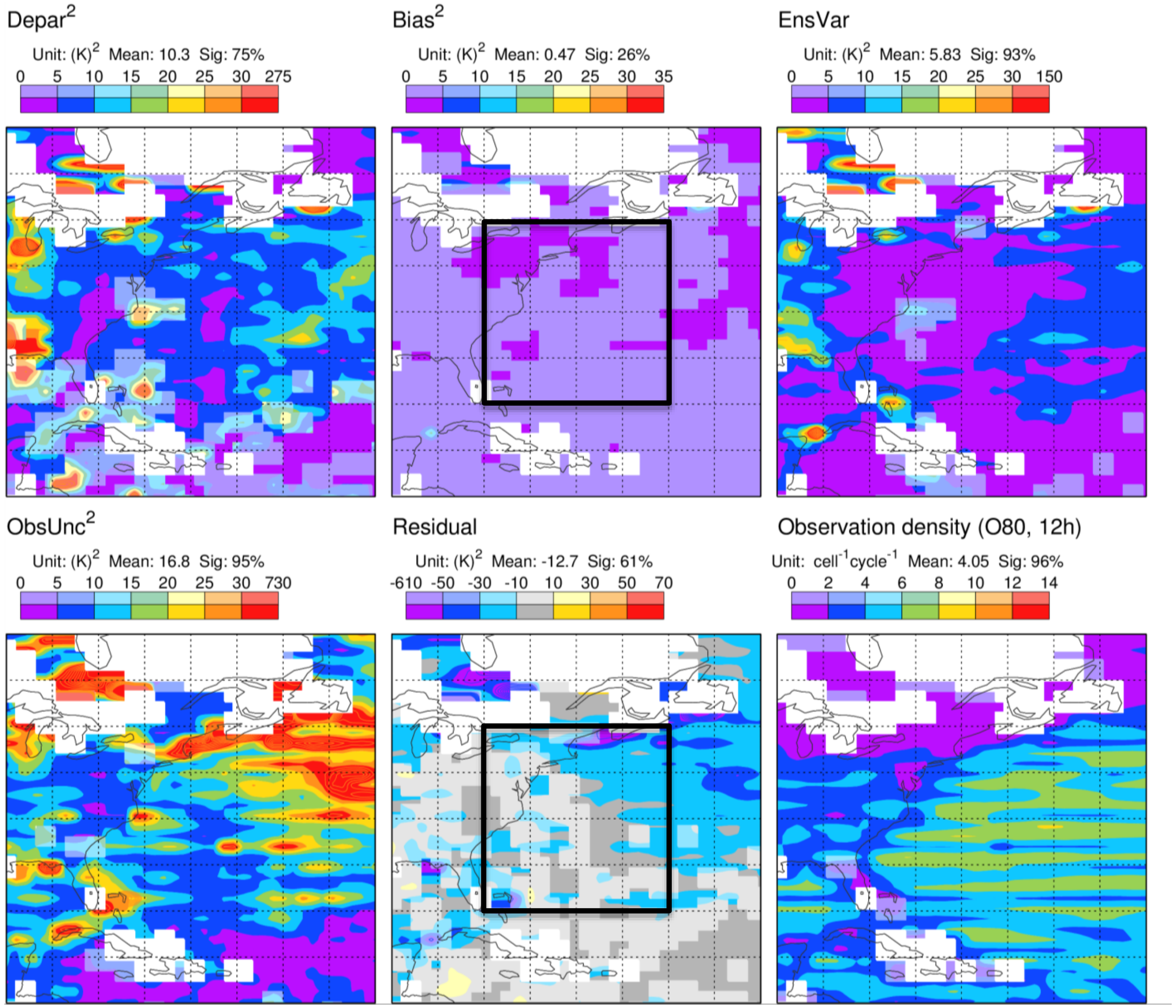


From Heini Werni. Based on trajectories ascending by more than 600 hPa in 2d

EDA variance assessment with 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

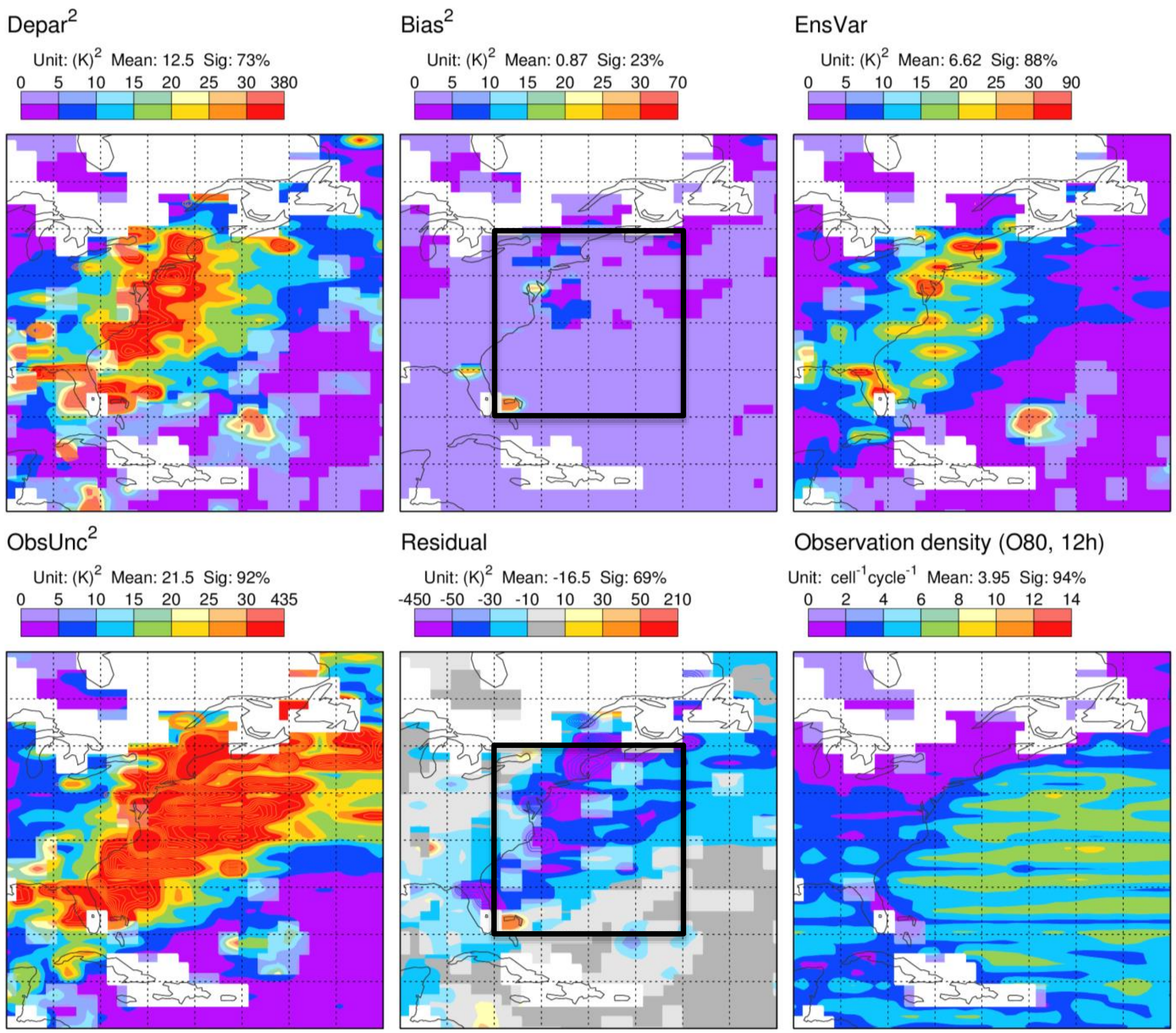
Increased Depar^2 and EnsVar in WCB situations

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

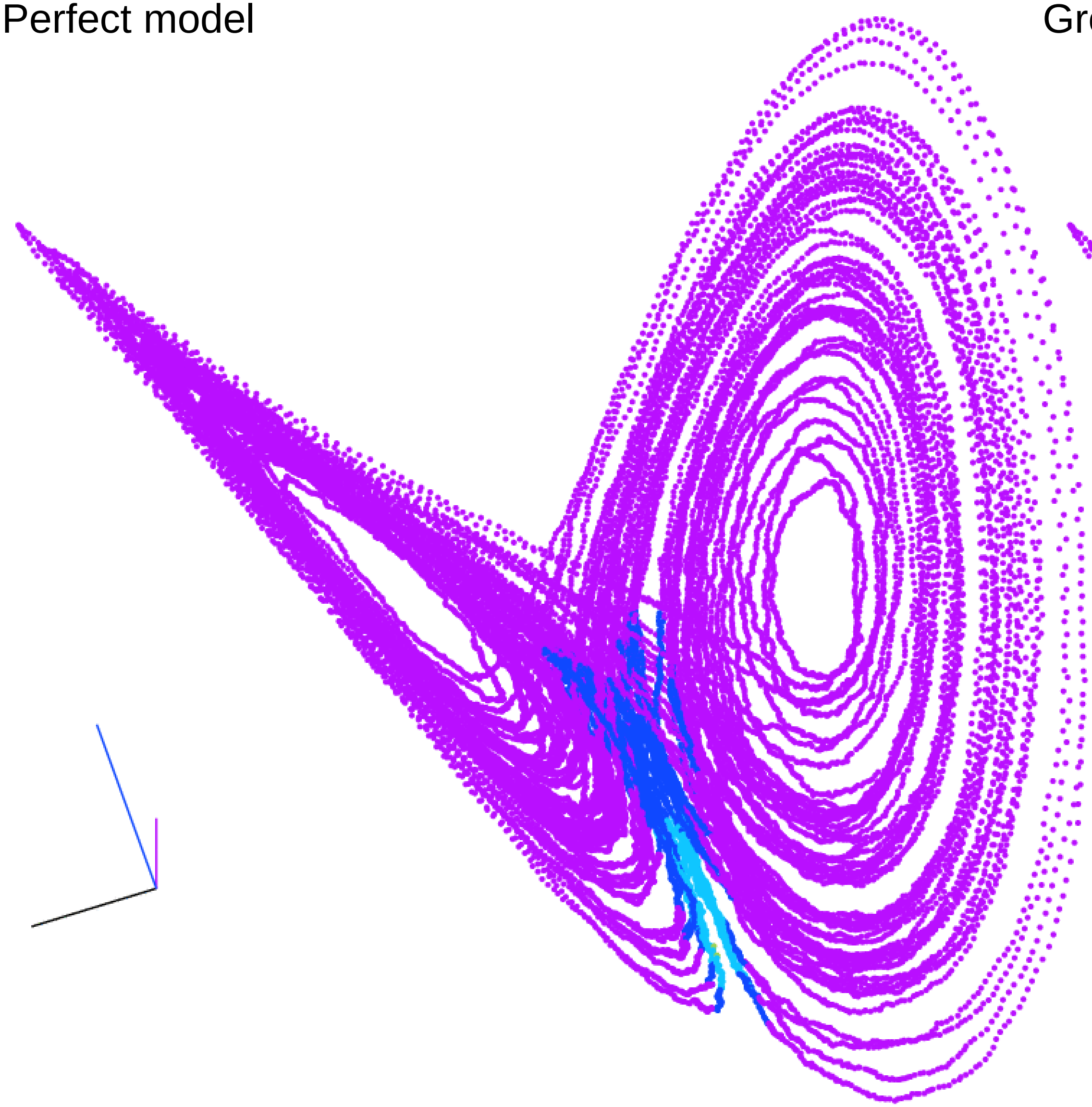
Microwave channel 5

Outline

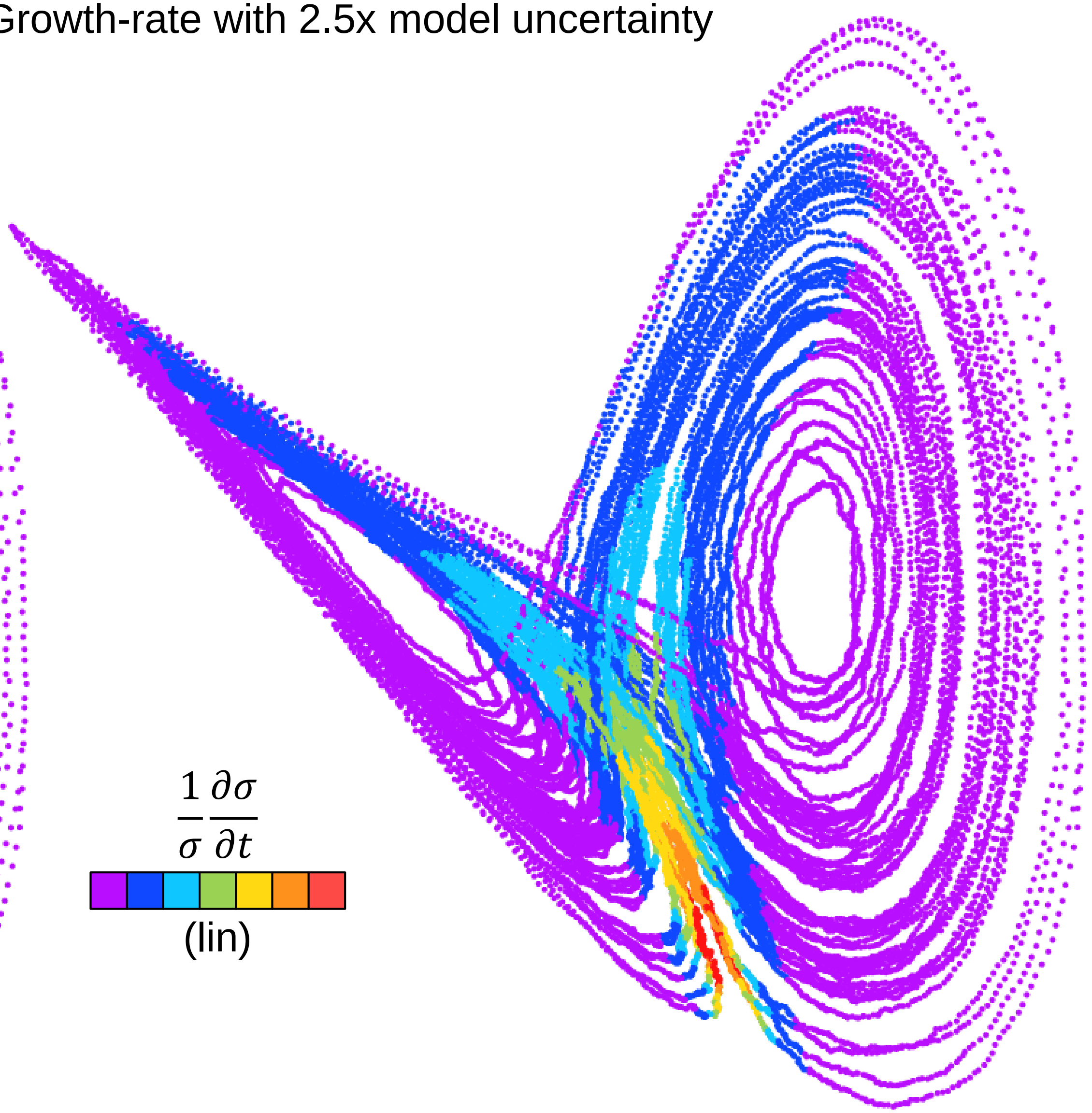
- The role of operational diagnostics
- Mean error
- Variance error (& predictability)
- A diagnostic framework for forecast system development

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

Perfect model



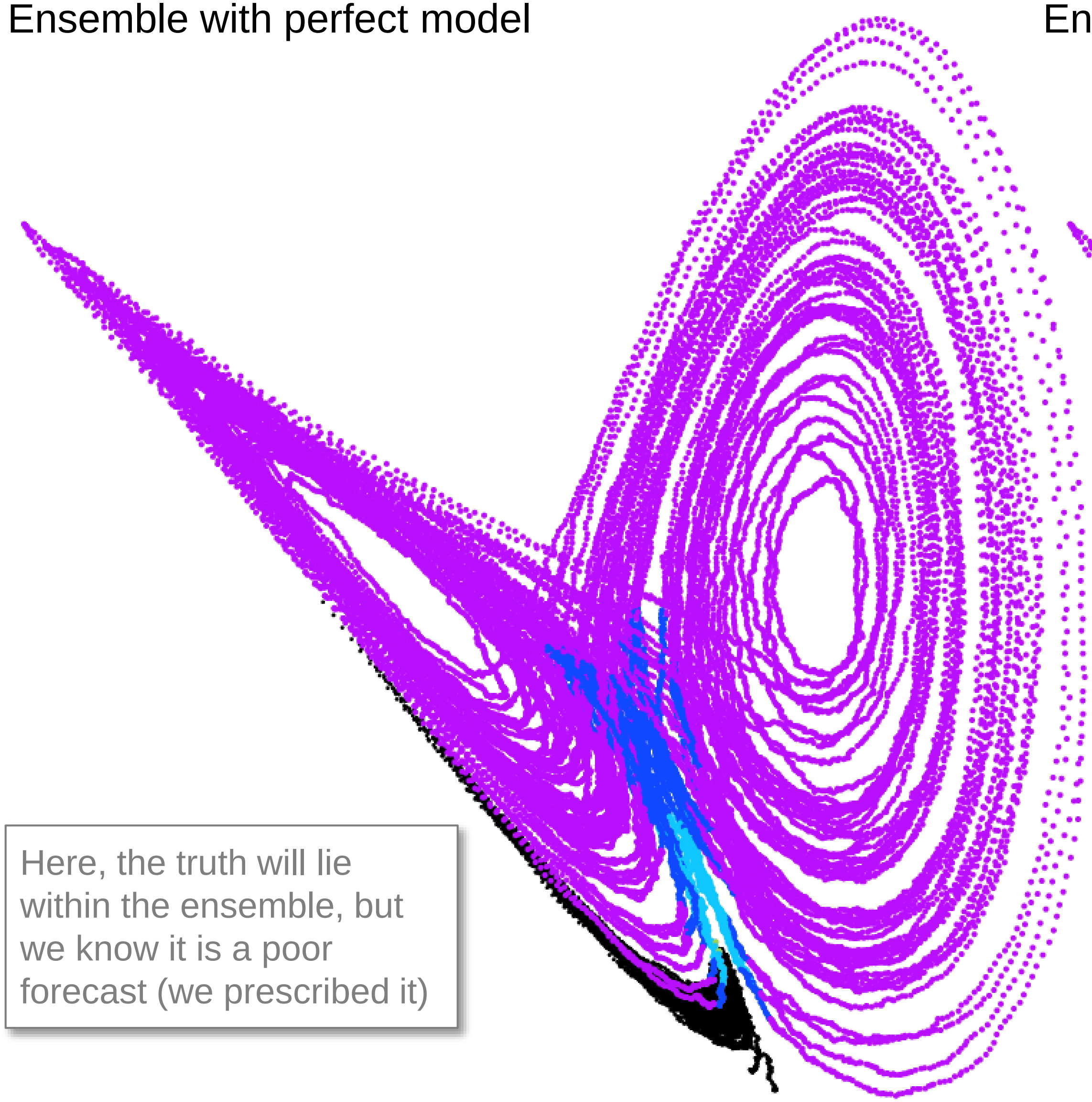
Growth-rate with 2.5x model uncertainty



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

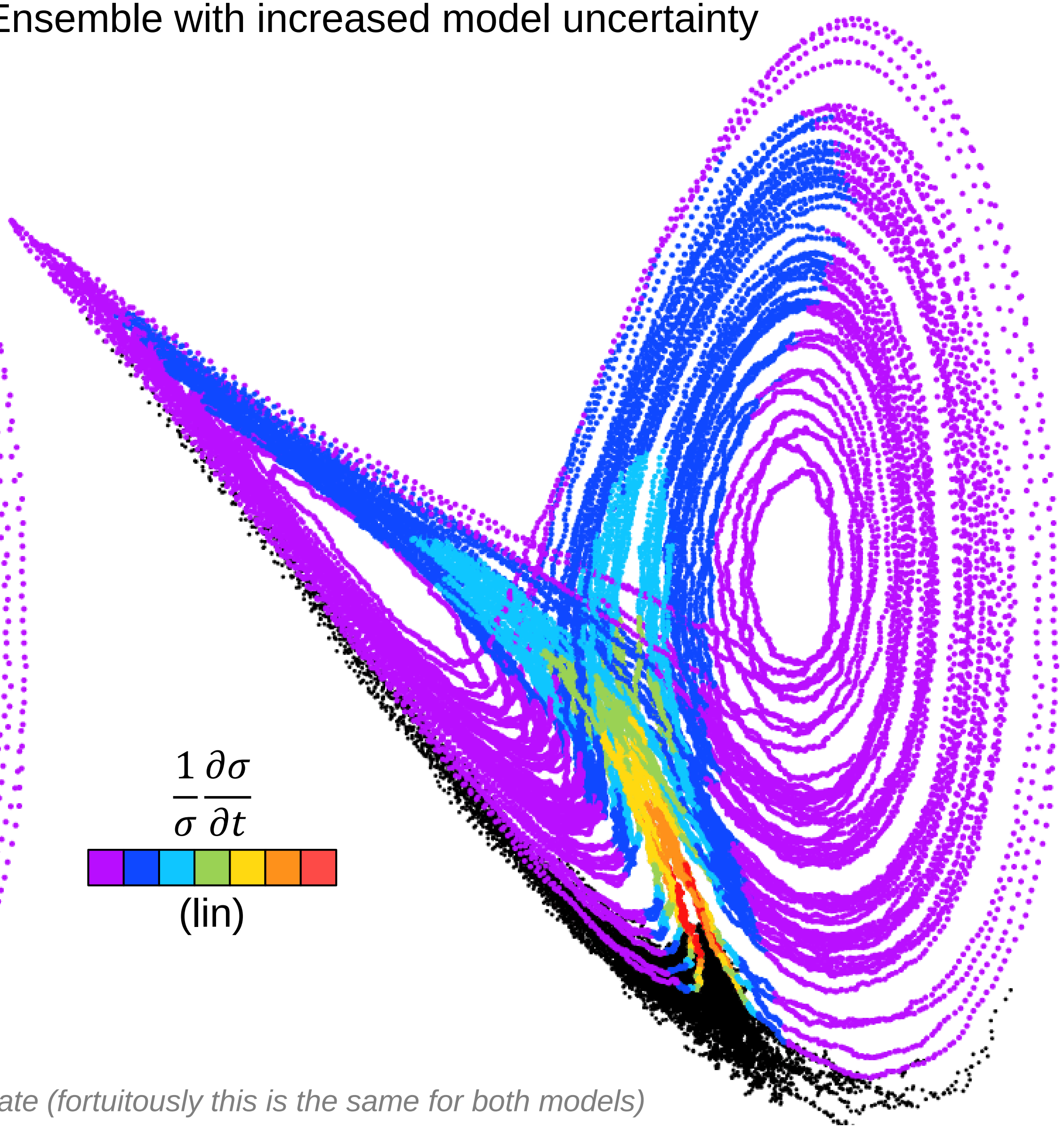
“van Lorenz” attractor: Forecast with fastest uncertainty growth-rate (black)

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



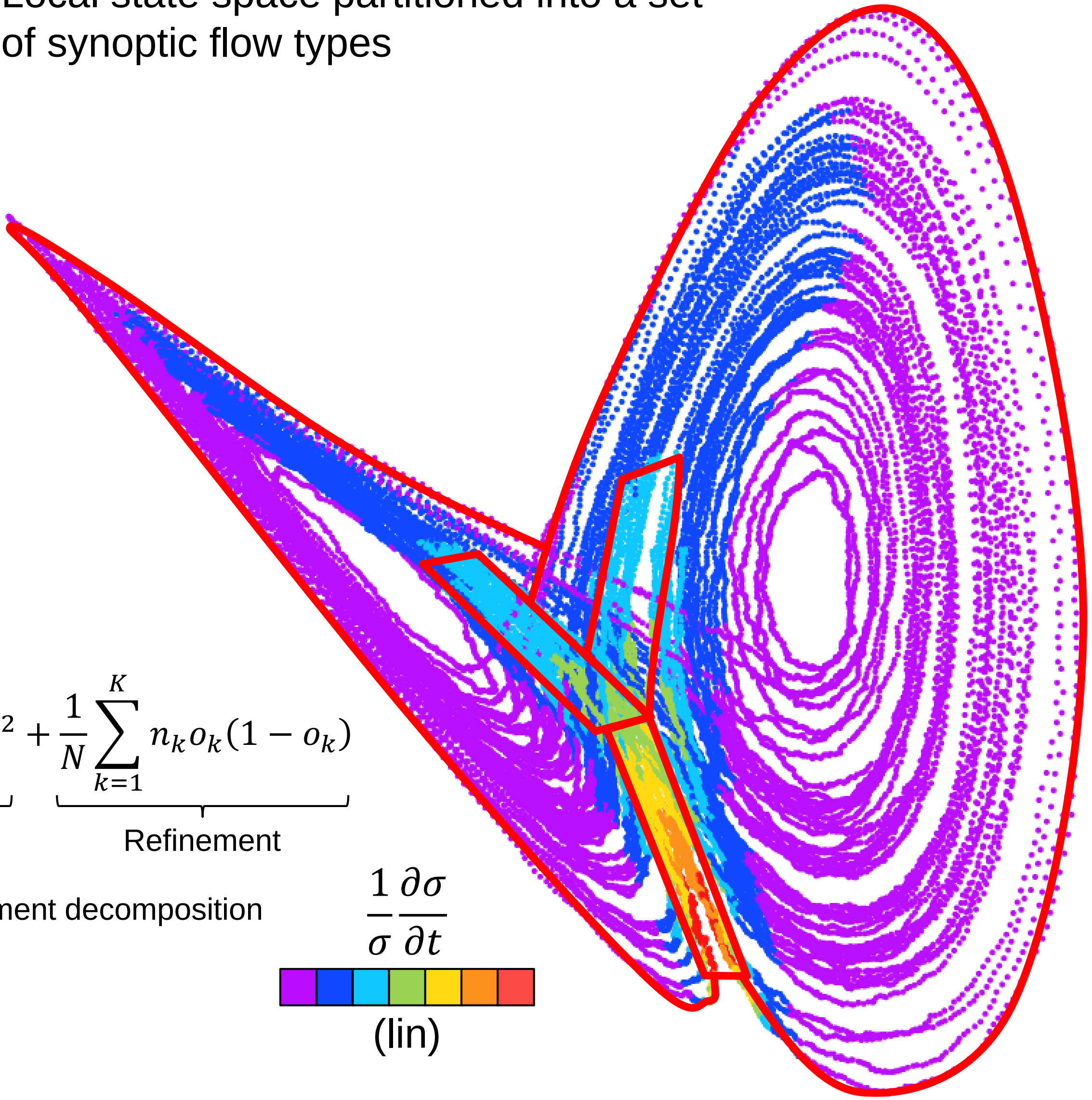
$$\frac{1}{\sigma}\frac{\partial \sigma}{\partial t}$$

(lin)

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

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

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

- Better skill at short-ranges (and thus into the medium-range)
- Better model and representation of uncertainty at all lead-times

Can prioritise efforts on flow-types that contribute most to reliability aspect of a proper score

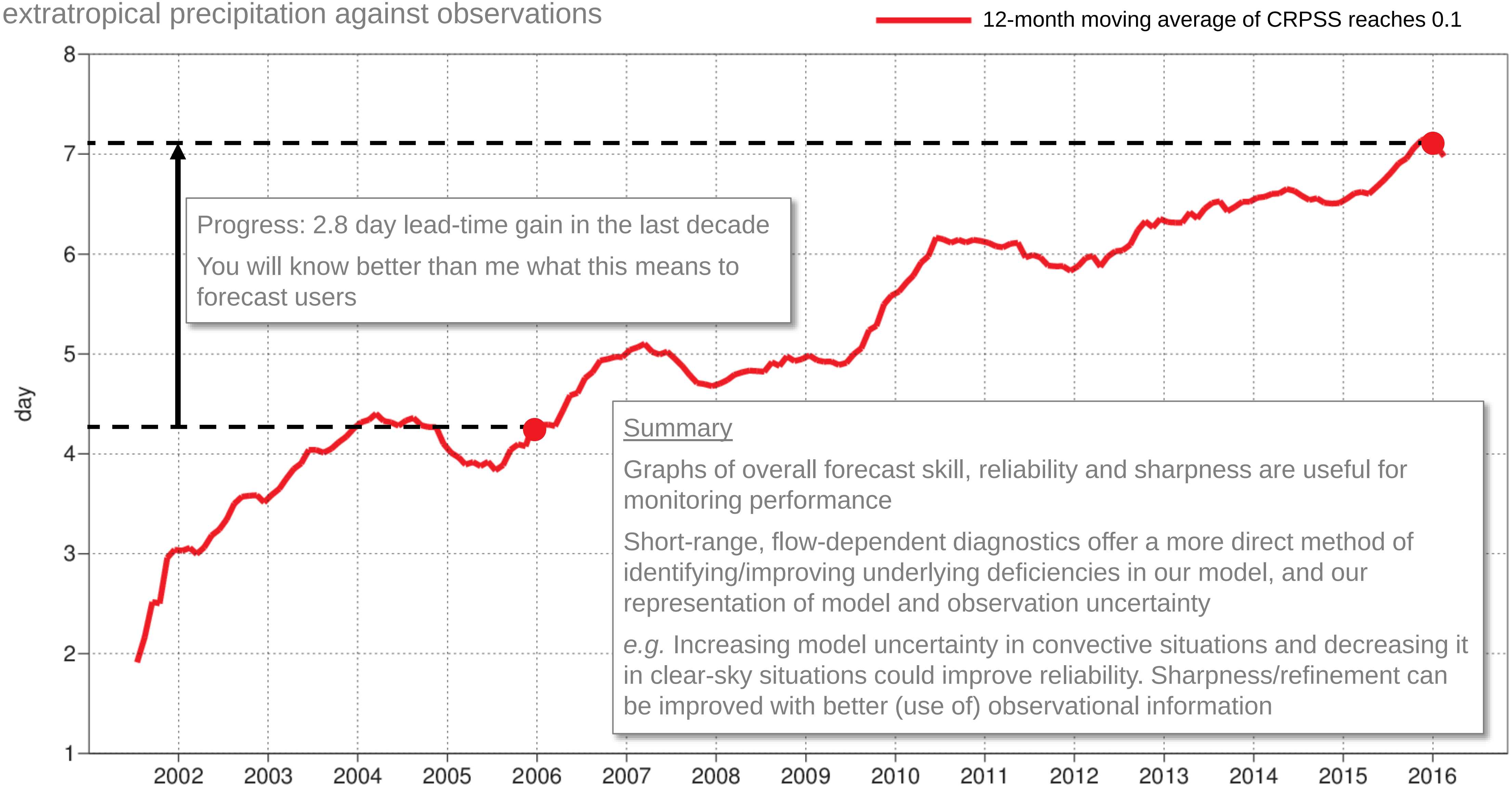
(Better observational information should improve the refinement aspect)

Thought experiment:

- Think of the $k = 1, \ldots, K$ as a partition of initial local flow types
- (Probabilities will be reasonably constant for a given flow type if the flow-types are defined tightly-enough, and the event is local and at short-range).
- 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.

Trend in probabilistic forecast performance & Summary

CRPSS, extratropical precipitation against observations



Progress: 2.8 day lead-time gain in the last decade

You will know better than me what this means to forecast users

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 model, and our representation of model and observation uncertainty

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

Thank you

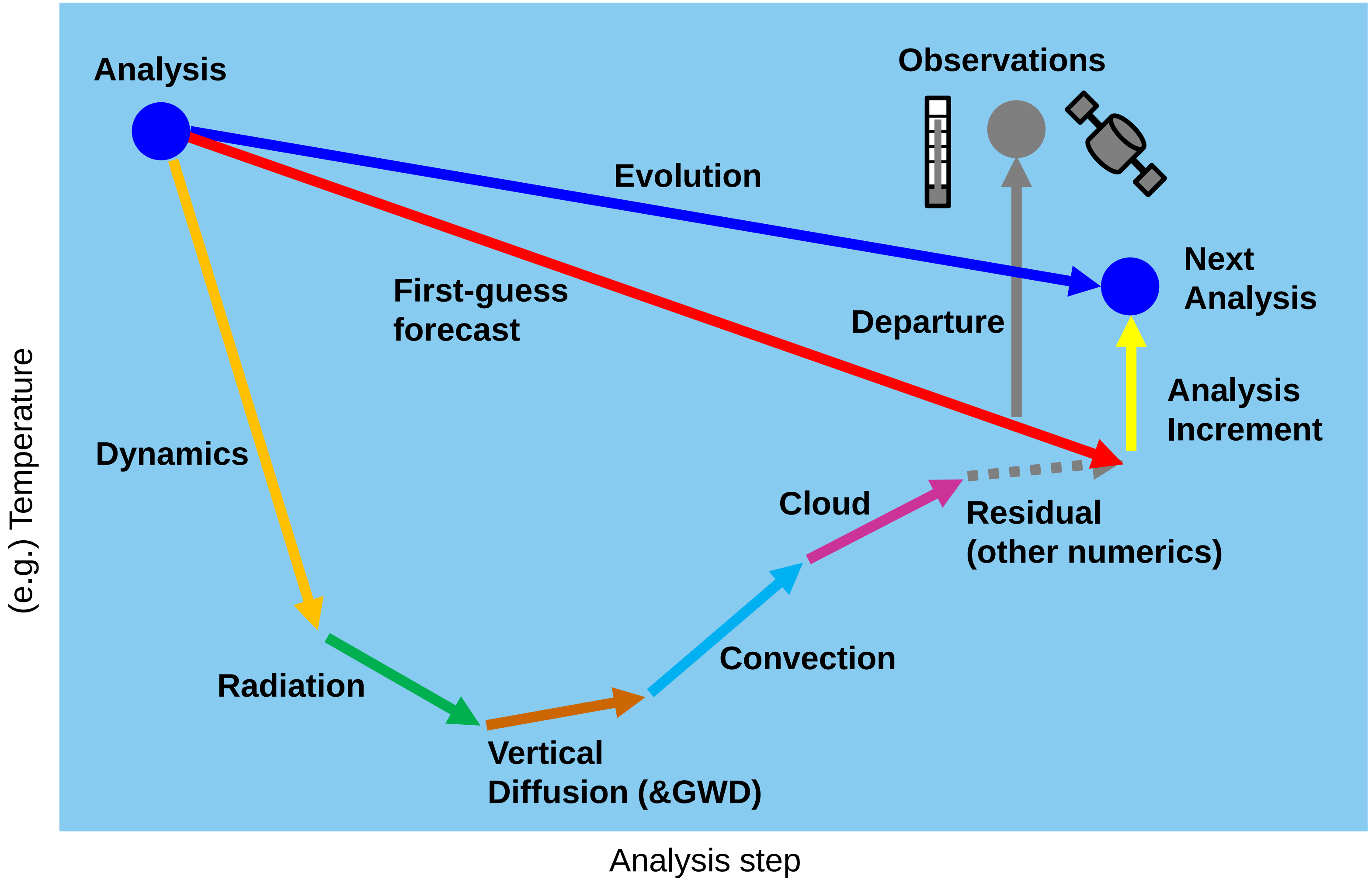
The Initial Tendency approach to diagnosing model error

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)