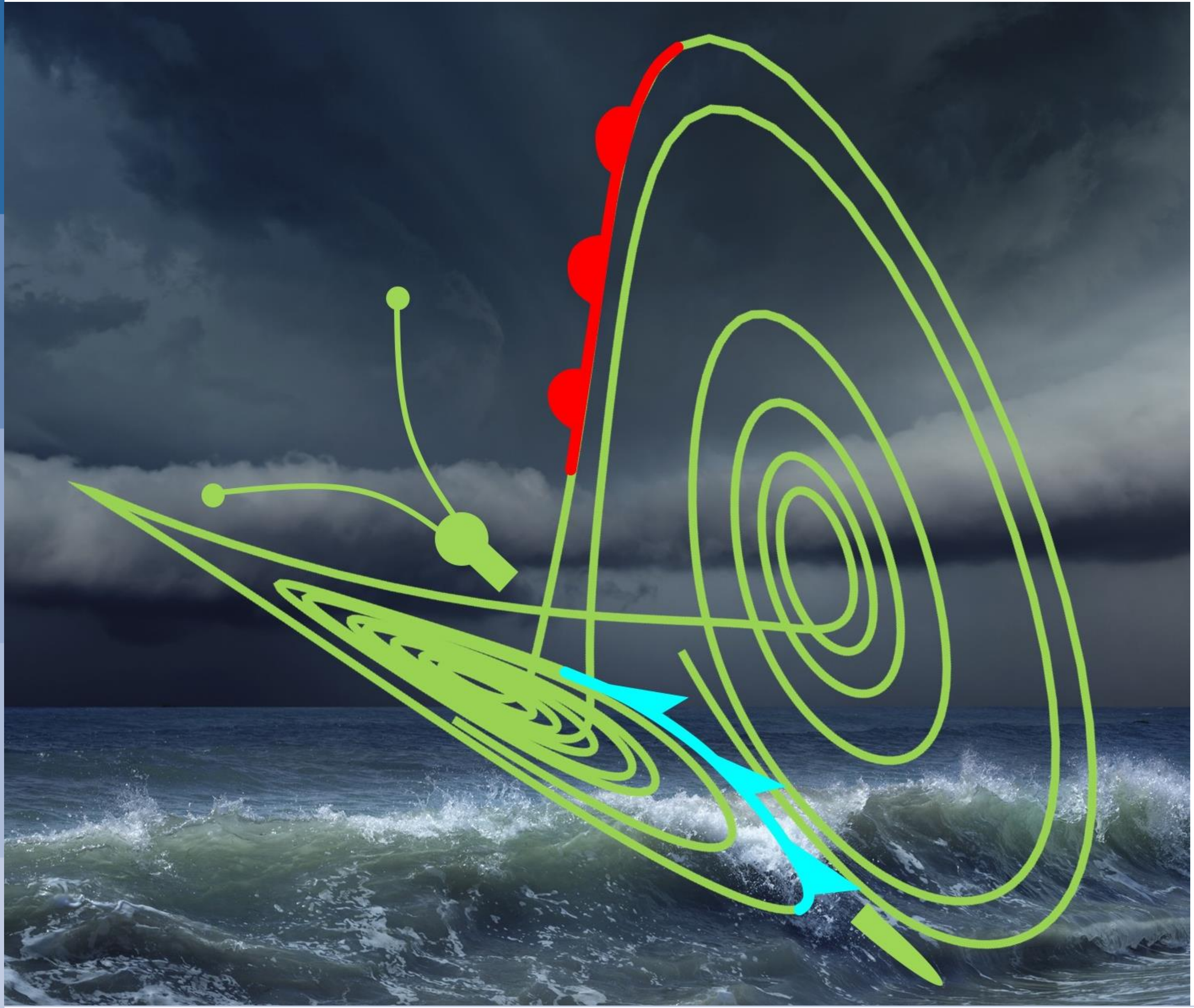


# Challenges and Limits in Ensemble Weather Prediction


Mark Rodwell

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



5th OpenIFS Workshop

19 June 2019, Reading University

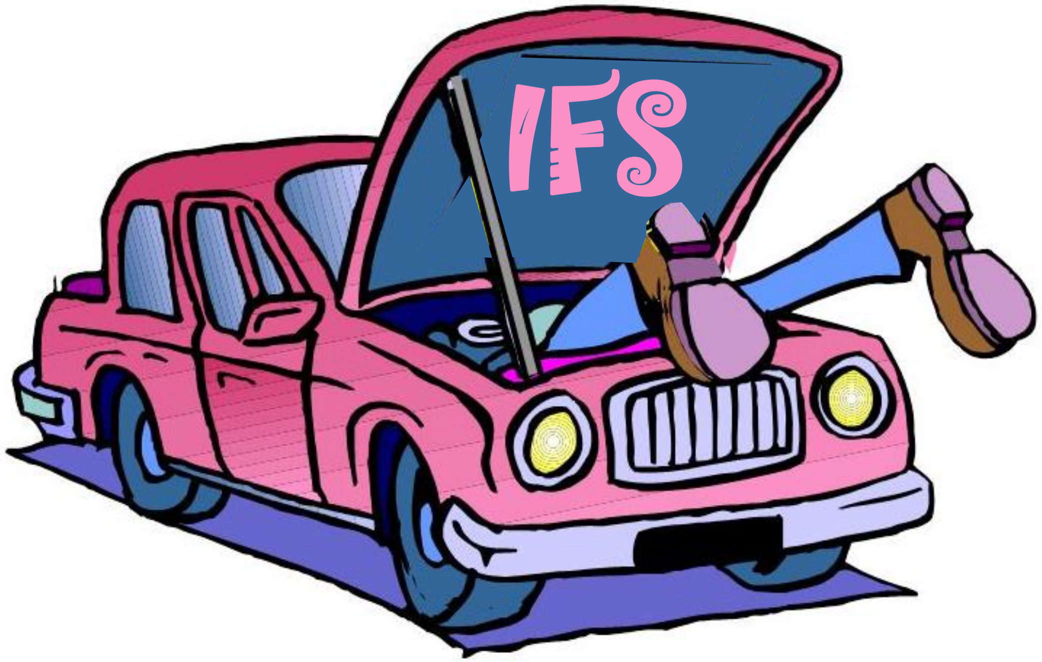
# Challenges and Limits in Ensemble Weather Prediction - Outline

- A good ensemble forecast system
  - Reliability and Sharpness
- The Perfect Storm
  - When the going gets tough …

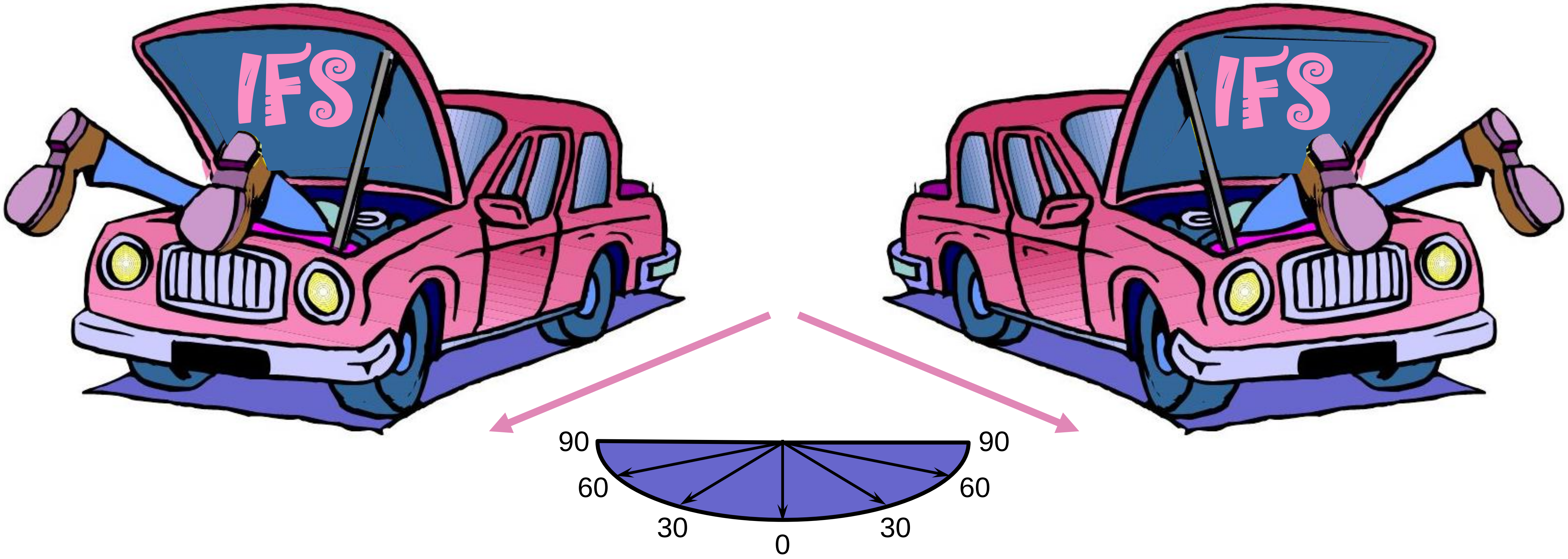
    … consult the diagnostics!!
- Extratropical transition of Tropical Cyclone Karl

# What makes a good ensemble forecast system?

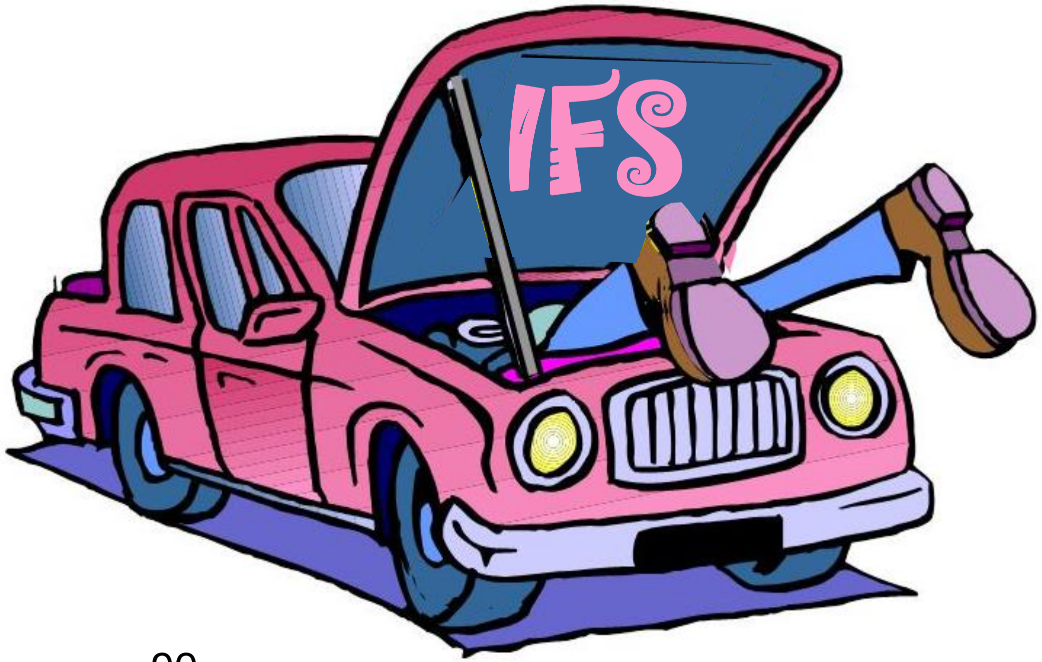
# What makes a good ensemble forecast system?

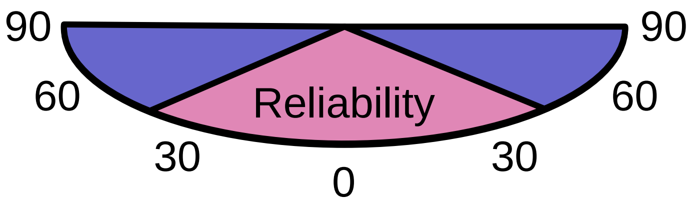



# What makes a good ensemble forecast system?

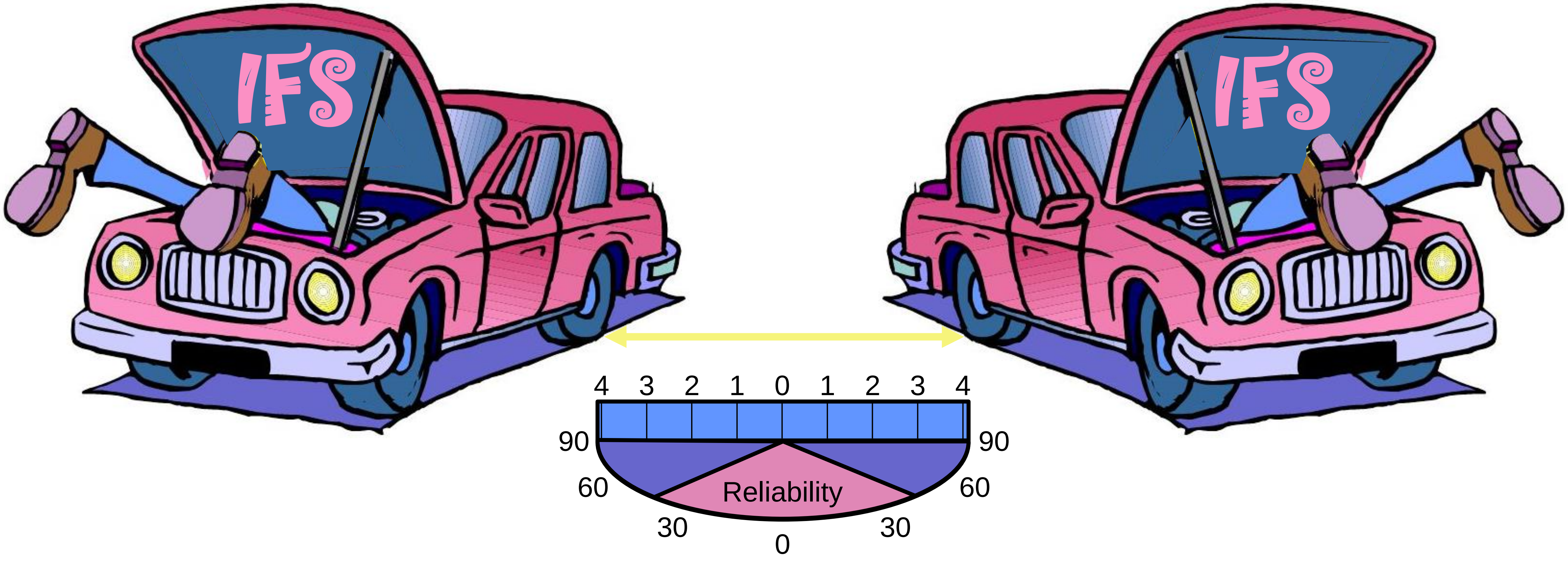


# What makes a good ensemble forecast system?

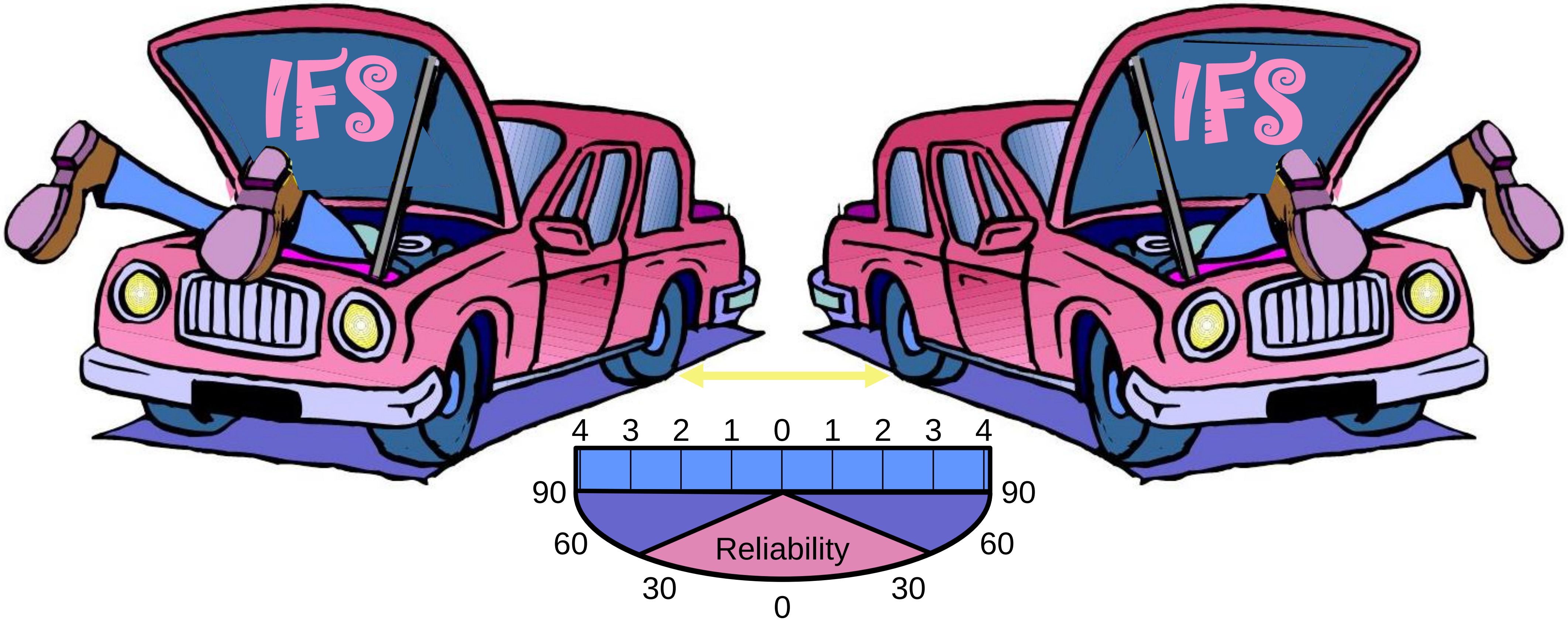



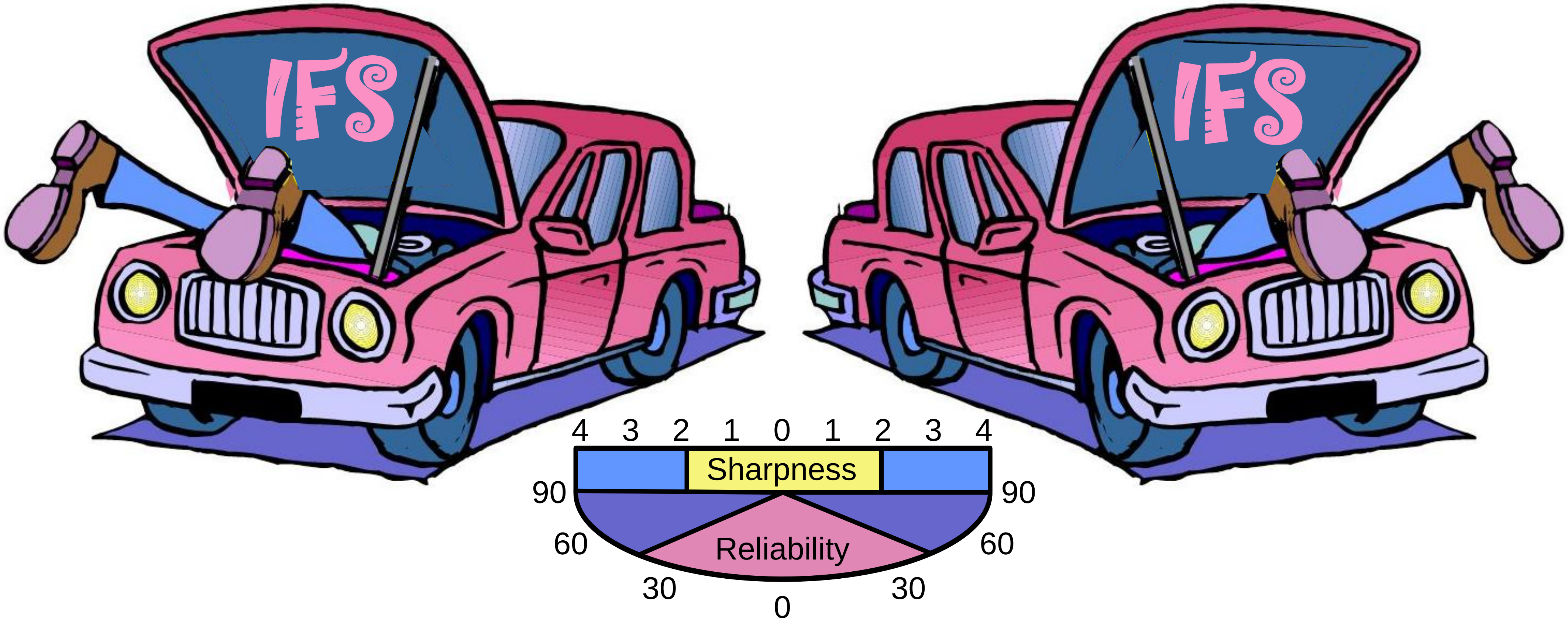
# What makes a good ensemble forecast system?



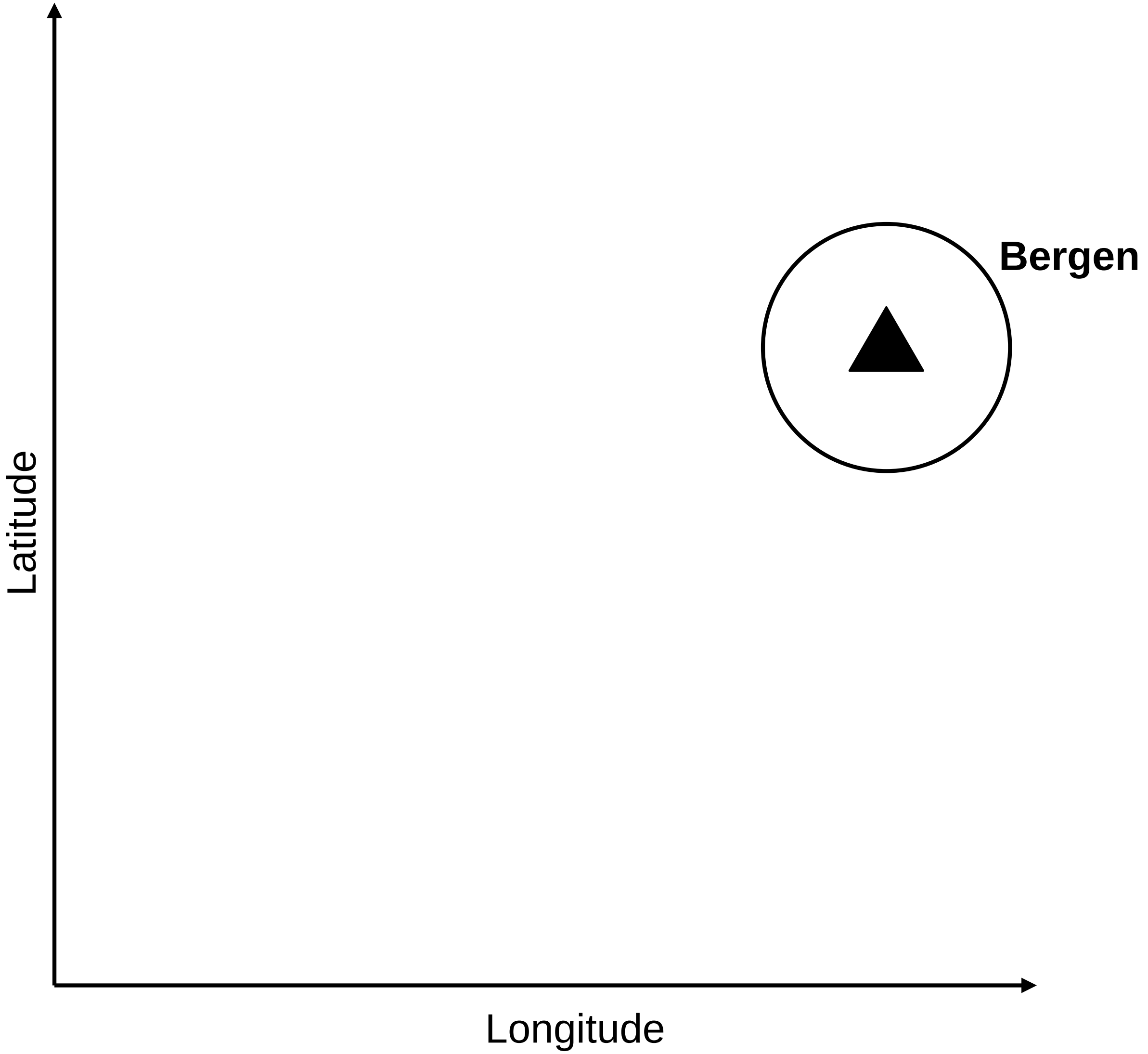
# What makes a good ensemble forecast system?



# What makes a good ensemble forecast system?



# Motivation: Reliability and Sharpness




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

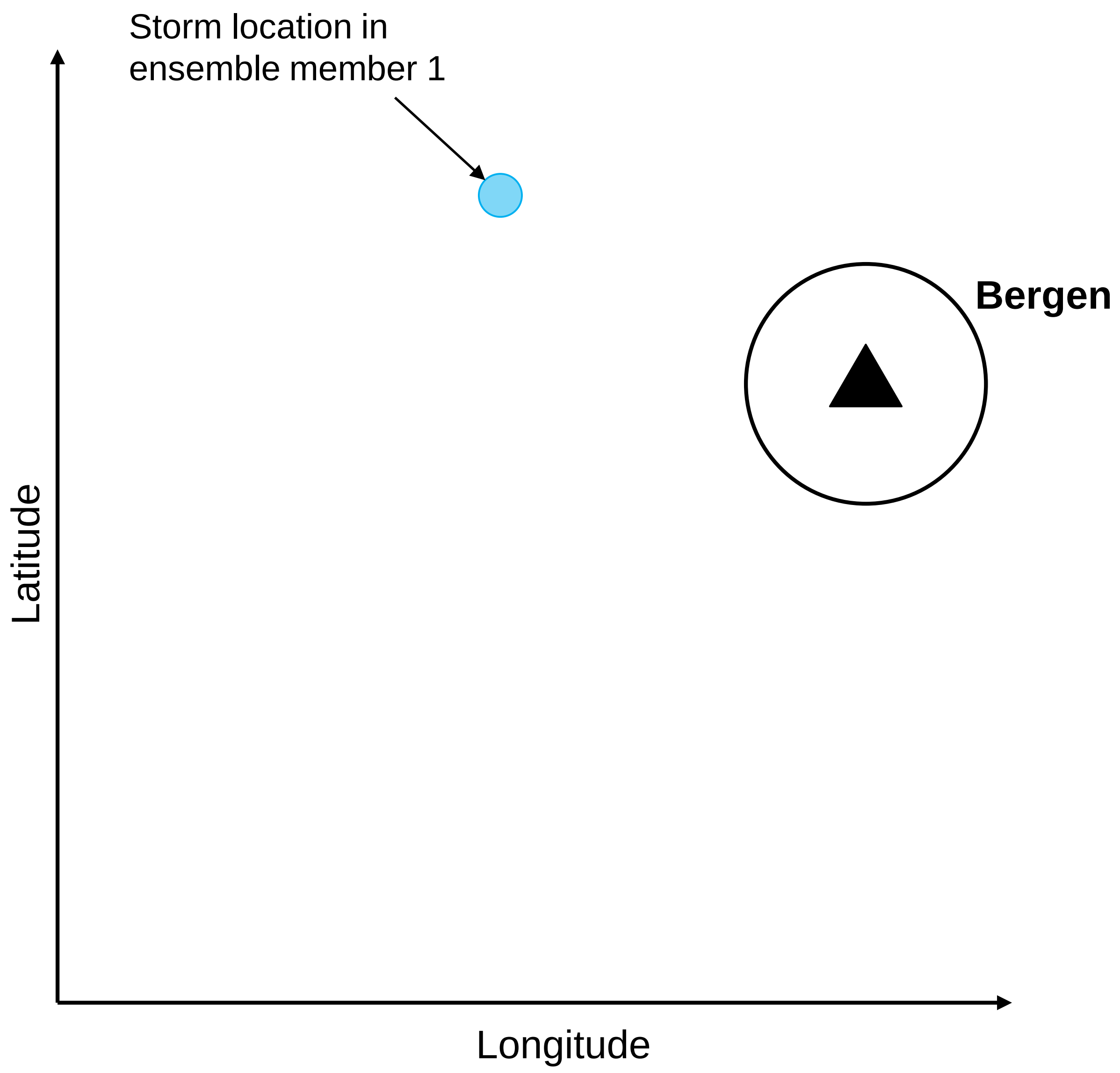
$$\overline{\mathbf{Error}^2} = \overline{\mathbf{Spread}^2}$$

(averaged over many forecast start dates)

"The task of NWP research is to maintain/improve reliability while decreasing spread (improving refinement)"

Q. Can we develop diagnostics which efficiently (optimally?) guide us in this task?

# Motivation: Reliability and Sharpness




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

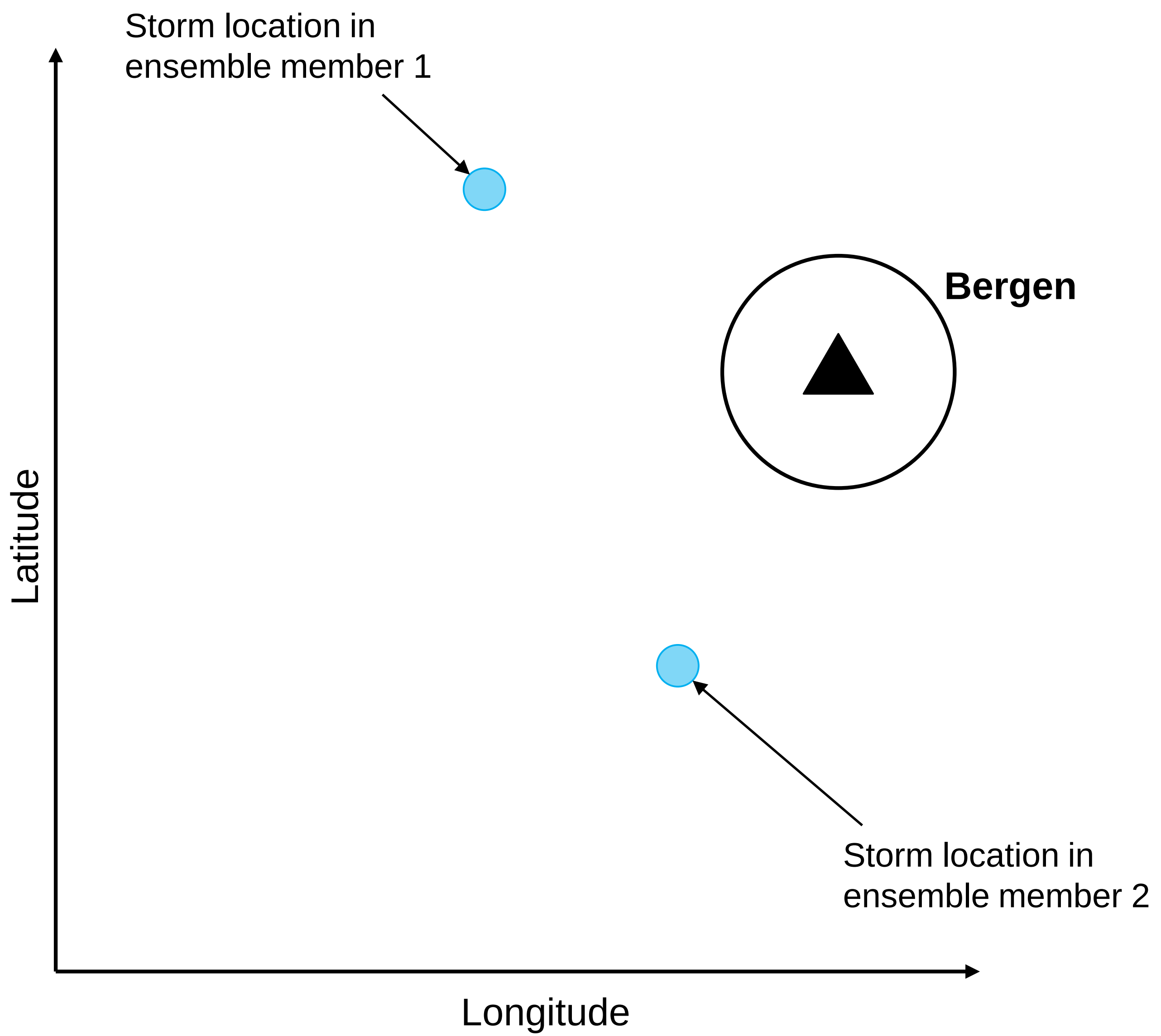
$$\overline{\mathbf{Error}^2} = \overline{\mathbf{Spread}^2}$$

(averaged over many forecast start dates)

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

Q. Can we develop diagnostics which efficiently (optimally?) guide us in this task?

# Motivation: Reliability and Sharpness




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

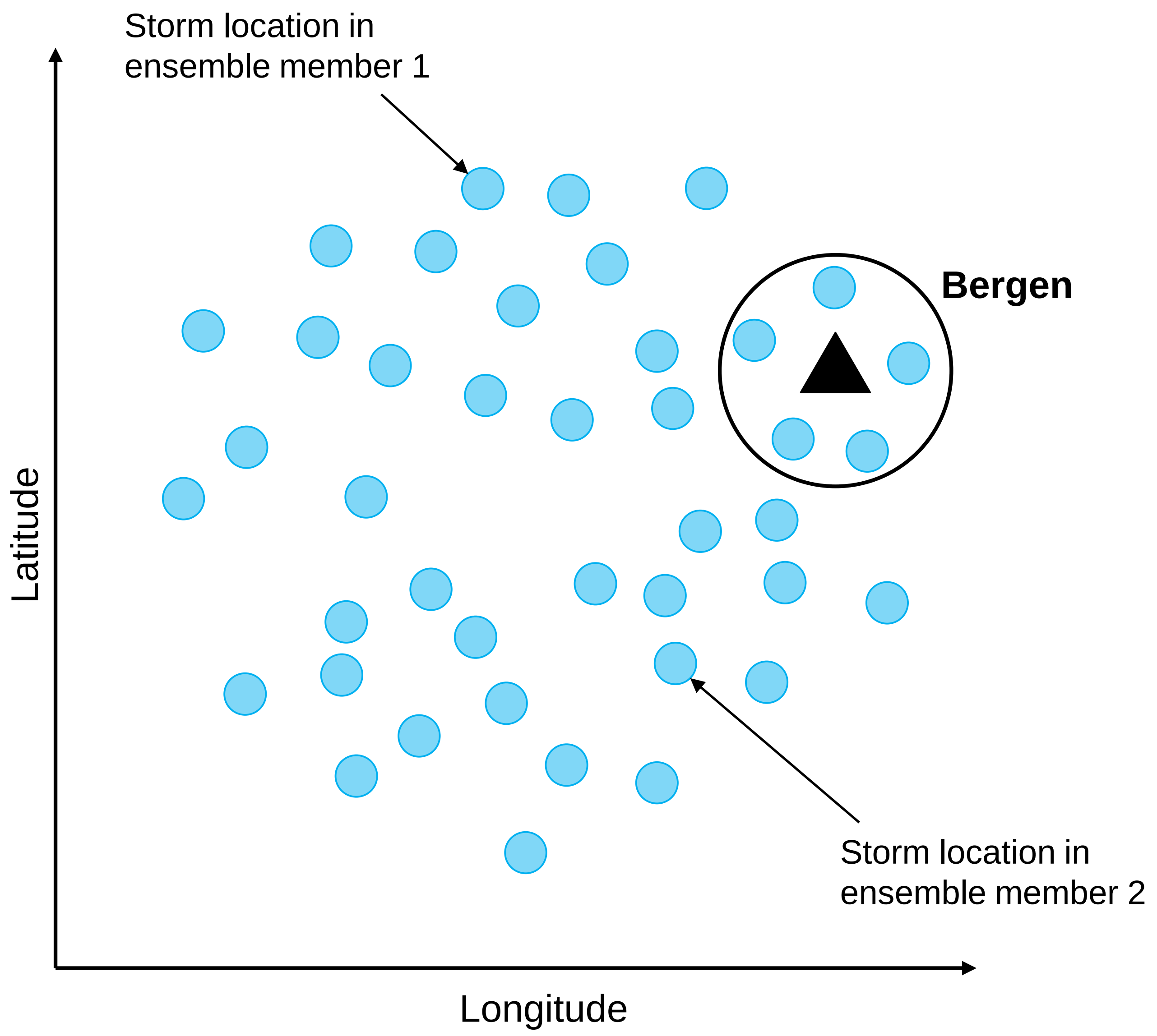
$$\overline{\mathbf{Error}^2} = \overline{\mathbf{Spread}^2}$$

(averaged over many forecast start dates)

"The task of NWP research is to maintain/improve reliability while decreasing spread (improving refinement)"

Q. Can we develop diagnostics which efficiently (optimally?) guide us in this task?

# Motivation: Reliability and Sharpness




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

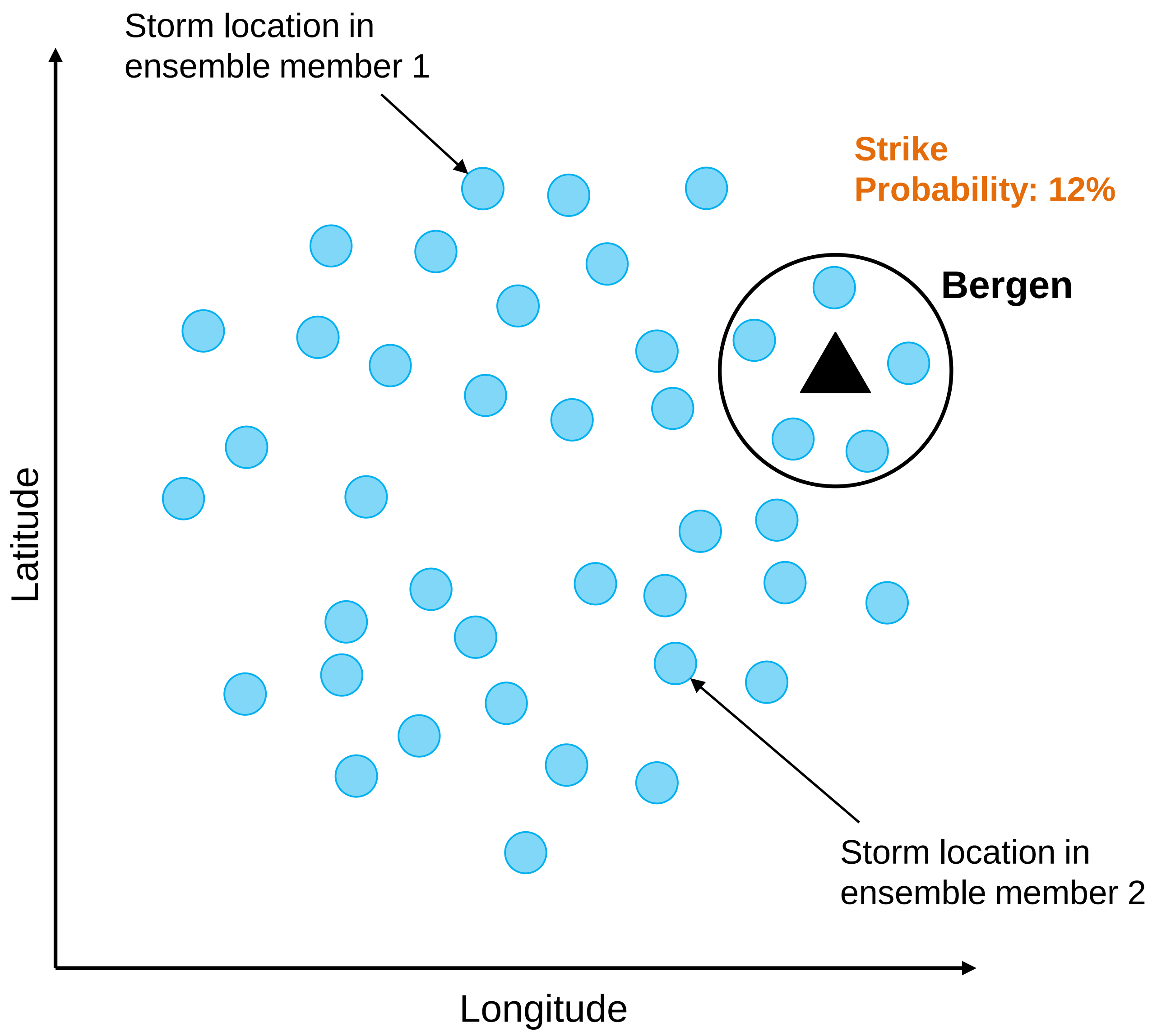
$$\overline{\mathbf{Error}^2} = \overline{\mathbf{Spread}^2}$$

(averaged over many forecast start dates)

"The task of NWP research is to maintain/improve reliability while decreasing spread (improving refinement)"

Q. Can we develop diagnostics which efficiently (optimally?) guide us in this task?

# Motivation: Reliability and Sharpness




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

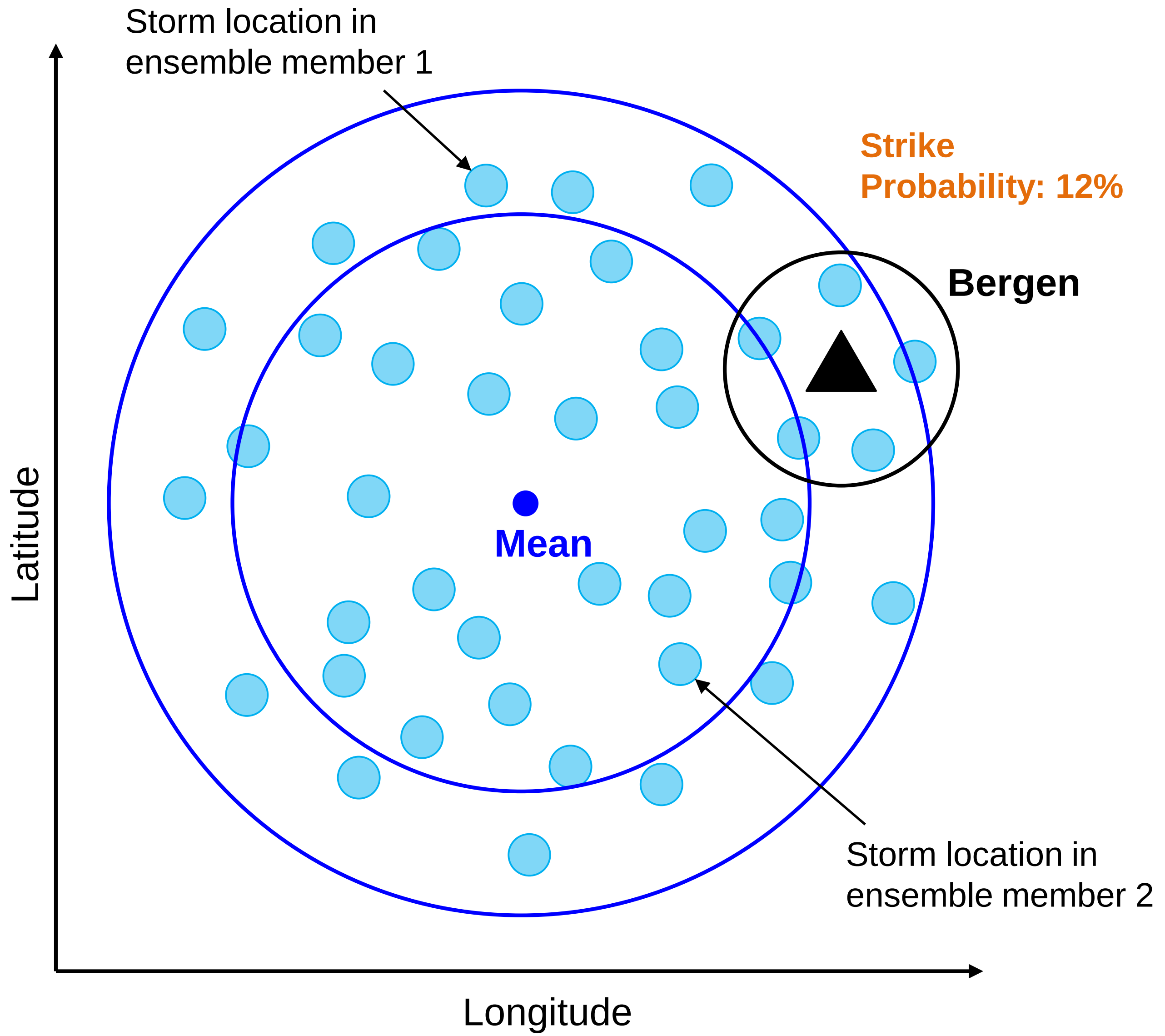
$$\overline{\mathbf{Error}^2} = \overline{\mathbf{Spread}^2}$$

(averaged over many forecast start dates)

"The task of NWP research is to maintain/improve reliability while decreasing spread (improving refinement)"

Q. Can we develop diagnostics which efficiently (optimally?) guide us in this task?

# Motivation: Reliability and Sharpness




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

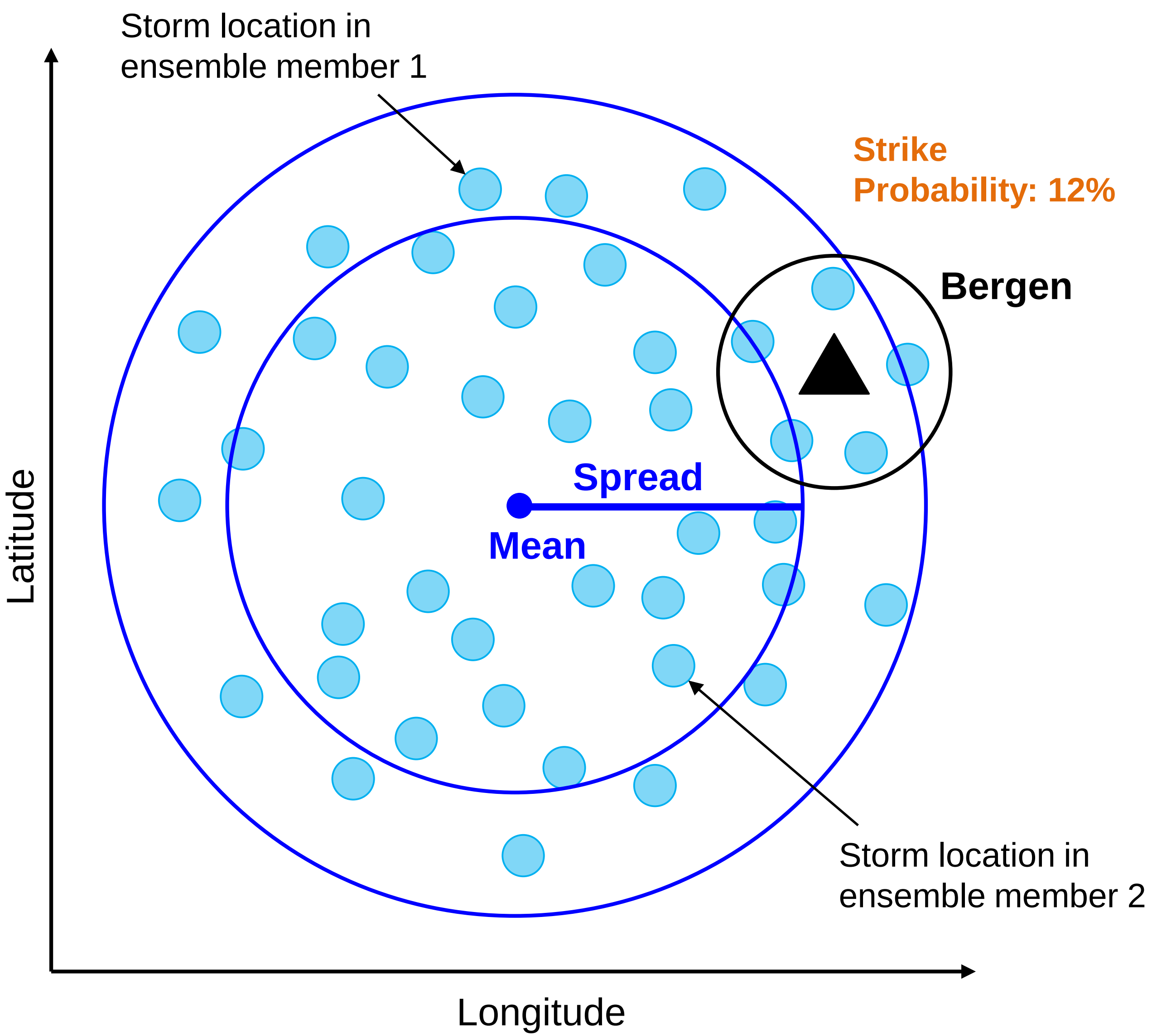
$$\overline{\mathbf{Error}^2} = \overline{\mathbf{Spread}^2}$$

(averaged over many forecast start dates)

"The task of NWP research is to maintain/improve reliability while decreasing spread (improving refinement)"

Q. Can we develop diagnostics which efficiently (optimally?) guide us in this task?

# Motivation: Reliability and Sharpness




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

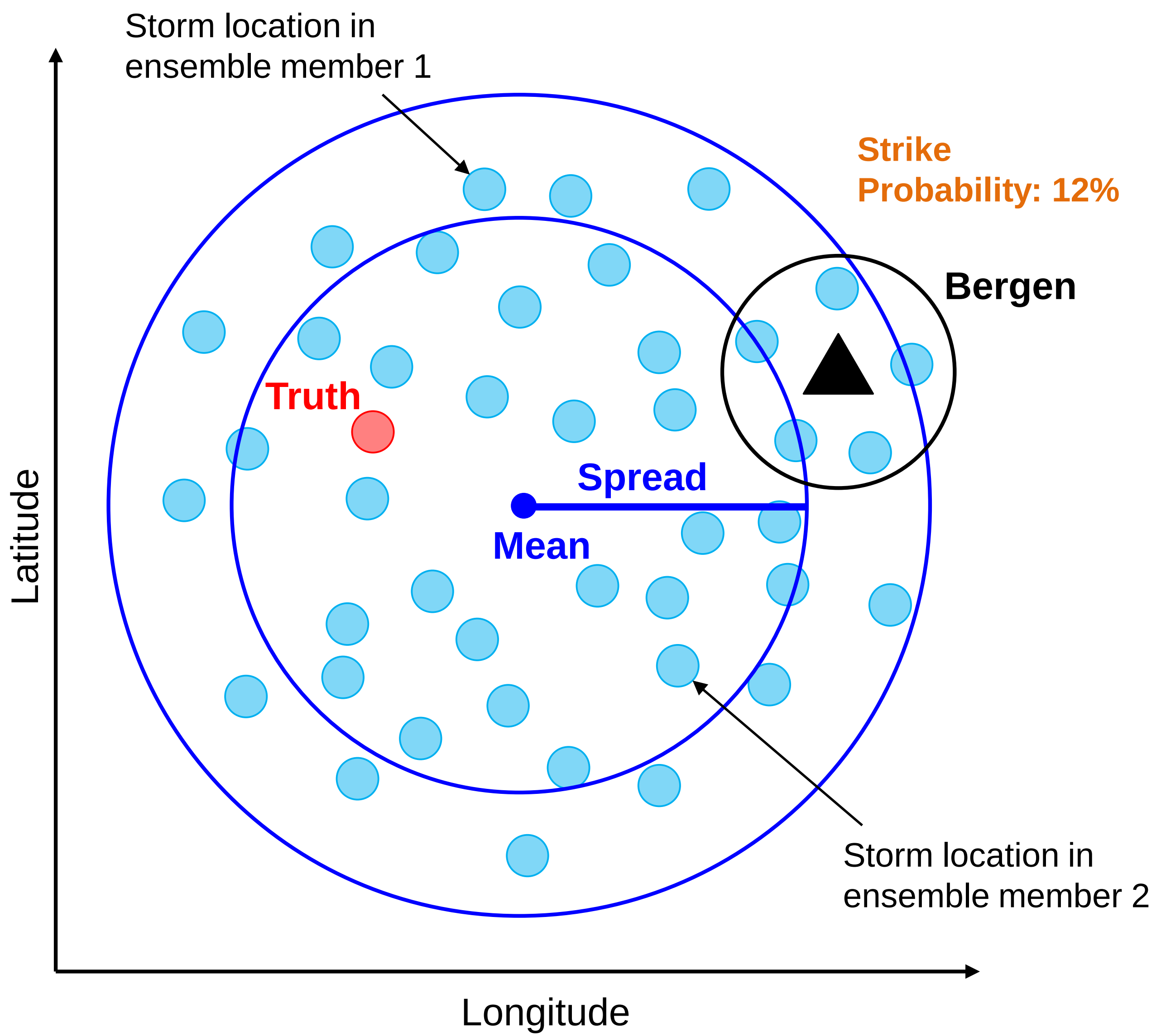
$$\overline{\mathrm{Error}^2} = \overline{\mathrm{Spread}^2}$$

(averaged over many forecast start dates)

"The task of NWP research is to maintain/improve reliability while decreasing spread (improving refinement)"

Q. Can we develop diagnostics which efficiently (optimally?) guide us in this task?

# Motivation: Reliability and Sharpness




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

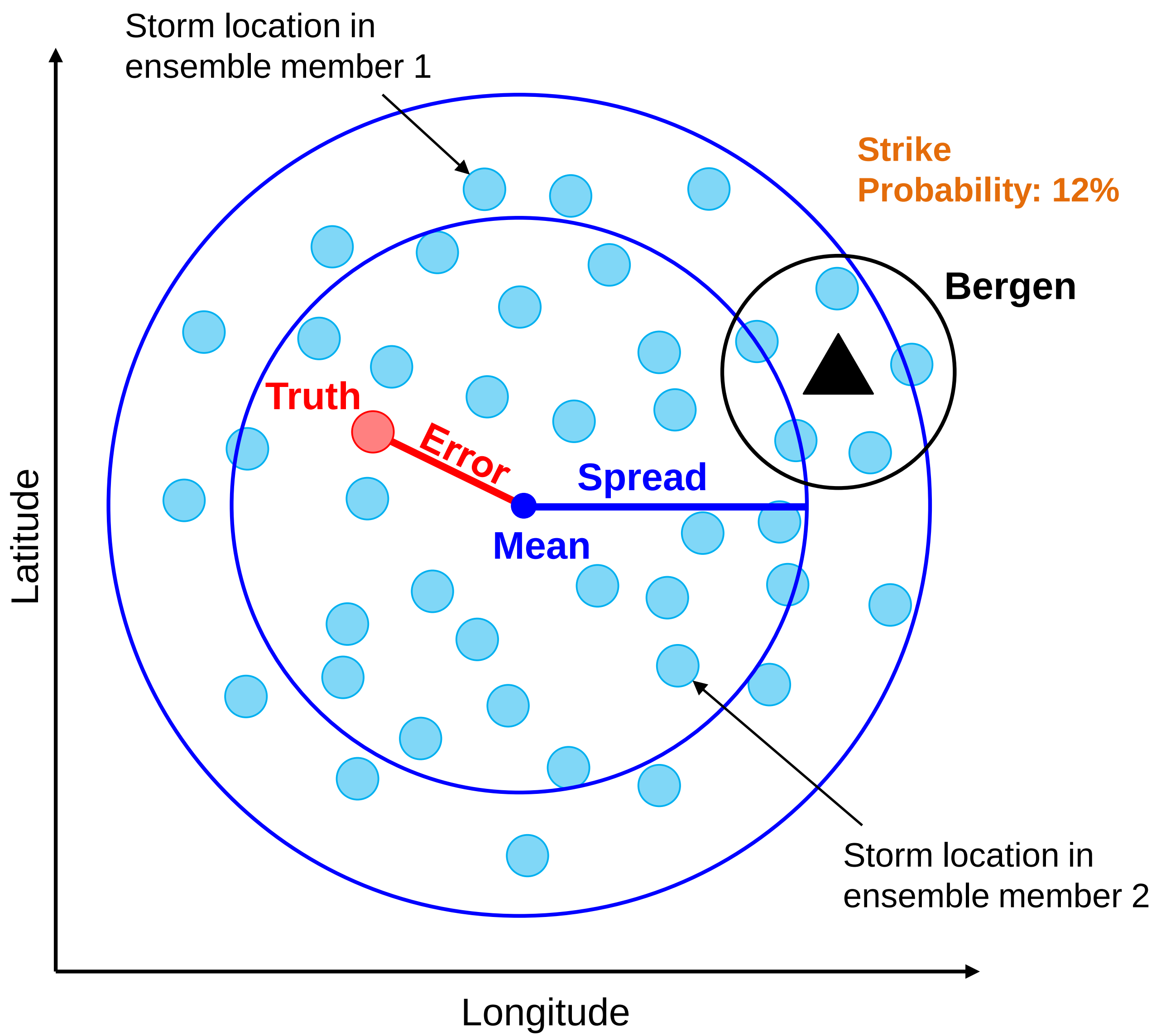
$$\overline{\mathbf{Error}^2} = \overline{\mathbf{Spread}^2}$$

(averaged over many forecast start dates)

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

Q. Can we develop diagnostics which efficiently (optimally?) guide us in this task?

# Motivation: Reliability and Sharpness




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

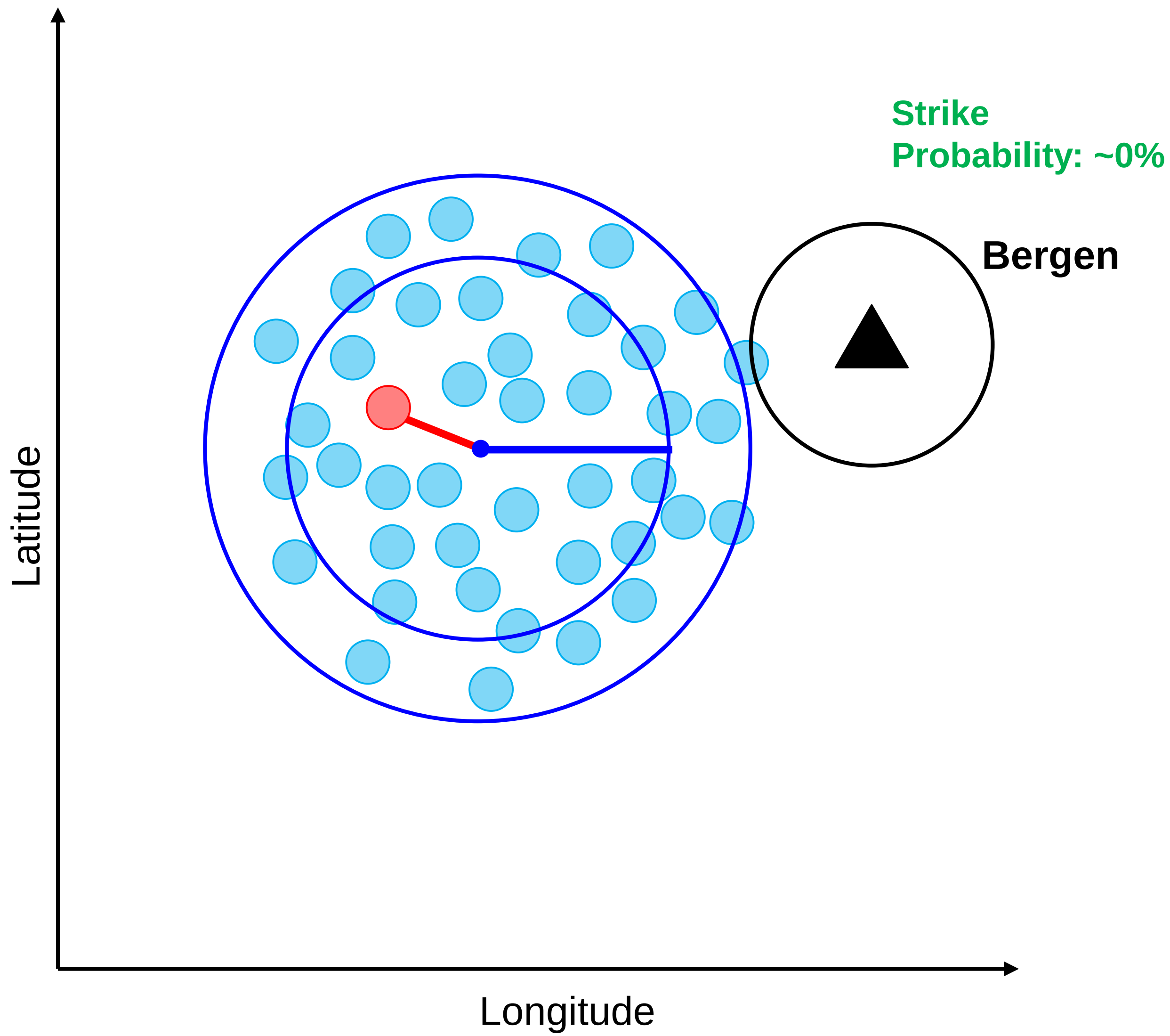
$$\overline{\mathbf{Error}^2} = \overline{\mathbf{Spread}^2}$$

(averaged over many forecast start dates)

"The task of NWP research is to maintain/improve reliability while decreasing spread (improving refinement)"

Q. Can we develop diagnostics which efficiently (optimally?) guide us in this task?

# Motivation: Reliability and Sharpness




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/improve reliability while decreasing spread (improving refinement)"

Q. Can we develop diagnostics which efficiently (optimally?) guide us in this task?

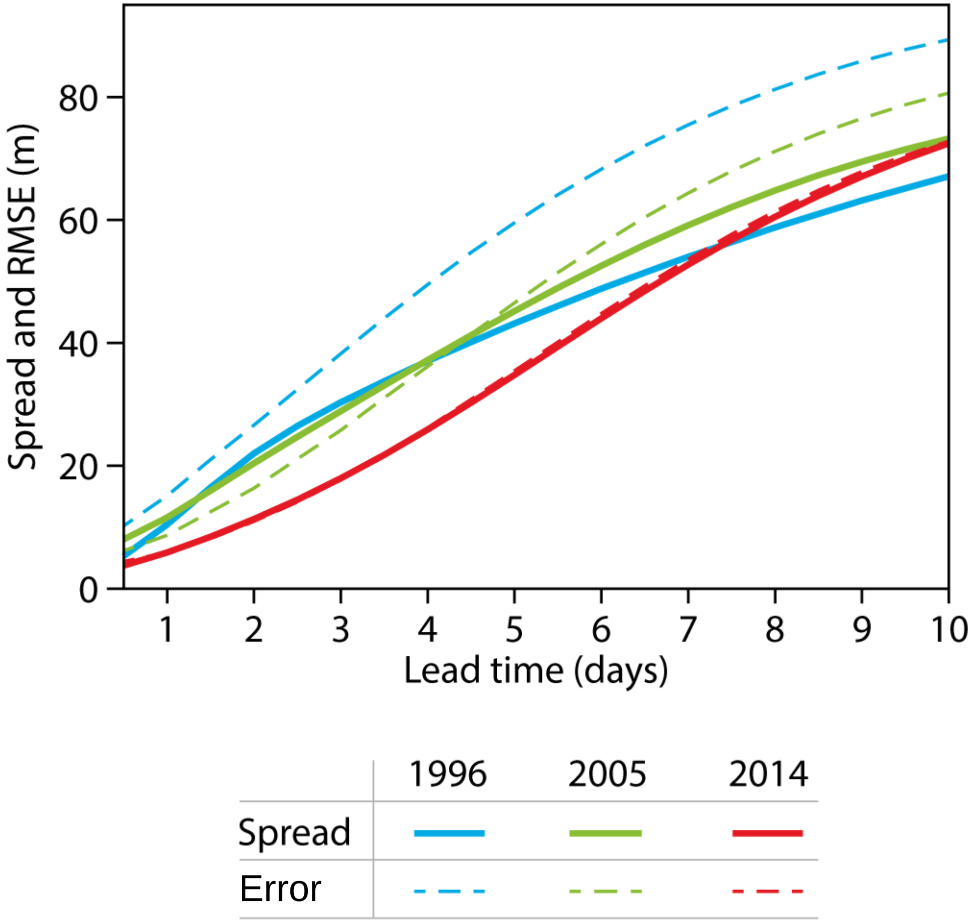
# Ensemble spread and error

Z500

*Rodwell et al. 2018, BAMS*

Annual means N.Hem. (ECMWF)

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




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

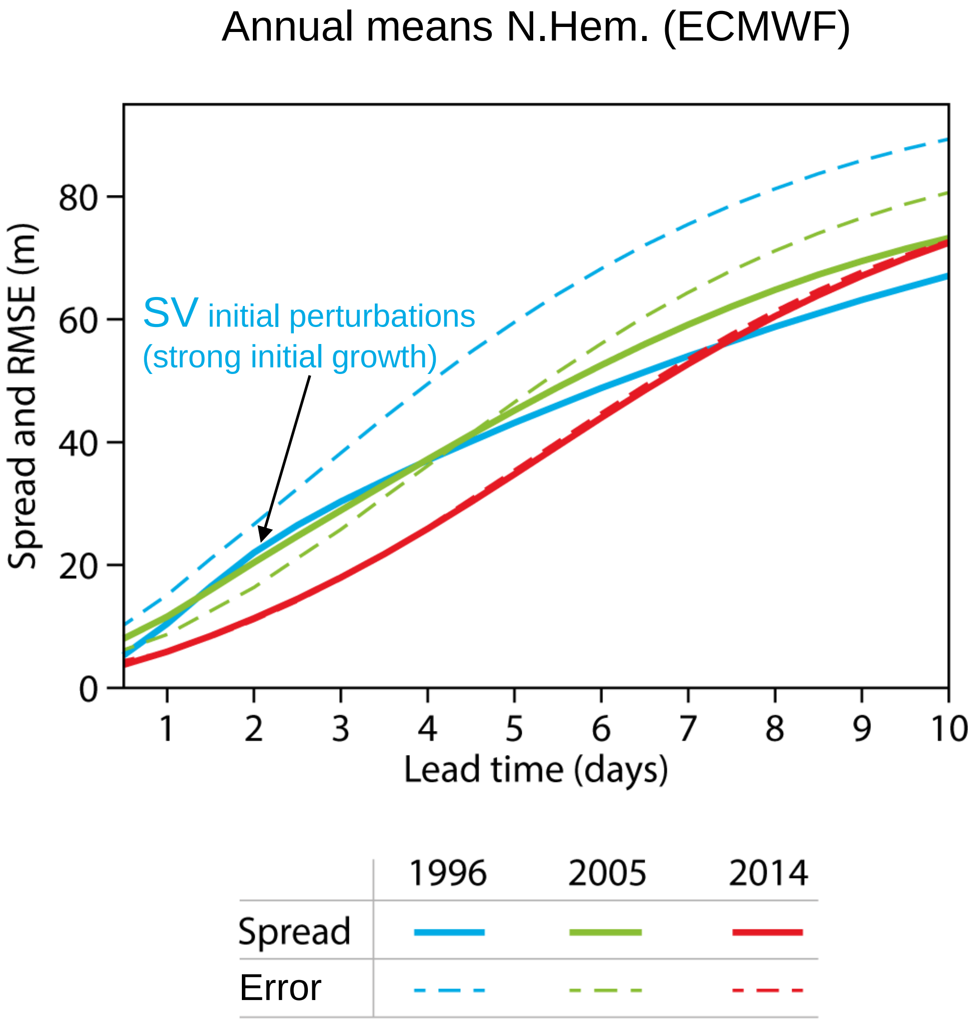
*Spread = ensemble standard deviation (scaled to take account of finite ensemble size)*

# 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




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

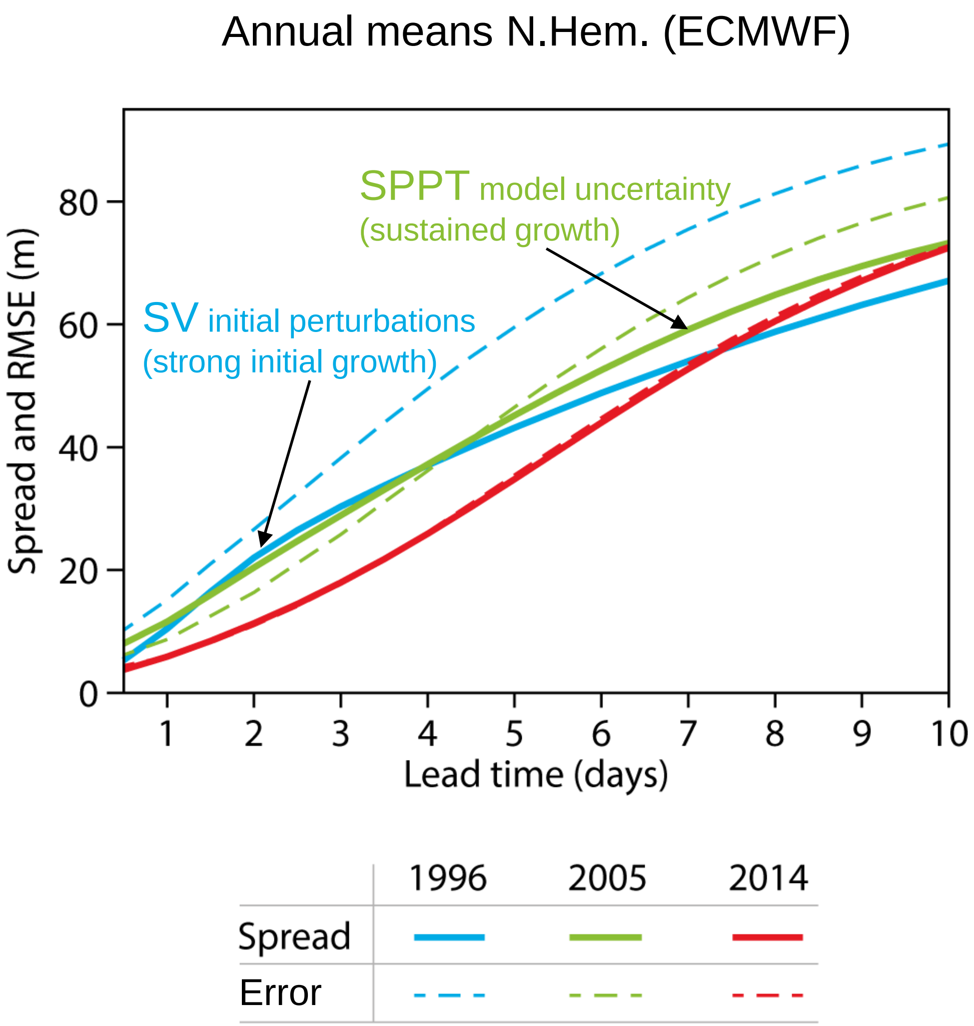
*Spread = ensemble standard deviation (scaled to take account of finite ensemble size)*

# 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




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

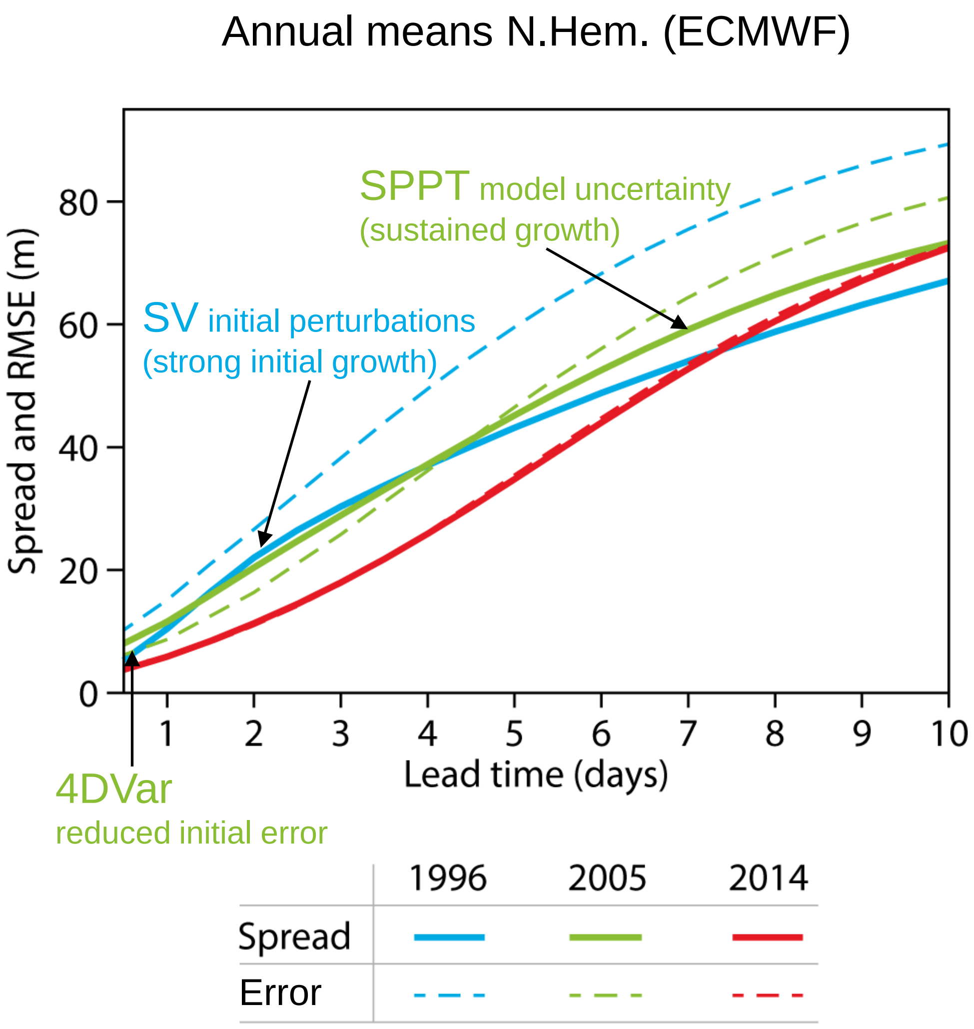
*Spread = ensemble standard deviation (scaled to take account of finite ensemble size)*

# 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




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

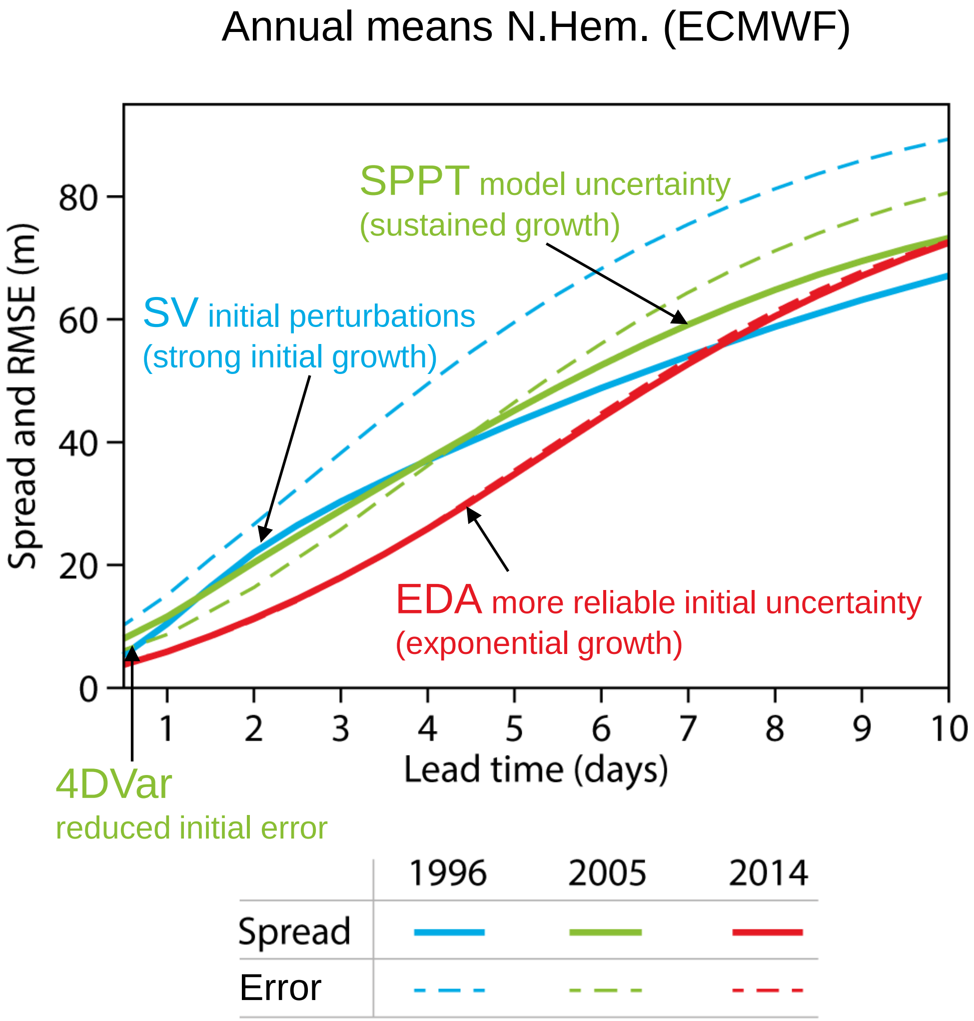
*Spread = ensemble standard deviation (scaled to take account of finite ensemble size)*

# 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




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

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

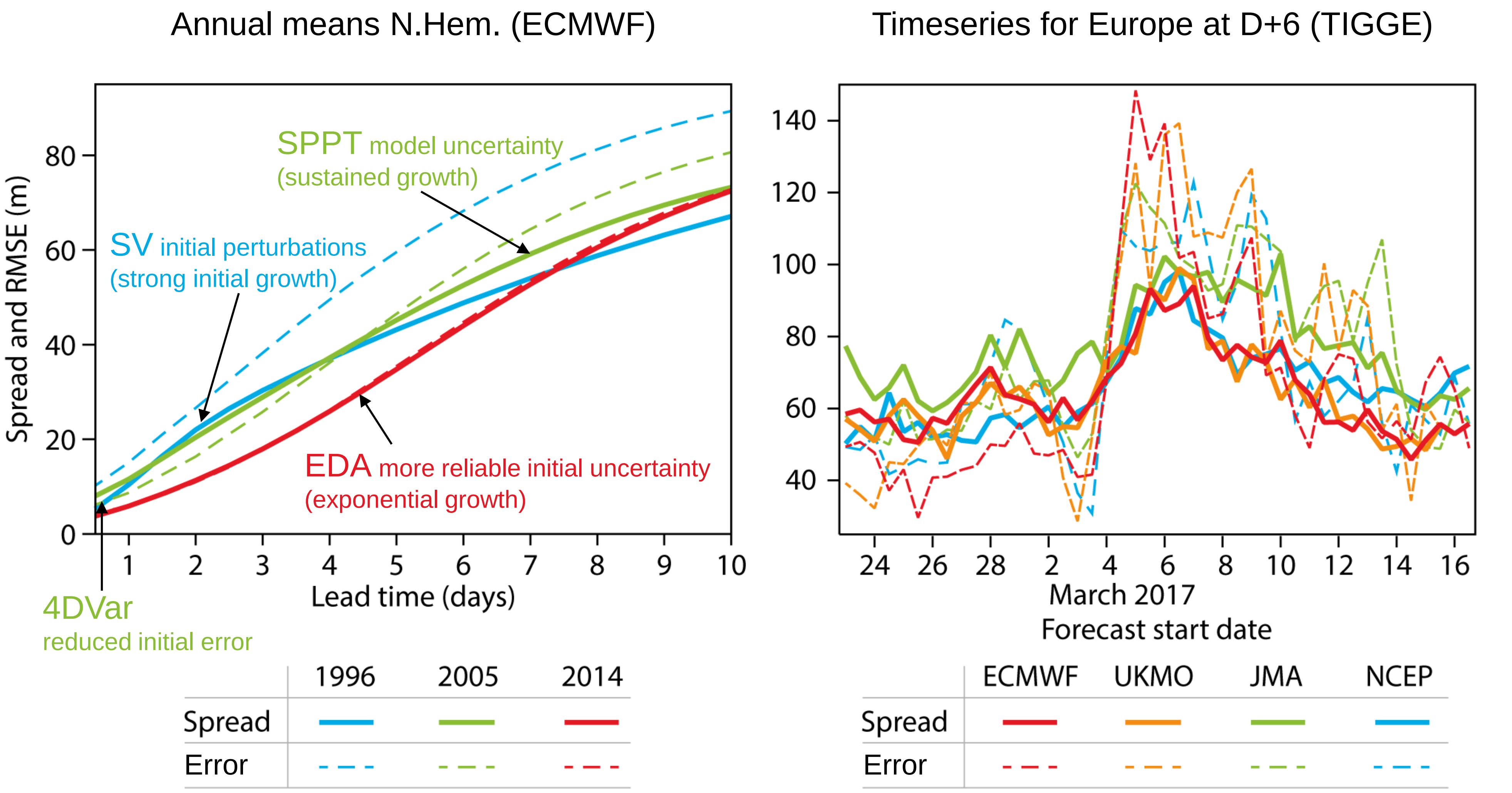
# Ensemble spread and error

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?

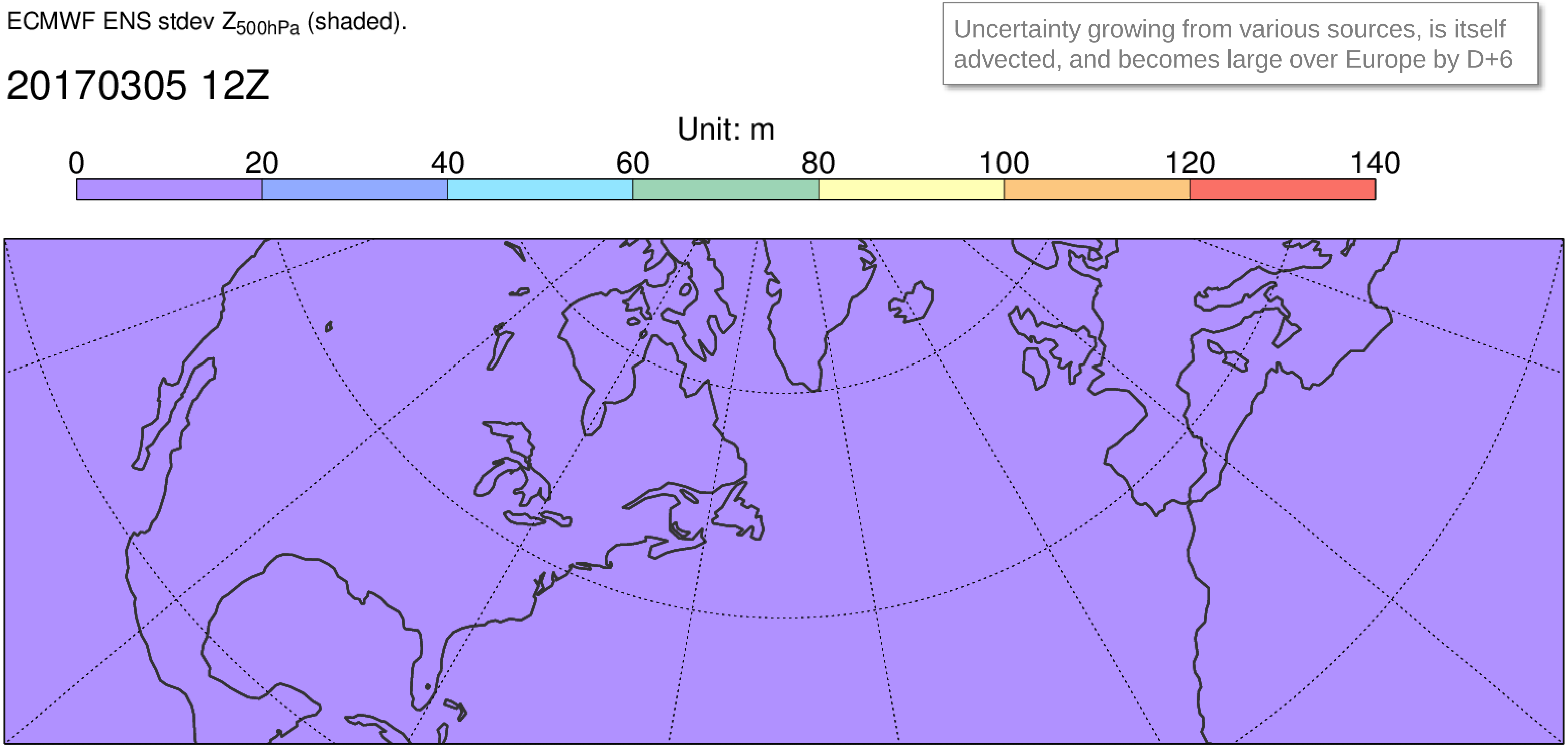



*500 hPa geopotential height (Z500). "Error" is RMS of ensemble-mean error*

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

# Animation of ECMWF ensemble forecast spread 20170305 12Z D+0 to 6: $\sigma_{Z500}$

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



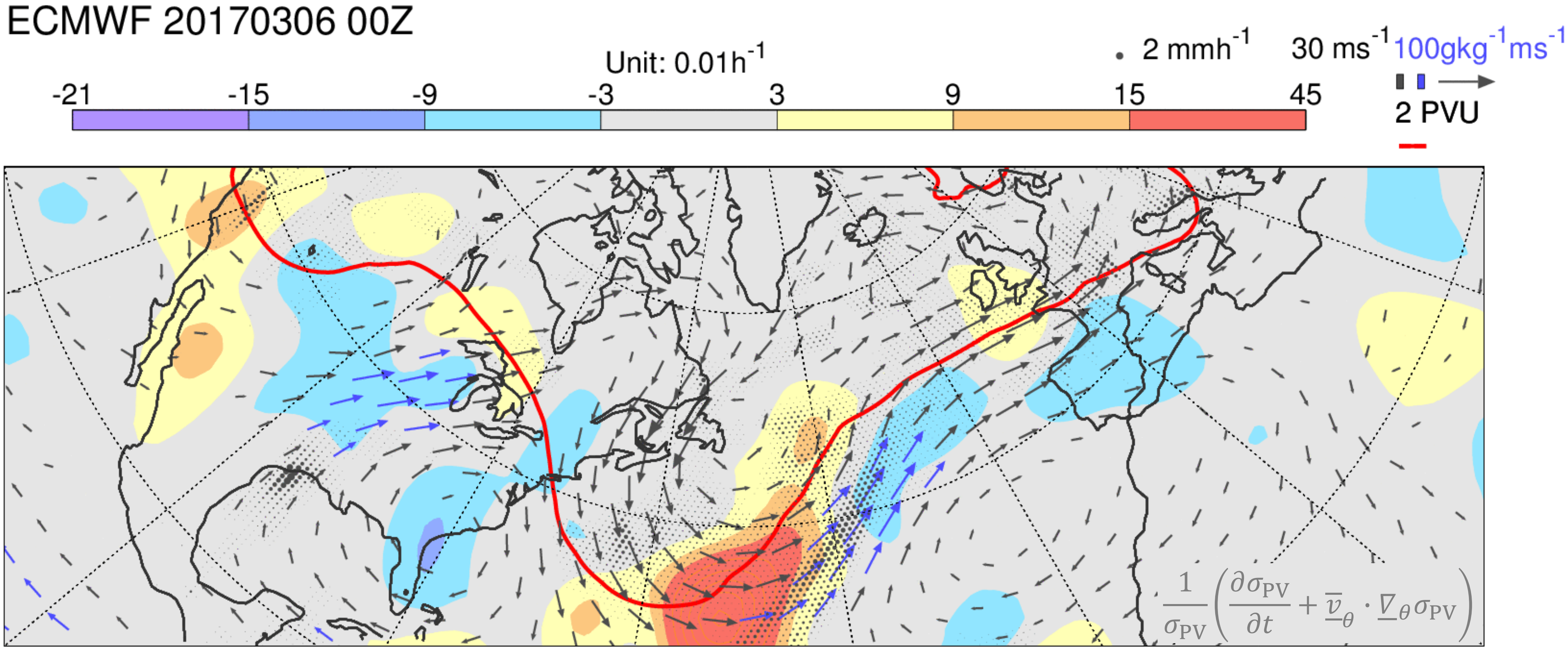
# Uncertainty growth-rate along the truth trajectory – Based on EDA background $\sigma_{PV_{315}}$

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?




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

*Control forecast $PV_{315}$=2, $\underline{v}_{850}$ and $q|\underline{v}|_{850}$, Ensemble-mean precipitation. 1d running-mean gives 12h-integrated growth rate with any diurnal cycle removed. T21 smoothed*

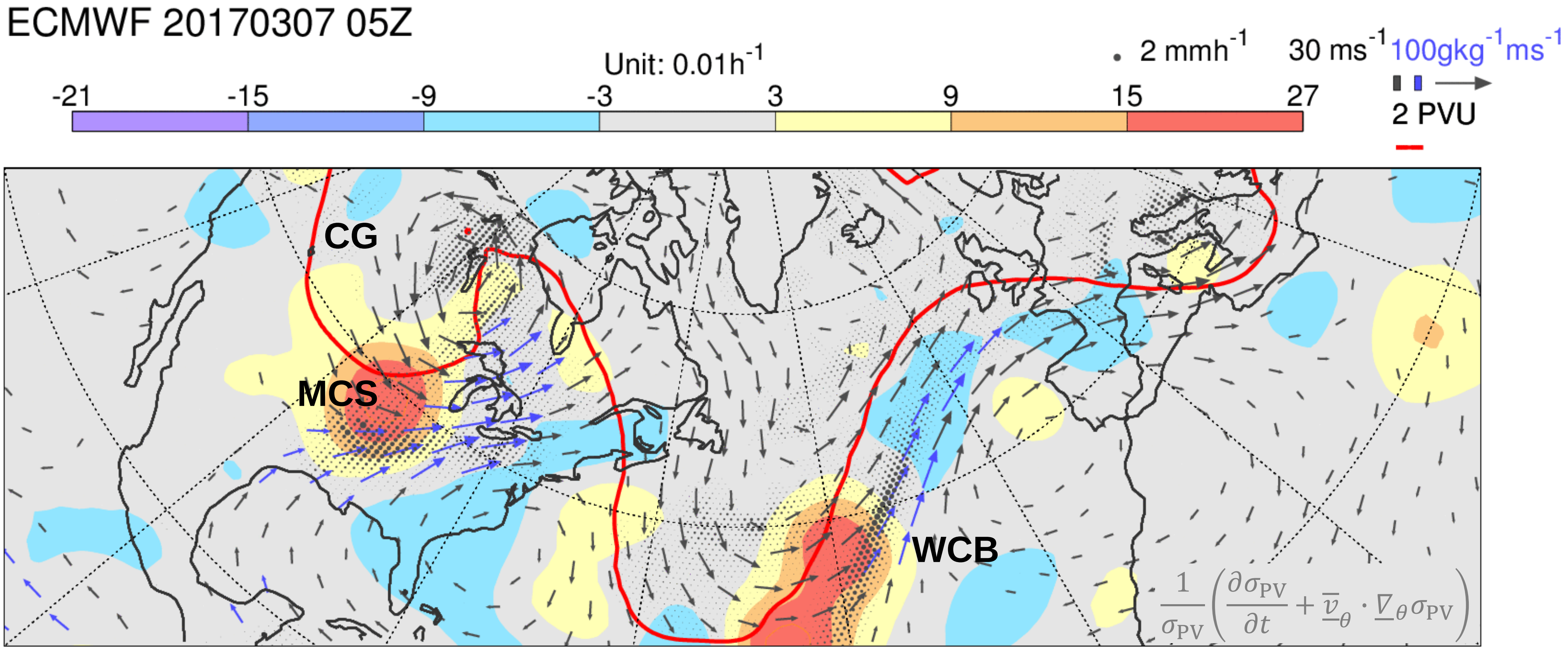
# Uncertainty growth-rate along the truth trajectory – Based on EDA background $\sigma_{PV_{315}}$

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

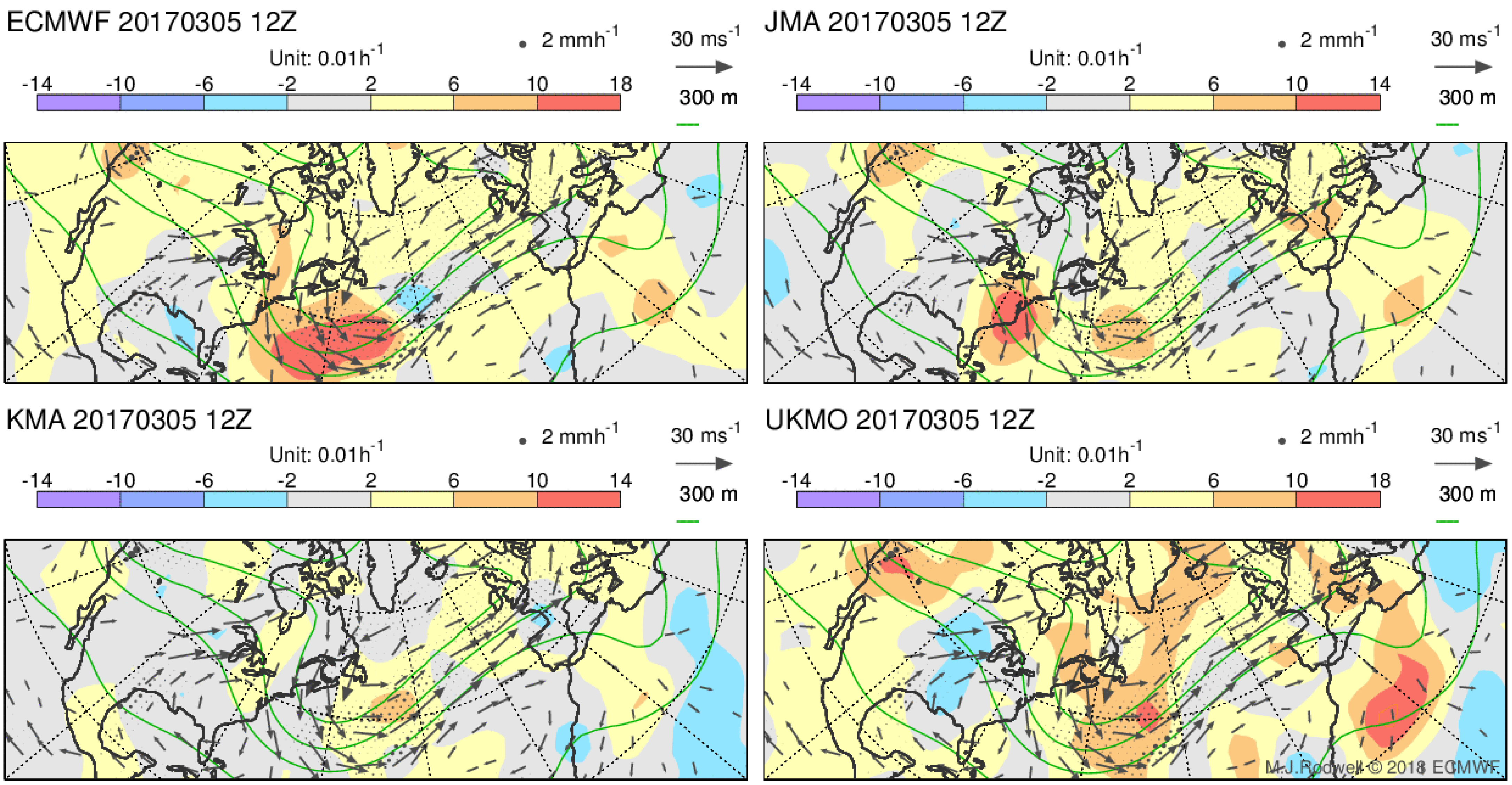
Aim: Evaluate short-range synoptic flow-dependent representation of uncertainty

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




*Control forecast $PV_{315}$=2, $\underline{v}_{850}$ and $q|\underline{v}|_{850}$, Ensemble-mean precipitation. 1d running-mean gives 12h-integrated growth rate with any diurnal cycle removed. T21 smoothed*

# Uncertainty growth-rate along the truth trajectory - Based on 12h ENS $Z_{250hPa}$ TIGGE




**ECMWF:**

EDA($PV_{315K}$) ≈ ENS($Z_{250hPa}$) ≈

**JMA:**

≈ ECMWF

**UKMO:**

Stronger growth-rates over Europe/Africa

**KMA:**

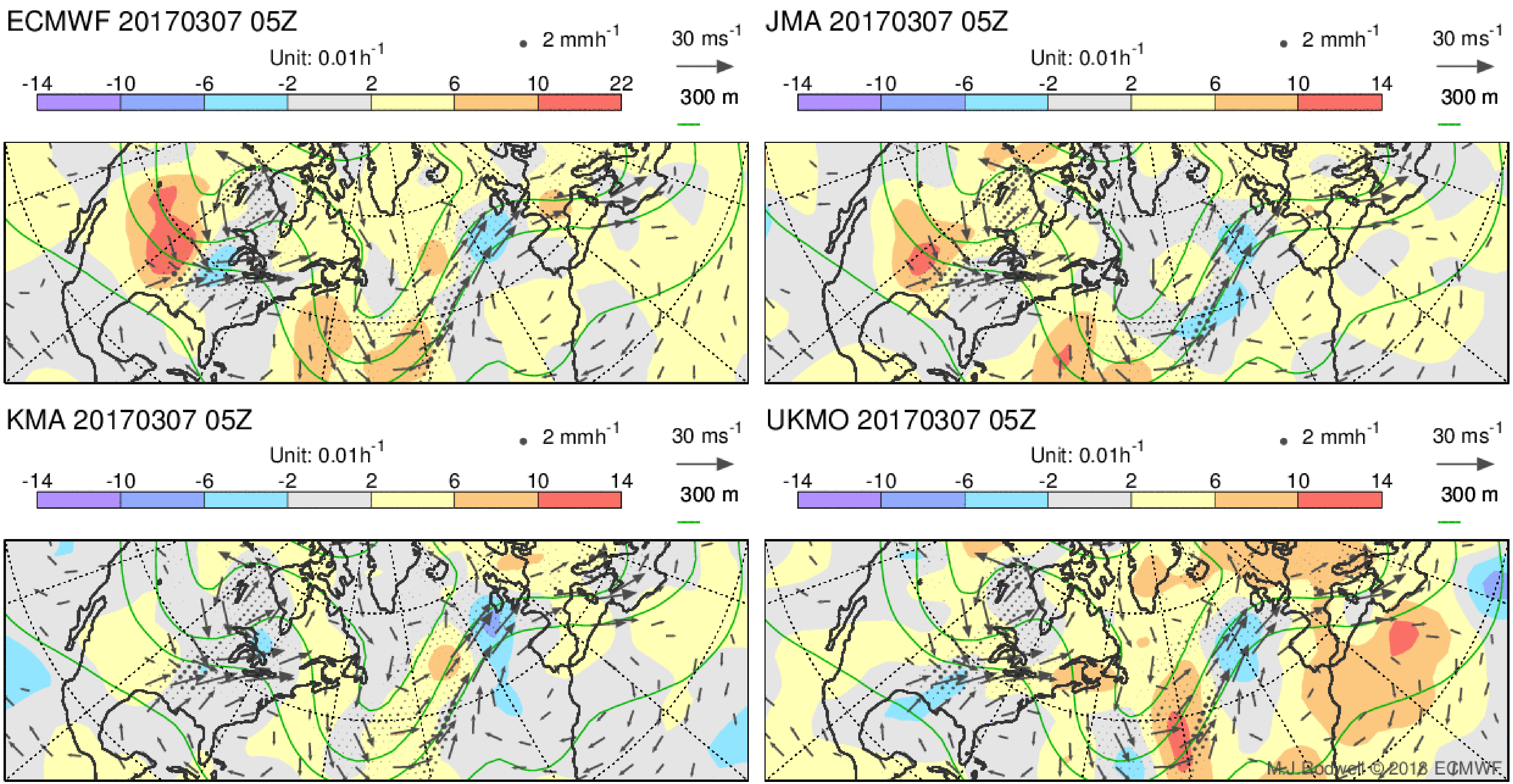
Weaker everywhere

Which is best?

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

*Control forecast $Z_{250}$ (CI=300m) and $\underline{v}_{850}$, Ensemble-mean precipitation. 1d running-mean gives 12h-integrated growth rate with any diurnal cycle removed. T21 smoothed*

# Uncertainty growth-rate along the truth trajectory - Based on 12h ENS $Z_{250hPa}$ TIGGE




**ECMWF:**

EDA($PV_{315K}$) ≈ ENS($Z_{250hPa}$) ≈

**JMA:**

≈ ECMWF

**UKMO:**

Stronger growth-rates over Europe/Africa

**KMA:**

Weaker everywhere

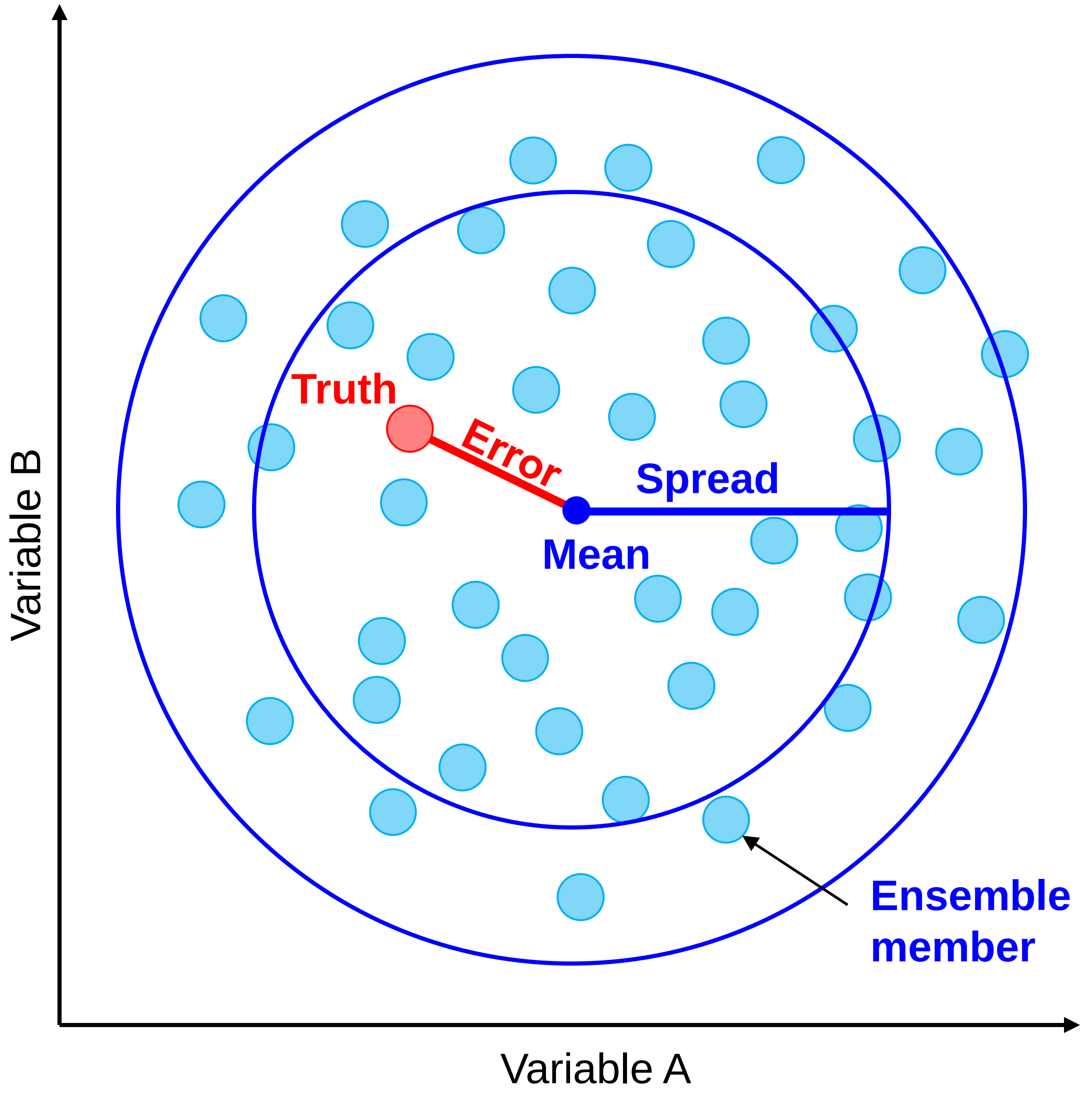
Which is best?

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

*Control forecast $Z_{250}$ (CI=300m) and $\underline{v}_{850}$, Ensemble-mean precipitation. 1d running-mean gives 12h-integrated growth rate with any diurnal cycle removed. T21 smoothed*

# 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‡

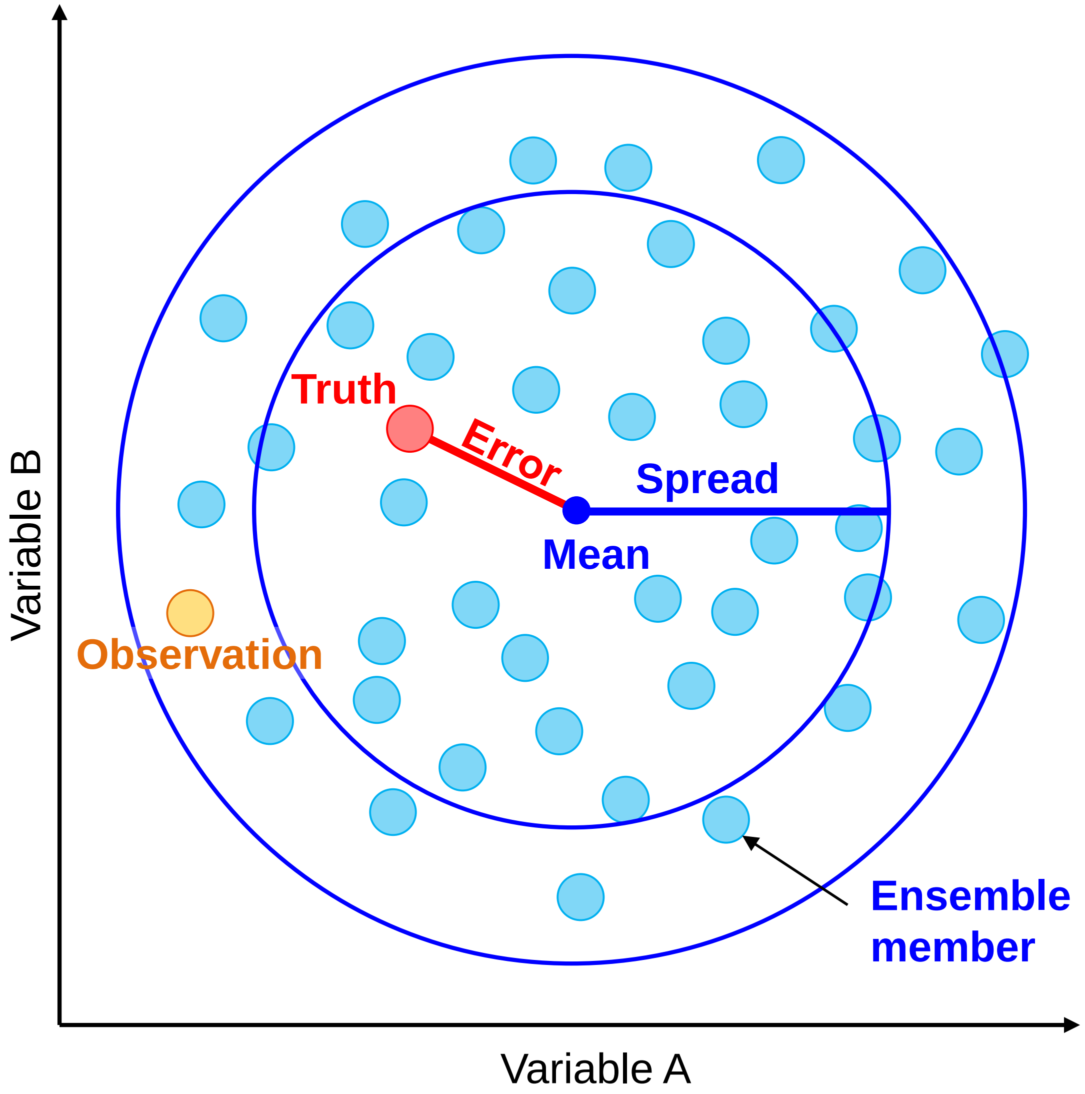
$$\overline{\mathbf{Departure^2}} = \overline{\mathbf{EnsVar}} + \overline{\mathbf{Obs.Unc^2}}$$

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

# 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‡

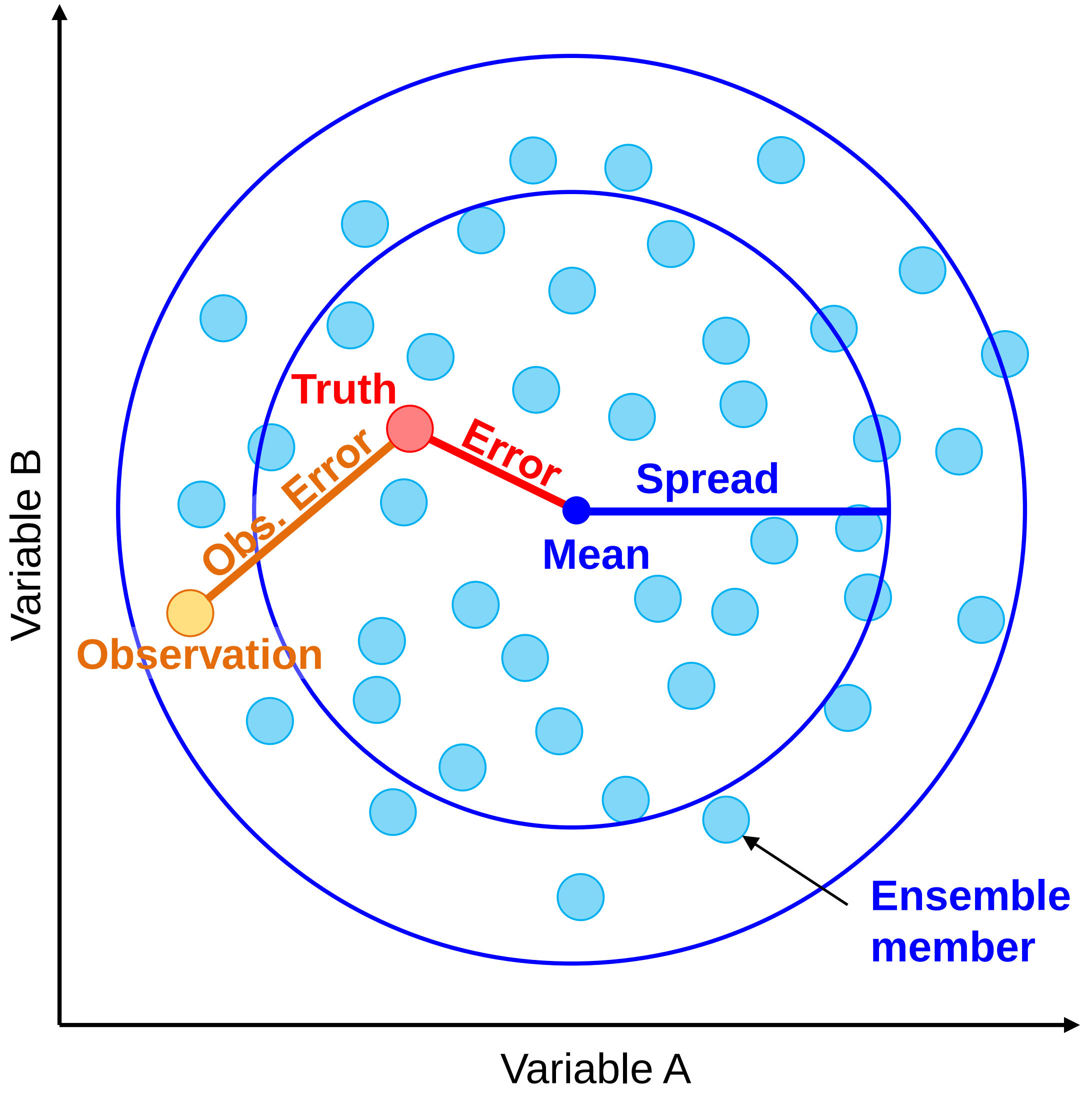
$$\overline{\mathbf{Departure^2}} = \overline{\mathbf{EnsVar}} + \overline{\mathbf{Obs.Unc^2}}$$

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

# 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‡

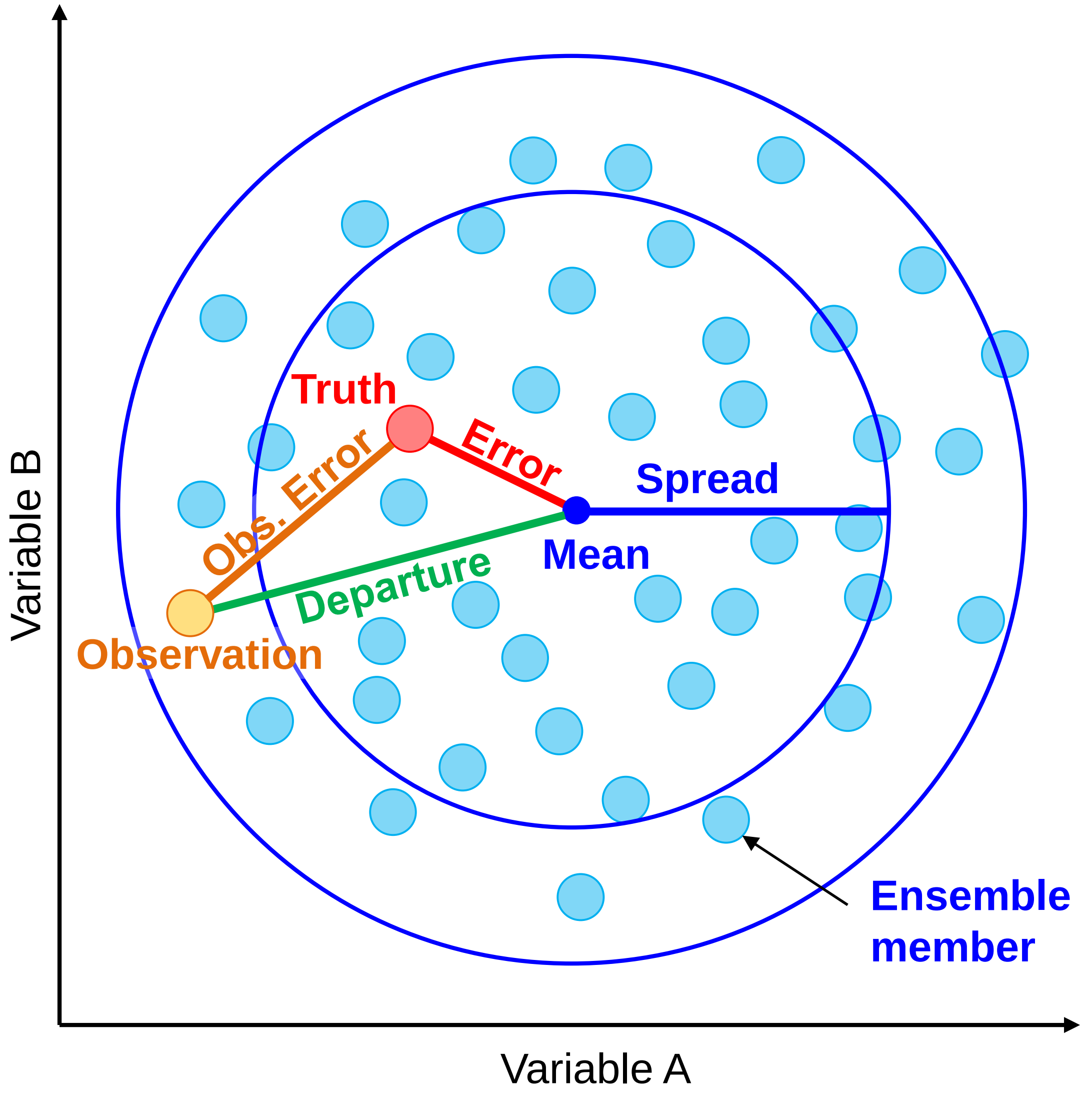
$$\overline{\mathbf{Departure^2}} = \overline{\mathbf{EnsVar}} + \overline{\mathbf{Obs.Unc^2}}$$

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

# 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‡

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

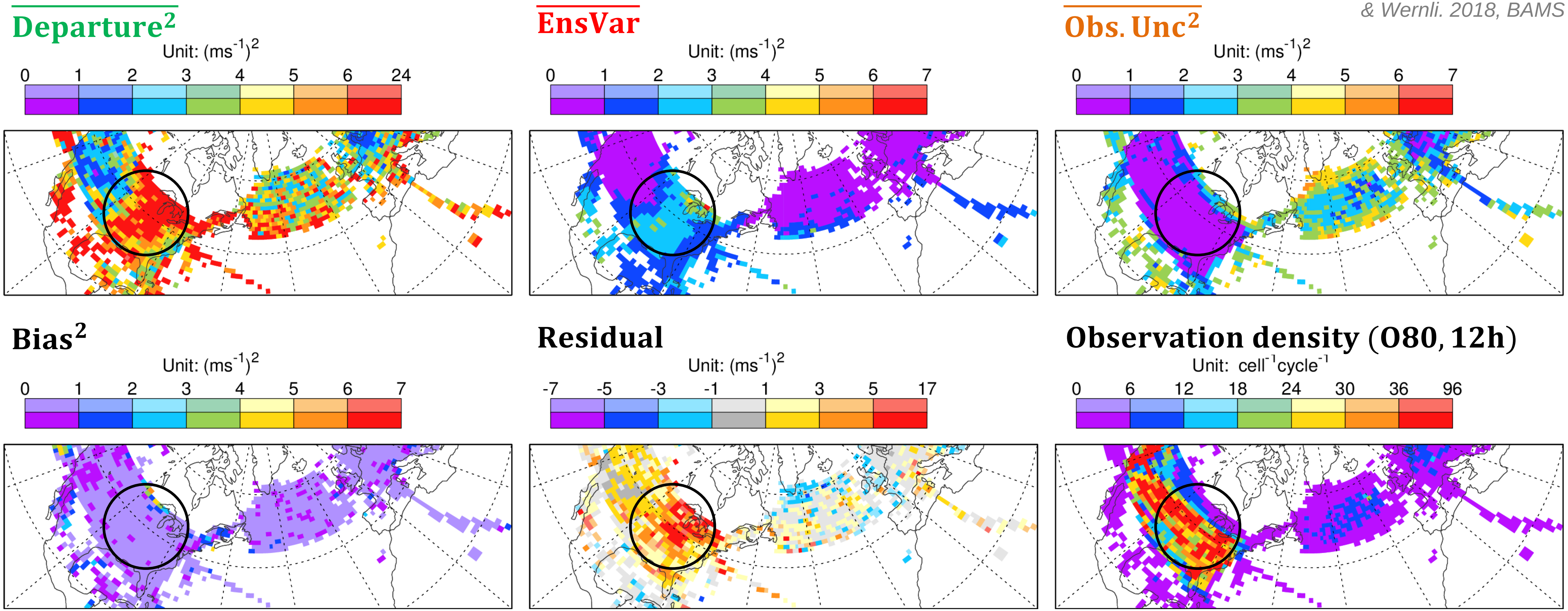
Any imbalance in this equation indicates that the (initialization of) the ensemble forecast is unreliable

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

# Uncertainty growth evaluation. “Rocky trough/CAPE” composite. EDA $u_{200}$ aircraft obs

54 cases, 12h window

*Rodwell, Richardson, Parsons & Wernli. 2018, BAMS*



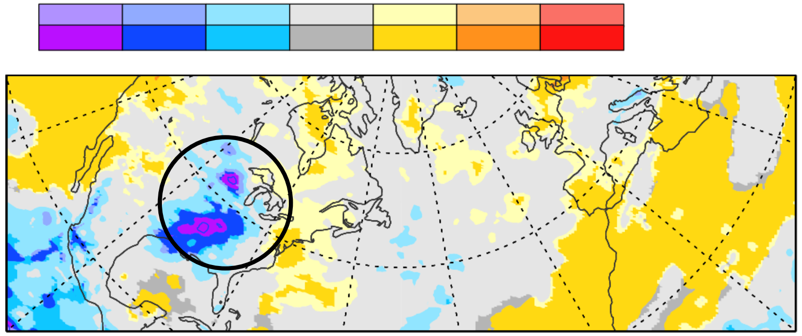

$\mathbf{Departure^2} = \mathbf{EnsVar} + \mathbf{Obs.Unc^2}\ (+\mathbf{Bias^2} + \mathbf{Residual})$. Enhanced background variance in Great Lakes / Mississippi River region. Even larger Departures. Bias$^2$≈0, but Residual ≫ 0 indicates insufficient background variance (since estimated observation error and density are similar over north-western North America where Residual is smaller – *i.e.* well balanced). **Uncertain forecasts for Europe may still be over-confident!**

# Systematic process error. “Rocky trough/CAPE” composite. EDA control $T_{300}$

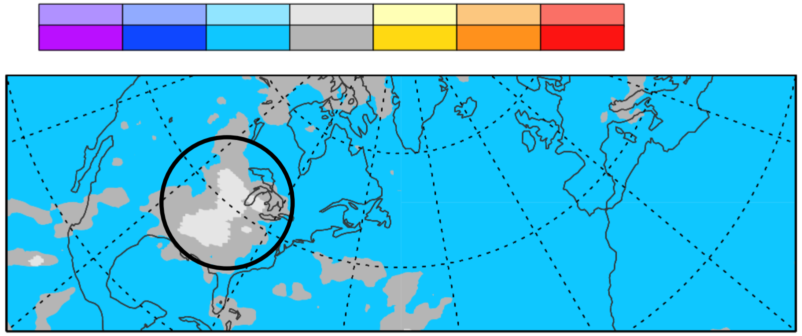
54 cases, 12h window

*Rodwell, Richardson, Parsons & Wernli. 2018, BAMS*

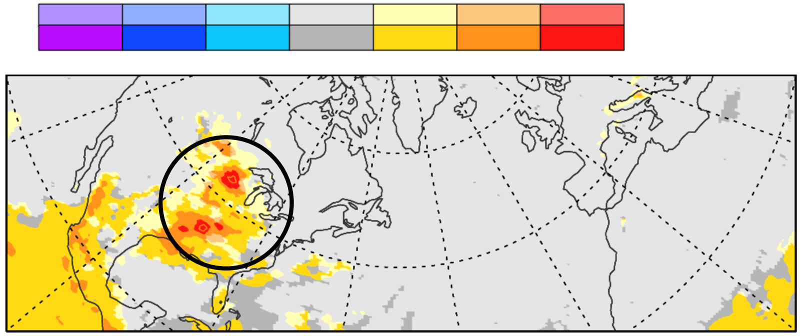
## Dynamics



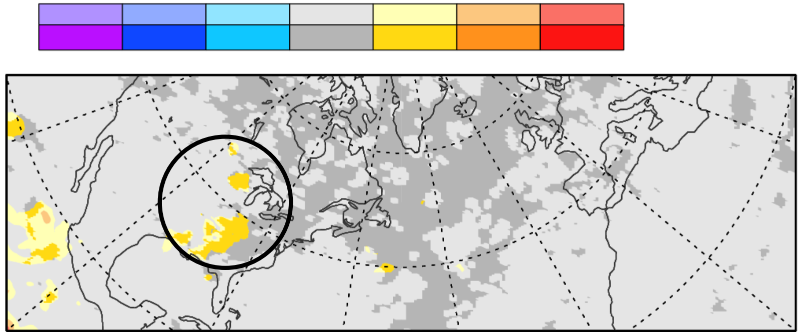

## Radiation



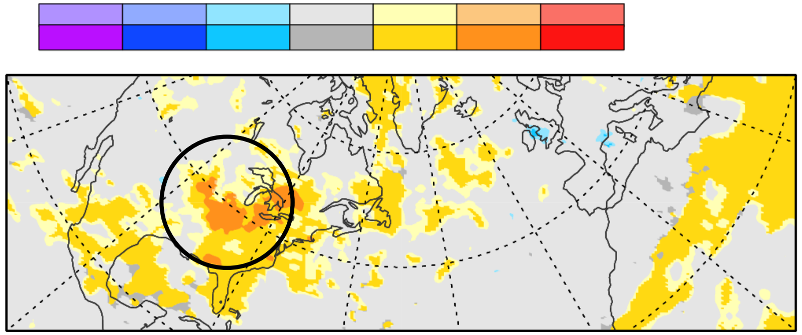

## Convection



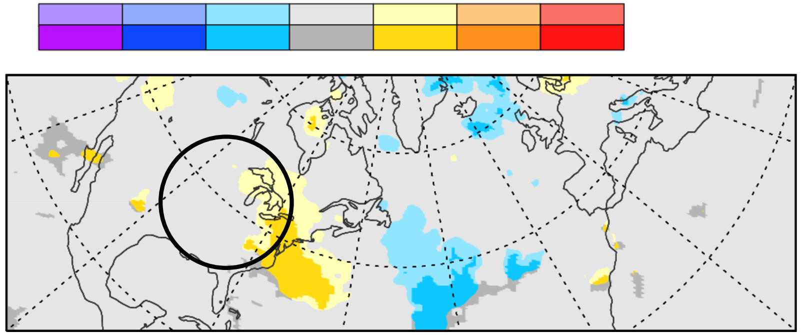

## Cloud




## Increment



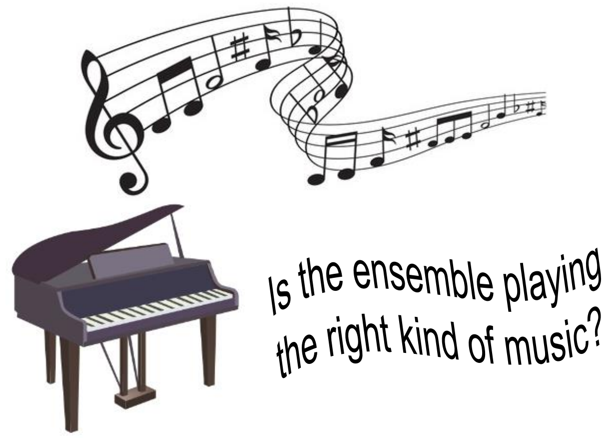

## Evolution




**Dynamics** + **Radiation** + **Convection** + **Cloud** + **Increment** = **Evolution**. Budget shows how the model represents dynamics and physics of MCS. Positive (and statistically significant) analysis increment suggests that the background forecast is too cold near the top of the convection. Hence, m**odel bias (as well as model uncertainty) may be an issue.**

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

54 cases

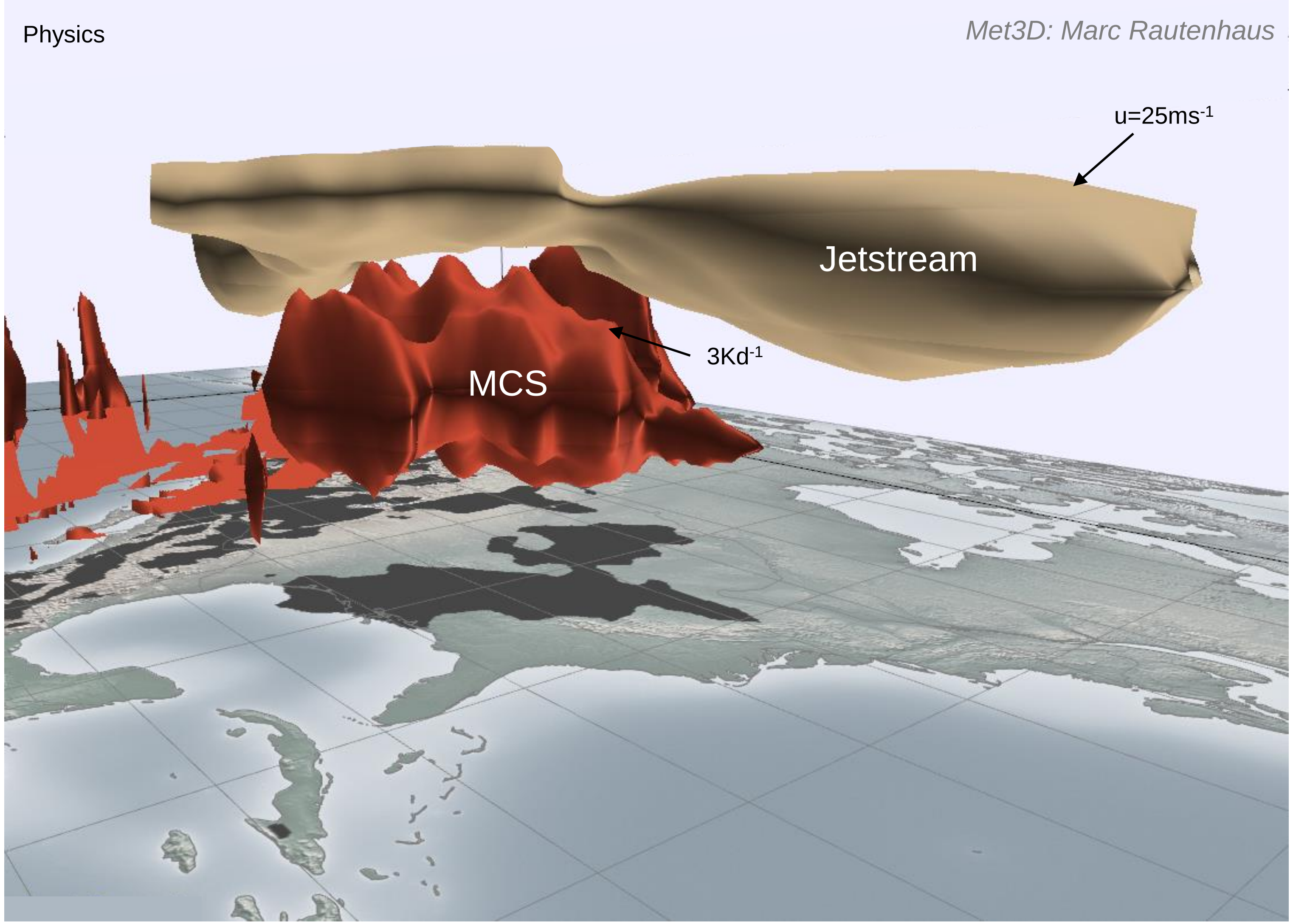



If we don’t hit the string hard enough, the wave in the string will be too weak

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

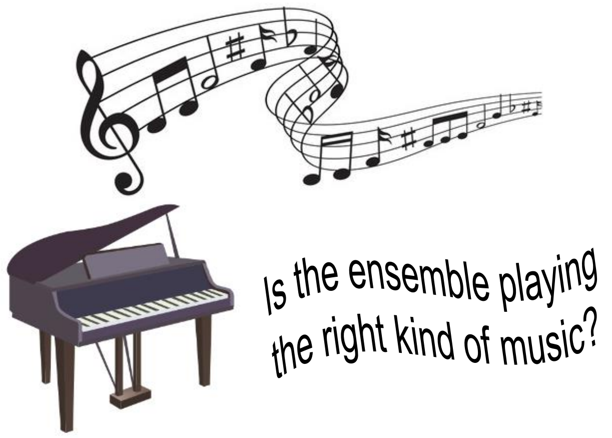
We do not know when to press the key (mesoscale convection itself involves chaotic uncertainty)

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



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

54 cases

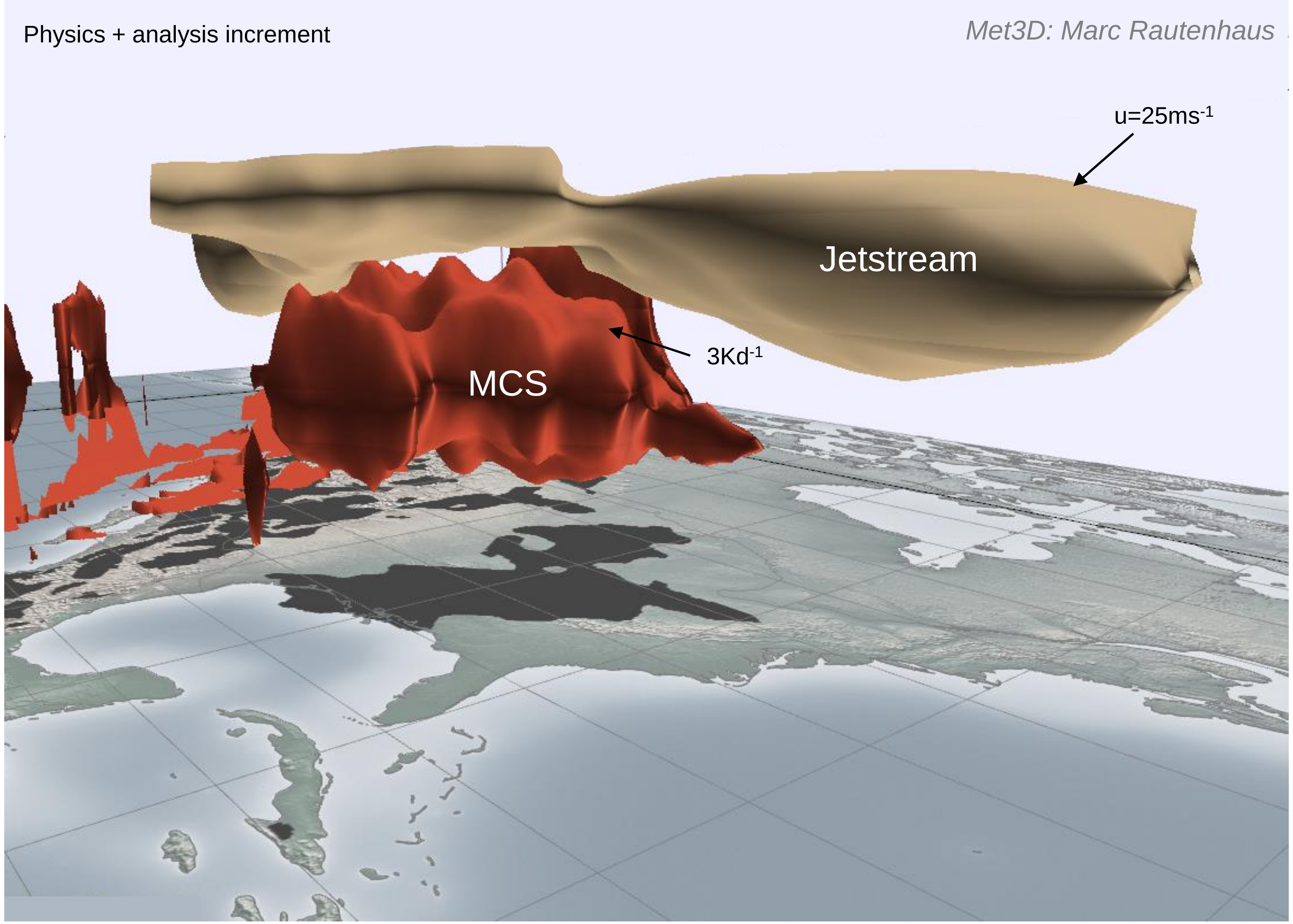



If we don’t hit the string hard enough, the wave in the string will be too weak

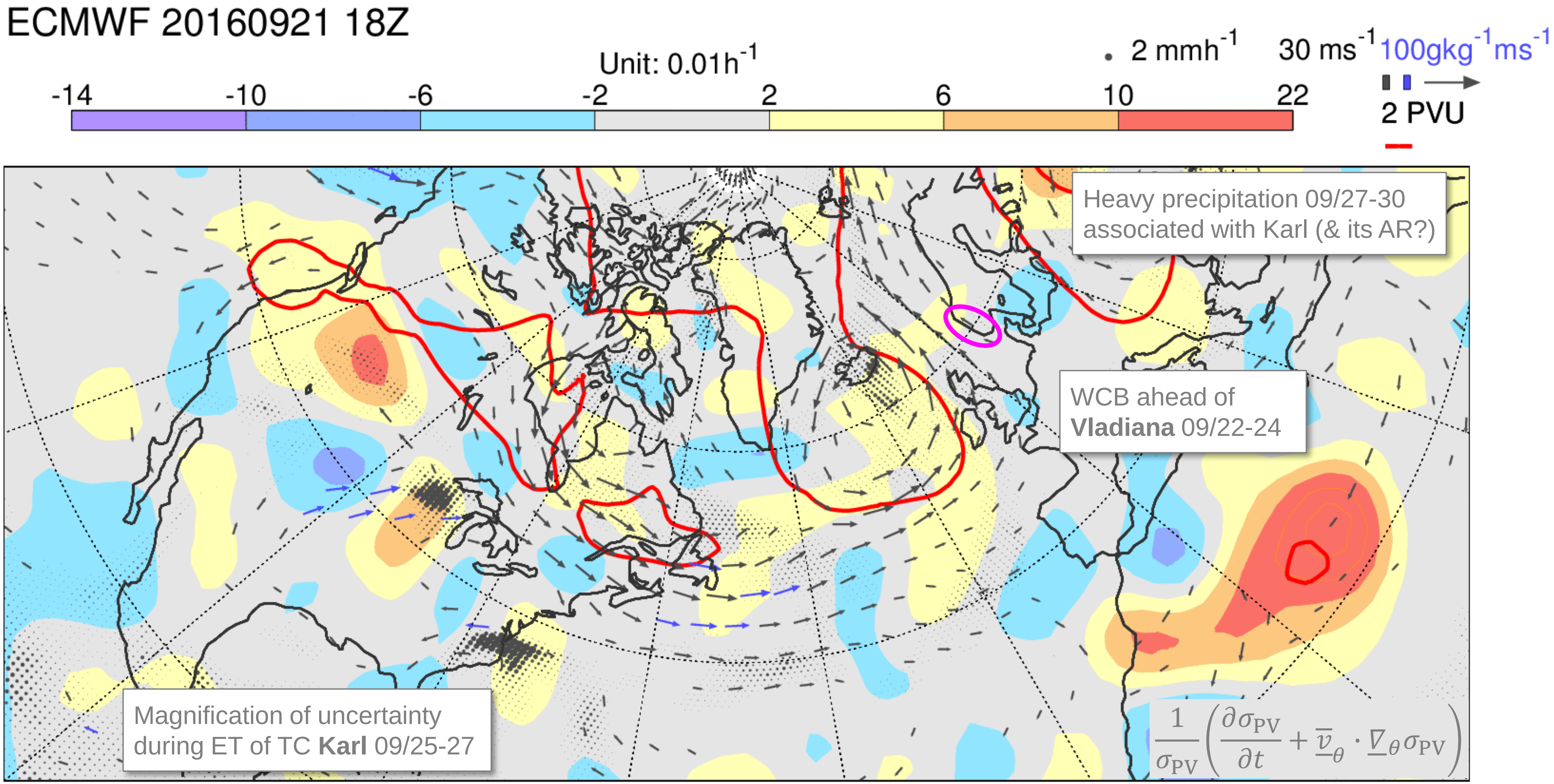
If we hit the string at the wrong time, the wave will arrive over Europe at the wrong time

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

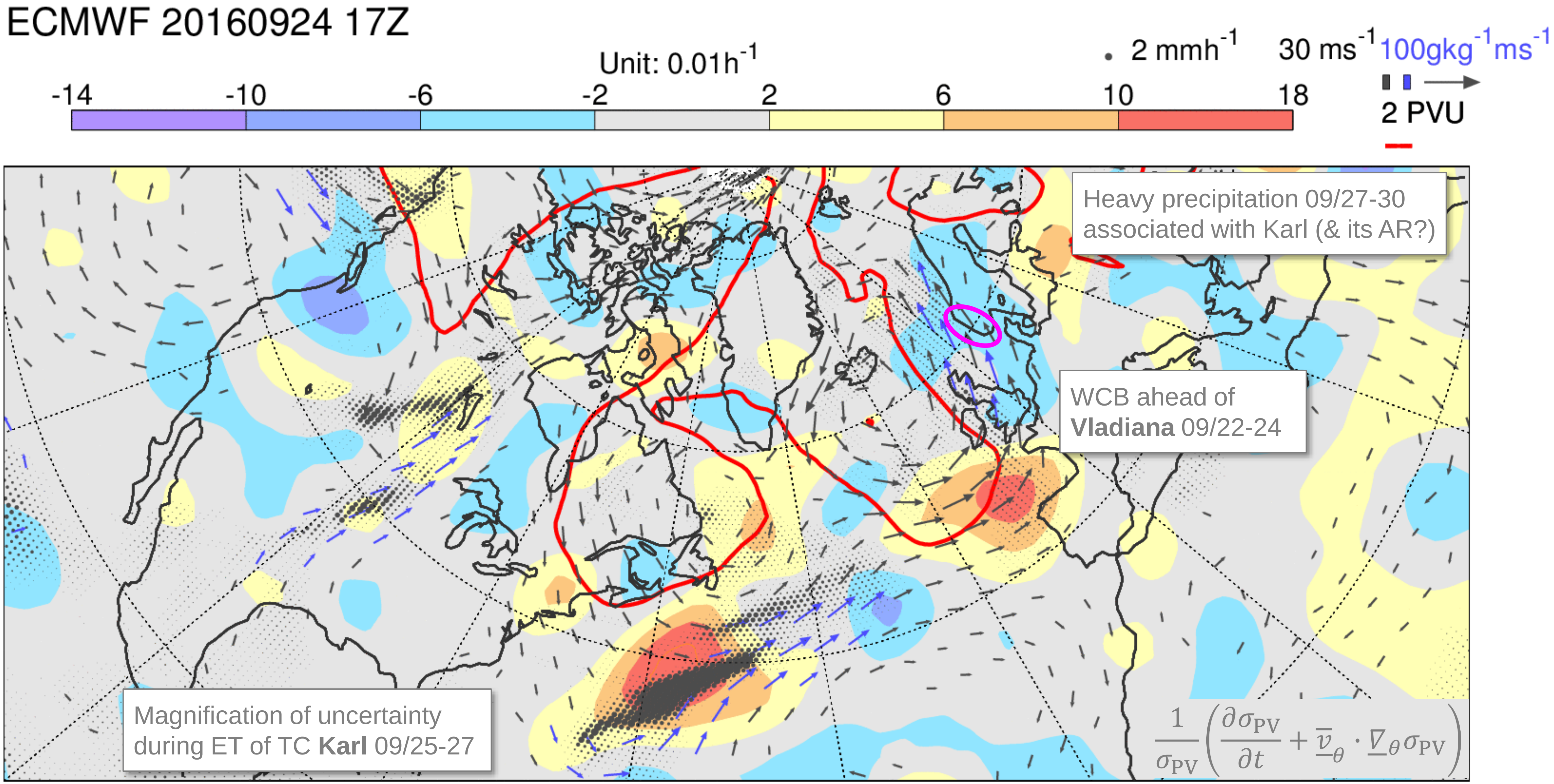


# Uncertainty growth-rate along truth trajectory – EDA $\sigma_{PV_{315}}$: NAWDEX Case




*$PV_{315}$=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*

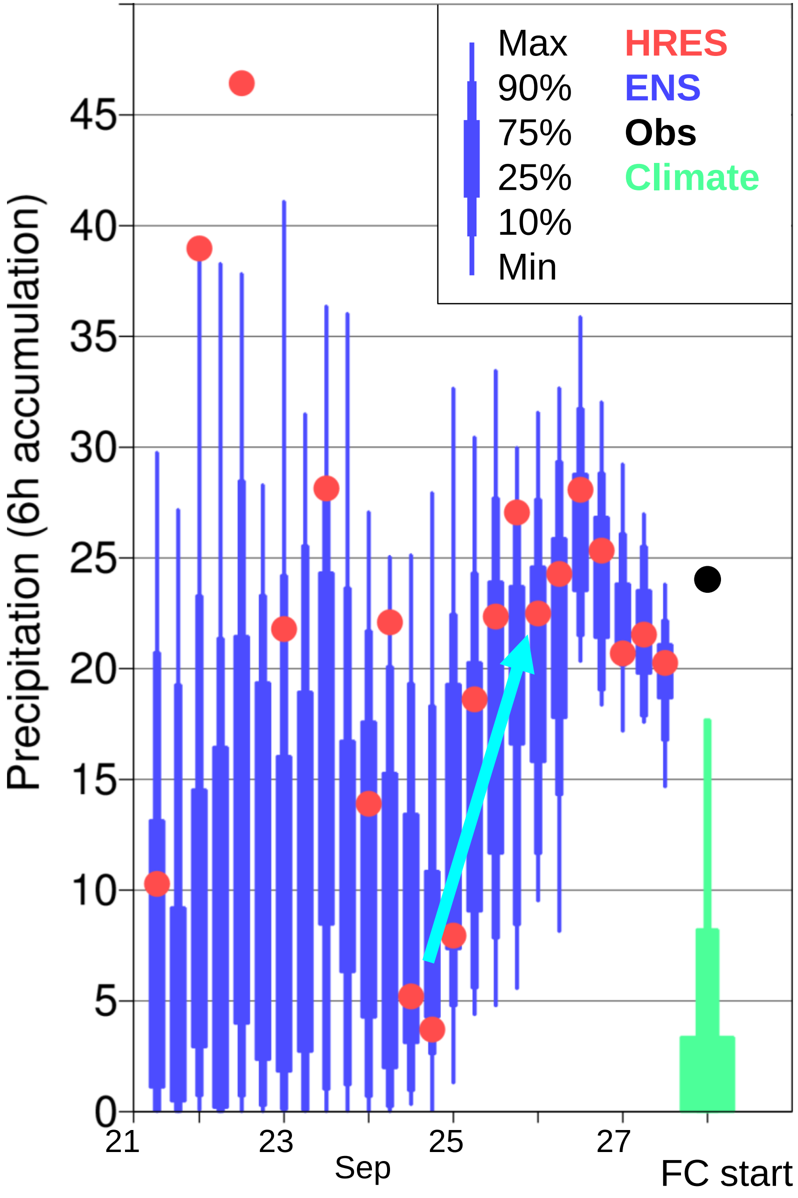
# Uncertainty growth-rate along truth trajectory – EDA $\sigma_{PV_{315}}$: NAWDEX Case




*$PV_{315}$=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*

# Precipitation forecast for Bergen, Norway on 27 Sep 2016 (12-18Z) following TC Karl

*Plot from Linus Magnusson*




Once uncertainties associated with the extratropical transition of Karl are resolved, the probability for strong precipitation firms-up

Note the observation is at the top of the last forecast distribution: Fine or reflecting issues with model representativity of point observations?

# Challenges and Limits in Ensemble Weather Prediction

- Reliability and Sharpness $\Rightarrow$ Skill
  - Faithful representation of uncertainty growth-rates (which are flow-dependent) – LIMIT
  - Better estimation of observational error – CHALLENGE

    … and correlated observation error
  - Assimilation of better observational information – CHALLENGE

# Challenges and Limits in Ensemble Weather Prediction

- Reliability and Sharpness ⇒ Skill
  - Faithful representation of uncertainty growth-rates (which are flow-dependent) – LIMIT
  - Better estimation of observational error – CHALLENGE

    … and correlated observation error
  - Assimilation of better observational information – CHALLENGE
- "Perfect storms"
  - Large inherent growth-rates – LIMIT

    … ⇒ downstream deterministic forecast "Busts"
  - Model physics (and stochastic physics) working near their limits – CHALLENGE

    … beyond their limits ⇒ Ensemble Jumpiness (& Det. Busts) – CHALLENGE
  - Difficulty to observe the truth. Non-linear observation operators – CHALLENGE

    … use of targeted observations? – CHALLENGE

# Challenges and Limits in Ensemble Weather Prediction

- Reliability and Sharpness $\Rightarrow$ Skill
  - Faithful representation of uncertainty growth-rates (which are flow-dependent) – LIMIT
  - Better estimation of observational error – CHALLENGE

    … and correlated observation error
  - Assimilation of better observational information – CHALLENGE
- "Perfect storms"
  - Large inherent growth-rates – LIMIT

    … $\Rightarrow$ downstream deterministic forecast "Busts"
  - Model physics (and stochastic physics) working near their limits – CHALLENGE

    … beyond their limits $\Rightarrow$ Ensemble Jumpiness (& Det. Busts) – CHALLENGE
  - Difficulty to observe the truth. Non-linear observation operators – CHALLENGE

    … use of targeted observations? – CHALLENGE
- Flow-dependent diagnostics of data assimilation
  - DA process tendency budget: mean increment $\Rightarrow$ process bias
  - EDA variance budget: mean residual $\Rightarrow$ wrong growth-rate or poor modelling of observation error

Thank you