Predictability Diagnostics 2

Mark Rodwell
(with the support of ECMWF and external collaborators)

ECMWF Training Course on Predictability

11 May 2017, ECMWF Reading
“All in a spin”
The Vorticity Equation

Motivation (2D flow):

\[
\zeta_z = \frac{\partial v}{\partial x} - \frac{\partial u}{\partial y} \quad (\equiv \hat{k} \cdot \nabla_z \times \mathbf{v})
\]

\(\hat{k}\) is the unit “vertical” vector and \(\nabla_z \times\) is the horizontal curl operator.

Curl of the 3D momentum equation in absolute frame of reference:

\[
\frac{d\zeta}{dt} = -\zeta (\nabla \cdot \mathbf{u}) + (\zeta \cdot \nabla) \mathbf{u} + \frac{1}{\rho} \nabla \rho \times \nabla p + \nabla \times F_u
\]

Lagrangian Divergence Tilting Baroclinic Friction
tendency in absolute vorticity

Shallow atmosphere approximation & assuming horizontal, barotropic, frictionless flow:

\[
\frac{\partial \zeta}{\partial t} + \mathbf{v} \cdot \nabla \zeta = -\mathbf{v}_x \cdot \nabla \zeta - \zeta \nabla \cdot \mathbf{v}_x
\]

\[= -\nabla \cdot (\mathbf{v}_x \zeta) \quad \text{"Rossby Wave Source"} \]
Based on operational analyses for the period DJF 2015/16, with terms integrated between 100-300 hPa.
Based on operational analyses and forecasts for the period DJF 2015/16, with terms integrated between 100-300 hPa.
Extra-tropical Rossby waves

100–300 hPa $v_\psi$, $v_\chi$, RWS

Group Speed (of wave-packet)

$\mathbf{c}_g = \frac{\partial \omega}{\partial k}$

Phase Speed

$\mathbf{c}_x = \frac{\omega}{k} = \bar{u} - \frac{\beta}{k^2}$

$\approx 17 - 7 = 10 \text{ ms}^{-1}$
Flow dependent predictability and reliability
Animation of a very poor medium-range single forecast

Potential Vorticity on the Potential Temperature = 320K surface. 20110410 0 UTC. Step (days, hours) = 0 00.0

Observed

Forecast
We see the mixing of air masses. The eventual block (high pressure) over Northern Europe is not well predicted with a single forecast, it is easy to quantify the error (pointwise differences, pattern correlations etc.)
All forecast centres suffered. Suggests an issue of reduced "predictability".
Ensemble members start from very similar conditions. Differences account for our uncertainty in the truth and are almost imperceptible to the eye here.

Differences then grow with lead-time and the members become completely different beyond about day 4.

Member 32 agrees well with the observed outcome. Simply a case of low predictability? How do we make progress?
Potential Vorticity on the Potential Temperature = 320K surface. 20110404 0 UTC. Step (days, hours) = 0 00.0
For this start-date, there were many more good\* forecasts
\(^*\text{RMSE less than second best member in the bust case}\)
Reliability and Sharpness – Example based on forecast of storm location

In a “reliable” forecast system, the truth can be considered as another ensemble member.

Reliability is very useful: if we predict an event with probability 70%, it will happen with frequency 70%.

A testable consequence of reliability is that:

average Error = average Spread
(averaged over many forecast start dates)

Given we had a reliable system, progress would be …

Predicting “sharper” (tighter) distributions while retaining reliability
(A more predictable day should also have a sharper distribution)
Reliability in ensemble forecasting

Error^2 = EnsVar + Residual

Error ≈ Spread
(when averaged over enough start dates - a consequence of statistical “reliability”)

Progress: While maintaining this relationship, we wish to reduce both error and spread (i.e. “shrink diagram in x,y”)
Proper scores (Brier, CRPS, Ignorance etc.) help ensure that development decisions lead us in this direction

(Cross-terms on squaring have zero expectation. EnsVar is scaled variance to account for finite ensemble-size)

Adapted from Rodwell et al. (2015) QJRMS
Ensemble spread and error

500 hPa geopotential height (Z500). “Error” is RMS of ensemble-mean error
Spread = ensemble standard deviation (scaled to take account of finite ensemble size)

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

... but uncertainty varies from day-to-day. The real reason we make ensemble forecasts. What causes this, and how can we evaluate it in our forecasts?

To make progress, we must avoid too much chaos, and look at the growth of uncertainty at very short lead-times

European Centre for Medium-Range Weather Forecasts

Rodwell 2016, ECMWF Newsletter

Mark J Rodwell
Average initial conditions of 584 single forecast “busts” over Europe at day 6

Trough over the Rocky mountains, with high convective potential ahead
Conducive to the formation of mesoscale convection
Can average over such cases to evaluate flow-dependent reliability and thus our model uncertainty
(Subsequent evaluation requires independent data to avoid misleading results)

‘CAPE’ = Convective Available Potential Energy
Mesoscale convection over Kansas

Systems grow to typically 500km in scale, with embedded convective cells and tornados.
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.
Reliability in ensemble data assimilation

Chaos makes it difficult to identify problems in the medium-range using the spread-error relationship
Go to much shorter lead-times – within ensemble data assimilation process
Need to take account for observation error
Obtain diagnostic equation to evaluate “instantaneous” growth of uncertainty

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

(Cross-terms on squaring have zero expectation. EnsVar is scaled variance to account for finite ensemble-size)
Evaluating model uncertainty in upper-level winds during “Rocky trough” situations

54 cases

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

(a) Departure\(^2\)  Unit: (m/s)\(^2\)

(b) Bias\(^2\)  Unit: (m/s)\(^2\)

(c) Spread\(^2\)  Unit: (m/s)\(^2\)

(d) ObsUnc\(^2\)  Unit: (m/s)\(^2\)

(e) Residual  Unit: (m/s)\(^2\)

(f) Obs. density  Unit: (2\(^8\)x2\(^8\)\(^{-1}\))\((12h)^{-1}\)

Ensemble spread highlights enhanced uncertainty around the Great Lakes
Large errors ensue

Errors are relative to observations that are also uncertain but, even if we take this into account, there appears to be too little spread (and model uncertainty) in this flow situation

Depar\(^2\) = Bias\(^2\) + Spread\(^2\) + ObsUnc\(^2\) + Residual
Reliability \(\Rightarrow E[\text{Residual}]=0\)
Initial tendency budget from control forecast during “Rocky trough” situations

T300, 54 cases

- Decomposing EDA control forecast into process tendencies shows how the model represents dynamics and physics of MCS
- The positive (and statistically significant) increment suggests observations are warmer than the model

Process tendencies accumulated over 12hr background, the analysis increment, and evolution of the flow
MCS – Jetstream interaction (composite)

- Increments emphasize model systematic error: MCS does not interact enough with Jetstream
- Also need to strengthen stochastic physics to increase background variance?
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 representation of model uncertainty and observation uncertainty
E.g. Increasing model uncertainty in convective situations and decreasing it in clear-sky situations could improve reliability. Sharpness can be improved with better (use of) observational information
Optimising forecast configuration and usage
How do we optimally combine information from the ensemble and the high-resolution forecast?

Is this dependent on lead-time?
In the example, the weight for the high-resolution forecast is $w_{HRES} = 3$ and the probability of 1mm precipitation = 9/13.

In the real case, find $w_{HRES}$ that maximises (e.g.) Brier Skill Score or Ignorance score.

Can do analytically by solving $\frac{dBSS}{dw_{HRES}} = 0$. 

Mark J Rodwell

European Centre for Medium-Range Weather Forecasts
At short lead-times, high-resolution system is very valuable. At longer lead-times weight → 1. Based on years 2001-2005.
The combined system is more skilful (on average) at all leadtimes and for each threshold.

Results are cross-validated so no artificial inflation of skill. Based on years 2001-2005.
• Rossby waves and the “Rossby Wave Source” – simple models make useful diagnostics
• Flow-dependent reliability is key – EDA reliability budget seems useful
  – Effective, efficient, focuses on reliability - not sharpness
  – Development and diagnosis of ensemble data assimilation likely to be key to future NWP progress
• Instabilities as magnifiers of uncertainty
• Approaches to optimising system configuration and combining multiple sources of information

Previous talk
• Waves and spatio-temporal variability
• Initial tendencies approach