Model Errors & Diagnostic Tools: Mark Rodwell
Diagnostics in weather prediction

Which model is best & why?

What are the common errors?

Do we have the correct uncertainty?

“The ensemble spread was too large!”

Smart & easy
Development decisions & publications

“Diagnostics Explorer”
Potential Vorticity on the Potential Temperature = 320K surface. 20110410 00UTC, VT = 20110410 00 UTC, step = 000 hr

Analysis

High Resolution Forecast

-30 0 0.3 1 2.3 5 10 20

Unit: PVU Area Mean: 0.07
Diagnostics

Deterministic forecasting (flow evolution to day-6)

It is difficult, by day-6, to disentangle model error from the natural growth of initial condition uncertainty (chaos)
Ensemble forecasting (initial conditions)

Potential Vorticity on the Potential Temperature = 320K surface. 20110410 00UTC, VT = 20110410 00 UTC, step = 000 hr
Ensemble forecasting (flow evolution to day-6)

Potential Vorticity on the Potential Temperature = 320K surface. 20110410 00UTC, VT = 20110416 00 UTC, step = 144 hr
The complexity of today’s models, with numerous interactions between physical processes and the resolved flow (including teleconnections), can make it very difficult to isolate the offending process(es). Single column and LES models can help, but these do not take into account the evolution of the resolved flow.

Figure from Peter Bechtold
An approach to over-coming the issues of predictability and model complexity: Look at initial process tendencies
First-guess = sum of all processes.
Analysis increment corrects first-guess error, and draws next analysis closer to observations.
Relationship between increment and individual process tendencies can help identify key errors.
Initial temperature tendencies and D+10 error


Dynamics
Radiation
Diffusion

Strong upper-tropospheric increments (where radiation is not balanced by dynamics)

Error grows x10 by D+10 (due to poorly constrained humidities?)

Note that increment and residual plotted with smaller contour interval. D+10 error also has different interval.
1st example: Method questions 12K warming

Temperature tendency profiles over the Amazon (300-320°E, 20°S-0°N)

Data assimilation using control model

Data assimilation using reduced entrainment model

Mean first guess tendency, red, (the sum of all processes) is ‘quite small’: A reference value for the realism of the model’s physics

Greatly increased first-guess tendency: Perturbation leads to poorer physics. Reject this perturbation from climate ensemble?

Rodwell and Palmer (2007). 6hr tendencies. 31 days (January 2005) X 4 forecasts per day. 70% conf.int. T159, L60,1800s.
Atmospheric motion vector wind (infrared, visible, and water vapour)

Sometimes the increments (or departures) may reflect observation issues
Mean zonal wind tendency (60-180°E) during MJO

Period: 20130201-27 (MJO convection active over warm-pool)

Better balance with dynamics when convective momentum transport is halved

Work with Peter Bechtold, Anton Beljaars, Jian Ling, Philippe Lopez, Frederic Vitart & Chidong Zhang

From Madden and Julian (1972)
Model climate response to Sahara aerosol change

Precipitation, 850hPa winds and 500hPa heights

- Tropical Rossby-wave response to aerosol change
- Local monsoon response explained by tendencies
- Kelvin-wave response - triggering secondary convection
- Forcing of extratropical Rossby-waves

Rodwell and Jung (2008). JJA season response to (primarily) a reduction in aerosol over the Sahara
‘Stretching’ and vorticity advection from Tropics

Rossby Wave Source: shading unit = $10^{-11}$ s$^{-2}$. Streamfunction: contour interval = $2 \times 10^7$ m$^2$s$^{-1}$. Divergent wind vectors

‘Stretching’ and advection account for $\frac{1}{2}$ to $\frac{1}{3}$ of RMSE of vorticity forcing at day-1

Reducing in this mean error should improve prediction of stormtracks
10 April Rockies trough with CAPE & MCS ahead

Z500 and CAPE anomaly

- Trough
- CAPE
- Contour Interval = 100m
- Shade unit = 100Jkg$^{-1}$
- Unit = mm
- 30ms$^{-1}$

12-hr Radar-observed precipitation

- Mesoscale convective systems (MCSs)
- Unit = mm
- 0 to 50

Z500 at 0UTC, CAPE at 6UTC (T+6), 12hr ‘NEXRAD’ Radar precipitation accumulated to 9 UTC

Rodwell et al. (2013)
Skill of single forecasts (Europe, leadtime = 6 days)

Score is the spatial Anomaly Correlation Coefficient (ACC) x 100 for 500 hPa geopotential height (Z500) over Europe (12.5°W – 42.5°E, 35°N–75°N). The date shown is the forecast start date.

‘Bust’ around 10 April
- Initial condition error?
- Model error?
- Reduced predictability?
The ensemble of first-guess forecasts develops spread over the first 12 hours associated with uncertainties in the prediction of a mesoscale convective system. The incorporation of new observations by the ensemble of data assimilations results in a contraction of the spread. Key question: Is the final analysis spread too large or too small to correctly reflect the predictability of the subsequent flow? Data: Temperature at 200 hPa from 10-member EDA, valid at 6UTC.
Composite ensemble spread & error (Z500 at day 6)

Background

‘Trough/CAPE composite’

Spread

Error

Error ≈ spread (system ‘reliable’ in the mean). e.g. stormtrack

30% increased error. Spread not fully predicting the reduced predictability?

Composite over all 84 events 10 November 2010 – 20 March 2012 (0 or 12UTC) with a strong trough over the Rockies and positive CAPE ahead. ‘Error’ is RMSE of ensemble-mean (dominated by random component), ‘Spread’ is ensemble standard deviation, scaled for finite ensemble.
Z500 and Precipitation: 23/08/2008

Contours: Z500
Shading: Precipitation (24hr accumulation)
Calculation of ‘SEEPS’ precipitation score

Failure to predict heavy precipitation ahead of Low over northern Europe, too much frontal precipitation to the south. A station’s climatology is used to define threshold between ‘Light’ and ‘Heavy’. In (f), the box size indicates a station’s relative contribution to the area-mean score. Forecast is day 4 24-hr accumulation on 20080823
Lead-times have a 365-day running-mean applied. Clear trends in (single) high-resolution forecast skill. A strong argument for continuing to monitor deterministic forecasts of precipitation. The slightrend in lead-time for the (fixed system) ERA Interim forecasts must be due to improvements in the observing system.
Precipitation forecast performance trend

“1 day-per-decade performance gain”

Extra-tropics

Lead-times have a 365-day running-mean applied.
Diagnostics

Category frequency (blue) & SEEPS error (red)

Diagonal panels show category frequencies \((\div 2)\)
Off-diagonal panels show SEEPS contributions for each error-type

DJF 2009/10. 24-hr precipitation accumulations, Extra-tropics

Too few dry days

Penalised for drizzle problem

Too much drizzle

Heavy precipitation has good frequency

2-category errors rare at short lead-times. High penalties later-on.

1-category errors more common at short lead-times. Can we improve these?

Ensemble and high-resolution information

EPS Meteogram
Madrid 40.33°N 3.6°W (EPS land point) 612 m
Deterministic Forecast and EPS Distribution Friday 17 January 2014 00 UTC

Total Precipitation (mm/6h)

EPS Control (31 km) High Resolution Deterministic (16 km)
In the example, weight$_{HRES}^{}$ = 3 and the probability of 1mm precipitation = 9/13
In the real case, find weight$_{HRES}^{}$ that maximises (e.g.) Brier Skill Score or Ignorance score
Can do analytically by solving $\frac{\partial BSS}{\partial w_{HRES}^{} } = 0$
The weight to give the high-resolution system

At short lead-times, the high-resolution system is very valuable. At longer lead-times weight $\rightarrow 1$. Based on years 2001-2005.
Combined system is more skilful

Brier Skill Score for the event that $P_p > x$ mm day$^{-1}$

<table>
<thead>
<tr>
<th>Lead-time (days)</th>
<th>CPS</th>
<th>EPS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.35</td>
<td>0.30</td>
</tr>
<tr>
<td>5</td>
<td>0.25</td>
<td>0.20</td>
</tr>
<tr>
<td>10</td>
<td>0.15</td>
<td>0.10</td>
</tr>
</tbody>
</table>

5% Significance
CPS > EPS

Results are cross-validated so no artificial inflation of skill. Based on years 2001-2005

~10hr improvement
It’s often sunny this time of year – shall we go hiking?

Weatherman says it’ll be sunny on Saturday

Weathergirl says Saturday doesn’t look so good now!

What a lovely time …
I actually think the countryside does look best in the rain …

Then & Now

MON

I’ll keep the weekend free

Weather app says there’s an 80% chance of a dry scenario this weekend

Let’s keep it free for mountain biking!

THU

I’ll run to the station and buy the train tickets

Its 50:50 on Saturday due to local showers, but 90% chance dry on Sunday

Let’s aim for Sunday, and invite friends over for lunch on Saturday

FRI

Let’s hope for the best!

Nearly 100% probability it’ll be sunny on Sunday!

I’ll book the train online for Sunday

SUN

Awesome time in the hills

I’ll run to the station and buy the train tickets

Nearly 100% probability it’ll be sunny on Sunday!

I’ll book the train online for Sunday
Summary

- **Diagnostic issues**: Chaos & complexity
- **Deterministic model error**: Initial tendencies within data assimilation cycle
  - Assessing models (e.g. Perturbed climate ensemble)
  - Identifying errors (Upper-tropospheric cold bias, convective momentum transport)
- **Ensemble distribution**: Spread & error of EDA (observation-space) & ENS
  - Quantifying flow-dependent uncertainty (What is an ‘ensemble bust’?)
  - Key processes that magnify uncertainty (e.g. MCSs, baroclinic instabilities, etc.)
  - Key initial condition errors & sub-grid-scale physics uncertainty
- **Understanding the circulation**
  - Targeting and monitoring key sources of predictability (e.g. Rossby wave source)
- **Diagnostics for users**
  - The deterministic and probabilistic weaknesses of the forecast system
  - Exploiting information from high-resolution and ensemble (combined approaches)