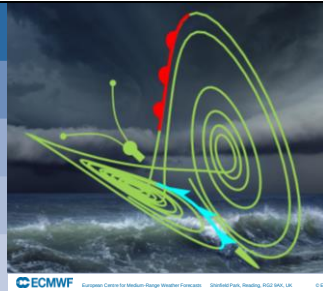


Slide 1



Diagnostics 2

Mark Rodwell

ECMWF Training Course on
Predictability

6 November 2024, ECMWF Reading

ECMWF

European Centre for Medium-Range Weather Forecasts

Sharncliffe Park, Reading, RG2 9AT, UK

© ECMWF November 6, 2024

mark.rodwell@ecmwf.int

1

Slide 2

Outline

- Tropical forcing of midlatitude Rossby waves
- Predictability, reliability and sharpness

ECMWF

European Centre for Medium-Range Weather Forecasts

Sharncliffe Park, Reading, RG2 9AT, UK

© ECMWF November 6, 2024

mark.rodwell@ecmwf.int

2

In this lecture, the aim is to first link back to the previous lecture by showing how tropical anomalies can influence the extratropical flow and could lead to extended range predictability for the midlatitudes. After this, the lecture will focus on the variance aspect of the ensemble, and how this can be evaluated. I will highlight the improvement of flow-dependent reliability as a practical path to increased ensemble forecast skill.

Slide 3

Outline

- Tropical forcing of midlatitude Rossby waves
- Predictability, reliability and sharpness

ECMWF

European Centre for Medium-Range Weather Forecasts

Sharncliffe Park, Reading, RG2 9AT, UK

© ECMWF November 6, 2024

mark.rodwell@ecmwf.int

3

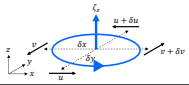
Slide 4

The Vorticity Equation

Motivation (2D flow) :

$$\zeta_p = \frac{\partial v}{\partial x} - \frac{\partial u}{\partial y} \quad (\text{or } \hat{k} \cdot \nabla_h \times \mathbf{v}) \text{ is the vorticity}$$

$\hat{k}$ is the unit vertical vector
 $\nabla_h \times$ is the horizontal curl operator



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

$$\frac{d\zeta}{dt} = -\zeta(\mathbf{V} \cdot \mathbf{u}) + (\zeta \cdot \mathbf{V})u + \frac{1}{\rho^2} \nabla p \times \nabla p + \nabla \times \mathbf{F}_a$$

Lagrangian tendency Divergence Tilting Baroclinic Friction



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

$$\frac{d\zeta}{dt} + \mathbf{v}_g \cdot \nabla \zeta = -\mathbf{v}_g \cdot \nabla \zeta - (\mathbf{V} \cdot \nabla) \zeta = -\mathbf{V} \cdot (\nabla \zeta) \quad \text{“Rossby Wave Source”}$$

ECMWF European Centre for Medium-Range Weather Forecasts Slide 2 (Revised)

Vorticity is a key attribute of the atmospheric circulation. This slide introduces the concept of vorticity and the figures hopefully give more physical insight into the mathematical terms.

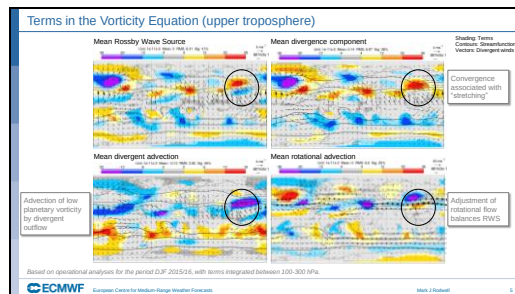
Note that there is some cancellation between the divergence and tilting terms in the vorticity equation. This cancellation is associated with “stretching”. Indeed, stretching only affects vorticity if it is accompanied by convergence into the axis of rotation: ballerinas spin faster because they bring their hands in, not because they stand up. The figures under the vorticity equation illustrate the four terms after this cancellation has been removed. The baroclinic term is non-zero if the gradients of pressure and density are not parallel (if they are parallel then the flow is barotropic). This term also explains the “sea-breeze” effect when the daytime air density over the warmer land is less than over the cooler ocean. The frictional term may simply act to damp the flow (as illustrated).

Interestingly, the same form of the vorticity equation also applies if we measure the 3D wind and the change in absolute vorticity relative to the rotating Earth. (The component removed from the material derivative to take account of the Earth’s rotation exactly cancels the component removed from the tilting term, and other terms are unaffected).

Having discussed the mechanisms in which the vorticity of a parcel of air can be modified, we will now assume

that the last three terms are negligible! (as a first-order investigation of mid-latitude synoptic scales). This is the case if the flow is horizontal, barotropic and frictionless. The equation at the bottom of the slide shows the remaining terms in the vorticity equation. Basically, the material derivative has been split into an Eulerian time derivative and advection parts (associated with the rotational wind $\vec{v} \cdot \nabla \psi$ and divergent wind $\vec{v} \cdot \nabla \chi$). This later term can be combined with the divergence term to produce the “Rossby Wave Source”. By placing this on the right-hand-side of the equation, we have in mind a situation where tropical divergent anomalies can force extratropical non-divergent response – tropical-to-extratropical forcing.

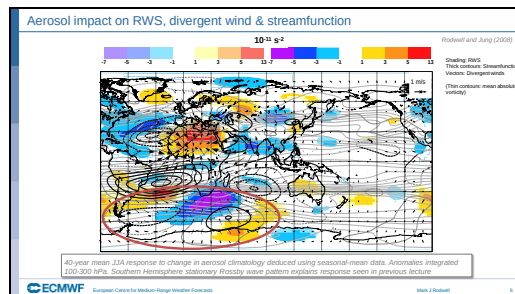
Slide 5



Here we see some terms of the barotropic vorticity equation averaged over the season DJF 2015/16. The circle in the top-left panel highlights the mean Rossby Wave Source (RWS) down-stream of Tibet. We see that both components are important in the RWS. There is a strong positive vorticity forcing due to convergence associated with stretching as the jet passes the peak in orography (top right). There is also a strong negative vorticity forcing due to advection by the divergent outflow associated with “Warm Pool” convection over Indonesia (bottom left). One can imagine how the MJO might alter this term as it passes-by. Since we have

averaged over a whole season, the local time derivative in the barotropic vorticity equation is negligible and so the mean RWS must be balanced, on average, by vorticity advection by the rotational flow; with implied adjustments to the streamfunction field (bottom right).

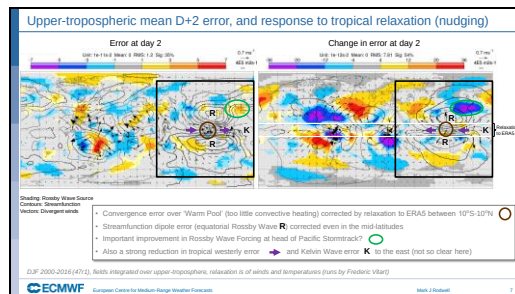
Slide 6



This plot summarises how the seasonal-mean upper-tropospheric anomalous divergent flow (vectors) can interact with the mean absolute vorticity (thin grey contours) to produce a Rossby wave source (coloured). The anomalous streamfunction associated with the induced extratropical stationary Rossby waves is also shown (thick contoured). There are, in fact, two Rossby wave trains in the Southern Hemisphere that occur on subtropical and extratropical jets, respectively. We can now understand the extratropical response seen in the previous lecture. This demonstrates the way tropical anomalies in general can force extratropical responses, and could lead to extended range predictability for the midlatitudes.

Note that other non-stationary Rossby waves are also likely to be excited but these do not appear in the seasonal-mean precisely because they are non-stationary and thus get averaged away.

Slide 7



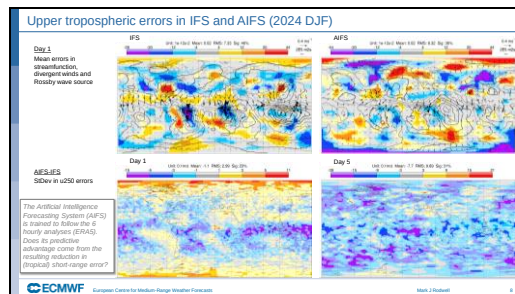
These plots show the Rossby Wave Source (shaded), Streamfunction (contoured) and divergent winds (vectors). All fields shown have been smooth spatially and integrated between 100-250hPa.

We focus on the East Asian longitudes. The left panel shows (re-) forecast errors at D+2, averaged over 16 years and 9 start dates per year within the December-January season. An upper-tropospheric divergence error is visible over the 'Warm Pool' region of the Maritime Continent. This indicates a deficit in convective heating. Comparison with rain gauge data appears to confirm this (c.f. Lavers et al 2021), although the data mainly samples land areas. We also see a dipole error in streamfunction straddling the equator. Between these two cyclonic streamfunction centres exists a strong westerly wind error along the equator. Although not so visible here (we need rotational and divergent winds to see it), there is also an easterly wind error extending out into the equatorial Pacific.

In parallel re-forecast experiments winds and temperatures are 'relaxed' (or 'nudged') towards ERA5 re-analysis data between 10°S and 10°N. The difference between the two sets of forecasts is shown in the right panel. Notice how the convergence error is removed. This is not surprising because it is basically imposed by the relaxation. However, the heating implied by this relaxation acts to force an equatorial Rossby mode which straddles the equator and extends way beyond the region of relaxation. (This is an $n=1$ Rossby mode as discussed in the previous lecture). The fact that the Rossby mode acts to correct the streamfunction dipole error demonstrates that the error is itself the Rossby mode response to reduced equatorial heating. The westerly wind errors along the equator, although corrected by the relaxation, are also consistent with the Rossby mode interpretation. The improvement in winds to the east is also consistent with a Kelvin mode response to the equatorial heating, although this is less clear in these plots. Combined, the Rossby and Kelvin modes match very well Gill's (1980) equatorial wave solution to equatorial heating.

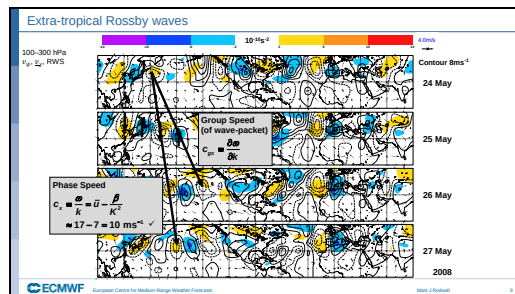
Finally, there is a tantalising mean improvement to the Rossby Wave Source at the head of the Pacific stormtrack which could demonstrate a causal link (teleconnection) from the tropics into the extra-tropics. Possibly a pathway to gaining extended-range extra-tropical predictability from the Madden Julian Oscillation, for example.

Slide 8



With the rise in data-driven forecasting, here I am showing (top) a comparison between the IFS and AIFS of their systematic errors in upper-tropospheric streamfunction, divergent winds and Rossby wave source. By being forced to follow successive 6 hourly atmospheric analyses, the AIFS appears to show reduced systematic errors in the tropics. In particular, the ITCZ does not weaken like we saw it do in the IFS, and deep convection over tropical land areas is better sustained. The bottom panels show the differences in the standard deviation of upper tropospheric wind errors for days 1 and 5. Here we see the AIFS makes improvements at day 1 in the tropics but, by day 5, errors are reduced globally. An interesting question is whether the extra-tropical improvements at day 5 are due to improved forcing from the tropics. The ability to learn from machine learning, and to even calculate sensitivities of the non-linear ML models to (e.g.) initial conditions is a new avenue for diagnostics to pursue.

Slide 9



Often we see wave-trains within the extratropics that tend to propagate eastward. Here, the contours highlight meridional flow anomalies in the upper troposphere (strictly, what is contoured is the rotational flow component of the total meridional flow. The shading is the RWS and the vectors are the divergent wind)). By drawing a line through some of these features it is clear that they are propagating eastward. What we are seeing is a Rossby wave train. In the tropical waves discussion in the previous lecture, you will recall that Rossby Waves propagate west. This westward (upstream) propagation is still occurring here (see the β -term), but the wave is now also being advected downstream by the Jet ($\bar{u}$). (The equatorial wave equations were linearized about a resting basic state, but this is even less appropriate here).

As an aside, for large-scale waves, k can be small-enough (e.g. in the range 2-5) that the upstream propagation balances the downstream advection and the wave becomes quasi-

stationary. Such stationary waves (or teleconnections) can be important for creating seasonal-mean anomalies that may be predictable if the waves are being forced by slowly-varying features such as tropical sea-surface temperature anomalies.

Notice that, in the figure, the head of the wave-train appears to move more quickly than the individual anomalies move themselves. This faster effect happens at the 'group velocity' of the wave packet.

When such waves arrive over the Rockies, they can lead to convective instabilities over the Eastern US. Forecast errors in these waves and magnification of these errors by baroclinic and convective instabilities can lead to a "perfect storm" whereby forecasts for Europe a few days later can be very uncertain (leading to so-called Forecast "busts").

Slide 10

Outline

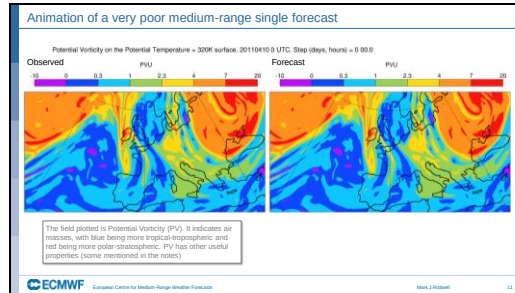
- Tropical forcing of midlatitude Rossby waves
- Predictability, reliability and sharpness

 European Centre for Medium-Range Weather Forecasts

Mark J. Roberts

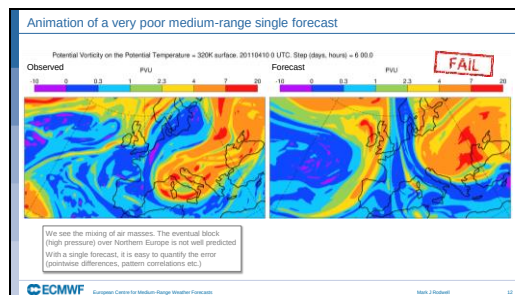
10

Slide 11



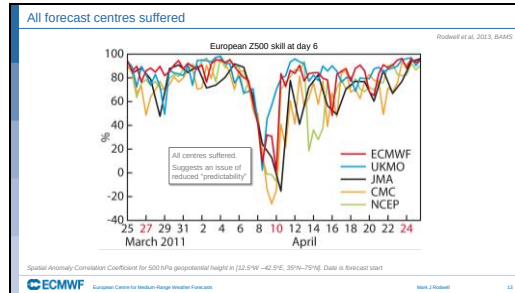
Here is an example that led to a very poor deterministic forecast at day 6. I am showing Potential Vorticity (PV) on the 320K isentropic surface. The orange colours highlight high PV air, primarily from the polar stratosphere. The blue colours highlight tropospheric air from lower latitudes. The green shading in-between highlights where the tropopause intersects this surface. PV is materially conserved in adiabatic, frictionless situations. Hence PV highlights nicely the atmospheric flow and the physical process (convection etc.) that are modifying it. In the next slide, we see an animation of the subsequent analysed and forecast flow.

Slide 12



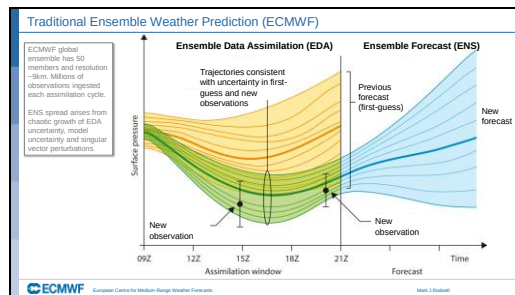
Over the first 6 days of the forecast, we see how the flow evolves. Terms that we previously dropped in the vorticity equation, such as the baroclinic term, are relevant here. We see examples of anticyclonic wave-breaking that, in the analysis, lead to a blocked-type flow over Europe at day-6. In this particular case, the single high-resolution forecast is particularly poor by day-6. What led to this forecast failure? Was it a model error? Was it that the initial conditions were unusually poor? Was it that the flow was more un-predictable than usual?

Slide 13



Here we see day 6 scores for several of the world's forecast centres. The score shown on the y-axis is the spatial anomaly correlation coefficient for 500 hPa geopotential over Europe. The x-axis indicates the forecast start date. In general, all centres score between 80 and 100% at this lead-time but, around the day of poor forecast shown in the previous animation, the scores for all centres drop dramatically. Although it is possible that there could be a common model error for all centres, it is more likely that this is a case of reduced predictability. How does the ensemble cope with this flow situation?

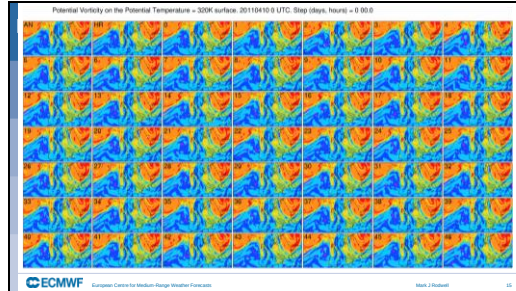
Slide 14



This is a schematic of the Ensemble of Data Assimilations (EDA) and Ensemble Forecast (ENS) at ECMWF. In the EDA, first-guess forecasts from 50 ensemble members (yellow) help quantify our (flow-dependent) uncertainty in the background. These first-guess forecasts are combined with new observations (black dots) in a way which is consistent with background uncertainty and observation error estimates. The result is a set of new model trajectories (green), which are then used to initialise* the ENS.

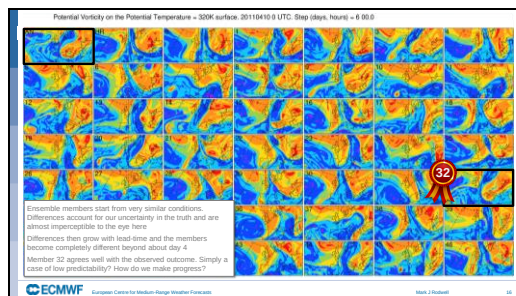
* Some modifications are currently made to the EDA in order to obtain the ENS initial conditions – these include re-centring and the addition of singular vector perturbations.

Slide 15



Here we see the same forecast start date, but I have also included the ensemble forecast system. For clarity, I am only showing the first 46 of the 50 ensemble members, along with the analysis (AN), high-resolution single forecast (HR) and the unperturbed ensemble control (0). Looking carefully, it is possible to see slight differences in the initial conditions of the perturbed ensemble members. These differences represent our uncertainty in the observations and also errors that are likely to grow most quickly in the early stages of the forecast. The initial conditions for the control forecast match more closely those of the high-resolution forecast (although this is not easy to see differences here).

Slide 16

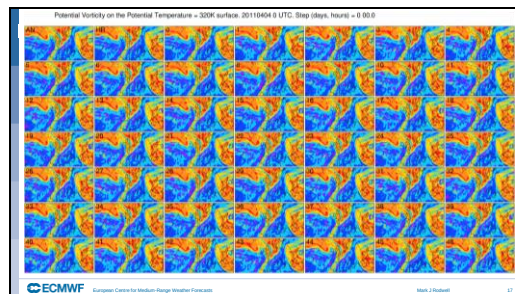


Again we see how the forecasts evolve over the first 6 days. In addition to the differences in initial conditions, the perturbed ensemble forecasts also include 'stochastic physics' (stochastically-perturbed physical tendencies – SPPT) which purports to represent, amongst other things, the effects of sub-grid-scale uncertainty (or non-determinism) for a given resolved flow field.

At day-6, there is a lot of 'spread' in the ensemble – reflecting our uncertainty in the outcome. But do any members capture the real outcome? One member (32) matches quite well the analysis. It is clearly possible that the issue is purely associated with low inherent predictability, but how do we

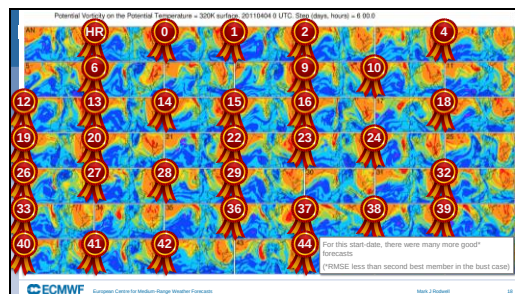
know if we could do better? Would a 'better calibrated' 50-member ensemble forecast have included more (or less!) good members? Ultimately, of course, we would like to tighten the ensemble distribution and increase the number of good forecasts. It may be constructive to try to identify the reasons behind the low predictability.

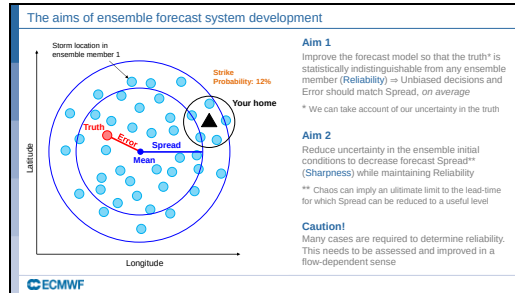
Slide 17



Here is an animation of the ensemble forecast from 6 days prior to the very poor forecast. In this case, the ensemble is both sharper and has many more good members.

Slide 18



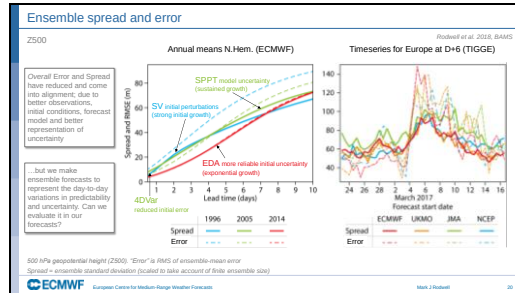


A key requirement of an ensemble forecast system is that it is “reliable”. By this, we mean that the eventual truth can be considered as another ensemble member. If this is the true then, when averaged over enough cases where we predict an event with probability p , the event should actually occur with frequency p . Reliability is a very important property for forecast users because it enables them to make unbiased decisions. If a decision could be based purely on the forecast then, for a reliable forecast system, one would decide to take action if the cost to do so was less than $p \times$ the loss that would otherwise occur.

One of the simplest consequences of reliability is that the mean difference of an ensemble member from the ensemble-mean should, on average, match the mean difference of the truth from the ensemble-mean (an adjustment needs to be made to account for the ensemble size, particularly for small ensembles). This is sometimes known as the “spread-error” relationship. It is a necessary condition for full reliability.

It is important to note, that reliability is not a guarantee of high skill – a constant forecast probability equal to the climatological frequency for an event is perfectly reliable, but not particularly skilful. We also require our forecasts to be sharp – with probabilities closer to 0 and 1.

Slide 20

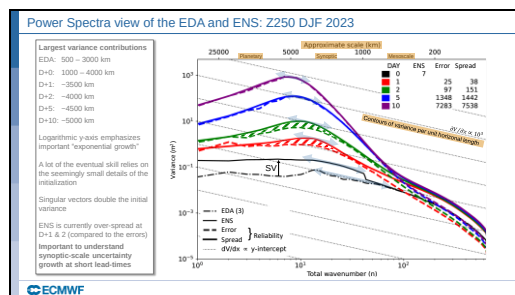


In the left plot, we look to see how the annual-mean agreement between spread and error has improved over the years. This is based on operational forecasts of 500 hPa geopotential heights (Z500) for three different years. In 1996, errors increased sharply from the initial time (blue dashed) and, in an attempt to match this, “singular vector” (SV; Molteni and Palmer 1993) perturbations were applied to the ensemble members to boost the initial growth of ensemble spread (blue solid). Nevertheless, at longer lead-times the spread failed to keep pace with the growing errors. By 2005 salient developments had been the introduction of the representation of model uncertainty (“stochastic perturbation to physical tendencies”, SPPT; Buizza et al. 1999) and 4-dimensional variational data assimilation (4DVar; Rabier et al. 2000). These developments allowed SV magnitudes to be scaled down, while improving the agreement between the error (green dashed) and the spread (green solid). Note also the smaller initial error. By 2014, a key development had been the introduction of the Ensemble of (4DVar) Data Assimilations (EDA; Isaksen et al., 2010). It is difficult to distinguish the spread (red solid) and error (red dashed) curves (for Northern Hemisphere Z500 at least). Notice also that the initial exponential shape of the error growth curve matches what one might expect based on theory. This remarkable achievement has involved many other incremental improvements to the model, the observations, and the modelling of observations and their

errors (e.g. English et al. 2000; Dee and Uppala, 2009; Jung et al., 2010, Geer and Bauer, 2011).

However, this is not the whole story. A perfect match of mean spread and error is only a “zero’th-order” requirement for forecast reliability. What we really need is flow-dependent reliability – we wish to “predict the predictability” that will vary from day-to-day. The right-hand plot shows daily variations in spread and error at a lead-time of 6 days for four different NWP centres. The agreement between centres of an increase in spread around 6 March suggests a flow-situation that is inherently less predictable. The partial agreement with the error curves indicates a degree of flow-dependent reliability. A key question is whether this can be improved further.

Slide 21



The degree to which spread and error agree (or disagree) can be shown as a function of the spatial scales in the atmosphere. In models like the IFS, state variables are sometimes stored as grid-point fields, and sometimes decomposed into a sum of 2D wave-like patterns covering the globe. These wave patterns are known as spherical harmonics. Ideally, the spread-error relationship should be satisfied for each such harmonic. In the power spectrum plot, harmonics are grouped by, and plotted against, their total wavenumber – this is the sum of the number of waves zonally and meridionally. Assuming a degree of

isotropy, the approximate spatial scale is indicated in the top x-axis, with planetary scales to the left, reducing to synoptic scales in the middle, and then to mesoscales and smaller towards the right. The thin dashed lines are contours of equal variance-per-unit-horizontal-length. These lines slope diagonally because the logarithmic x-axis squashes together more total wavenumbers as we move to the right.

For the EDA variance (black dash-dot curve at the bottom), the largest contribution to variance is at scales around 500-3000km. The addition of the singular vectors, which are computed at scales > 1000km, acts to shift the largest variance contribution of the ENS initial conditions (black solid curve) to slightly longer scales (1000-4000km).

The forecast variance (plus the variance of the verifying EDA analysis*, 'Spread', thick solid) is shown for lead-times 1d (red), 2d (green), 5d (blue), and 10d (purple). As lead-time increases, the largest uncertainty contribution shifts to longer scales ~5000km.

The logarithmic y-axis emphasizes the theoretical exponential growth, which we saw in the previous slide for 2014 – after the introduction of the EDA. Although initial uncertainty is 'small', the details of this uncertainty are key to the eventual skill at longer lead-times.

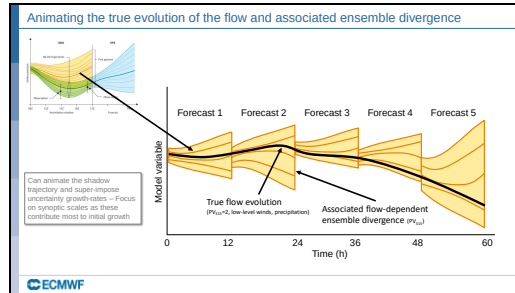
Comparison with the ensemble-mean error variance (thick dashed) suggests

that the ensemble is over-spread in the synoptic scales at days 1 and 2 (red and green hatching highlights the difference). In the next slides we will look to understand this over-spread. For several potential reasons, the agreement is much better by day 5, but it can be argued that the damage to predictive skill out to day 10 has already been done.

These results motivate the diagnosis of synoptic-scale uncertainty growth at short lead-times (0-2 days).

For scales smaller than about 400km, the diagnosed agreement between error and spread (+analysis uncertainty) appears poorer. These scales are over-emphasized by the logarithmic y-axis, and do not contribute much to the total uncertainty (shown in the top right table). One potential explanation for the disagreement is a reduced ability for current observations and data assimilation systems to constrain the analysis at these small scales.

* it can be important to consider the uncertainty (variance) in the verifying analysis, particularly at short lead-times.

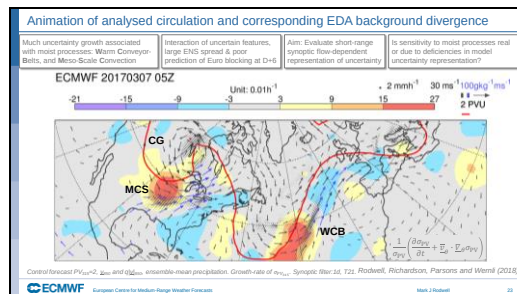


A great advantage of looking at short-range uncertainty growth is that all ensemble members will be representing essentially the same synoptic situation. This means that we can go beyond the statistics of spread growth, and investigate the physical/dynamical reasons for the growth in a flow-dependent way. Can we diagnose which flow-features act to strongly magnify uncertainty, and why?

Going back to the schematic of the EDA/ENS system, we can think of using the background/first-guess forecasts within the EDA system (yellow curves). By concatenating sequences of such 12h forecasts, we can produce an animation of the 'true' evolution of the flow (to within the small uncertainty of the background) as represented by the black curve, and super-impose the divergence of the members over each 12h interval. For the IFS, this divergence will be associated with chaotic growth of uncertainty from the previous EDA analysis and from the impacts of model uncertainty representation (no singular vector perturbations are included in the EDA).

When comparing different models – for example within the TIGGE archive – we can simply do something very similar with the first 12h of each ensemble forecast. Comparisons could reflect differences in the spectra of initial uncertainty, the effects of the deterministic chaos in each model, model uncertainty and any other initialisation procedures.

Slide 23



Spread and error curves at, say, day 6 represent the 6-day integral of the models' instantaneous (and flow-dependent) uncertainty growth-rates, and it would be preferable if we could diagnose deficiencies in these instantaneous growth-rates. Here we go some way to achieving this by diagnosing deficiencies at synoptic spatio-temporal scales. To illustrate this synoptic-scale growth-rate, and with cases of large forecast uncertainty in mind, we focus on the jet stream level where uncertainties are likely to propagate most rapidly and be magnified by subsequent down-stream cyclogenesis. A useful quantity for this purpose is potential vorticity, PV, on the $\theta=315\text{K}$ isentropic surface. We will estimate the growth-rate by calculating tendencies in the ensemble standard-deviation σ_{PV} in the 12h background (first-guess) forecasts of the EDA. Because of the strong advection by the jet stream, flow features associated with strong amplification of uncertainty are better identified by calculating the growth-rate following the horizontal ensemble-mean flow,

$\frac{d}{dt}$, along the isentropic surface (see figure for equation). The time derivative term, which represents the local growth-rate, is estimated using a 1h timestep. A spatio-temporal filter is then used to highlight uncertainty growth that projects onto the synoptic scales. The shading shows an example of the resulting field which is representative of the situations investigated later in a more systematic way. The red region over North America indicates a strong source of uncertainty. The 850hPa winds in this region (blue arrows) indicate relatively warm moist southerly flow which is associated with high values of Convective Available Potential Energy, CAPE (not shown) and an increased likelihood for intense convection and mesoscale convective systems (MCSs); the black dots show ensemble-mean precipitation - and thus the general location where precipitation is likely. The large uncertainty growth-rates thus suggest that the ensemble places considerable uncertainty in the location(s) and magnitude(s) of this convection; possibly reflecting inherent dynamical predictability limits or the model's limited ability to represent such convection with a ~18km grid resolution. The red line indicates the tropopause position (2PV units on the 315K surface) and this highlights a trough feature over the Rocky Mountains (in the animation you can see the Rossby wave character of its evolution over the North Atlantic). The jet stream (generally just to the south of the red line) will be affected by the uncertainty generated by this baroclinic and convective situation,

and it will propagate this uncertainty (as well as the signal) down-stream. The figure also indicates enhanced uncertainty growth-rates over the North Atlantic which appear to be associated with a Warm Conveyor Belt. WCBs are associated (in the ECMWF model) with more slantwise ascent than the MCSs discussed above, although they also include embedded convection. As the flow evolves over the next few days (as seen in the animation), the uncertainty associated with the MCS activity is advected northwards into a developing strong ridge, and the uncertainty associated with the WCB is co-located with a developing strong trough. The subsequent interaction of the ridge and trough, and the pooling of two major sources of uncertainty, seems likely to be a major factor in the very large uncertainties seen at day 6 over Europe for forecasts starting between 5 and 7 March 2017 on the slide, a few slides previous.

