PART I

Diagnostics and prediction of high-impact weather

1

Global prediction of high-impact weather: diagnosis and performance

Mark Rodwell and Alan Thorpe

1.1 Introduction

Weather that has a high impact on society is generally rare, but when it occurs it often has devastating impacts; being rare it lies in the tail of the climatological distribution of weather events. Impacts can be as a result of a short-lived extreme (such as very strong local winds) or a more persistent regime (such as a period of drought). Usually these phenomena are embedded within larger-scale weather systems and owe their existence to patterns of flow and sources of heat and moisture that are associated with the larger-scale circulation. At forecast lead-times beyond a few days, this large-scale flow can only be predicted by global models. It is because of this embedded nature of high-impact weather systems, and the authors' location at the European Centre for Medium-Range Weather Forecasts (ECMWF) that we focus here on global predictability and forecasting aspects. Of course for a more complete picture of high-impact weather, it is necessary to overlay the detailed structures evident in regional forecasting systems.

There has been huge progress in weather science over the past half century and more. This has led to significant advances in our ability to make skilful predictions of weather even as far ahead as two weeks. These advances have been as a result of scientific research, and increasing observational and computational capabilities. Modern numerical weather prediction (NWP) relies on international cooperation to exchange observations made across the globe and to share scientific findings. It also relies on advanced numerical models that allow the laws of physics to be 'solved', starting from an initial description of the current state of the atmosphere (and increasingly of the ocean and land surfaces). These global models require supercomputers of incredible speed – today 200 trillion floating point operations per second is not untypical – to enable high resolution predictions to be made in real time so that society has access to timely forecasts.

As high-impact weather events are relatively unusual, it is important to draw on long observational records. The past 30 years or more have seen a rapid increase in the availability of satellite observations, and this means we have a much better description of weather phenomena to draw upon. It is now possible, and it is routinely done, to carry out re-analyses of past weather and use today's science and models to re-forecast this weather. Such re-analyses and re-forecasts are an invaluable source of information on the frequency, properties, and our current skill in the prediction of high-impact events.

A key advance has been in the resolution at which these global models can operate. Today at ECMWF, the high resolution global forecast has a grid mesh of around 16 km by 16 km. This means that smaller-scale phenomena can now be explicitly resolved, and even predicted. For example the super-storm Sandy, that affected New York in late October 2012, was predicted relatively well by global forecast models even as far ahead as eight days; see Magnusson *et al*. (2014). Nevertheless, further improvements in resolution continue to be needed if the details of high-impact weather are to be resolved.

As longer and more accurate data records become available, and as model resolution increases, the scientific focus on high-impact weather is accelerating. This chapter explores new diagnostic tools and methods of forecast evaluation that help to reveal how and why predictions of high-impact weather can be improved as part of the overall pursuit of accurate and reliable weather forecasts.

In this chapter, we focus mainly on instantaneous extremes (short-lived weather extremes). We highlight current forecast system performance and demonstrate some useful diagnostic techniques. The chapter is structured as follows. We discuss systematic model errors in Section 1.3, and discuss estimates of uncertainty (observation, background, and model) within the probabilistic (ensemble) context in Section 1.4. A summary is given in Section 1.5 and then, in Section 1.6, we identify future avenues to pursue in order to improve the forecasting of high-impact weather. Firstly, however, it is useful to discuss

how NWP is done today – the framework within which extreme weather is predicted.

1.2 Global NWP: how it is done today

Since, as we have noted, high-impact weather is usually embedded within larger-scale weather systems, and these systems can only be predicted beyond a few days with global models, we start with a brief introduction to global NWP.

Numerical weather prediction is largely an initial value problem. As Fig. 1.1 shows schematically, it involves an estimation of the current state (analysis) of the atmosphere, and the integration of a numerical model of the laws of physics from this analysis forward in time (as indicated by the medium-range forecast). Error growth with lead-time (e.g. due to chaos: Lorenz 1963, and biases: e.g. Jung *et al*. 2010) is restrained by the periodic introduction (assimilation) of new observations. The assimilation process combines short-forecast (background) information from a previous analysis with these new observations in a way that is consistent with estimated errors and uncertainties in the observations (Isaksen *et al*. 2010), background (Bonavita *et al*. 2011), and sometimes the model (Trémolet 2007; Fisher *et al*. 2011). A good data assimilation system is crucial for the success of NWP. Turned on its head, data assimilation also represents fertile ground for the diagnosis of errors and uncertainties – as we will discuss here in the context of severe weather.

One component of ECMWF's Integrated Forecast System (IFS; for a recent listing of further details see, e.g. Jung *et al.* 2012) is a single 'high resolution' forecast, known as the HRES. Figure 1.2 highlights trends in HRES performance, as measured by spatial anomaly correlation of 500 hPa geopotential height. For the Northern Hemisphere, the rate of improvement has been at about one day's lead-time per decade. The relative lack of conventional observations (from Synop, radiosonde, aircraft, etc.) of the Southern Hemisphere, and the increase over the last 30 years in global satellite data coverage, explain the Southern Hemisphere's faster improvement from a lower starting value.

One contributing factor to the improvement in forecast performance over the last 35 years has been the increase in model resolution (see, e.g. Simmons *et al*. 1989) – permitted by the increase in available computing power. HRES horizontal mesh size has roughly halved every eight years. At present, the HRES model stores the state of the

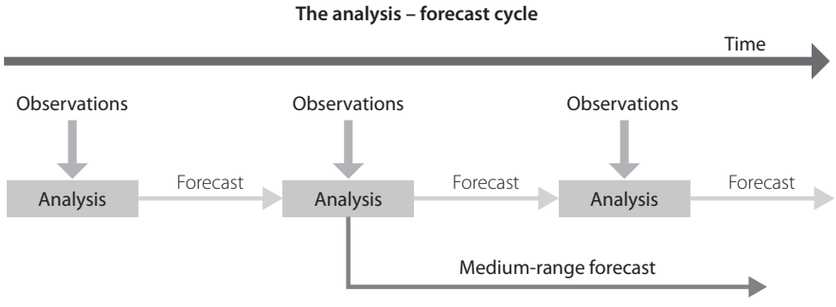

Fig. 1.1 Schematic of the analysis–forecast cycle. See main text for explanation.

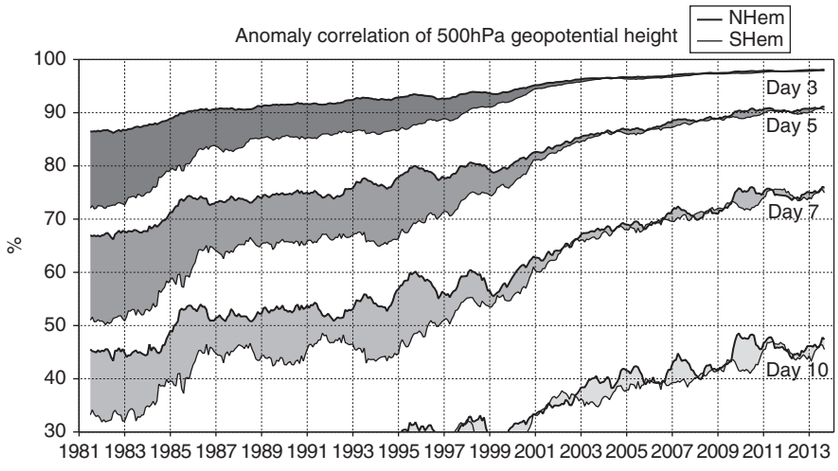

Fig. 1.2 Trends in ECMWF HRES performance, as measured by spatial anomaly correlation of the extratropical 500 hPa geopotential height field. Anomalies are from the climatology of the ECMWF 'interim' re-analysis (ERA-Interim, Dee *et al*., 2011). Scores are shown for the northern and southern extratropics (poleward of 20° latitude) and for four lead-time ranges.

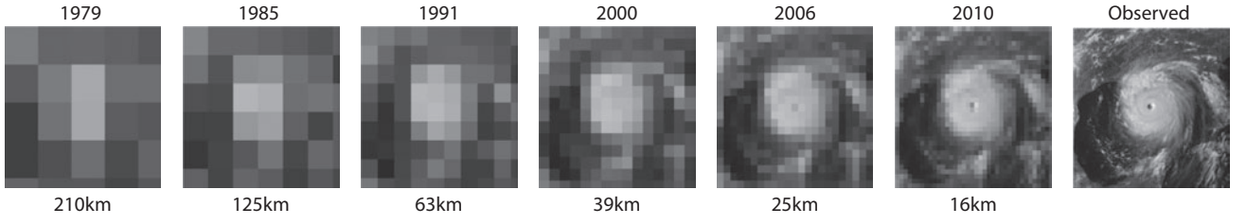

Fig. 1.3 Diagrams showing how well a particular tropical cyclone would have been resolved over the years as HRES model resolution has increased.

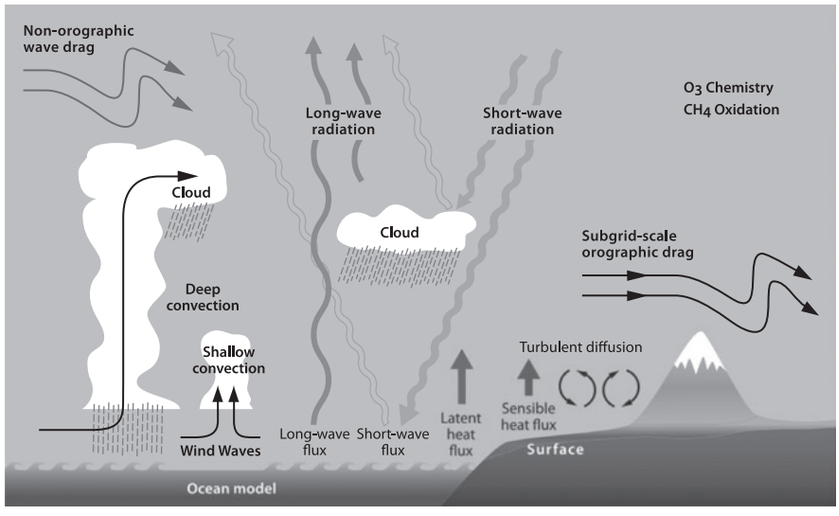

Fig. 1.4 Schematic diagram showing some of the physical aspects represented in the current ECMWF forecast system.

atmosphere in around five billion variables. Figure 1.3 demonstrates the impact of resolution on how well a given tropical cyclone would have been represented in the forecast model. It is only by being able to resolve such severe weather features that we stand any chance of being able to predict them; and it is only relatively recently that we have been in that position.

In addition to resolution effects, fundamental to the increase in forecast skill has been the improvement in observation, assimilation, and model aspects. At ECMWF, the number of assimilated observations has increased greatly over the years, reaching about one million per day by the year 2000, and now exceeding 40 million per day (still orders of magnitude less than the 30 billion state variables). There have also been improvements in quality and usage of observations, for example with the introduction of variational bias correction in 2006 (Dee and Uppala 2009; see also Derber and Wu, 1998) and the direct assimilation of 'all-sky' microwave radiances in 2009 (Bauer *et al*. 2010, Geer *et al*. 2010). The ECMWF assimilation system has also improved; for example changing from 3D variational data assimilation (3D-Var) to incremental 4D-Var (which allows interpolation in space and time; see, e.g. Courtier *et al*. 1994) in 1997, and with the introduction of flow-dependent background errors in 2011 (Isaksen *et al*. 2010). 4D-Var currently takes as much computer power as the ten-day forecast – a sign of the importance attached to the data assimilation process. The ECMWF forecast model has also been extensively developed over the years. The physical aspects currently represented (parameterized) in the model are shown in Fig. 1.4. Improvements to parameterizations have included, for example, the introduction of a prognostic cloud microphysics scheme in 2010 (Forbes *et al*. 2011). At the heart of model development is a drive to 'get things right for the right reasons'.

1.3 The need for dynamical understanding: getting things right for the right reasons

Statistics of the forecast system are essential for monitoring progress and for assessing the impacts of possible system upgrades. However, if we are to initiate new developments, there is also a need for dynamical understanding; and dynamical paradigms can be helpful in diagnosis. For

example, an understanding of the importance of Rossby waves – their triggering and propagation – motivates the development of diagnostic tools based on the barotropic vorticity equation; the Rossby wave source for example. Sometimes one can think of small-scale weather being forced from the larger scale – perhaps associated with regime behaviour. Potential vorticity (PV), which is essentially the product of the vorticity and the stratification, provides a natural and meaningful way of connecting the dynamical and physical components. 'PV thinking' (Hoskins *et al*., 1985; Thorpe, 1986; Joos and Wernli, 2012) can help us in designing diagnostics to understand such connections in reality and in forecast error.

Figure 1.5 shows the analysed surface pressure and (isentropic) PV on the 330 K isentropic surface for 21 July 2012. These fields highlight a strong extra-tropical cyclone in the North Atlantic. This cyclone was 'high-impact' not only from being strong, but also because it occurred shortly before the opening of the London 2012 Olympic Games. Within the cyclone, high PV values are predominantly associated with air from further north and higher up, including from the stratosphere. In general, as such air descends, it can form a 'tropopause fold' and sometimes leads to damaging sting jets at the surface, or to the triggering of secondary frontal cyclones. To the east, low PV values are partly associated with air from further south and lower down, and partly as the result of a reduction in upper-tropospheric stratification due to mid-tropospheric latent heat release and upper-tropospheric radiative cooling. Errors in this physics may lead to forecast errors in the developing downstream trough (see, e.g. Madonna, 2013).

We can use the 'initial tendencies' method to obtain a better understanding of the salient processes involved in the development of this cyclone, and possible errors in the parameterization of these processes. Figure 1.6 explains the basis of this method – which is centred on the data assimilation system. As we have noted above, the data assimilation provides the 'next analysis' by combining information from a previous short-range first-guess forecast with the new observations. The short forecast provides a complete set of model fields, but these will include forecast errors. The observations are more scattered but are bias-corrected (in general). The data assimilation combines these sources of information in a way that is consistent with the estimated observation and short-forecast errors. The difference between the new and previous analyses is an estimate of the true evolution of the flow over the assimilation window. The analysis increment is the difference between the new analysis and the first-guess forecast, and can be thought of as a correction to the first guess. The fact that the first-guess forecast can be expressed as the sum of the tendencies of each of the processes represented in the model – dynamics, radiation, convection, etc., provides a means of relating increments to model process errors.

Figure 1.7 demonstrates application of the initial tendency method to the case of the strong cyclone shown in Fig. 1.5. Figure 1.7 (a) to (e) show the initial tendencies on temperature at 500 hPa (T500) from the various processes represented in the IFS model, each accumulated over the first 12 hours of the HRES forecast initialized from the 6UTC analysis on 21 July 2012. This is the forecast used as the first guess for the 12UTC analysis depicted in Fig. 1.5. Figure 1.7(g) shows the corresponding analysis increment, and Fig. 1.7(h) shows the analysed evolution of the atmospheric state. The total budget is closed with the

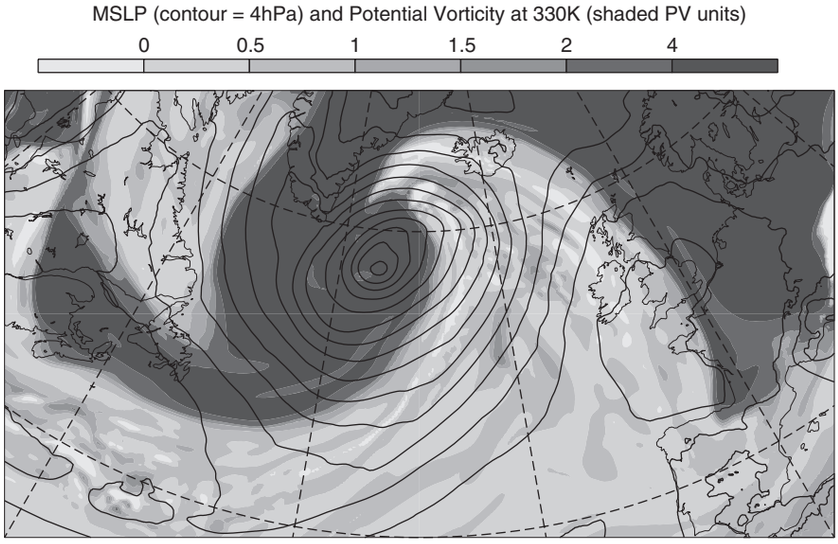

Fig. 1.5 Mean sea-level pressure (contoured) and potential vorticity on the 330 K isentropic surface (shaded) from the HRES analysis at 12UTC on 21 July 2012 (from 12-hour window data assimilation).

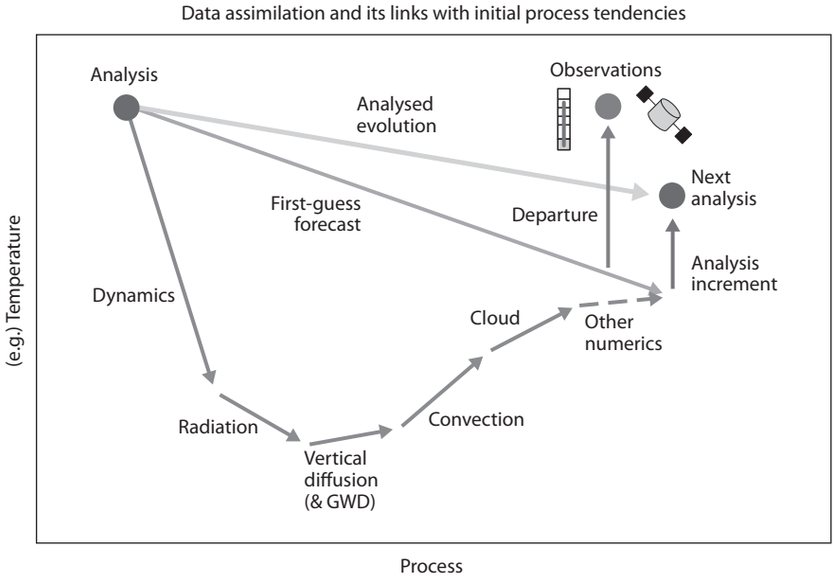

Fig. 1.6 Schematic showing the links between data assimilation and initial process tendencies. Please see the main text for more discussion.

residual term shown in Fig. 1.7(f), which reflects other minor processes and numerical aspects. Hence the sum of terms (a) to (g) is equal to (h). Note that the increment and residual terms are smaller in magnitude and plotted with a smaller shading interval.

Figure 1.7(h) highlights the movement of the cold and warm fronts associated with the cyclone. Much of this evolution is purely dynamical, as shown by Fig. 1.7(a), but other terms are also important. In particular, notice the positive temperature tendencies associated with the convection (Fig. 1.7d) and cloud (Fig. 1.7e) within the cold front, and the strong warming at the cloud-head (Fig. 1.7e). These terms are comparable in magnitude with the dynamical tendencies (Fig. 1.7a), and suggest that moist physics can play an appreciable role in the overall evolution of 'extreme' events – via the modification of PV for example. The negative analysis increments around the northern perimeter of the cloud-head (Fig. 1.7g) suggest that this cloud-head warming leaves the first-guess forecast a little too warm. Although further investigation would be needed, this is an identification of a possible model error – and highlights the power of this technique. Figure 1.7(i) shows the 'departure' of one observation field (AMSUA channel 5 microwave radiance observations) from the same radiance field obtained from the first-guess forecast. This microwave channel observes mid-tropospheric temperatures. Notice the similarity in patterns between the departures and the analysis increments. It is clear that this is one observation source that is 'supporting' the increments. Notice that the departure field in Fig. 1.7(i) is missing (white) in some key areas because it is still difficult to use this data in cloud-affected regions. This is indicative of a key issue for forecasting – the (moist) instabilities that act to magnify forecast error (including severe cyclones) are often the most difficult to analyse, and can lead to particularly large forecast errors at longer lead-times.

The technique of 'initial tendencies' is a refinement of the methodology proposed by Klinker and Sardeshmukh (1992). Recently, it has been applied to averages of several analysis cycles, to highlight more definitively model issues associated with convective momentum transport and upper tropospheric humidity drift (Leroy and Rodwell, 2014), as well as aiding the understanding of the beneficial impacts of a change in model aerosol climatology (Rodwell and Jung, 2008) and for assessing perturbed climate model ensembles (Rodwell and Palmer, 2007). A variation of the approach is discussed by Mapes and Bacmeister (2012). Klocke and Rodwell (2014) show that initial tendencies can identify an introduced model error, when diagnosis at longer lead-times cannot. Indeed, the ability of the method to identify potential errors lies in its application very early on in the forecast – which minimises complications associated with interactions between processes and the resolved flow, the growth of chaos, and variations in predictability.

1.4 The need for uncertainty information: assessing the degree of confidence

Incremental reductions in initial condition and model errors over the years have led to reduction in the growth of error from chaos and bias throughout the forecast. Figure 1.8(a), which shows annual means of the root-mean-square error

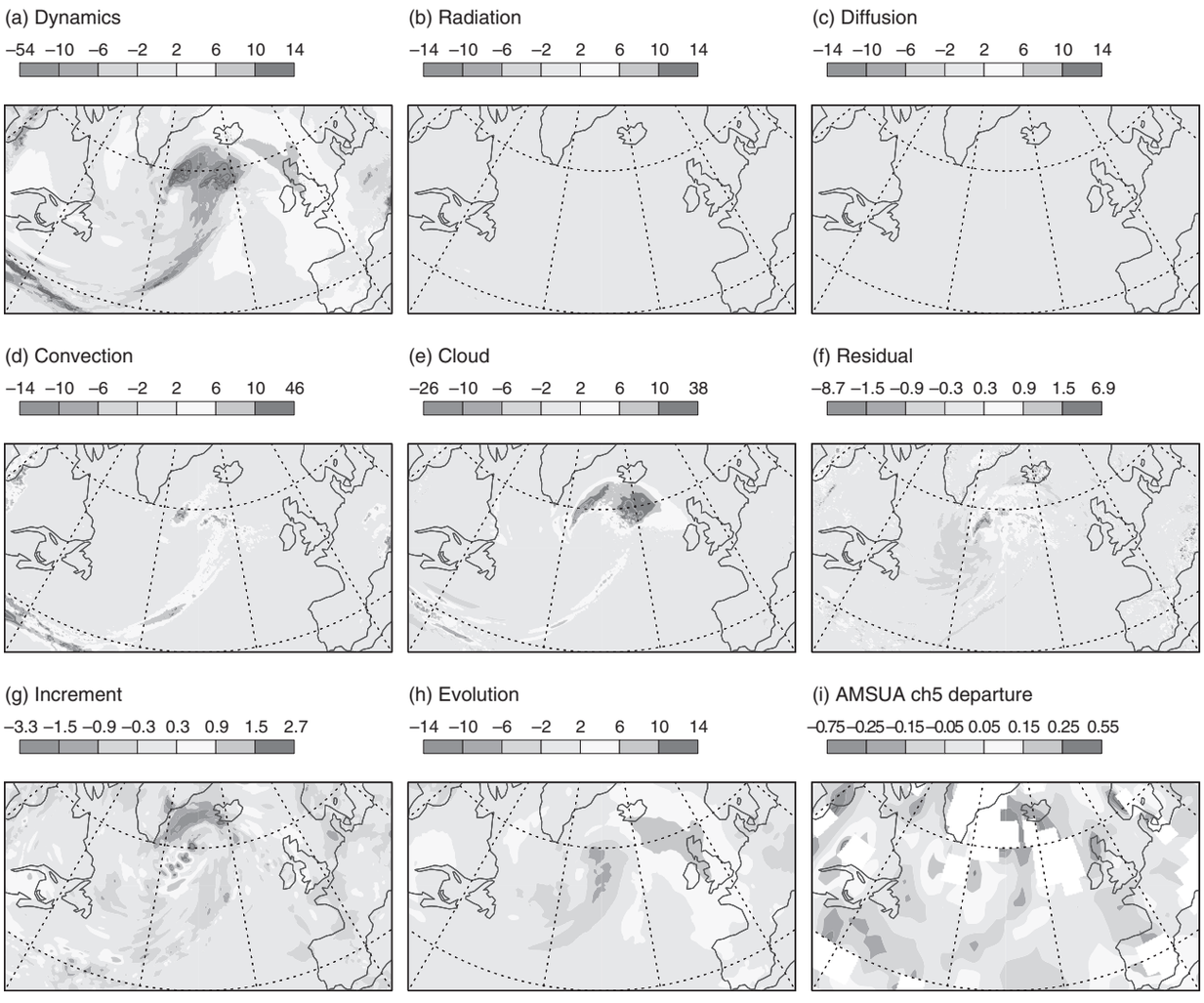

Fig. 1.7 For temperature at 500 hPa, (a) to (f) show initial tendencies from the various processes represented in the IFS model (and a residual term associated with other processes and numerics), each accumulated over the first 12 hours of the HRES forecast initialized from the 6UTC analysis on 21 July 2012. (g) Analysis increment at 18UTC. (h) Analysed evolution of the flow between 6 and 18 UTC. (i) Departure of AMSUA channel 5 microwave radiance observations (which reflect mid-tropospheric temperatures) relative to the same radiance field obtained from the forecast. (Please see the main text for further explanation).

(RMSE) of HRES 500 hPa geopotential heights in the Northern Hemisphere for selected years, highlights the integrated effects of these changes. Note the reduction in 12 hr errors and the (greater) reduction in forecast errors at day 10. Notice also that the shape of the curve at days 1 and 2 is flatter in 2013 (black) than before (light and dark grey).

These days, quantifying uncertainties is mainstream in operational NWP. Another component of the IFS is the 50-member ensemble (ENS; see, e.g., Leutbecher and Palmer, 2008) of perturbed, lower-resolution forecasts. Perturbations (historically singular vectors; Buizza and Palmer, 1995) are applied to the initial conditions to represent our uncertainty in the present state, and stochastic perturbations are applied to the physics tendencies to represent model uncertainty (Buizza *et al.*, 1999). The fundamental reason for the ENS is the desire to 'predict the predictability', which is dependent on the initial state, our uncertainty in this initial state, the lead-time, and what 'event' we are interested in predicting.

The lead-time aspect is highlighted in Fig. 1.8(b), which shows annual-means of (solid) the RMSE of the ENS ensemble-mean and (dashed) the ensemble standard deviation (spread). (Note that, because Fig. 1.8b quantifies anomalies from the ensemble-mean, values should typically be about a factor $\sqrt{½}$ smaller than those in Fig. 1.8a – which effectively quantifies anomalies from

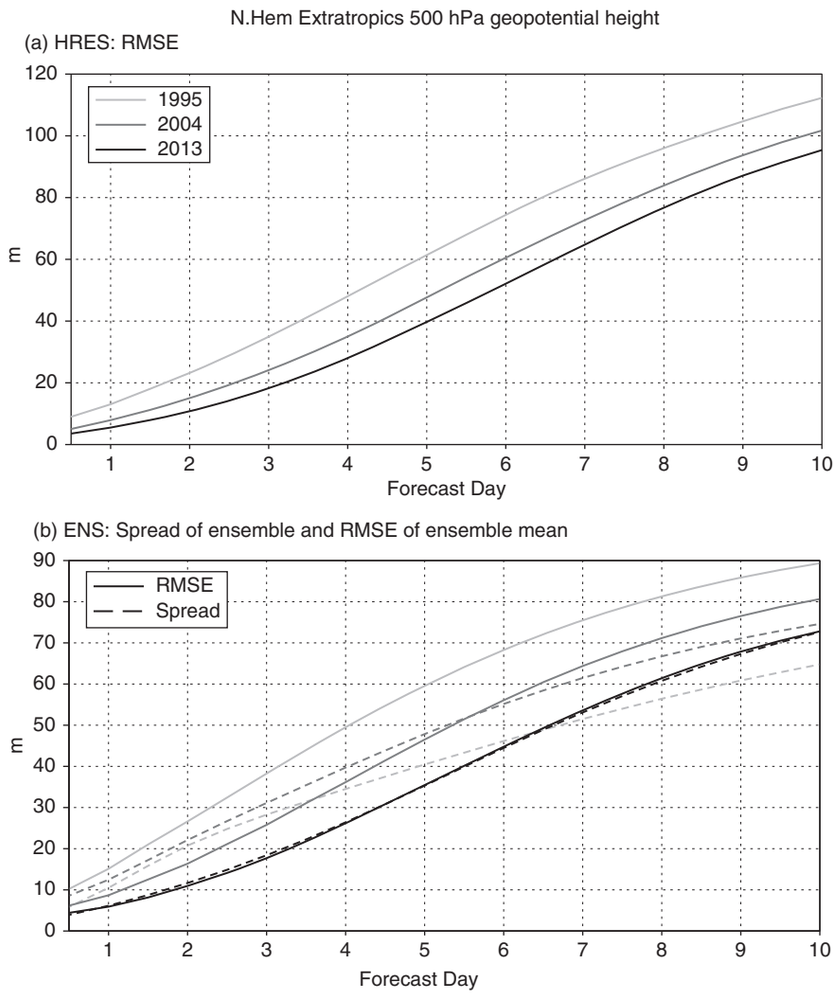

Fig. 1.8 For geopotential height at 500 hPa in the Northern Hemisphere, (a) annual mean of HRES RMSE as a function of forecast lead-time, and (b) annual means of ENS ensemble-mean RMSE (solid) and ENS spread (dashed) as a function of forecast lead-time. Both panels show results for the years 1995, 2004, and 2013.

the analysis.) For a well-balanced ensemble system, the mean spread should equal the mean error (Leutbecher and Palmer, 2008). Figure 1.8(b) shows that the ensemble was very under-spread in 1995 (light grey). In 2004 (dark grey), initial condition uncertainty was unduly large, partly to ensure reasonable spread at longer lead-times. In 2013 (black), the better balance of initial uncertainty and model uncertainty leads to a very good agreement between mean spread and mean error at all lead-times. This agreement ensures that the spread decreases with decreasing lead-time, and hence that the probability of an event occurring should ideally converge to either 0% or 100% as the lead-time decreases.

For example, we might define an event to be extreme precipitation (greater than 50 mm in 24 hours). Figure 1.9 shows the ENS day 1 probability for such an event to occur on 21 July 2012. Very high probabilities (≈90%) occur over Beijing at this time and, indeed, extreme precipitation did occur.

It is clearly of interest to know how the longer-range forecasts predicted this extreme precipitation in Beijing. Figure 1.10 shows cumulative distributions of precipitation that can help explain the performance of the ENS over a range of lead-times. The observed climate curve (grey dotted) is the fraction of July days during the period 1980–2009 for which the observed 24 hr precipitation was more than the given threshold on the x-axis. It starts at the climatological frequency of a wet day (57% of days in Beijing in July have precipitation >0 mm) and 'asymptotes' to the maximum recorded value (1% of days in Beijing in July have precipitation >63 mm). This curve represents the distribution of a baseline probability forecast using only climatological information. The model climate curve (grey dashed) is derived from 100 ensemble re-forecasts, with five members each, initiated from analyses spanning the previous 20 years and within two calendar-weeks of the real-time forecast. The (minor) differences between the observed and model climate distributions

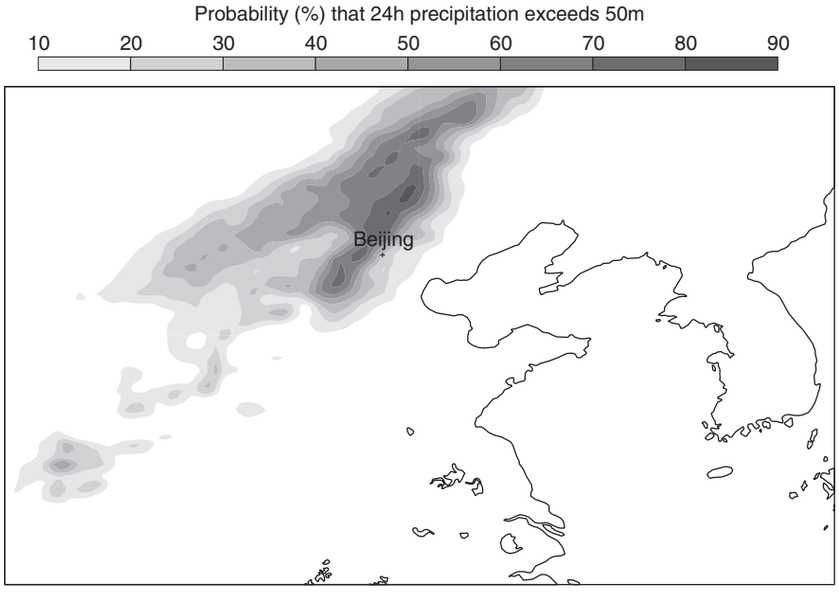

Fig. 1.9 Probability that precipitation would exceed 50 mm on 21 July 2012, based on ENS accumulations over lead-times 0–24 h ('day 1 precipitation').

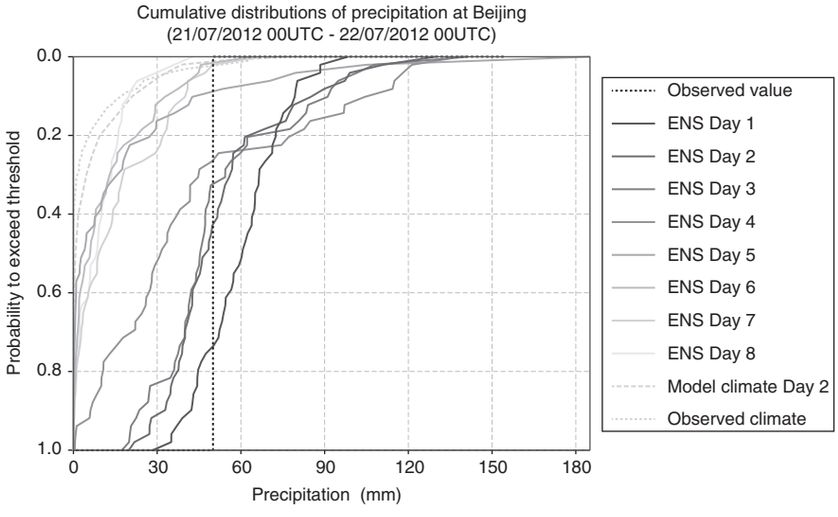

Fig. 1.10 Cumulative distributions of precipitation at Beijing based on the observed climate (from Synop observations), the model climate at Day 2, the operational ENS at lead-times 1–8 days, and the observed outcome. Please see the main text for more explanation of these curves.

reflect forecast system deficiencies (for diagnostic purposes, the forecasts are not calibrated here), along with discrepancies between the periods used to derive these climatologies. The cumulative distribution for the observed value (black dotted) is a step-function centred at the observed outcome. This curve represents the distribution of a perfect probability forecast (the vertical line at the observed outcome reflects a delta-function in the probability density function). Each ENS curve (solid) shows, for the indicated lead-time, the fraction of 50 real-time ensemble members that predict more than a given threshold on the x-axis (the darker the shade of grey, the shorter the lead-time). The main point to emphasize is the gradual progression from curves at long lead-times (e.g. day 8) that are close to the observed climatology towards curves at short lead-times (days 3–1) that are close to the step-function of the observed value. Such convergence is not only desirable from a meteorological modelling point of view, but also from a user perspective. A smooth ramping up (or down) of probabilities as the possible extreme event approaches is much more desirable for emergency planning than contradictory information from one day to the next.

Having mean-spread equal to mean-error is clearly a desirable property for an ensemble prediction system, but this is not sufficient for it to be a perfect ensemble