Met O (APR) Turbulence and Diffusion Note No. 264

Dealing with uncertainty in convection

by

S.H.Derbyshire

14th February 2000

Met O (APR)
(Atmospheric Processes Research)
Meteorological Office
London Road
Bracknell
Berks, RG12 2SZ

Note

Met O (APR) TDN-264

Dealing with uncertainty in convection

S.H.Derbyshire

Dealing with uncertainty in convection

S.H.Derbyshire

14^{th} February 2000

Abstract

Predictions of precipitating convection are limited by several forms of uncertainty and variability. Because convection can have severe impacts, these uncertainties need to be quantified, but reliable quantification of convective probabilities is a challenging task. A first essential step is to make explicit which uncertainties are being allowed for, and which are excluded from our 'ensemble'.

Numerical modelling of 'ensemble convection' is increasingly discussed, but with various possible meanings. The ensemble terminology and its relation to forecast requirements are reviewed. Clarity of terminology is needed to ensure that probabilistic techniques are more than an *ad hoc* means of hedging our bets, and will stand critical scrutiny.

Current plans to represent convection more effectively as an ensemble-average in the UM highlight the different requirements for convection diagnosis in forecasting. It may not be possible for one scheme to be optimal in both prognostic and diagnostic roles. The recommended operational strategy is to redevelop the UM convection scheme to be a more fully 'ensemble' scheme, along the lines of current proposals in MP section, and at the same time systematically promote use of the Convection Diagnosis Project output in the forecasting community, together with appropriate use of large-scale model ensembles (e.g. from ECMWF) at medium range. Further research is also needed to quantify convective variability associated with a number of causes.

Probability or 'event-based' thinking (as is natural to forecasters) is in principle entirely compatible with objective statistical verification, but may be viewed as motivated by the nonlinear impacts of meteorological variables, which are perhaps not fully captured by conventional skill scores.

Contents

1 Introduction

Precipitating convection is notoriously difficult to predict reliably, yet erroneous forecasts are very visible and unwelcome to the public and the customer. We naturally turn therefore to probabilistic or ensemble terminology.

The aims of this note are: to review probability and ensemble forecasting of precipitating convection, in the light of current convection research and possible convection-scheme developments; to clarify the terminology; to list the main technical questions and indicate how they can be addressed; and to make recommendations for strategy in this area.

Technically a probability is a measure, defined over an *ensemble* of possible atmospheric states. Under certain assumptions the probability may be identified with 'long-run frequency'. Examples of the application of probability to convection include:

- Increasing use in NMC of probabilities based on large-scale model ensembles (Mylne 1999)
- A desire for the UM convection scheme to behave in a more fully ensemble-averaged manner (even though it is already intended to be an ensemble scheme - Gregory and Rowntree 1990)
- An increasing trend for convection research based on Cloud-Resolving Models to pursue convection as an ensemble problem (e.g. Shutts and Gray 1999), which offers some hope of quantifying the ensemble properties.

But how do these different uses of 'ensemble-thinking' in convection fit together (or are they essentially separate)? If the UM convection scheme became more fully an ensemble-averaged scheme, what would this imply for forecasting? Over which ensemble exactly should it be averaged? How adequately do large-scale model (e.g. ECMWF) ensembles cover the uncertainties associated with convection? How should Cloud-Resolving Models or other research techniques be used to support both ensemble-averaged and other (possibly non-deterministic) forecast requirements?

Feynman *et al.* (1963) describe the theory of probability as ' a system for making better guesses ... We make guesses when we wish to make a judgment but have incomplete information or uncertain knowledge', and define the probability of a particular outcome of an observation as 'our estimate for the most likely fraction of a number of repeated observations that will yield that particular outcome'.

But what does 'repeated observation' of the 'same problem' mean in the atmosphere, as opposed to a controlled laboratory environment? The atmosphere never exactly repeats itself, but we can hope to categorize the information available in 'real time' and collate, estimate or predict the conditional frequencies based on that categorization.

We might for instance speak of the probability of precipitating convection over Southern England in a specific synoptic situation (say, a cyclonic polar maritime airstream) in daytime in June. Usually though it is not pure climatology we want, but rather a probability *conditional on other information*, i.e. using some knowledge of model output and its error-characteristics. What then is the probability of convection given the above meteorological situation *and* that some specified automated forecast-product is predicting convection? Or at all, in Southern England during the day in question? The conditionality on other information poses of communication, since a forecaster may have very few words with which to put across the uncertainties.

Examples of typical problems found by forecasters with current automated forecasting products are described in the Annex. There appears to be 'room for improvement' – but are the errors inevitable, owing to the 'intrinsic randomness' of convection, rather than reflecting correctable systematic flaws in our techniques?

Some uncertainty, especially about the timing or location of convection, can be reflected qualitatively in the wording of a forecast. But the development and general improvement of numerical products, including automated site-specific forecast products which are intended to be more precise and less 'broad-brush' than regional or national forecasts, has increased the need to quantify the uncertainties associated with convection. If we do try to quantify convective probabilities then they should be based on a clear methodology to avoid confusion about what is meant.

This risk of confusion is not just an academic concern. The ambiguity of quoting say 'Lightning Risk 3', without saying how much allowance is being made for likely types of model error, could have safety implications. Internal reviews of forecasting practice (B.Callander, personal communication) have emphasized the importance of consistent definition and application of probability forecast techniques.

In this note we shall try to summarize the probability and ensemble issues in convection, and how (in time) they can be systematically addressed. First of all we briefly review the specific relevance of probability to convection; the pros and cons of probability-forecasting; and the verification of probability-forecasting and its connection with customer-value.

1.1 Specific relevance of probability to convection

The need for probability-thinking in convection arises from (i) the instability and short predictability horizon of convective events and (ii) their potential severity.

The atmospheric instability of which showers and thunderstorms are manifestations leads to a rapid divergence in the possible 'trajectories' of atmospheric evolution. Beyond the lifetime of an individual convective cell (say half an hour) any initial uncertainties are greatly amplified. So any attempt to predict the movement or evolution of an individual convective cell may be termed 'nowcasting'. At longer forecast periods, the prediction of location, time and strength of convection obviously fits Feynman *et al.*'s description of 'incomplete information or uncertain knowledge'.

The severity of convective weather can take various forms (heavy rain, hail or snow; lightning; gusts, squalls or even tornadoes), and (as with any severe weather) a forecaster (or automated forecasting tool) may be expected to indicate even a small risk.

In numerical modelling too, the question of probabilistic or ensemble representation of convection covers some important but difficult issues.

In general an ensemble representation could be:

(i) fully probabilistic, i.e. quoting probabilities of a given outcome

(ii) stochastic[1] (random or Monte-Carlo) – cf. Mason & Thomson (1987)

(iii) ensemble-averaged

Note that these three approaches, although all probabilistic or statistical, are actually very different, and pose questions about what we really want from a model convection scheme. Do we want the scheme to produce a speckled pattern of showers statistically similar to a radar snapshot? That would be the stochastic option. Do we want a relatively bland convection field that is substantially averaged over the convecting region (the ensemble-average option, which may be good for model performance generally)? Or do we embrace the fully probabilistic option (which probably needs to be diagnostic rather than prognostically integrated)?

We shall return to these questions in §2.

1.2 Pros and cons of probability forecasting

The Met.Office has historically been cautious in using probability in public weather forecasting, as compared say with practice in the USA, where probability weather forecasting has been so familiar for decades that Feynman *et al.* (1963) use weather forecasting to introduce their discussion of the whole concept of probability.

Differences in practice between UK and USA partly reflect objective differences in the meteorology of these regions (e.g. the more common occurrence of severe convection in North America), but perhaps also reflect historical or cultural influences that can be

[1]from a Greek root meaning 'to guess'

questioned. Let us therefore try to set out objectively the pros and cons of probability forecasting for convection.

The key advantage of probability forecasting (properly used) lies in conveying both what we know (e.g. that conditions are favourable for thunderstorm development) and what we don't know (e.g. precisely where or when a thunderstorm will occur). This advantage is particular significant in relation to short-lived events of great subjective[2] importance.

The main potential disadvantages quantitive probability forecasting in public broadcasts are

(i) the risk of *confusing the public* with too much numerical information

(ii) the risk of *misleading precision* in probability forecasts; e.g. talk of '33% chance of a shower' is surely over-precise, whilst (more seriously) any forecaster who quotes high (say, 99%) confidence levels had better make quite sure they are justified

(iii) the *temptation to hedge one's bets* excessively [we should remember that a reliable deterministic forecast conveys more information than a medium-probability forecast; the latter may convey little more than climatology] leading to ...

(iv) ... some public *suspicion about verification* of probability forecasts [3] - can forecasters be held to account for a probability forecast?

Murphy *et al.* (1980) suggest that with care and some education these perceived problems can be largely overcome.

In relation to point (i), we may note that the qualitative use of probability is not new in UK public forecasting, and indeed some understanding of likelihood is essential to any effective use of convective weather forecasts. We must surely expect that the more specialized or 'intelligent' customers (including the public-service weather forecaster as an internal customer for model products) will increasingly require formal risk assessments.

Similarly point (ii), essentially an issue of the integrity or 'quality' of probability forecasting, can be addressed by developing a clear technical methodology that is aligned with customer needs and meets the spirit of corporate quality standards.

The opposite risk of 'excessive hedging' (iii) can be addressed by ensuring that skill scores give an incentive towards definite (or high-probability) predictions where appropriate. The issue of verification (iv) is complex and will now be discussed.

[2] i.e. to the user, which is what counts (cf. also the use of probability weather-forecasting for the solar eclipse of August 1999

[3] cf. a recent tabloid cartoon, captioned: "they're playing it safe again - cold with occasional showers and some bright sunny spells!"

1.3 Probability, verification and nonlinearity

To define probability forecasting satisfactorily, we need to define our target, i.e. state objectively what constitutes a good forecast. Such a target should penalize over-confidence or excessive hedging but reward use of a forecaster's 'best judgment', taking into account all available information (Murphy 1993).

The approach taken here to probability-verification (essentially following Murphy) is that the aim of probability forecasting is to maximize customer[4] 'utility' in an uncertain situations, especially when there are *many different risk-thresholds* for action. Multiple thresholds may arise either because there are many customers, or because one customer has many actions. This view of probability-forecasting in terms of utility leads relatively straightforwardly to a methodology for verifying customer forecasts.

'Utility' is a concept used in economics to quantify the acceptability of an outcome to individuals or other decision-makers. To each possible outcome one may notionally assign a (net) *utility* U, taking into account financial loss, loss of life or limb, or loss of enjoyment *according to the customer's own subjective evaluation.* Utility is not the same as financial gain or loss, but may be estimated by evaluating the risks which an economic agent is prepared to accept. Even in the case of loss of life, one may investigate the level of risk which individuals consider acceptable, and thus estimate the value[5] effectively assigned to life.

Consider now a customer (or range of customers) using a weather forecast as a basis for certain actions. For a given meteorological event, with hindsight, we may evaluate the net utility of any actions taken to protect against that event. Responses to a convection forecast might include: carrying an umbrella; cancelling a sporting or other event; avoiding certain lightning-sensitive activities; diverting a river or taking action to protect a dam; advising or instructing the general public to avoid travel or to take shelter from a severe storm.

Let us define m to be the actual meteorology at verification time, m_0 to be the meteorology (including observations and NWP products) available at forecast time and the operator $\langle\rangle$ to denote the 'expectation' (i.e. ensemble-averaged) value conditional on knowing m_0. Here 'knowing m_0' does not imply that the NWP forecast is correct, just that the forecaster has received the product concerned.

First note that if (to a sufficient approximation) the utility function U_a (for a given

[4]Here 'customer' denotes the person or group whose requirements are to be met (whether or not there is any financial transation)

[5]Those uncomfortable about 'value' might prefer to regard 'utility' as a shorthand for discussing the risks people are prepared to accept, and how meteorologists can best help them to manage those risks

action a) is a *linear* function of the meteorology, then

$$\langle U_a(m) \rangle = U_a(\langle m \rangle)$$

and hence we only need to know the mean $\langle m \rangle$! In this situation probability-forecasting is not necessary for the customer's decision-making.

Let us illustrate this in sporting terms. Some sports are more weather-sensitive than others. E.g. professional football *is* affected by convective precipitation but in a relatively gradual way, and usually without sharp thresholds for action. It is plausible that the utility of any actions would be essentially linear in m - if so, the customers might be well-served by being told the 'expectation' meteorology $\langle m \rangle$.

In contrast the running of a cricket or tennis event liable to interruption by any significant rainfall will be a more nonlinear function of the meteorology, and there is a more obvious requirement for forecasting the *probability* of such interruptions.

Next consider a hydrological customer concerned about flooding if rainfall exceeds a certain threshold. This scenario (cf. Fig.1) provides another example of a nonlinear utility function, in which the probability of exceeding the threshold is clearly the forecast objective.

If we now imagine a *wide range of customers* with different thresholds then evidently we need to predict the probability of each threshold, leading ultimately to a requirement to predict the convective PDF. More generally, but leading to the same conclusion, we may say that for given known m_0 we need to be able to forecast the expectation of arbitrary utility functions U.

1.3.1 Verification and utility

So probability forecasting is effectively forecasting the *utility* of a range of courses of action. Some customers might explicitly ask the meteorologist to make the decision[6] for them ('should we cancel the event?'). Other customers might prefer to retain decision-making responsibility, with input from the probability-forecast, but the objective task of probability forecasting is similar in each case. Roughly speaking, low probabilities p correspond to customers with a high sensitivity (of order p^{-1}) to the event concerned.

If then we are effectively forecasting the utility $U(m)$, for some known function U, we can apply all the normal verification techniques to $U(m)$ rather than m.

We can for instance evaluate whether the forecasts are unbiased, i.e. the conditional

[6] although the meteorologist may reasonably disclaim any formal responsibility

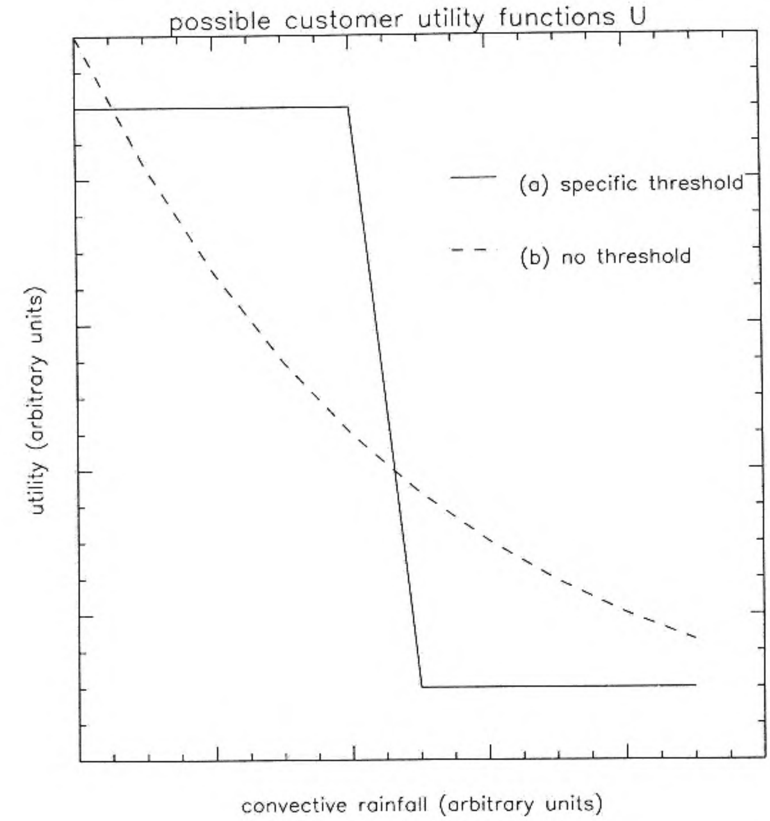

Figure 1: Examples of possible utility functions U for impacts of convective rainfall. The utility function of customer (a) implies the need for a forecast of the probability of exceeding his specific threshold. In contrast, customer (b) has a more general aversion to rainfall and would probably be well served by a conventional forecast of 'average' (or 'expected') precipitation.

ensemble-average $\langle U(m) \rangle$ matching forecasts in long-run trials, and similarly we can evaluate the r.m.s. errors.

In particular if U is a *threshold* function, with outcomes bad or good according as the outcome falls into category i or not, then we may use the indicator function χ_i (1 if m is in category i, 0 otherwise) as the basis of verification.

The mean-square error $\overline{[(\chi_i)_f - (\chi_i)_o]^2}$ is then the Brier score (using the notation $\overline{()}$ to denote a climatological average). The 'reliability' $R(m_0)$ (for this category of m_0) is $\langle [(\chi_i)_f - (\chi_i)_o]^2 \rangle$. The overall reliability $\overline{R}$ is a measure of how effectively the forecast is unbiased for each m_0, weighted according to the climatology of m_0.

1.3.2 Meteorological classification

The known meteorology m_0 (i.e. the information available in advance to the forecaster or automated system) may be divided into 'classes', to provide some direct verification of frequency conditional on m_0. (Each class need to be broad enough to contain a statistically significant sample.) For instance we might attempt to evaluate the frequency of convection given some value (or band) of a specified convective instability index, or more basic meteorological categories, such as 'Northerly flow in June at Bracknell at 12 Z'.

We may now apply 'analysis of variance' (or covariance) techniques to understand how well we are doing. These techniques are based on the simple rule that for sums of uncorrelated random variables the variances (or covariances) are additive. We can therefore analyze the variances (or covariances) as sums of contributions within and between certain classes of our choice.

The *resolution* of a classification (in the sense of Murphy 1993) corresponds to the fraction of the overall variance explained by the variance between classes. The 'uncertainty' corresponds to the total observed variance.

It can then be shown that the Brier score is the sum of the 'reliability' variance (i.e. the error in predicted frequency within each class) and the 'unexplained' variance (the part the classification does not pick up). In principle, from a very large observational dataset we could develop a finely-resolved classification and enforce 'reliability' by taking the empirical values within each class. In practice of course (given the multiplicity of relevant factors) we are unlikely to make much progress without using all our scientific understanding of convection.

1.4 Relationship to other forms of verification

The view of probability-forecasting as related theoretically to nonlinear utility-functions can help us understand how its verification relates to standard verification procedures in meteorology.

It is sometimes said that forecasters are concerned with 'events' (meaning severe or otherwise high-profile[7] events), whereas conventional verification picks up the skill in 'ordinary' weather. Is the forecasters' approach in some sense *more probabilistic*?

First note that all systematic verification of any forecasting 'track record' is necessarily statistical. Whether we are forecasters or model developers we need to look for reproducible [8] results. There would be no point in retuning our models or forecast procedures to a past case unless we expected similar issues to recur. If we are trying to forecast, say, for this year's Wimbledon (an important high-profile event for UK weather forecasters) it is of limited use to know about our skill for last year's Wimbledon – instead we need to consider all relevant information about model performance in *meteorologically* comparable situations.

So though the concern about 'events' is indeed a matter of 'utility', we can gain reproducible benefit from considering utility in objective verification only where the utility is correlated with the meteorology – the events/utility argument is not an argument for focusing objective verification on a small sample of high-profile events but does give some motivation for weighting objective verification towards severe (or other important types of) weather, e.g. major storms, low cloud, or major changes.

There is also a need for *detailed case-study* of individual events. Such case-study is commonly referrred to as 'subjective verification' but in the author's opinion that phrase can be misleading. Are not case studies really a form of *research* to identify types of model errors and their impacts on forecasts rather than a 'verification' of reproducible forecast accuracy? Such case-study research complements the more statistically significant, but less detailed, picture obtained from routine verification.

(Mathematically one might express this need by saying that conventional objective verification, using mean and r.m.s. scores, is fundamentally allied to linear thinking about mechanisms, whereas case studies remind us that many model errors are highly conditional on particular circumstances and cannot be considered as linear perturbations from climatology.)

In summary

[7] 'public profile' can be viewed as one measure of utility-significance

[8] including statistically reproducible information – i.e. estimates of the likelihood of certain errors

- objective verification is about how well we are doing, and should act as a target. This should be based on large samples of data sufficient to expect reproducible conclusions (though these conclusions themselves may concern not just biases but typical random errors).
- case-based verification should aim to identify in more detail the mechanisms of model error[9], and may stimulate the development of objectively verifiable diagnostics.
- probabilistic verification following Murphy's methodology is both objective and non-linear (in the sense of reflecting the potential nonlinearity of weather impacts). Use of Brier scores for instance can be viewed as a special case of a nonlinear utility-verification (in this case a simple threshold step-function), but one could imagine other nonlinear functions. In general nonlinear verification (with usually a bias towards more extreme events) is likely to increase the requirements on observational data for useful results (both the quantity for reproducibility and also the quality of 'weather' data).

In general the current objective verification of our skill at convective precipitation seems rather limited and perhaps there is scope for improvement (we are supposed to be weather-forecasting not pressure-forecasting!). However improvements in convective verification will require attention to the statistical issues.

In work linked to the Convection Diagnosis Project, about which more later, Will Hand (personal communication; see Metnet, fp0100/~fpwh/ CDP/verification/main.html) has developed a verification system based on a neural-network classification from Meteosat, combined with Nimrod rainfall analyses. This work is also being used to develop a convection climatology. This Metnet references also includes further discussion of Brier scores and other technical aspects of verifying probability forecasts.

2 Convective variability and ensemble representations

Convective clouds vary in their depth, structure, duration, updraught and gust strengths, precipitation and in their microphysical and electrical properties. They may occur in a field of 'air-mass convection' (perhaps with some degree of organization) or as isolated large events.

[9] e.g. if model precipitation came too early, was the synoptic evolution or frontal movement too fast, or did the convection scheme 'jump the gun'?

As well as some intrinsic variability (no two clouds being identical), we may expect them to be influenced by variability in the surface and in the upper flow. Surface variations may create thermals (as well known to glider and hang-glider pilots), influencing the location and perhaps other properties of the clouds.

Convection may also be influenced by large-scale dynamics, e.g. by an upper trough, which can (i) directly affect the profile stability to convection and (ii) initiate vertical motion (increasing relative humidity) through the 'potential vorticity vacuum-cleaner' effect discussed by Hoskins *et al.* (1985).

Depending on the extent to which we can predict them in a given context, these influences may be treated as deterministic or probabilistic. For instance one might feel that a mesoscale model should be able to resolve mesoscale troughs deterministically, and make less allowance for unforecast features. Such a decision should of course be based on evidence of what the model can actually predict, rather than just prior expectations.

2.1 Ensemble representations

An *ensemble* representation of atmospheric flow (as the term is used in turbulence theory, or in the ECMWF 'ensemble' prediction system, Palmer *et al.* 1997) allows for *many possibilities* or 'realizations'. This ensemble of possibilities may or may not be sampled effectively by a given spatial or time-average. E.g. the spatial distribution of air-mass convection in a cold-air outbreak over the sea, with many cells within a flow that is almost statistically homogeneous, may well approximate the ensemble of realizations, whereas an isolated summer thunderstorm over land should be regarded as only one realization of the ensemble.

In practice ensemble thinking can be used in three ways, as:

- PDF prediction – explicitly quoting probabilities
- Stochastic (Monte Carlo) methods, where any single prediction is highly unreliable but the distribution (or at least mean) of many predictions may be correct
- Ensemble average – explicitly representing the ensemble mean (conditional on certain information)

For instance the Unified Model convection scheme (Gregory and Rowntree 1990) is described as an 'ensemble' (meaning ensemble-averaged) convection scheme. However, for a combination of numerical and physical reasons, it tends in practice to operate in a partially stochastic way. It certain does not on its own generate any information about PDFs.

2.2 Manual techniques and probabilities

Established manual forecasting techniques contain elements of probabilistic thinking. E.g. the Forecasters' Reference Book (Meteorological Office 1993) summarizes traditional tephigram-based assessment of convection (Fig.2), and hints at a probabilistic interpretation. It notes the variability in the convective cloud-top heights (which are relevant e.g. to lightning risk), and suggests that

- most tops are around the 'slice'[10] level (the level at which the profile ceases to be locally conditionally unstable)
- occasional large clouds reach the 'parcel' level (the level of neutral buoyancy of a moist-adiabatic parcel)
- exceptionally in favourable conditions the clouds may penetrate to the 'overshooting' level

It seems a reasonable simplification to identify two key steps in the formation of a large (precipitating) cumulus or cumulonimbus cloud:

(i) the formation of convective cloud when a parcel reaches its lifting condensation level (LCL) – or perhaps more critically the level of free convection (LFC) where it becomes buoyant

(ii) the growth from shallow cumulus to deep convection

These two steps are governed by somewhat separate processes and criteria

(i) convective inhibition in and just above the boundary layer (is the boundary layer warm and moist enough for parcels to reach LFC)

(ii) the moist instability of the free troposphere (CAPE)[11]

The significance of both (i) and (ii) for forecasting can be shown by simple cloud observations. Observations of altocumulus (e.g. contrails turning into altocumulus castellanus) show significant instability aloft, whose release is inhibited by the difficulty of forming cloud. The case of such instability waiting to be triggered should be reflected by a probability-forecast, e.g. as a small chance of intense convection. On other occasions, observations of fair-weather cumulus show that the step (ii) can be the limiting step, and

[10] strictly J.Bjerknes' slice method includes a compensating subsidence term but this can usually be neglected because the mass compensation is over a wide area

[11] Here we define CAPE from the LFC upwards as the 'positive area', regarding any non-buoyant sections as part of the convection inhibition

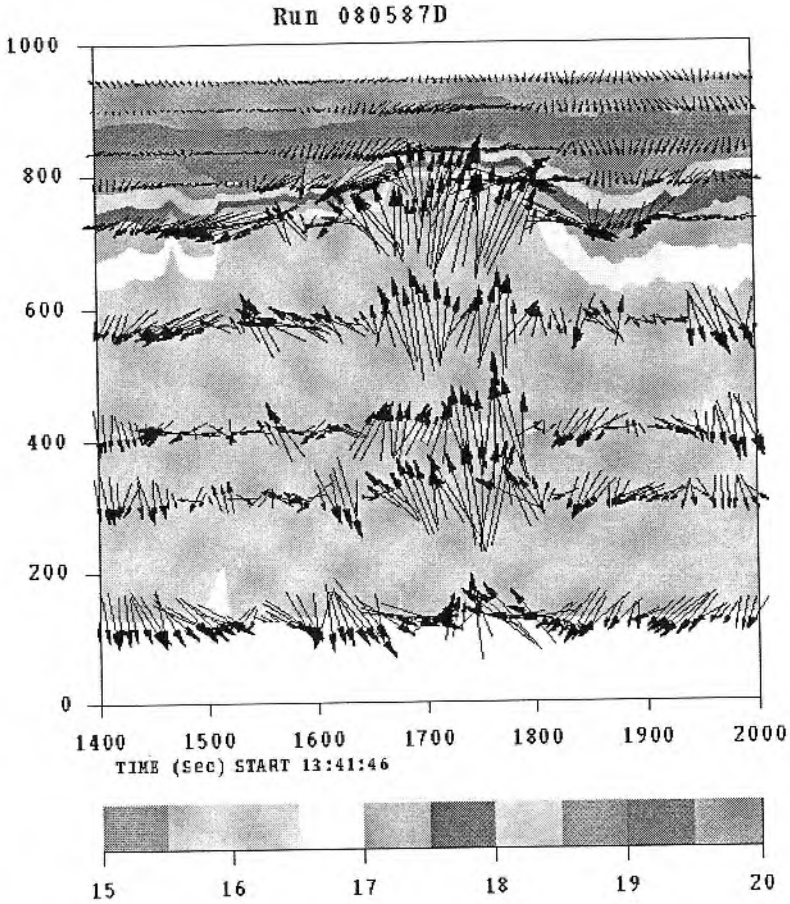

Figure 2: Example of boundary layer convection, measured using the Cardington tethered-balloon facility. The height-scale on this time-height diagram is in metres. Arrows show the wind-vector in the plane of the vertical and the mean wind-vector. Shaded contours are potential contours (marked in Celsius). Note the pronounced thermal (updraught up to about 2ms^{-1}) pushing up the capping inversion but failing to penetrate to heights where it would form cloud. Courtesy B.M.Claxton, A.J.Lapworth

in some cases the presence of a strong capping inversion may enable us to rule out deep convection.

In an ensemble-averaged convection scheme one wishes to know the total amount of mass extracted by convection from the boundary layer and pumped into the free atmosphere (i.e. the mass flux). This mass flux is currently diagnosed either from the stability of low-levels or from the CAPE of the profile as a whole. Considerable uncertainty remains: probably the area fraction of convection is governed more by (i) and the local intensity by (ii), but this remains to be clearly established.

Several process questions remain only partially answered. How deep and how long does a near-surface temperature anomaly need to be to be relevant to criterion (i)? And

Tool	Rationale and probabilistic nature
CPS	broad-brush convective impact on large-scale model *ensemble-averaged*
EPS	allowance for *synoptic* uncertainty *stochastic ensemble used to generate probabilities*
CDP	allowance for *local variability* not captured in large-scale model *probability-based*
VSHRP	short-range prediction of advection and development *quasi-deterministic*

Table 1: Generic rationale and probabilistic nature of different types of convection tools

to what extent should the moist adiabatic stability criterion (ii) be modified by mixing with the environment, which leads to a less buoyant updraught that does not rise so high (especially in a dry environment)? How should criteria (i) and (ii) be combined?

2.3 Automated tools: CPS, EPS, CDP and VSRFP

The basic requirement for operational convection software in a weather-forecasting organization similar to the Met.Office may be summarized as:

(i) a Convective Prognostic-Integration scheme (CPS) to update the *large-scale fields* of temperature, humidity (and perhaps wind) to prevent these becoming unrealistic

(ii) a Convection Diagnosis Package (CDP) to diagnose the location, intensity (and perhaps probability) of convection from the large scale model (at forecast periods > 6 hours, say)

(iii) a Very Short Range Forecasting (VSRFP) Package which assimilates observations of convective cells, and advects and develops them (for up to 6 hours, say)

Their rationale and probabilistic nature is summarized in Table 1.

These modules may not be completely independent as software, and may use certain common elements.

The CPS needs to operate (potentially) on every gridpoint globally at every timestep, and will therefore be bound by storage and run-time constraints, leading to limitations on its performance in detail. Thus although direct output from the CPS is used to a considerable extent by forecasters, there may be fundamental limitations to the performance of a CPS in its response, say, to mesoscale topography.

Some of the forecasters' typical problems with convection (as in Annex A) may be beyond the scope of a CPS. (This is always a debatable point – one cannot usually prove the impossibility of improving the CPS – but some limitations are suspected to be fundamental.)

It is therefore desirable from a general perspective to develop a separate, or partially separate, CDP that is more sophisticated and can be applied to a limited area and run only when diagnostic output is required, i.e. not usually every time-step. There is however a consistency issue between CDP and CPS. If these two schemes were to differ systematically in, say precipitation over the UK, then this difference would need to be interpreted as an error in one scheme or other, and (if significant) removed by modification.

In turn a CDP may not deal with individual convective events, as they are observed by radar, satellites or on the ground. A CDP may or may not have a cell model, and may or may not be able to advect those cells. Such limitations motivate the development of a separate, or partially separate, nowcasting package.

The current convective software in the Office may be summarized as follows. The CPS is an entraining-plume mass-flux scheme (including downdraughts and convective momentum transport) which does not handle or advect individual cells. The very short range forecasting system Nimrod (Golding 1998) is mainly a blend of advection or extrapolation of the nowcast with the NWP model. GANDOLF (Hand 1996) is an object-oriented code that includes models of cells or advection. A CDP (Hand 1999) now incorporates elements from the CPS and GANDOLF. The probability of convection, its likely depth and precipitation are diagnosed essentially from the CPS with allowance for additional surface variability.

Both between CPS and CDP and between CDP and VSRFP there are difficult questions of unification. Some consistency is desirable where possible, not only to minimize software development, but also because inconsistencies might feed into systematic errors. At the very least our scientific understanding of the relevant aspects of convection should be unified, so that the simplifications made for specific operational purposes are conscious and well-judged.

2.4 Mass flux convection schemes

A mass flux is a measure of the total amount of ascent in the updraughts, and may be written as the area fraction of the updraughts multiplied by the mean velocity in the updraughts. (Of course the updraughts will be compensated by subsidence elsewhere, possibly very remotely).

If the term 'mass flux' appears an esoteric piece of convection jargon, it may be helpful to note that (in deep convection) the mass flux is fairly closely related to the total precipitation.

It should be noted that current mass-flux convection schemes (including the Gregory-Rowntree scheme used in the Unified Model) do not normally quantify the actual vertical velocities or updraught area, i.e. they do not really distinguish between widespread moderate convection and isolated strong convection.

One of the key parameters in current mass-flux schemes is the rate of *entrainment* of environmental air into the updraught. The Gregory-Rowntree scheme assumes that the updraught is diluted by the environment at a rate $\epsilon \sim 30\%$ per km of ascent (varying somewhat with height).

If, say, a moist plume ascending from the boundary layer is just saturated with a specific humidity of 5 g/kg, implying latent heat content equivalent to 12.5K in temperature, and the environmental relative humidity is 50%, then the above value of ϵ contributes (through mixing and evaporation) a cooling of the plume by 1.3K per km. Such effects are very significant for convection.

Another term commonly used in mass-flux convection schemes is the 'closure'. This represents the amount of mass-flux (and parcel properties) at the bottom of the convecting profile. In the Unified Model two convective closures are available: a CAPE closure (in which, roughly speaking, the vertically integrated mass-flux profile is proportional to the adiabatic CAPE), and a 'buoyancy' closure, where the mass-flux at the level where convection is initiated is effectively linked to the instability at that level.

Neither of the current closures is based on detailed process-research, and so the closure is regarded as one of the main targets for convection-scheme redevelopment.

2.5 Ensembles in Cloud-Resolving Models

Cloud Resolving Models are a research tool that is increasingly bearing fruit and supporting the improvement of NWP convection schemes. They are helping to develop a quantitative view of convection as a turbulent flow, with certain statistical dependence on 'external' factors (e.g. stability, moisture).

In CRMs we normally want *many clouds* which *forget their initial conditions* (arbitrary imposed buoyant parcels), as opposed to 'ballistic simulations' whose trajectories are highly dependent on initial conditions.

One approach (e.g. Craig & Tompkins 1998) is to run long quasi-equilibrium simulations (for up to 3 weeks). However the CRM, as well as forgetting its arbitrary initial eddy structure, also forgets its initial mean profiles, which we therefore lose any ability

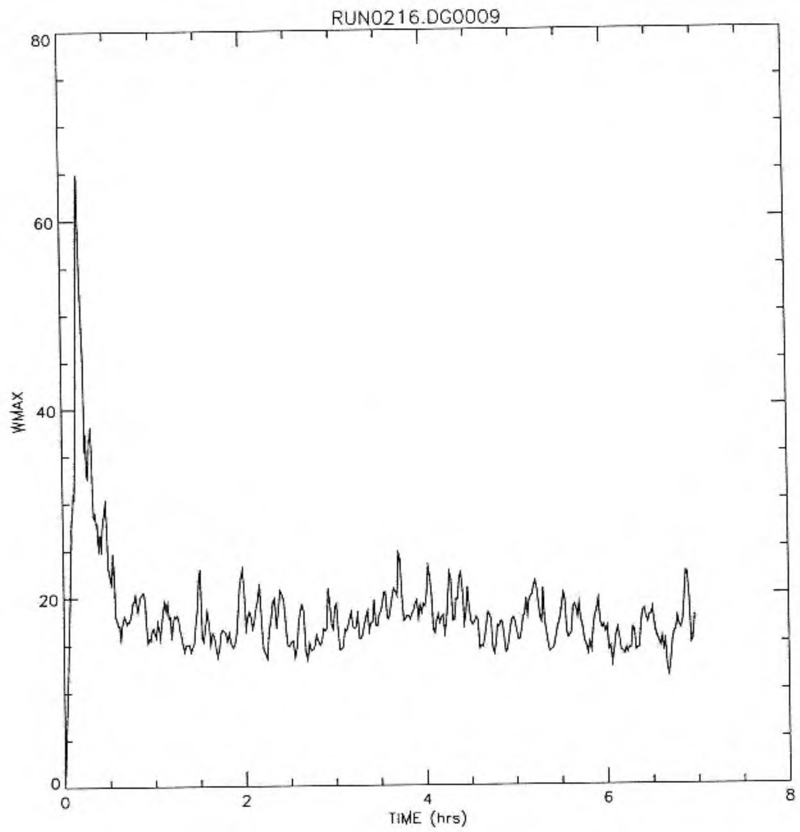

Figure 3: Illustration of ballistic and quasi-equilibrium behaviour in a Cloud-Resolving Model. Plot shows maximum vertical velocity (ms^{-1}) over a domain 100km$\times$100km in the horizontal, at 1km horizontal resolution. Note the strong transient or 'ballistic' convection in the first half-hour, effectively determined by the initial buoyant parcel, but note also the evolution towards statistical-equilibrium convection.

to prescribe.

The plots shown in Fig.3 and Fig.4 were obtained by imposing strong relaxation of the mean profiles to prescribed θ and q profiles. (NB the fields are not relaxed *to the mean!* but rather a horizontal uniform increment is applied to bring the horizontal mean back to the target). Fig.3 illustrates some short-term ballistic behaviour evolving rapidly into a statistical quasi-equilibrium. Fig.4 shows typical cloud diagnostics from such a simulation.

It should be noted that 'ensemble' convection in CRM runs does *not* rule out some organization into open-cell or other well-known patterns (Fig. 5).

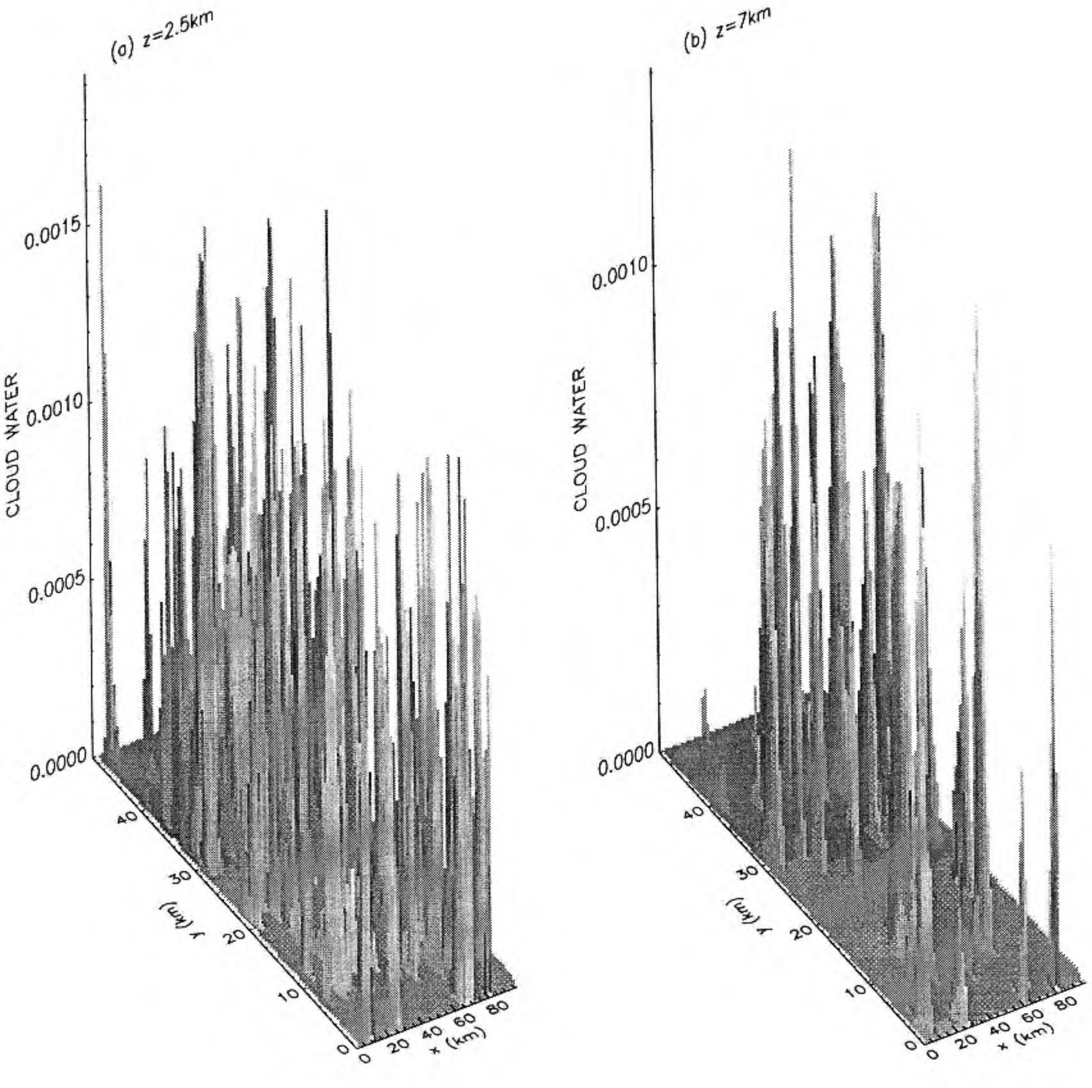

Figure 4: CRM cloud-field as shown by cloud-water content q_l at (a) 2.5km and (b) 7km (close to the 'parcel tops' level), in an idealized simulation of vigorous convection. Note that the peak q_l values are almost as large at 7km as at 2.5km, but that the peaks are notably fewer.

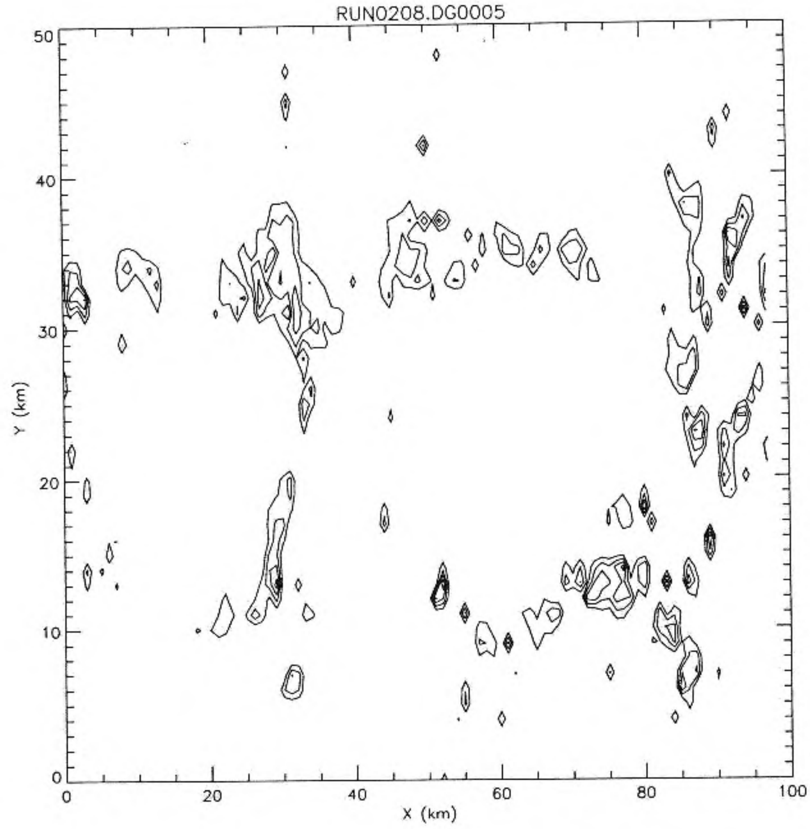

Figure 5: Examples of cloud structure generated by a Cloud-Resolving Model. Plot shows contours of liquid water q_l at height 2.5km in a constrained quasi-equilibrium simulation. Note the development of ring-shaped structures resembling classical 'open-cell' convection. The development of convective structures is considered further by Shutts & Gray (1999).

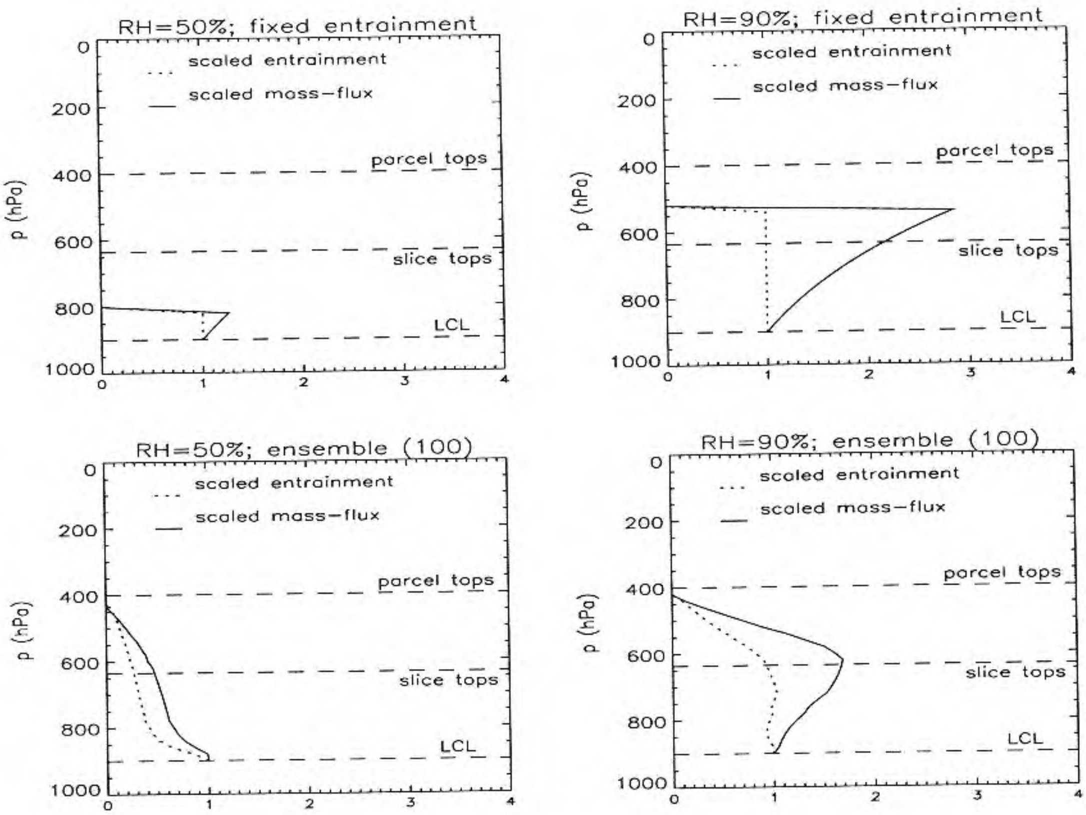

Figure 6: Illustration of the impact of mid-tropospheric relative humidity in diagnostic evaluation of idealized mass-flux schemes. These are (i) a basic entraining-plume model and (ii) an 'ensemble' scheme with a wide-range of different entrainment rates. Dashed lines show 'significant levels' (from the classical adiabatic theories) for the test profile.

3 Current problems and future convection strategy

Let us now take stock of issues with the UM convection scheme, its forthcoming redevelopment and wider implications.

3.1 The current CPS scheme and its limitations

The current UM convection scheme is based on Gregory & Rowntree (1990). It is an ensemble scheme, in the sense that it does not attempt to represent individual clouds. But though it represents many clouds, it treats them (in many ways) as though they were all the same, e.g. ascending parcels of air are all diluted to the same extent by entrainment. Limited allowance for variability is made in detrainment but this does not cover the full impact of ensemble variability.

Fig. 6 illustrates in a highly simplified mass-flux scheme the impact of mid-tropospheric relative humidity. In a mass-flux scheme with a prescribed entrainment rate (e.g. the Gregory-Rowntree scheme) such dryness tends to prevent convection (unless much greater conditional instability can build up). In contrast a more fully ensemble-averaged scheme will respond even to small amounts of conditional instability. Effectively the ensemble scheme assumes that a few convective elements (possibly in isolated major storms) are able to ascend almost adiabatically.

Other significant assumptions are that each gridpoint contains some updraughts, downdraughts and significant clear air. For purposes of calculating radiation the convective cloud amount is diagnosed in a manner that does not exceed about 30% on any gridpoint. By definition therefore such a scheme will never match a detailed satellite image which can resolve image pixels as essentially cloudy or clear.

3.1.1 Numerical considerations

In general the interactions of convection parametrization with model numerics are notorious. Problems can occur if convection is either too slow or too fast. Switching off or reducing subgrid convection can lead to damaging and even fatal gridpoint storms in NWP models (cf. Derbyshire *et al.* 1999). On the other hand, high rates of subgrid convection can also cause numerical problems. The vertical structure of convective increments is also important.

Some of the numerical problems appear to originate from the existence of sharp thresholds in most convection schemes, and it is plausible that that a 'true ensemble convection scheme' should be smoother and numerically better-behaved. Let us now consider that argument.

An 'ensemble' convection scheme, in the sense usually intended in NWP, implies an averaging which is at least roughly equivalent to some spatial and/or temporal filtering. The rationale for such filtering may be summarized as

- numerical: advection schemes are inaccurate for grid-scale modes and time-stepping is inaccurate for inter- (or intra-) timestep fluctuations. So some filtering over gridlengths and timesteps might be viewed as a design consideration for physics schemes.
- dynamical: convection schemes should (arguably) be designed suppress unpredictable self-organization (even if on relatively large scales)
- initialization: it may be inappropriate to allow the NWP model to develop structure finer than the data used for initialization

- physical: the assumptions of convection schemes are not valid when a model starts to resolve (or partially resolve) individual clouds (at say 10km resolution)

Apart from spatial or temporal averaging, an ensemble (-averaged) scheme can be expected to blur various thresholds (e.g. for convective initiation), according to the relevant uncertainties and variabilities, and hence give smoother numerical behaviour.

There are however subtleties to the concept of filtering. Suppose for instance (as an example of filtering) that convective increments were calculated on a deliberately coarsened grid (either by averaging or as selected points - or indeed at selected timesteps) and then interpolated. (Similar procedures are indeed sometimes used with radiation calculations.) Such a convection calculation would then fail to 'see' certain finescale components of the NWP field, and might be unable to remove instability. Ultimately the model might 'blow up' on fine scales.

There is therefore an important difference between (i) *filtering convective increments* and (ii) *using convection to filter the large-scale field.* Filtering increments essentially acts to slow down its response to the finest spatial and temporal scales, which is in contrast to the grid-point storm considerations discussed by Derbyshire *et al.* (1999).

There may be a specific physical reason for some of the problems at high-resolution (A.Grant, pers.comm.), namely that current convection schemes do not represent horizontal transports at all. In fact some representation of horizontal frictional transports is in principle justified, especially in a non-hydrostatic model. Nevertheless it is clear that we also need to consider convective-dynamical interactions from a numerical-analysis perspective.

In summary, both on numerical and physical grounds the CPS should not be taken literally at the limits of model resolution. The uncertainty associated with this 'non-literalness' may be viewed as contributing to generating the relevant conditional ensemble.

As noted above, use of the term 'ensemble convection' begs the questions 'which ensemble?' and 'how treated?'. But if a convection scheme is to be considered ensemble-averaged in any sense then it should be capable of running steadily under steady forcing, a requirement that the current scheme does not generally meet.

3.1.2 Clouds amount, radiation and impact

Convective cloud amount is of course observable but as with precipitation we need to decide what we are trying to verify, i.e. point values or averages of some kind.

From our discussion of ensemble principles it should now be clear that if a convection scheme is expected to describe ensemble-averaged properties then fundamentally it cannot describe the variability in convective cloud amount on scales less than, say, 100 km. If we want greater detail then it will almost certainly need to be via a probabilistic diagnostic package (cf. our above discussion of the CDP).

Other scientific limitations of the present convective cloud amount diagnosis also make it unwise to put much weight on these variables for forecast purposes.

But are there aspects of convective cloud detail that cannot be treated diagnostically?

It has recently been shown (S.Cusack, personal communication) that more accurate treatment of radiation can have significant benefits to mesoscale model skill-scores, confirming that radiation is not just a climate issue. Do the limitations of convective radiation treatment then contribute significantly to uncertainty in convective precipitation?

During the day, over land, the strongest radiative feedback is the impact on surface heating. In clear daytime conditions over land coming solar radiation of order 1000Wm^{-2} will be redistributed as sensible and latent heat, mainly over a boundary layer of order 1000m depth. If the ground is very dry (so the heat-flux is sensible rather than latent), this implies a boundary layer heating of 1Wm^{-3}, or approximately 1K/1000s (=3.6K/hour).

Of course if there is convection then the incoming radiation will be reduced, and heat will be vigorously transported. Such estimates suggest that this basic convective-radiative feedback is a significant factor in the diurnal cycle over land, and could play some role in the performance of a CPS. More subtle questions about cloud-radiative interaction (e.g. the possible maintenance of convection by cloud-top radiative cooling) are current research topics. Although in general feedbacks between radiation and convection do require prognostic modelling rather than offline diagnosis, and from the above estimates these feedbacks are not negligible, it is not clear that they are currently among the leading-order problems in convection prediction.

3.2 Plans in Model Parametrization for convection scheme re-development

As demands rise for accuracy in all parts of the UM, the semi-stochastic behaviour of the current CPS seem increasingly unacceptable in numerical terms. Such behaviour almost certainly introduces 'noise' and otherwise degrades various aspects of model performance. The recent development of an operational Site-Specific Forecast Model has further highlighted the problems of semi-stochastic behaviour (requiring the convection scheme even to be switched off). (Note that in some operational diagnostic products the convective precipitation is now being time-averaged over 1 hour in partial recognition of this problem.)

The problems and plans of the Model Parametrization group are documented under Roy Kershaw's home page on Metnet (see www-hc/~hadyb/con.html) but are briefly summarized here.

Many of the problems identified are 'numerical' but have potentially serious impacts on the forecast. The current coupling between deep convection and the boundary layer is poor; the Gregory-Rowntree scheme was designed to overlap with the boundary layer scheme, but it is now widely thought that a convection scheme should be interfaced to the boundary layer at cloud base.

In the current scheme convection tends to be intermittent; to interact undesirably with the boundary layer scheme; to show various systematic problems in the tropics with significant impact on global circulations; and to show poor diurnal variation overland.

These numerical problems seem to reflect weaknesses in the physical basis of the current scheme, and therefore are unlikely to be overcome just by 'better numerics'. For instance it is probably inevitable that a basic entraining-plume model with fixed entrainment must fluctuate in the convection depth, if it is to balance a given tropospheric cooling. In effect such a scheme 'tries' to generate a stochastic ensemble, whereas a fuller incorporation of the ensemble properties into the scheme itself would be numerically and physically preferable.

The immediate development objectives are to produce a convection scheme which is 'less sensitive to vertical resolution, less intermittent, and which interfaces to the boundary layer scheme in a realistic way'. This is seen as an essential first step in producing a more controllable scheme which within the present terminology will be more 'ensemble-averaged' and capable of subsequent improvement. This development will require scientific support from sources such as CRM work.

3.3 How much convection, and how strong?

As we seek to improve our forecasting of convection, we will need to be clear whether 'moderate precipitation' from a mass-flux scheme (such as the Unified Model CPS) really represents a likelihood of moderate precipitation, or say a possibility of intense precipitation (with very different practical implications).

One simple hypothesis suggested by our discussion of manual techniques is that

(i) the probability of forming convective cloud is controlled by boundary layer stability and triggering

(ii) the intensity of convective updraughts and storms reflects the conditional instability of profile above BL (e.g. $w \sim \mathrm{CAPE}^{1/2}$ or the scheme described in Swann 1999)

Essentially this hypothesis assumes three stages of convection

I		II		III
BL convection	→	shallow Cu	→	deep Cu, Cb

and that the transitions I→II and II→III are in certain senses independent. In reality this may not be quite correct, and there may be a need for closer coupling between the treatment of the boundary layer and deep convection. For instance the secondary triggering of convection by downdraughts and gust fronts from previous storms suggests that the probability of forming convective cloud is not controlled entirely from 'bottom up'.

Nevertheless it is widely agreed both for microphysical reasons and for forecasting the possibility of intense convection that we need to go beyond simple mass fluxes to quantify the updraught velocities (e.g. Swann 1999).

Grant & Brown (1999) developed a similarity approach to shallow convection where a parameter m_b/w^* (cloud-base mass flux over convective velocity scale) can be interpreted as a measure of the fractional area occupied by active updraughts. The application of this approach to deep convection is currently under investigation.

The simple convective-ensemble model illustrated in Fig.6 would imply that a dry mid-troposphere might permit occasional large storms, whereas a moist mid-troposphere would allow more frequent but less intense convection. Further work is needed to determine whether such an idea can be used reliably.

3.4 Quantifying uncertainties and variabilities

An appropriate general strategy for development of convective probability-forecasting would seem to be the following.

Identify (at a research level) all relevant forms of convective variability. Estimate their effects (again at research level), and decide which are operationally significant. Where appropriate, seek to implement operationally as PDF, ensemble-averaged or stochastic techniques.

An initial inventory of convective variability might be as follows:

- intrinsic variability in convective cloud-fields
  Cloud-Resolving Models should be able to help quantify this. Some of the uncertainty reflects the entrainment of dry environmental air (and hence depends on humidity). For instance a dry profile might favour an isolated big storm as opposed to 'popcorn' convection. Some fundamental work on convection ensemble statistics and variability

is being conducted at Reading University under G.Craig.
Research aim: identify and quantify main influences on variability.

- variability induced by surface temperature and humidity
  Impacts of such variability are already estimated in CDP.
  Research aim: more detailed understanding and quantification of these impacts. E.g. given that boundary-layer horizontal variations are strongly height-dependent, what is the effective height at which they are relevant to deep convection?

- orographically triggered
  Orographic triggering is still rather poorly quantified.
  Research aim: improve quantification using combined understanding of flow over orography and convection

- diurnal cycle
  The performance of the current models in the diurnal cycle of convection is still a subject of active research. To identify any weaknesses in the convection scheme, rather than e.g. problems in the surface or boundary layer schemes, requires detailed observations and/or modelling. The diurnal cycle of convection will be influenced by factors including the boundary-layer temperature, the response of convection (or convection schemes) to instantaneous temperatures and the lifecycle of convective cloud systems.
  Research aim: clarify the respective roles of these factors and the scope for improving the diurnal cycle. At present the diurnal cycle has to be regarded as contributing to uncertainty in convection forecasts.

- upper flow synoptic variability
  This may be largely covered by work linked to the ECMWF EPS.
  Research aim: identify convective contribution to forcing large-scale variability

- upper flow mesoscale
  Some uncertainties in the mesoscale will not be captured, owing to resolution, lack of data, or unpredictability of the mesoscale flow. From information available through data-assimilation it should be possible to estimate climatological errors in θ and q, and perhaps attempt to predict these from to synoptic-scale gradients or otherwise.
  Research aim: quantify mesoscale variability and develop techniques to use this information in predicting convective uncertainty

- potential instability
  Many textbooks define an atmospheric profile as 'potentially unstable'[12] whenever

[12] or even 'convectively unstable', a potentially misleading term in this context

θ_e (or equivalently θ_w[13]) decreases with height. Note that θ_e and θ_w depend strongly on the actual moisture content q (as opposed to θ_{es} where the actual is overridden by the saturation value). This definition is motivated by the argument that a moist layer capped by a dry layer aloft can become destabilized by large-scale ascent. This is true but seems at most a theory of embedded convection. In any case with modern NWP models the possible large-scale ascent should be quantifiable rather than purely speculative.

An alternative argument (with some support from TOGA-COARE observations, Godfrey *et al.* 1999) is that dry intrusions essentially inhibit convection but that (i) inhibition of weak convection (cf. the ensemble mass-flux scheme of §3.1.1.) might in certain circumstances allow instability to build up to permit severe storms (ii) dry upper intrusions can be correlated with cyclonic PV anomalies. See Browning and Golding (1995) for a UK case-study.

Research aim: investigate whether potential instability is really a useful concept in this context.

- Mesoscale Convective Systems

  Mesoscale Convective Systems are large long-lived agglomerations of thunderstorms which are a dominant feature in tropical convection, and also sometimes affect the UK, where they are important owing to their severity (Gray 1999; Marshall and Gray 1998; McCallum and Waters 1993). Recent informal discussions between APR and NMC aviation forecasters (P.Davies) have confirmed the need to improve their representation. Because MCSs are relatively long-lived, even a 12-hour forecast in the tropics might benefit some form of VSRFP technique.

  Research aim: use recent research results to guide potential development of operational techniques

- spatial sampling, especially in slow-moving convection

  It is well recognized that the risk of local flooding is particularly acute with slow moving convection, and forecasters are able to take some account of this. To quantify these effects properly we need to know the length-scales (both the size and spacing of peak-precipitation regions), speed of movement and lifetimes.

  Research aim: support this task using Cloud-Resolving Models and observations

- UK-type tornadoes

  The UK does not normally suffer the major tornadoes common in the continental USA. However smaller tornadoes (e.g. Pike 1998 and references) are well documented

[13] θ_e and θ_w are alternative ways of labelling a moist adiabat, by the θ value at the top or at the bottom. For unsaturated parcels the relevant moist adiabat is found by Normand's construction.

and can cause structural damage. Whilst the priority of this issue is less in the UK than in the USA, an improved predictive understanding is desirable.

Research aim: identify conditions for isolated severe convection and relation to other tools for convection

Where the science is sufficiently clear, many of these factors could potentially be included in a CDP. For instance, research evidence now casts doubt on the classical dilute-plume approach that the entrainment rate ϵ is a quasi-universal parameter. The future development of the UM convection scheme is likely to include (explicitly or implicitly) a recognition of this variability in ϵ. In the CDP therefore, there might also be some merit in explicitly allowing for the intrinsic variability in entrainment rate ϵ, and particularly for some probability of a substantially lower ϵ, which would allow convection to penetrate much nearer to the adiabatic level.

3.5 Towards a comprehensive ensemble strategy

Weather forecasting is a matter of communication as well as science, and inevitably users will effectively 'recalibrate' their interpretations of a given forecast product over time. However it is highly desirable to have clear, fixed and agreed definitions. At present there is a noticeable tendency for divergent interpretations of what the current UM convection scheme is supposed to represent, especially at higher resolution.

Further progress in convection to improve the UM performance seems to require a *limitation* of the CPS 'job description'. This means accepting the CPS is not primarily a tool for diagnosing convection in the senses required for forecasting, but rather a means of computing the impact on the large-scale atmospheric state (and hence should be optimized to that purpose).

Hence then the continuing development and operational status of the Convection Diagnosis Project and other such tools will be also an essential plank in the whole strategy.

CDP and CPS do not of course operate independently, since the CPS will affect the mean atmospheric state (stability and humidity) which will then be used by the CDP. It is important therefore that both CPS and CDP should respond correctly to that state.

There might be a case for some form of combination of the CDP with the large-scale EPS, if there was found to be a significant time-range in which the errors from intrinsic (or surface-linked) convective variability were comparable with those from large-scale evolution.

4 Conclusions and recommendations

This document is intended to promote an informed debate about our terminology, practice and strategy in the probabilistic or ensemble representation of convection.

Convective cloud-fields and their impacts will presumably always defy fully deterministic prediction, so we can expect that statistical issues in convection are here to stay. In particular there is a need to estimate probabilities of severe impacts even when those probabilities are small.

Historically the unpredictability of convection (and indeed large-scale evolution) has often been covered by vague phrasing in forecasts. But as models improve, with greater detail and greater reliability in many respects, there is an increasing need to quantify convection probabilities more objectively and systematically.

In practice all such probabilities are conditional and to speak of '50% probability' without saying what is assumed is (strictly) meaningless. We need for instance to be careful about probabilities derived from synoptic ensembles (e.g. the ECMWF system). Such ensembles (now deservedly popular) may well cover the *synoptic* uncertainties, which at longer range may indeed be the dominant uncertainties. But (especially at shorter range) they may not fully reflect the true uncertainty in our prediction of convection or other severe weather.

To make best use of probability and ensemble techniques, we need to be clear about the detailed meanings of these statistical concepts. In practice all probabilities are conditional so the 'conditions' must always be defined (otherwise talk of probabilities is meaningless). Similarly, if we speak of an 'ensemble technique' we need to ask (i) which ensemble? (i.e. what do we consider known?) and (ii) how treated? (via quoted probabilities, stochastically, or ensemble-averaged?).

We may summarize our discussion by saying

- Accurate deterministic prediction is best, but not always possible
- Systematic errors should be corrected as far as possible (either in a model or diagnostically, e.g. by a forecaster), but we may not yet know how to do this
- We should strive to estimate and allow for the remaining uncertainties, whether or not we regard these as ultimately removable

Research tools such as Cloud-Resolving Models are increasing our ability to quantify the properties of convection, including its variability. Such tools should be able to help improve probability predictions, if we can properly define the questions we are asking.

Already CRMs (Swann 1999) have shed light on the issue of entrainment parametrization, and shown that entrainment parameter ϵ is not a fixed quantity but adapts to the environment through a process of variability and natural selection. Indeed some of the problems of the current UM CPS can be traced to prescribing a parameter which is in fact variable. Such issues will be addressed by the current redevelopment plans.

4.1 Recommendations for strategy

1. Plan on the basis that the future of convection forecasting will involve increasingly sophisticated diagnostic techniques and quantified probabilities, based on clear definitions and supported by a clear scientific methodology which would stand independent scrutiny, e.g. in the spirit of quality standards such as ISO9000.

2. In particular, recognize that in practice all probabilities are conditional on specific assumptions, and that there must be clarity and transparency about which uncertainties are allowed for and which are not.

3. Plan to make diagnostic schemes such as CDP gradually replace direct forecast use of the UM CPS. As much as possible should be done in advance of redevelopment of UM CPS, expected to bear fruit in 2002.

4. Clarify the extent of the nowcasting requirement for deterministic or semi-deterministic convection tools - an ensemble approach is not always best. Consider how far this requirement extends e.g. to tropical Mesoscale Convective Systems.

5. Redevelop the UM CPS along the 'turbulent ensemble' lines proposed by MP section, noting the requirement to give give accurate stability profiles, otherwise for instance any diagnostic scheme (CDP) would have to be tuned to 'wrong' profiles.

6. The reformulated UM CPS should be able to disaggregate mass-flux into peak updraught (closely related to peak precipitation) and area-fraction

7. All other uncertainties should be systematically noted and estimated. Different kinds of uncertainties should be kept conceptually separate but ultimately feed into probabilities. Users should know what is allowed for. NB risk of tuning to reflect all probabilities.

8. Use the increasingly quantitative information from Cloud-Resolving Models to help guide our steps in this area, recognizing that quantifying variability is a more demanding challenge than estimating an ensemble-average

9. Try out various convection options in high-resolution New Dynamics model, with intention of finding schemes that can reliably suppress inappropriate resolved convection. Also investigate whether resolved convection can ever be satisfactory for NWP purposes at resolutions of 1-10km.

4.2 Implementation and forecaster confidence

This paper has focused on practice, terminology and strategy. But sometimes, of course, the hard part is the detailed implementation of a prediction system that wins the confidence of forecasters and others.

We have argued above that an automated diagnostic tool for generating convection probabilities is a necessary part of operational strategy. However forecasters have pointed out to the author that it is difficult for a forecaster (at least initially) to trust an automated package which does not 'show its working'. It seems also that forecasters like to interact with a model (when time permits) so that they can assess its sensitivity to assumptions.

Whilst the 'human factors' of implementation and confidence may lie outside the strict scope of this paper, nevertheless they are partly related to our conclusions about the need to spell out assumptions. Thus the final recommendation is:

10. Provide the forecasters with as much information as possible about how the diagnostic package arrives at its probabilities, including reminders of the error-types it does not currently allow for, and (if possible) real-time information about its diagnosed surface temperatures etc. Consider allowing the forecaster to change those assumptions, at least by selecting precalculated alternative scenarios.

Future progress in forecasting convection does surely depend on grasping the nettle of quantified probability and ensemble properties. We should be encouraged by the realization that in this difficult endeavour we are genuinely addressing some of the long-standing difficulties in representing convection to meet the many different requirements of our customers.

5 Annex: typical forecaster problems with convection

The following comments from forecasters at Norwich Weather centre, circa July 1997, were solicited as 'impressions' of standard convective-precipitation output in the Limited Area Model. NB that model is no longer used operationally, and these comments (which are in any case not derived from systematic statistics) should be viewed only as general indications of how model limitations in convection can affect forecasters.

(i) Model considered reasonably good at maritime showers (though decay too quickly when move inland - because of crude treatment of land-sea differences?)

(ii) Errors in diurnal cycle: model showers too early in the day and decay too soon in evening/night

(iii) Altocumulus castellanus often modelled as heavy showers

(iv) Embedded showers and storms handled rather well

(v) Initialized model ascents not verifying well with observed moisture

(vi) Upper troughs often not resolved (probably a dynamics/resolution problem)

(vii) Inability to advect convective clouds frustrating

(viii) Problem of feedbacks between conv. scheme and large-scale motion

(ix) Model good indicator of areas of instability

(x) From experience, forecasters consider cloud-top cooling overnight as significant in maintaining showers/storms - model representation?

Other questions raised later include: how literally should we interpret the balance between 'convective' and 'dynamic' precipitation, e.g. in a front?

It can be seen that many of the perceived problems relate to fundamental questions about what a convection scheme can or cannot represent, in a given type of model. (Cf. our discussion in §3.)

References

Bader, M.J., Forbes, G.S., Grant, J.R., Lilley, R.B.E. and Waters, A.J. (1995): Images in Weather Forecasting. Cambridge University Press. 495pp.

Browning, K.A. and Golding, B.W. (1995): Mesoscale aspects of a dry intrusion within a vigorous cyclone.

Derbyshire S.H., Martin,G.M. and Webster, S. (1999): New Dynamics, time-advancement and convection – an interim briefing. Report to Chief Scientist, available at *Metnet*, fr0200/~apsd/nd_interim.ps.gz.

Feynman, R.P., Leighton, R.B. and Sands, M. (1963): The Feynman Lectures on Physics (vol. 1, ch. 6). Addison-Wesley, Reading,Massachusetts.

Golding, B.W. (1998): Nimrod: A system for generating automated very short range forecasts. *Meteorol. Appl.* **5**, 1-16.

Hand, W.H. (1996): An object-oriented technique for nowcasting heavy showers and thunderstorms. *Meteorol. Appl.* **3**, 31-41.

Hand, W.H. (1999): The Convection Diagnosis Project (CDP). 4th European Conference on Applications of Meteorology, Sweden, 13-17/9/1999, pp.274-279.

Hoskins, B.J., McIntyre, M.E. and Robertson, A.W. (1985): On the use and significance of isentropic potential vorticity maps. *Quart. J. Roy. Meteorol. Soc.*, **111**, 877–946.

McCallum, E. and Waters, A.J. (1993): Severe thunderstorms over south-east England, 20/21 July 1992: Satellite and radar perspective of a mesoscale convective system.

Marshall, C. and Gray, M.E.B. (1998): Mesoscale convective systems over the UK, 1981-97. *Weather* **53**, 388-395.

Mason, P.J. and Thomson, D.J. (1987): Large Eddy Simulation of the neutral-static-stability planetary boundary layer. *Quart. J. Roy. Met. Soc.* **113** 413-443

Meteorological Office (1993): Forecasters' reference book. Meteorological Office, Bracknell.

Palmer, T.N., Barkmeijer, J., Buizza, R. and Petroliagis, P. (1997): The ECMWF Ensemble Prediction System. *Meteorol. Appl.* **4**, 301-304.

Pike, W.S. (1998): The overnight tornadoes of 7/8 January 1998 and a coastal front. *Weather* **53** 244-257.

Shutts, G.J. and Gray, M.E.B. (1994): A numerical modelling study of the geostrophic adjustment process following deep convection. *Quart. J. Roy. Meteorol. Soc.*, **120**, 1145–1178.

Swann (1999): Evaluation of the mass flux approach to parametrizing deep convection. Submitted to *QJRMS*

Acknowledgments

Thanks to all participants in the Convection Workshop of June 1998, including the Convection Group, Model Parametrization, Forecasting Products and a number of forecasters. I am grateful for subsequent discussions with Ken Mylne, Will Hand and others on ensembles and probability-forecasting, and particularly for comments by Alan Grant, Will Hand, Paul Davies and Pete Clark on this note.