Limited Area Modelling in CRM mode

Peter Clark

***Met Office** Joint Centre for Mesoscale Meteorology (JCMM)*
Meteorology Building, University of Reading
United Kingdom *

1. Introduction

This paper discusses parametrization issues within models designed for short range forecasting (NWP) with resolutions and parametrizations chosen to allow deep convective clouds to develop explicitly. In practice this currently means resolutions in the range 1-3 km. Many operational centres have recently implemented or are in the process of implementing NWP models with this resolution for short-range forecasting. A pre-requisite is a non-hydrostatic dynamical core, in contrast with the hydrostatic core at the basis of many global NWP and climate models (though not all, the Met Office's Unified Model (MetUM) being an example). As we shall discuss further below, this resolution is insufficient to resolve deep convective clouds at all well, so the term 'Cloud Resolving Model' (CRM) is a little misleading – the author prefers 'Deep Convection Permitting Model' (DCPM) or 'Cloud System Resolving Model' (CSRM), which emphasises the organisation (and, in practice, horizontal size) of the models we are discussing. For the remainder of this paper we shall use the term 'DCPM'.

Both CRM and CSRM are terms originating in the 'process study' community, and many of the models which are being adapted for NWP use have been developed and used for such studies. However, process models may not contain the full range of physical processes needed for everyday forecasting, may not be sufficiently efficient and they generally do not contain data assimilation (DA) facilities. Other models have therefore been designed at the outset to be used for NWP (though they may also be very useful for process studies).

In addition to the need for DA, the other subtle but important distinction between process and NWP models is the method of large-scale forcing. Process models are often, though not exclusively, run with periodic lateral boundary conditions and homogeneously forced with large-scale tendencies. Apart from providing a degree of idealisation, this also ensures a scale separation between forcing and model response, while allowing the model response to develop over a long period of time. High-resolution NWP models, in contrast, are run forced by lateral boundary conditions from coarser scale NWP systems. This brings with it issues due to the provision of consistent lateral boundary conditions, and the 'spin-up' of scales not present in the lateral boundary data. On the other hand, however, information usually propagates out of the domain in a relatively short period of time (hours to days, depending on domain size and synoptic conditions), limiting the impact, good or bad, of upscale transport and ensuring (hopefully) that on reasonably long timescales the model remains consistent at large scale with the driving model.

This paper will not discuss issues concerning initial and boundary conditions further beyond making a few comments about verification and predictable scales in the next section. Following that, the main topic of the paper, the parametrization of processes within such models, will be addressed. Since the processes operating are the same as within larger scale models, we shall concentrate on highlighting major differences.

To some extent, parametrization can be thought of as an aggregation issue, i.e. representing the integrated effect of a distribution of detailed processes. The integration is often over dimensions other than space, such as water particles of varying size and phase, aerosol particles with varying size and composition, radiation wavelengths, surface features such as plants etc.. In going to higher spatial resolution, it is hoped that some spatial aggregation problems become easier or even unnecessary. A good example is cloud heterogeneity which (may) become less important and, of course, deep convection which, it is hoped, will not require parametrization at all. The parametrization process, however, often makes use of cancellations or quasi-steady state assumptions which become less valid at higher resolution (and associated shorter timescales), so that, in some cases, we need to do **more** rather than **less** in higher resolution models.

2. Why bother? How do we assess high resolution forecasts?

Producing DCPM forecasts is considerably more expensive than forecasts at around 10 km resolution. Ignoring changes to vertical resolution, a 2 km forecast should be around 125 times more expensive than a 10 km forecast for the same domain and forecast length, assuming the physics costs the same per gridpoint timestep. As we shall see, some physics may be cheaper, some more expensive but these ratios generally turn out to be a reasonable guide. In practice, smaller, nested domains are used, but the cost remains high.

A very large number of studies exist in the literature which demonstrate that DCPMs can produce much more realistic simulations of many phenomena, especially large convective systems and their organisation. It is less clear what the forecast skill of such simulations is. For example, Mass *et al*, 2004, show that decreasing grid spacing from 12 km to 4 km in a set of MM5 forecasts ‘has only a limited impact on traditional objective verification scores’. As they discuss, this, to some extent, highlights problems with traditional objective verification scores. One major problem is the impact of positional errors. Even a small error in the location of a feature can lead to degradation of scores compared with a smooth model which does not resolve the positional error. While it could be argued that such a positional error may be serious if the forecast is taken literally (e.g. if the wrong town is evacuated), in practice a forecast of extreme weather in the vicinity is very useful guidance, so intelligent use of forecasts should take into account likely positional errors.

These limitations have been recognised by many and a number of novel verification techniques developed which particularly address rainfall. This is a particularly active area and the reader is referred to Beth Ebert’s web page (see references). One such method was developed by Roberts and Lean (2008) and applied by Roberts (2008) to a year Met Office of forecasts over the UK using the Met Office’s Unified Model (MetUM) at approximately 12 km resolution. This is, of course, not a DCPM but it represents the scale of model in which a DPCM may be embedded. Figure 1 shows a measure of skill of rainfall forecasts as a function of forecast time and spatial scale. In this case, the skill which is taken to represent a useful forecast is about 0.5.

A number of features are clear from this:

1) The grid scale shows a skill signal but at a very low level which is not useful.

2) Even at roughly 5 times the gridlength (50 km) only the analysis shows useful skill, and one has to look at more than 10 times the grid length to find marginally useful skill at T+24h.

3) Skill decreases with forecast time with a fairly constant rate at a given scale after an initially more rapid decrease lasting 6-12 h. We interpret this as showing the benefit of mesoscale data assimilation.

4) The most rapid initial rate of decrease in skill is around 75 km, which is close to the horizontal scale of error covariance in most of the data types used in the DA system.

Figure 2 reinforces the final point; it shows the reduction in skill as a function of scale from T+0 to T+6 compared with T+18 to T+24; in this time the magnitude of peak skill reduction drops by a half and the scale moves from around 75 km to 300 km.

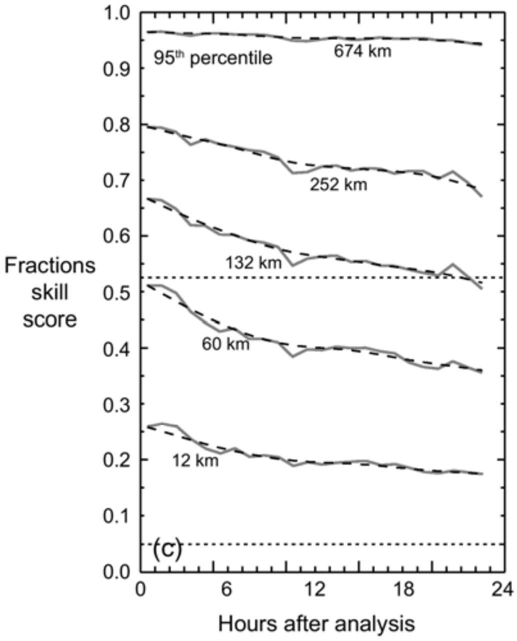

Figure 1 Fractions Skill Score for 95th percentile precipitation threshold as a function of forecast time and spatial scale for one year of Met Office 12 km forecasts (from Roberts, 2008)

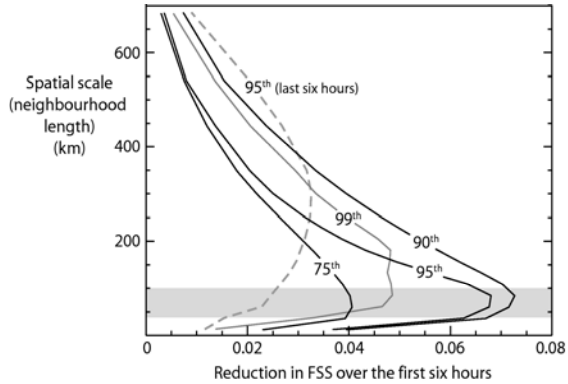

Figure 2 Graph of the decrease in FSS over the first 6 h for the 75th, 90th, 95th and 99th percentile thresholds (black). Grey shading indicates scales at which the loss of skill peaked over this range of thresholds. Dashed line shows decrease in FSS from T+18 to T+24 for the 95th percentile threshold. Results are for 1 year of Met Office 12 km forecasts (from Roberts, 2008).

Without further DA, DCPM forecasts will inherit the spatial errors from the host analysis. These will also rapidly propagate in from lateral boundary forcing. Though DA systems for DCPMs are under development, they have yet to demonstrate that they can radically improve the larger scales in the analysis. Lean et al, 2008, demonstrated that, after an initial spin-up period, a 1 km resolution forecast system could significantly improve rainfall forecasts over those from the 12 km system on spatial scales of 75 km and above, using the same (12 km) analysis in both (Figure 3). The spatial scale for improvement varies from forecast to forecast, but is controlled by the 12 km model and analysis.

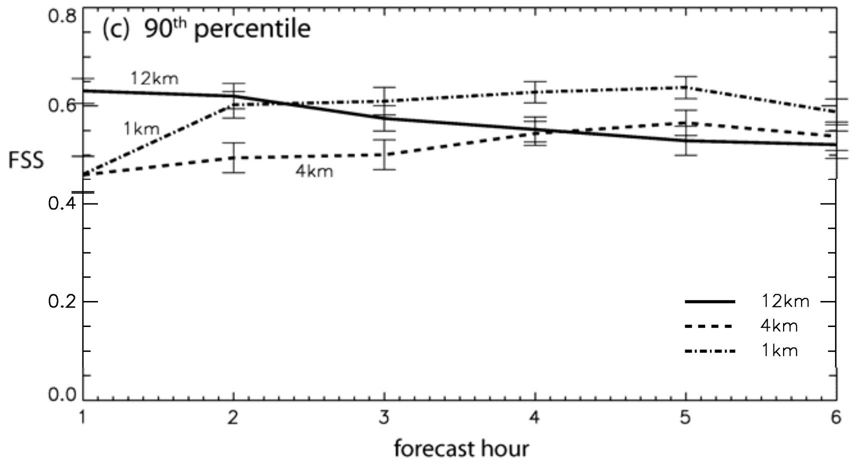

Figure 3 Fraction skill scores for 90th percentile hourly accumulations aggregated over a selection of convective rainfall cases over southern England cases from summer 2003,4 and 5. The hourly fraction skill scores are shown as a function of forecast time for a fixed sampling radius of 75km (each point is the score for the accumulation from that time to an hour later). The error bars are calculated by a bootstrapping. From Lean, et al*, 2008.*

This demonstrates that explicit representation of convection is benefiting the model at these relatively coarse scales, even though the forecast is not significantly improved at the cloud scale. We would expect therefore, some benefit at these scales in the analysis if DA is used in the DCPM. Subsequent work has shown this to be the case, so that forecasts throughout the period are better than the 12 km model. It should be noted that improvements are not so clear-cut in a 4 km DCPM, consistent with the idea that a 4 km model is not simulating the convective impacts well.

DA systems can improve the analysis at finer scales (though no system has yet demonstrated routine consistent analyses at the cloud scale). However, predictability at these scales is, in general, extremely short range (Lorenz, 1969). It is generally extremely unlikely that predictability beyond an hour or two is possible at the cell scale, and, in practice, achievable predictability is probably much less. The envelope of longer-lived organised systems is probably more predictable, but even these are unlikely to be deterministically predictable beyond a few hour (optimistically!). It is inevitable, therefore, that at longer lead times a probabilistic approach will be necessary, through ensemble systems when affordable but through a more probabilistic interpretation of 'deterministic' forecasts otherwise.

We can sum up the motivation behind DCPM as follows, i.e. why bother?:

- NOT to give accurate site-specific forecasts (e.g. there will be a thunderstorm over ECMWF at 2.45 pm tomorrow afternoon), though nowcasting and longer range probabilistic forecasts are more realistic objectives.
- To benefit from predictable small-scale forcing – downscaling (land/sea, orography, land-use). This may still be dominated by large-scale error. For example the air quality community rely on this and use models to generate high resolution wind fields to calculate atmospheric transport, often ignoring issues of predictability.
- To represent both small and large-scale effects of deep, moist convection (much) more accurately than convection schemes can achieve. Hence to develop organisation of convection more accurately. Similarly, to represent the effects of some gravity waves more accurately, either through inherent interest in themselves or through their effect on dynamics.
- If we are lucky, a combination of the above may occur, i.e. predictable small-scale forcing may dominate the predictability of convection, making some situations more predictable than usual. A good example is the Boscastle Flood in the UK which was shown to be predictable with a reasonably high level of skill 12 hours ahead, despite affecting a region of order 5 km x 5 km (Golding *et al,* 2005)

These ideas are not exclusive but they sum up the majority of applications.

3. Surface exchange

3.1. General surface exchange schemes and tiling

Most large-scale models use either a single set of aggregated surface and sub-surface characteristics or some variation on a 'tile' approach, in which a number of distinct surfaces are taken to interact independently with the atmosphere. A given surface is taken to produce vertical fluxes of momentum, heat, moisture and perhaps other species with no horizontal transport. A given surface may be represented more or less complexly, but a surface roughness and a single surface albedo and emmissivity, with long wave radiation and turbulent heat and moisture exchange associated with a single surface temperature is the norm. A variety of sub-surface models of heat and moisture transport exist, but again, these are, for the most part, 1D. For sparsely vegetated surfaces, for example, dual or multi source models may be needed, in which plants and soil surfaces can have different temperatures and associated fluxes.

The tile approach originates from so-called 'blending height' ideas (Wieringa, 1986; Mason, 1988; Claussen, 1990). An important concept is the 'land-use length scale', a measure of the characteristic size of patches of a given land-surface type. This can often be difficult to measure in practice but blending height techniques were developed to deal with the aggregated effect of patches of a similar size, larger than the typical length scale for adjustment of the surface layer of each surface to local equilibrium (typically 10-20 m for moderate changes in surface), and not so large that the internal boundary layer depth over a given surface is comparable to the actual boundary layer depth (a few km and certainly much less than 10 km). In fact, Claussen (1995) has shown that 'flux-blending' works quite well for larger areas and scales of heterogeneity, but it is still the case that 'tile' approaches are actually rather more ideally suited to resolutions of a few km (probably down to ~1 km) capturing heterogeneity on scales of 100 m or so than to coarser scale models, where heterogeneity may be expressing the difference between entire towns and surrounding forests or farmland. It must be emphasised, in passing, that the 'tile' approach does not, in any way, imply a particular spatial arrangement of the surface types, beyond a characteristic patch size, and even this is often hidden in practice as the lowest model level is used as the 'blending height', which is the main parameter determined by the land-use length scale.

Existing surface schemes may therefore, in the majority of cases, be very well suited to use in DCPMs, with the following caveats:

- There may be a case for using more tile types at higher resolution, as the blending approach is more accurate and may show up errors more sensitively.
- In principle, the sub-surface should also be tiled, but often it is not for simplicity.
- There is a case for also tiling surface or near-surface wetness, as this may have as large an impact as land-use. However, this also requires a sub-grid rainfall model.
- There may be a case for treating at least some surfaces in more detail as they may become larger tile fractions in some areas. In some areas, for example, small lakes may become an important land-use fraction and, indeed, a number of lake schemes have been or are being developed.

The most widespread surface type that can become dominant is urban areas; where at lower resolution an entire town might exist inside one gridbox, DCPMs may resolve the larger scale variability (suburban to 'downtown') in major cities quite well, and even a small village might fill a grid box. There is clearly a strong case for enhancing the treatment of urban areas.

3.2. Urban Surface Exchange Schemes

A very wide variety of urban schemes exist. However, they can broadly be classified by the following diagram.

Modified 'traditional' surface energy budget
'Rough, well drained concrete'

↓ How to specify parameters?
Difficult to get storage so nocturnal sensible heat flux right

'Street Canyon' surface energy budget
Roughness, albedo, sky view factor functions of geometric parameters

↓ Canopy too heterogeneous for roughness
Need wind within canopy layer

Explicit Urban Canopy
Distributed drag, heat and moisture exchange

3.2.1. Modified 'traditional' surface energy budget

The 'flat' surface can be modified to emulate some urban characteristics. The surface roughness can be enhanced, the albedo and IR emmissivity changed to reflect typical building materials, 'soil' properties changed and drainage enhanced to reduce moisture availability, along with changes to surface resistance to evaporation. This can be summed up as 'rough, well-drained concrete'. This can produce some urban effects, including leading-order ones such as reduced urban winds and some characteristics of the nocturnal urban heat island (UHI). However, with a single surface, using the heat diffusion equation, it is difficult to reproduce the heat storage and radiative effects of a complex 3D cityscape. Furthermore, it is very difficult to know how to specify the effective surface parameters in terms of measureable properties of cities.

The standard MetUM urban tile within the surface exchange scheme (MOSES II, Essery, *et al.* 2001, Essery, et al, 2003) basically conforms to this model. However, in an attempt to improve the heat storage characteristics (and so the phase of the diurnal cycle of sensible heat flux), Best, 2005, introduced a 'thermal canopy' in the surface energy budget. In essence, the traditional surface energy budget is split; the SEB drives the change of temperature of a homogeneous block, intended to represent the thermal inertia of building materials.

The system does a reasonable job of predicting urban heat islands, substantially better than the simple 'rough concrete' approach (Best, et al, 2006). However, extensive testing against surface data from various cities has revealed limitations. The first is the specification of the roughness length for heat.

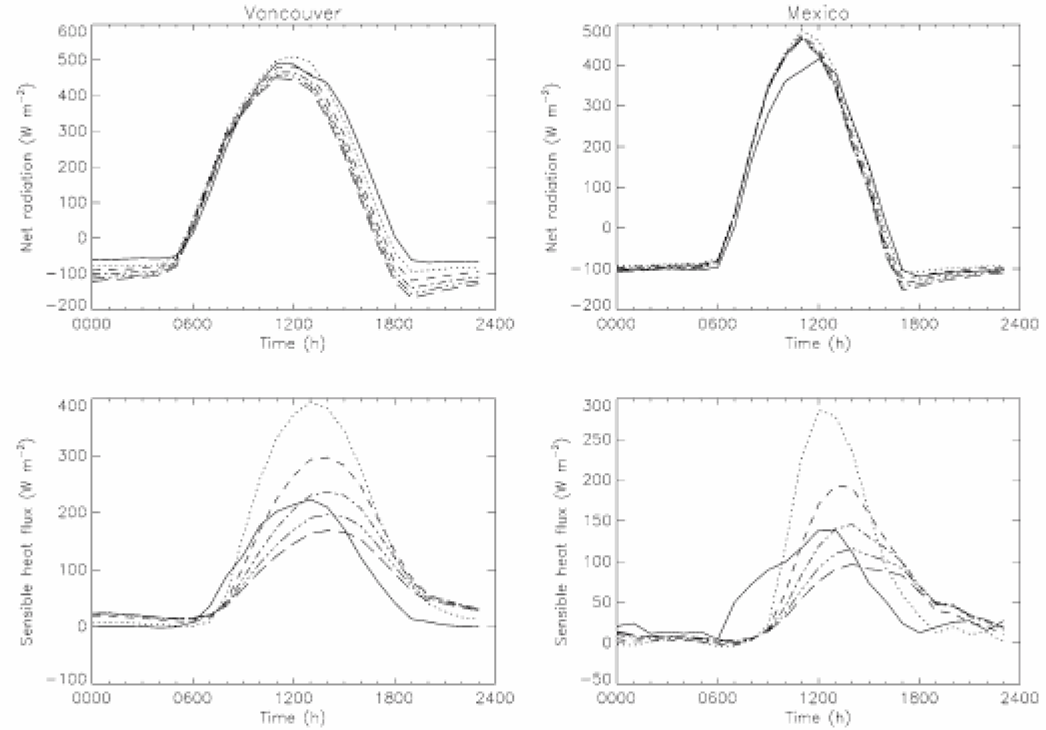

Figure 4 Sensitivity test to changes in roughness length for heat (m). Solid line: observed flux, dotted line: $z_{0t}/z_{0m} = 10^{-1}$, short dashed line: z0t/z0m $=10^{-3}$, dashed dot line: $z_{0t}/z_{0m} = 10^{-5}$, dashed treble dot line: $z_{0t}/z_{0m} = 10^{-7}$, long dashed line $z_{0t}/z_{0m} = 10^{-9}$.

Figure 4 (Fig. 3 from Best *et al* 2006) shows the sensitivity to changes in z_{0t}/z_{0m}. This is usually set to 0.1 over vegetated terrain. The smaller value in urban areas arises because of the large contribution of bluff body pressure forces to the momentum flux; this contributes to the amplitude of turbulence but not to the surface scalar flux.

It is evident from Figure 4 (and other results in Best *et al* 2006) that, while the model can be adjusted to give sensible magnitudes of the terms in the SEB, it is difficult to produce good phase behaviour, especially for the sensible heat flux around dusk and (to a lesser extent) dawn.

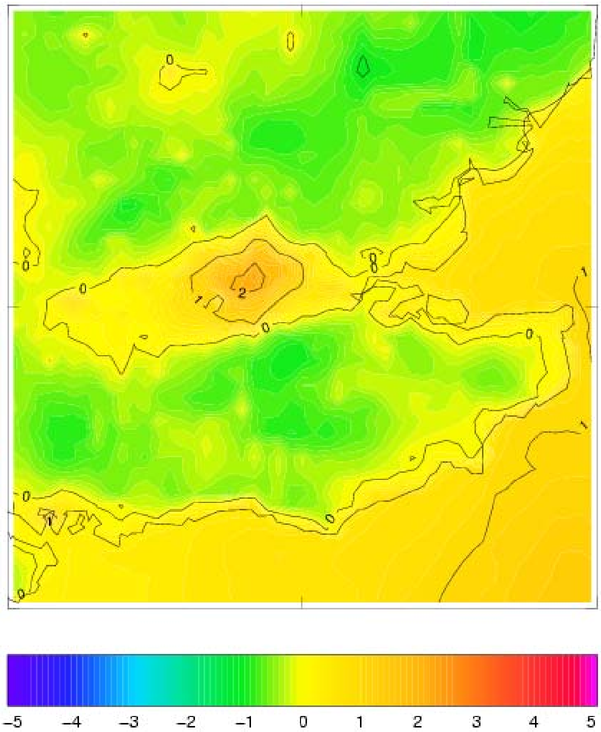

Figure 5 Mean 'screen' temperature anomaly from area average at 0000 UTC averaged from 1st June 2006 to 15th August 2006 from the Met Office operational 4 km forecast model.

Figure 5 shows the UHI over London at 00 UTC from the operational 4 km MetUM during the summer 2006 heat wave. A small estimated anthropogenic heat flux is included in this model, but it has been shown that much of the UHI arises from the surface properties. An average anomaly over 2 C is evident, and even a kink in the 1 C anomaly over Richmond Park to the SW of London.

3.2.2. 'Street Canyon' surface energy budget

The main deficiencies of the above approach have been addressed by adopting a simple geometrical model of the morphology of urban areas. A typical paradigm is the 'street canyon', a two dimensional idealisation of parallel buildings separated by streets. The system can be thought of as made up of four facets, roof, two walls and a canyon floor (or road). This is close to a typical suburban terraced housing pattern in the UK, though a flat roof is assumed. Making this idealisation allows a number of properties to be explicitly calculated. Figure 6 shows a schematic of such a scheme. An exemplar for this kind of scheme is that of Masson, 2000, which differs in detail from this schematic but which follows the basic design ideas.

Radiative exchange is calculable with few approximations, and the two major features can be quite accurately represented, namely the restricted 'view' of the atmosphere (the so-called 'sky-view factor) available to the canyon surfaces (and related surface to surface exchange of LW radiation) and the shading of walls (shown schematically in Figure 6 by the shadow cast by wall 2). The overall effect can be averaged over street directions in an area. If known, the distribution of real street directions can be used in calculating the average incoming direct short wave radiation, or alternatively a random orientation approach may be adopted.

Heat storage may be treated separately for each facet, which is represented by a single surface temperature for local radiation and turbulent exchange. The sub-surface may be treated by a multi-layer heat conduction model or a bulk single level treatment (e.g. force-restore, Blackadar, 1979; Yee, 1988).

Turbulent exchange of heat and moisture from surfaces is treated using a resistance network. That of Masson is slightly simpler than the schematic. Essentially, transport from each facet is calculated using an exchange coefficient which must be estimated somehow. Turbulent momentum transport must be treated for the surface as a whole as bluff-body pressure forces dominate, in general, compared with skin friction. Since a simple geometry has been chosen, empirical results for roughness length and displacement height in terms of building height and non-dimensional shape parameters can be used. Grimmond and Oke review these relationships for real cities; those due to Macdonald (1998) and Raupach (1994) are widely used.

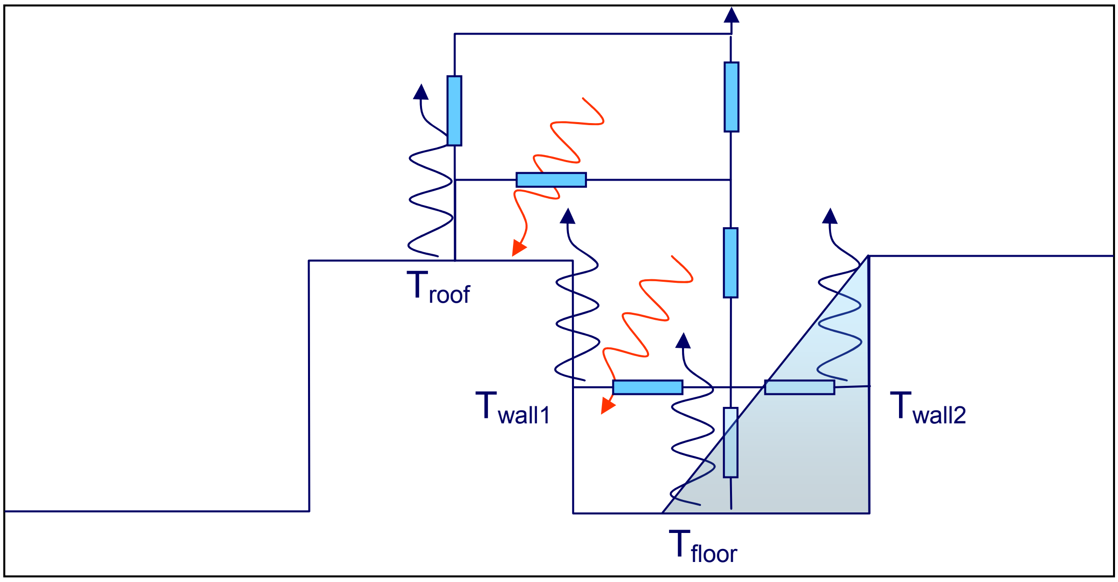

Figure 6 Schematic 'street canyon' surface exchange scheme.

A further simplification of this approach has been developed for the MetUM by the University of Reading and the Met Office. Analysis of a full ‘four facet’ approach reveals that the difference between wall and floor temperatures is much smaller than the difference between these and the roof temperature. While the accuracy of the assumption may depend somewhat on the material properties of the walls and floor, it arises primarily because of the strong coupling between walls and floor (via radiation and turbulence) and the impact of sky-view factor when compared with the roof tile. Secondly, the measurements of exchange coefficients with the various facets (Barlow and Belcher, 2002) showed that the roof and canyon are not strongly coupled. This means that the assumption behind tile schemes, of independent surfaces, is valid in spite of their proximity (though the same effective roughness for momentum must apply to each when coupled to the surface layer). The two approximations together lead to a ‘two-tile’ approach, the tiles being roof and canyon (Figure 7).

The two-tile approach, as well as making use of physically derivable parameters, improves on the single tile by recognising the considerable difference that can arise between roof and street temperatures as radiative exchange from the street is restricted by geometry and, in many cases, roofs show substantially different thermal inertia compared with streets.

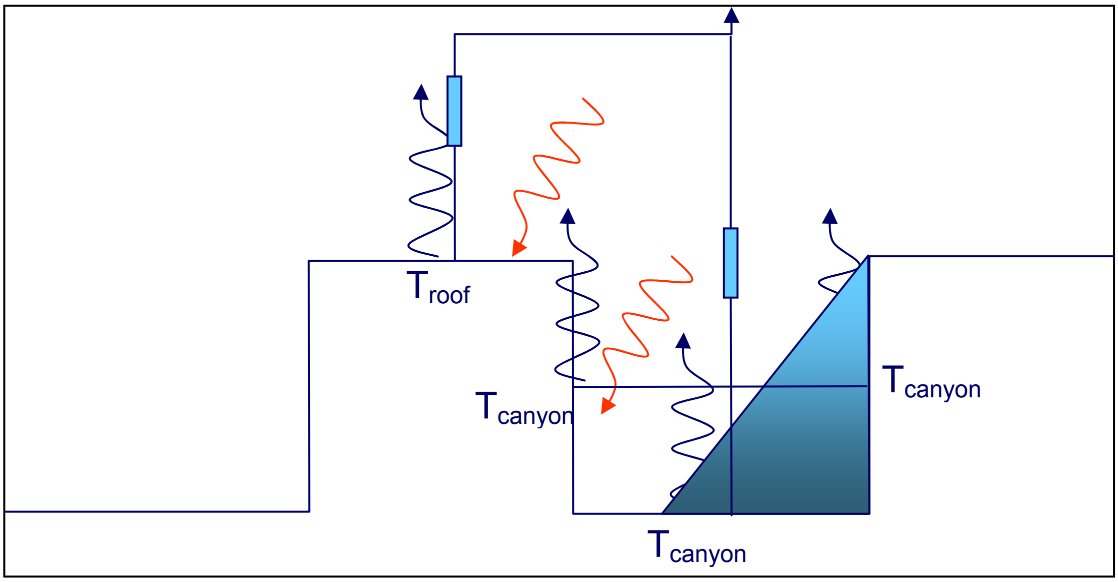

Figure 7 Schematic representation of the two-tile scheme.

A simple version of this was evaluated against surface flux data from various cities, (Figure 8 based on Fig.8 from Best, *et al*, 2006). This implemented a basic two-tile approach and a simple (and, perhaps, sufficient) treatment of effective albedo, but not the treatment of exchange coefficients and effective albedo and emmissivity developed by Harman (2003). A more complete implementation has recently been completed in the MetUM (Porson, 2008). A consequence of the modelled exchange coefficients is that the low effective roughness for heat discussed above arises naturally from the derived exchange coefficients.

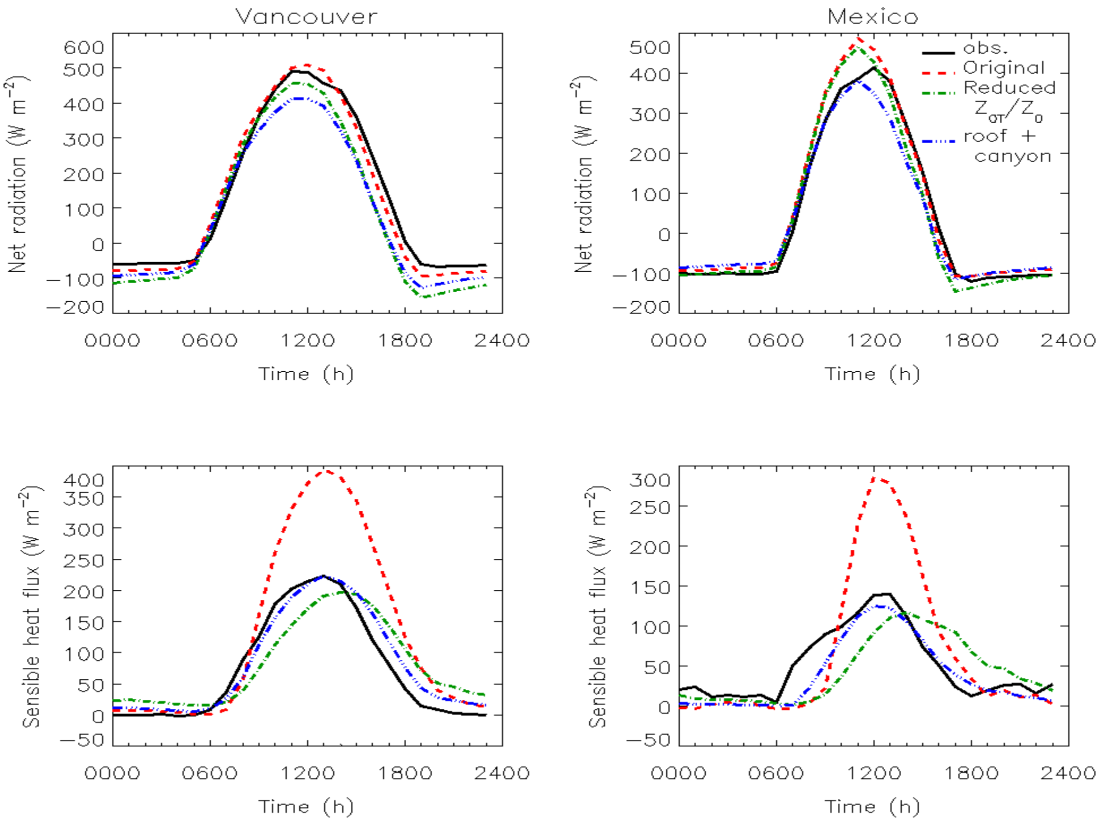

Figure 8 Summary of results from various versions of UM urban surface exchange scheme. Black solid line: observed flux; red dotted line: default model parameters; red dashed dot line: z_{0t}/z_{0m} = 10^{-7}; green dashed treble dot line: roof and canyon tiles. Left: Vancouver; Right: Mexico City.

3.2.3. *Explicit Urban Canopy*

The multi facet or two-tile schemes above have been found to do a reasonably good job of estimating surface fluxes over relatively uniform urban terrain. In practice, the schemes have two major deficiencies. The first is that only surface-layer fluxes are calculated. The flow in the so-called roughness sub-layer or canopy layer is not predicted. For some applications, such as pollution transport, the mean flow at building level is very important, so some idea of the canopy wind profile is needed. It is arguable that for a uniform urban surface this could be derived diagnostically. However, at this fine scale the details of the canopy are more important. Furthermore, some urban terrain is so variable, or some buildings so tall compared with the boundary-layer depth, that an effective roughness length has no meaning. The explicit urban canopy approach addresses these deficiencies.

The urban canopy model borrows from more traditional plant canopy models by modelling the flow effectively as a porous flow, with distributed drag elements. Different approaches exist; the approach of Martilli *et al* (2002) is exemplar. This uses a distribution of building heights, its cumulative form (the distribution of buildings taller than a given height) together with building and street widths, to model the distributed drag and scalar exchange from the buildings. The approach has shown very useful results, but is requires even more data than those above.

3.3. Concluding remarks on urban surface exchange

There are very many schemes in existence differing in complexity – the exemplars discussed above typify classes of scheme. Despite the enormous difficulty of the problem, considerable progress has been made in recent years. A number of issues remain:

- Availability and standardisation of urban parameters is a major issue. Schemes sometimes developed to suit available data rather than *vice versa*.
- A model intercomparison project is currently underway under Sue Grimmond (KCL) – see http://geography.kcl.ac.uk/micromet/ModelComparison/index.htm.
- Emphasis on performance with 'standard' parameters as much as on 'perfection'.
- Urban moisture is still a very difficult problem.

4. Radiation

4.1. Orography

The first radiation problem to address at high resolution has much in common with parts of the urban problem. The majority of large-scale models ignore both resolved and unresolved orography. It is straightforward to show that for shallow sub-grid orography the average effects largely cancel; during most of the day the main impact on incoming SW is to modify the effective surface area intersecting the incoming radiation. For a given sinusoidal variation, in orography, the surface SW is effectively modulated by the slope of the orography, i.e. a cosine. For shallow orography, the average effect is zero. LW radiation is modulated by the variation in surface temperature, which will also tend to average out to zero for shallow orography. As sub-grid orography increases amplitude, non-linear effects become more significant. However, the impact of resolved orography is likely to dominate at higher resolution.

Orography typically has a power spectrum which decreases as k^{-a}, with a around 2 or smaller (Beljaars *et al*). This means that the slope power spectrum is fairly flat with wavenumber, so as more orography is resolved, the variance in slope steadily increases. Furthermore, 3D effects become important. For SW, we need to consider the effects of slope and aspect as well as shadowing on the direct beam, slope and aspect and the 'sky view' on the diffuse radiation (Figure 9), and a new term arising from reflection from one hill side to another. When considering LW, we have to take into account the impact of 'sky view' on the atmospheric exchange and also the radiation that can be received from other parts of the terrain (hillsides) in sight, bearing in mind that that terrain may have differing land use.

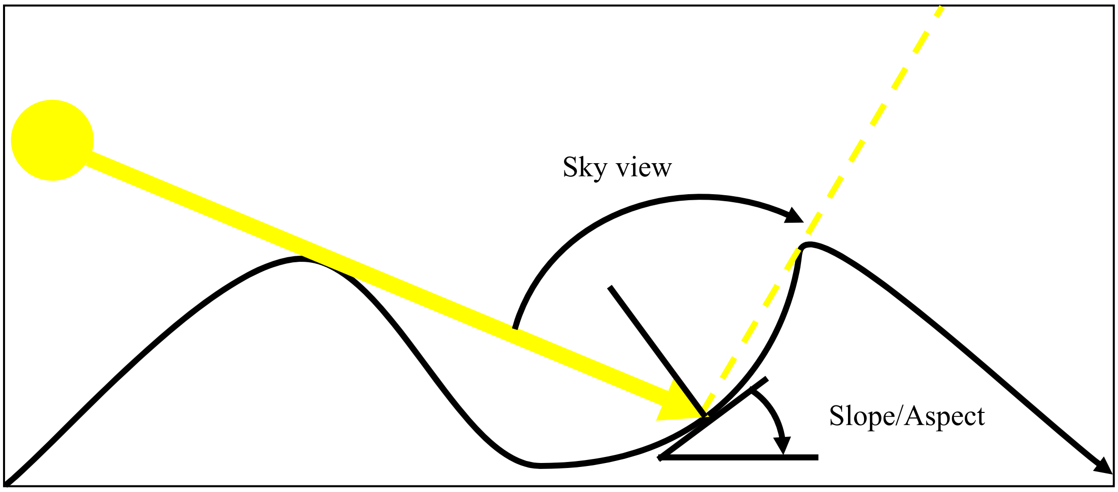

Figure 9 Schematic representation of some important aspects of the interaction of orography with radiation; the slope and aspect of the local surface interacts with the direct solar beam, which may be shaded by adjacent hills, while diffuse radiative exchange with the atmosphere depends on the 'sky view' factor.

Oliphant *et al* (2003) studied a particular case over the New Zealand Southern Alps in some detail and found a decreasing order of importance of slope aspect, slope angle, elevation, albedo, shading, sky view factor, leaf area index. The most important terms are relatively easy to understand. Resolved slope and aspect define the vector direction of the surface element and so have direct impact throughout day from the SW direct beam. South facing slopes receive more SW than north facing slopes. On the other hand, shading and sky view factor have their main impact on SW at dawn/dusk when SW is small anyway. In common with urban canyons, sky view factor is likely to have its largest impact on LW radiation.

Möller and Scherer (2005) reviewed the treatment of orographic impacts on radiation in a variety of mesoscale models. The majority do nothing (i.e. assume flat earth). A significant minority include slope and aspect, and only ARPS (in the survey) include shadowing as well. The impact of resolved slope and aspect is relatively easy to retrofit to schemes; the MetUM now treats slope/aspect.

It is not so clear what benefit is actually gained in terms of forecast skill. Möller and Scherer showed 0.5-1 C RMS improvement in 2 m T from explicit sub-grid model, but this was over the Alps and may be extreme. It is also likely that the primary benefit was in the local temperature diagnosis, rather than in forecast evolution. Roberts (2007) uses the scale-dependent verification technique discussed above to look at the forecast scales impacted by including slope and aspect in the radiation scheme (Figure 10). This shows some increase in forecast skill of rainfall forecasts, but also shows that much of this improvement is at small scales which have very little skill anyway. Though this is a relatively poor case, it does seem likely that the change in surface forcing due to surface slope will mainly be at small scales and systematic impacts may soon be lost in a forecast.

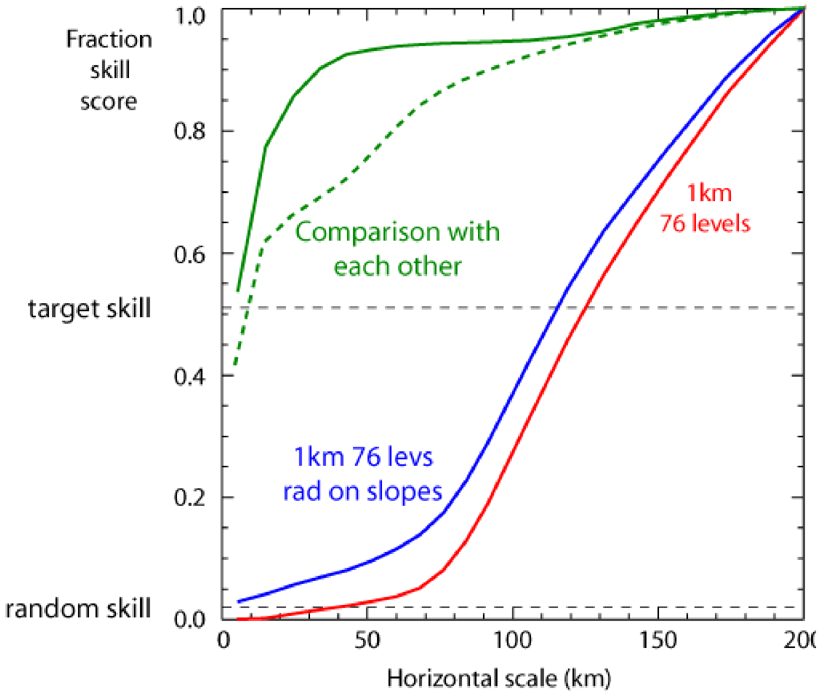

Figure 10 Impact of treating radiation on slopes in a single convective case in the MetUM. Red shows the control rainfall forecast skill as a function of scale, blue the skill with radiation on slopes included, and solid green the relative change in skill. The dashed green shows the relative impact of going from 38 to 76 levels for comparison. After Roberts, 2007.

4.2. 3D Radiation

It is very obvious that, in principle, the plane parallel approximation of operational radiation schemes falls down in DCPMs. The impact on SW is probably most important: a typical deep convective cloud may be of order 10 km high, and casts a shadow which may extend well over 10 km, i.e. many grid boxes (Figure 11). This shadow may have direct dynamical impacts, such as thermally-driven low-level convergence at its edge. For a thorough study of the impact of cirrus shadowing on convection triggering see Marsham *et al,* 2007. There are also impacts on LW from radiative exchange with the sides of clouds. 3D radiation schemes do exist, often based on Monte Carlo simulations, and these can be used to assess the importance of 3D effects. However, they are prohibitively expensive for operational use, and, in practice, they are probably not necessary. Existing models generally produce very useful forecasts without 3D schemes.

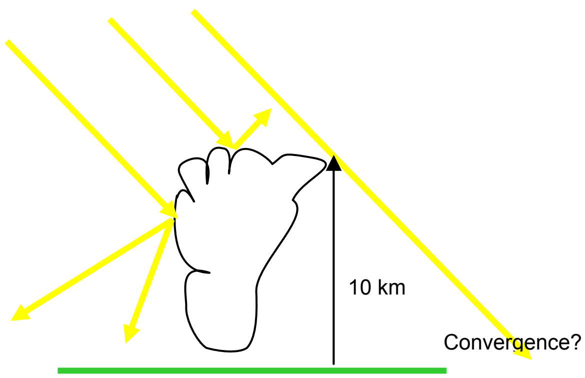

Figure 11 Schematic showing 3D interactions of SW radiation with cloud with possible thermally driven low-level convergence at cloud shadow edges.

The reasons for this success probably lie in the timescale for radiative forcing of the atmosphere and the integral nature of the problem. Direct SW heating and LW cooling of the atmosphere are relatively slow compared with cloud evolution, except perhaps at the top of layer clouds where the plane parallel assumption will work well. The main 3D effect, as discussed above, is likely to be SW shadowing and scattering by deep convective cloud. In practice, we get away without treating this probably because:

1. Individual clouds are poorly forecast anyway – we need to look for systematic errors to see 3D effects.
2. Clouds usually move and the timescale for atmospheric response to shadowing is roughly the timescale for BL overturning, since the main impact is through surface heating. In a convective boundary layer, this is roughly h/w*, where h is the boundary-layer depth and w* the convective velocity scale. This timescale is typically around 20 min, and clouds may well have moved 10 km in that time. Thus, the boundary layer acts as a low pass filter on the direct SW forcing. It is arguable, both for cost and consistency, that smoothed cloud fields should be used in the radiation scheme.

In practice, only a fortuitous organised response will produce systematic errors, e.g. cloud streets which happen to line up cloud shadows in a quasi-steady state. It is also worth remembering that a 3D scheme needs good 3D cloud properties, so test the model even further.

5. Cloud Microphysics

5.1. Fundamentals of parametrization schemes

The generality of microphysics parametrization has been discussed elsewhere in this seminar series (see paper by Forbes). The general problem is one of describing the behaviour of a large number of particles, covering a wide range of sizes, and, in the case of ice particles, shapes. The first level of simplification is to describe the particles in terms of a size and possible shape distribution. So-called ‘bin’ techniques are quite successful in predicting the evolution of such spectra. The techniques rely on discretizing size spectra into a finite set of representative particles. The approach requires a fairly large number of bins, the concentration in which has to be advected around the model. It is thus very expensive, and currently only practicable for research purposes. Operational (and many research) models generally use some form of bulk approach.

In the bulk approach, the aggregation over particles is achieved by representing the distribution in terms of a small set of classes of particle, each with a specified type of size spectrum (e.g. exponential, gamma, log-normal). In each class, a small number of parameters is allowed to vary; typically, these parameters are expressed in terms of moments of the distribution. Thus, for example, a single-moment scheme might allow the third moment (mean volume or mass) of particles to vary. Other moments must be parametrized in terms of this; for example, in the Marshall-Palmer rainfall distribution (an exponential distribution) the zeroth moment (the number concentration) is parametrized in terms of the third. More sophisticated schemes allow more moments to vary independently, though it must be remembered that chosen spectral shapes only allow a maximum set of parameters to vary independently (two in the case of exponential and log-normal). Transformations between classes are computed using appropriate integrals over the class spectra (either exact or approximate). Similarly, the average terminal velocity is computed as an integral, often representing individual particle fall speeds as a power-law function of size parameter.

5.2. How many and what classes?

It is important to recognise two features of bulk schemes. The first is that, in general, processes do not maintain a similar spectral shape, so imposing a shape effectively adds some effective process to the system which brings the size spectrum back to the specified shape. In general, this additional process is artificial and should be thought of as a truncation error of the scheme, though in the case of rainfall the actual process of large droplet breakup actually can tend to return spectra to exponential.

The second feature is that each 'class' has a single effective terminal velocity (though this may vary with e.g. bulk density of the class). If Q_x is the density of class x, it is assumed **all** to fall with speed V_x. Though a few schemes attempt to take account of the dispersion of fall speeds, most do not. In designing a scheme, the choice of the number of classes, amongst other things, determines to what degree condensate can be transported to different parts of a cloud system.

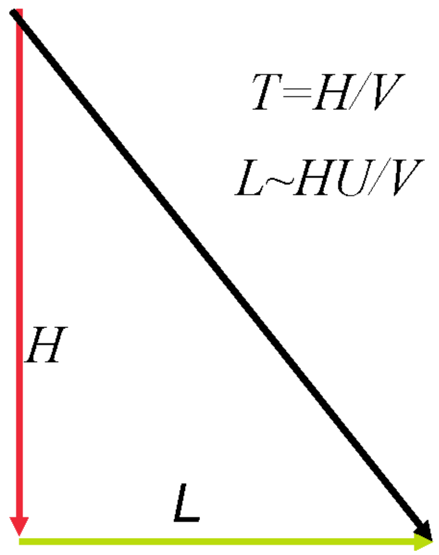

Figure 12 Simple schematic showing the relative importance of advection by horizontal wind U and fallout of hydrometeors with terminal velocity V

The number of classes needed and the complexity of treatment of each is an important question in the design of schemes. A convenient consequence of the properties of the processes that create rain means that there is a natural division between cloud liquid water and rain, as, in practice, a distinct gap tends to exist between their spectra. The same is not true for ice, where a distinct separation between small ice crystals and larger aggregates does not always exist.

A first consideration is based on simple kinematics. The sketch on the left (Figure 12). If we suppose that a given hydrometeor falls with bulk speed V from a height H where it is formed, a horizontal wind with speed U moves the hydrometeor a distance L~HU/V. Broadly, if L is similar to or smaller than the horizontal resolution of the model, we do not need to take this horizontal advection into account.

This idea can be expressed more mathematically by considering the conservation law for some bulk hydrometeor class x with density Q_x, terminal velocity V and sources and sinks S_x.

$$\frac{\partial Q_x}{\partial t} + \nabla . \left[\left(\mathbf{u} - V\hat{\mathbf{k}} \right) Q_x \right] = S_x$$

This may be re-written in advection form (using the continuity equation):

$$\rho \frac{D(Q_x/\rho)}{Dt} - \frac{\partial}{\partial z}(VQ_x) = S_x$$

Ignoring advection means ignoring the first term in this equation. This leaves a diagnostic relationship which is very easy to solve by (numerical) integration downwards from the top of the model (where $Q_x = 0$). Since $P_x = VQ_x$ represents the vertical (downward) hydrometeor flux, it is of direct interest at the surface – if Q_x is rain then $P_x = VQ_x$ is the rainfall rate. If the flux at layer boundaries is calculated, the simplest finite difference representation leads to:

$$P_{xk-1/2} = P_{xk+1/2} + S_{xk}$$

and hence, on integration:

$$P_{xsurface} = \sum_{k=1}^{top} S_{xk}$$

It is also instructive to write the above conservation law in the following form:

$$\rho\frac{D_H(Q_x/\rho)}{Dt}-\frac{\partial}{\partial z}[(w-V)Q_x]-\frac{Q_x}{\rho}\frac{\partial}{\partial z}(\rho w)=S_x$$

The first term represents horizontal advection. The third is minor, especially close to the peak updraught (zero at the peak upward mass flux). The second represents the divergence of net vertical hydrometeor flux, taking into account vertical air speed. Apart from cloud and very small ice particles where *V* is negligible anyway, it is clear that *w*-*V* can take either sign. In large-scale models, *w* is usually small even compared to snow fall speeds, but this is not the case in deep convection. The choice of classes can therefore determine whether precipitation falls out or can be carried elsewhere in a cloud (e.g. into downdraughts), and use of more classes may be necessary to ensure that both can happen.

Three points may be made here as an aside. The first is that the diagnostic relationship arrived at by ignoring advection is sometimes (mistakenly) described as the assumption that 'all the rain falls out in one timestep'. There is no such implication, as no timestep is involved in the relationship. The relationship effectively describes a quasi-steady state and, like all quasi-steady state assumptions (which occur frequently in parametrizations) a consequence is that should S_x *change* there will be an *instantaneous* response in the surface hydrometeor flux which *exactly* matches the change. If simple first order timestepping is used, the source in a timestep at level *i* is $S_{xi}\Delta t$; the contribution to the surface flux in the timestep is $S_{xi}\Delta t$, leading to the idea that all the source 'falls out' in one timestep. The notion of infinite fall speed is not helpful, but that of instantaneous response is, for example in the DA technique known as Latent Heat Nudging.

The second point is that, however complex the scheme, the hydrometeor budget is governed by the condensation rate, which is largely dynamically driven (especially in the case of cloud). Considering only 'warm' processes, as a simple example, with vapour density *Q*, bulk cloud water Q_l and rain water Q_r, we can write:

$$\frac{\partial Q}{\partial t}+\nabla.(\mathbf{u}Q)=-C$$
$$\frac{\partial Q_l}{\partial t}+\nabla.(\mathbf{u}Q_l)=C-S_r$$
$$\frac{\partial Q_r}{\partial t}+\nabla.\left[\left(\mathbf{u}-V_r\hat{\mathbf{k}}\right)Q_r\right]=S_r$$

where *C* is the condensation rate and S_r the conversion rate of cloud to rain.

Adding the equations for Q_l and Q_r then making the 'no-advection' approximation leads to the approximation:

$$\frac{\partial}{\partial z}(V_rQ_r)=-C$$

which indicates that rainfall production is determined by total column condensation, which is largely determined by the wet adiabatic temperature profile. Thus, where vertical hydrometeor flux dominates, we can conclude that the details of microphysics have little role to play in determining rainfall rates; on the contrary, the relationship is effectively reversed. The profile of hydrometeor flux determines the profile of rain water density (using the fallspeed *V*) and the balance of C and S_r then determines the cloud water density. These profiles have an important role in radiative exchange but are effectively slaved to the condensation rate. As a simple example, S_r is often represented by $kQ_l\,Q_r^a$, in which case Q_l=C/(Q_r^a), with Q_r determined from *C* as above.

Thus, we can make the generalisation that at *large scale*, where the steady-state, no advection approximation holds, cloud microphysics determines the hydrometeor concentrations needed to produce the dynamically-determined precipitation rate and condensation rate. At *small scale* microphysics determines the relative transport of hydrometeors and hence the spatial distribution of some of the diabatic heating associated with it. The above considerations enable us to define large and small in terms of L, and the following table gives some idea of typical numbers for a set of classes.

Cloud	L very large
Rain	L~5 km
Snow	L~50 km
Graupel	L~1-2 km

Table 1 Typical horizontal advection length scales for a selection of hydrometeor classes.

An objection can be raised to these arguments concerning cloud water, as these arguments imply that cloud water should always be treated prognostically. The problem arises because the above arguments do not include timescales for transformation. In the case of cloud, the condensation process from vapour is extremely fast, instantaneous for the purposes of most models. Ignoring the terminal velocity of cloud drops, it makes more sense to use 'total water' (non-frozen cloud liquid plus vapour) as a prognostic and diagnose the split between variables. In this case, we should consider total water as the class, in which case the above arguments still apply and we conclude that total water must always be treated prognostically.

The third point is that not treating a class prognostically is not the same as not treating it at all. The above arguments suggest that graupel and hail should be treated diagnostically in the majority of models apart from the very highest resolution. However, they are usually absent from large-scale microphysics schemes. This is because such schemes are designed for layer cloud outside convective cores. Graupel and hail are hiding in the microphysics schemes inside convection parametrizations.

Table 1 leads to the conclusion that prognostic cloud, snow and rain are necessary for sub-10 km models, and prognostic graupel is also needed for models around 1 km. Indeed, this very much represents the norm in DCPMs, with the addition of small ('pristine') ice and, of course, vapour, though other schemes have been proposed and a wide variation of detail exists. The next question is how many moments should schemes treat prognostically.

5.3. The need for higher order microphysics schemes

Single-moment bulk schemes in terms of mass are good for representing inter-conversion by collection processes if (and only if) particle number density is conserved by the process. A good example is accretion of cloud by rain, which conserves rain drop number density. This is the basis of the scheme by Rutledge and Hobbs, 1983, which was designed for studying the seeder/feeder process in which rain falls through cloud. This scheme has formed the basis of many others. The comments above regarding implied processes redistributing the rain drop sizes to maintain the spectral shape still apply and, furthermore, cloud droplet number density is, of course, not conserved by the process.

Other processes conserve (or approximately conserve) mass but change number density – self-collection or autoconversion being typical examples. The process occurs across the whole spectrum, modifying its shape, and only a relatively small proportion of mass is 're-classified' as a different class. Warm rain is a very good example. Stochastic coalescence conserves cloud water (q_c) but gradually broadens the size spectrum until sufficient large drops are available to generate rain. The number of cloud drops, N_c, thus decreases with time with the same q_c, implying larger drops. Similar considerations apply to ice to snow autoconversion. We expect single-moment mass-based schemes to have some problems with this kind of process, and two moment schemes are not uncommon e.g. the Cohard and Pinty (2000) warm rain scheme in Meso-NH. Graupel has further complications of 'wet growth' modes – this depends on the dissipation of latent heat of freezing so area or density becomes an important parameter. Triple moment schemes are thus sometimes used for graupel.

5.4. Main uncertainties in cloud microphysics

The main uncertainties in microphysics schemes can be summarised as:

- Ice/snow fall speeds, growth and collection depend on particle shape (habit) – a given set of parameters is likely to work well for a given type of cloud and temperature regime.
- Ice nucleation represents a constant problem:
    - Homogeneous rates are uncertain.
    - Heterogeneous rates rely on assumptions regarding ice nuclei
- Ice multiplication (Hallett-Mossop process).
- Coupling to aerosol schemes is a research question – sensitivity is easy to show but need more work using 'climatological' aerosol to demonstrate systematic impact.

5.5. Impacts of uncertainties

The impact of uncertainties in schemes is far too broad a topic to discuss here. The literature is full of papers investigating the sensitivity of deep convective cloud evolution (and other cloud types) to microphysics. An example using the MetUM is shown in Figure 13

There is a considerable difference, however, between sensitivity and forecast skill. The quality of forecasts at the cell scale is often not sufficient (yet) to enable a given scheme to be demonstrably better or worse than an alternative. Systematic errors, usually over many forecast cases, must be found to justify the use of a more complex scheme. An example is shown in Figure 14. This shows area average hydrometeor concentrations from two simulations of a semi-idealised GCSS case study, from the Large-Scale Biosphere-Atmosphere Experiment in Amazonia (LBA), Grabowski, 2006. This is a 'typical' diurnally forced continental case, starting with a stable boundary layer, moving through unstable, shallow cumulus to deep showers. The details are unimportant here; the most notable difference is that the showers last significantly longer into the evening using the full scheme, probably mainly because prognostic rain falls more into down-draughts, promoting more surface cold-pool generation and stronger updraughts. This is the kind of systematic error which is detectable in forecasts, though, in general, many cases would be needed to confirm the error as many other factors could lead to similar results.

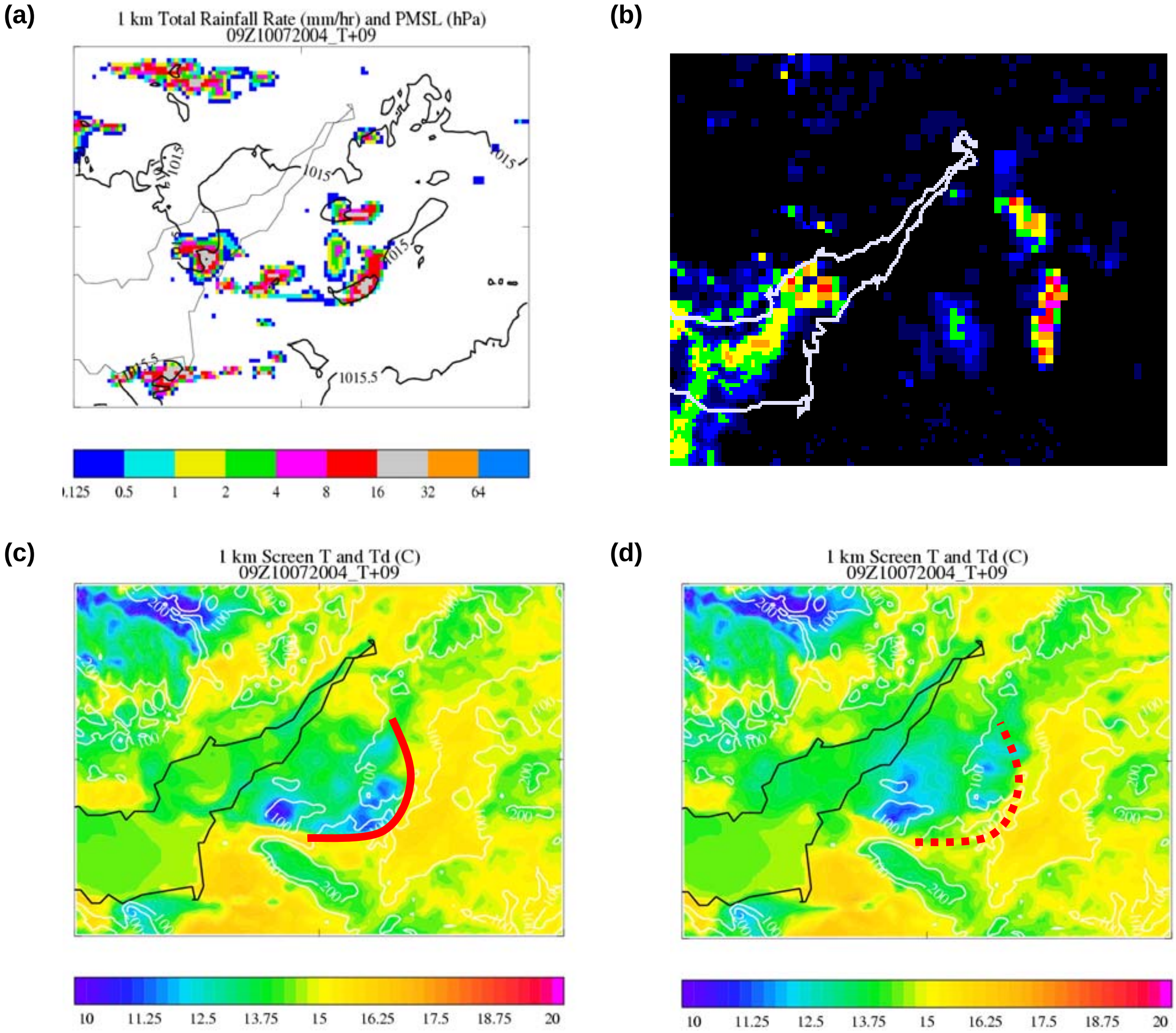

Figure 13 Example simulation from the CSIP pilot study over southern England, 10th July 2004. (a) 1 km resolution forecast from the MetUM (b) verifying radar rainfall rate (c) forecast 1.5 m temperature showing a surface cold pool generated by the outflow of a primary storm. Orography contours every 100 m in white. (a) Secondary showers have formed triggered by convergence at the leading edge of the outflow (the gust front, red line in (c)). The radar shows a similar organisation but somewhat faster propagation speed. (d) As (c) but changing the assumed snow particle shape to one that falls twice as fast and evaporates less rapidly. This changes the forecast dramatically, greatly reducing the cold pool and secondary shower formation. The red dashed line shows the gust front from (c).

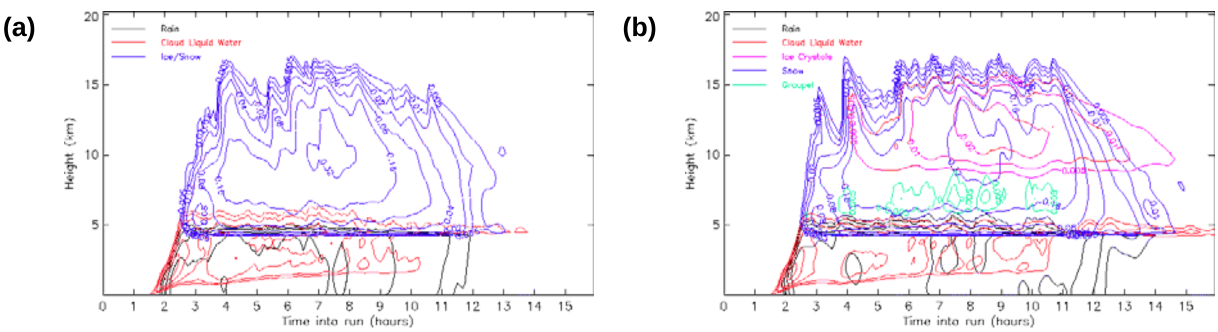

Figure 14 Area average hydrometeor concentrations for a MetUM simulation of the GCSS LBA diurnal cycle (see text for details). (a) Simplified microphysics with diagnostic rain, combined ice and snow and no graupel. (b) 'Full' single moment microphysics.

6. Turbulence and convection

6.1. Background

Before discussing turbulence and convection parametrization, it is worthwhile revising some basic concepts regarding averaging. Reynolds averaging has formed the basis of much work on turbulence for many decades. The Reynolds average is an ensemble average, i.e. an average over a very large number of realisations of the same flow each consistent with the boundary conditions (notionally started from initial conditions differing infinitesimally from each other). The term RANS is often used (‘Reynolds-Averaged Navier-Stokes’ equations). An essential feature of the equations is that they contain new terms representing the divergence of averaged ‘turbulent fluxes’ of the form $\langle u_i' y' \rangle$ where y represents any prognostic (e.g. a component of velocity, potential temperature, humidity), the prime represents deviations from the ensemble mean and angled brackets denote the ensemble mean. The parametrization process seeks to parametrize these fluxes (see below).

The turbulence parametrizations in NWP models are invariably based on the RANS approach. However, it is important to recognise that RANS does not, in itself, involve any volume averaging. The RANS equations yield non-turbulent solutions but these solutions may contain a wide range of spatial scales and may well contain scales un-resolved by a given model. There is thus no meaning to the term ‘sub-grid’, as no grid is implied. We may, however, consider a steady situation where the boundary conditions only contain scales much larger than the scale of the largest eddies which evolve in the turbulent flow. If (and only if) the spatial averaging length is much larger than the largest eddies in the flow then the volume average tends to the spatial average. Since the model must (presumably) resolve the applied boundary conditions (or, more correctly, the applied boundary conditions are those represented by the model), this implies a gap in the energy spectrum between the ‘resolved’ and the ‘parametrized’ flow – the well-known ‘spectral gap’. Much the same arguments apply for time-varying forcing, which must vary on timescales much longer than the turbulence for RANS to be successful. So RANS is supposed to contain **all** turbulence and produces ‘steady’ flow at turbulence scales irrespective of the model resolution (provided it is sufficient to resolve the forcing).

Deep convection schemes used in larger-scale models are probably based on Reynolds averaging. The word ‘probably’ is used because this is rarely stated explicitly. However, the usual requirement that a ‘grid-box’ contains many clouds is equivalent to the ‘spectral gap’ requirement above for equivalence of ensemble and spatial averaging. It is very clear that such schemes are not designed for DCPMs. Claims are made of some schemes that they are suitable for ‘mesoscale’ use, and some schemes are being designed which explicitly do not make this assumption (e.g. Gerard and Geleyn, 2005), but it is generally the case that we do not have convection schemes designed for DCPMs. Exactly when it is better not to use a deep convection scheme undoubtedly depends upon the model and case of interest, but we shall make the assumption that, at present, DCPMs operate without deep convection parametrization. As we shall see below, it is arguable that some form of deep mixing parametrization is still required beyond local turbulence parametrization at typical resolutions, but at present most models do not implement such a scheme.

If we volume average at scales smaller than the largest eddies then we are producing a ‘Large Eddy Simulation’ (LES). Volume averaged equations also produce new terms, here representing unresolved fluxes. The terms are superficially similar to those arising in RANS, but different because, unlike ensemble averaging, volume averaging does not commute with the differential operators forming the underlying equations. The parametrization scheme results in a flow which still evolves unsteadily to form a turbulent solution, albeit missing the smallest scales. This solution depends chaotically on the initial conditions in a way which RANS cannot.

The details of parametrization schemes will not be covered here. The terms to be parametrized contain a higher 'order' than the variable being predicted (i.e. the RANS turbulent fluxes are second order, while we only have equations for first order quantities). Using these equations we can write down new equations for the second order terms, but these now include third order terms, and so on. The parametrization process is one of developing new equations which close this set. This process was systematized (for RANS) by Mellor and Yamada (1973), and many schemes are based on this systematization. However, the difference between RANS and LES can still, to some extent, be thought of as primarily a difference in choice of length scale between those determined by the flow (the largest turbulent eddies) and those specified by the modeller. This idea is typified in the scheme of Cuxart *et al*, 2000, in which the same basic equation set is applied both to Reynolds-averaged flow and LES by changing the choice of turbulent length scale.

The basis of many parametrization schemes is Prandtl's mixing length hypothesis. First, we assert that the turbulent flux of y, $\langle u_i' y' \rangle$, may be written as the product of a characteristic turbulent velocity and a characteristic turbulent fluctuation of y. The latter is assumed to be given by the typical variation of y over the length scale of turbulent eddies, L. Thus

$$\langle u_i' y' \rangle = u_{turb} y_{turb} = u_{turb} \left(L \frac{\partial \langle y \rangle}{\partial x_i} \right)$$

A simple 'first order closure' takes the next step of applying the mixing length hypothesis to the characteristic turbulent velocity, which makes sense when one considers y to be a velocity component (see Figure 15). It is also natural to consider the turbulent velocity scale to be derived from the turbulent kinetic energy (TKE) and schemes which are based on simplifications of the prognostic equation for TKE, assuming homogeneity, are very common (e.g. Cuxart *et al*, 2000). Analysis of the TKE equation assuming steady state and neglecting transport terms leads to the result that $\mathrm{TKE}^{1/2} \sim |S|^{1/2} f(Ri)$, where S is the magnitude of the rate of strain tensor, and f(Ri) represents some function of stability expressed as the local Richardson number. This is essentially an extension of the form illustrated in Figure 15

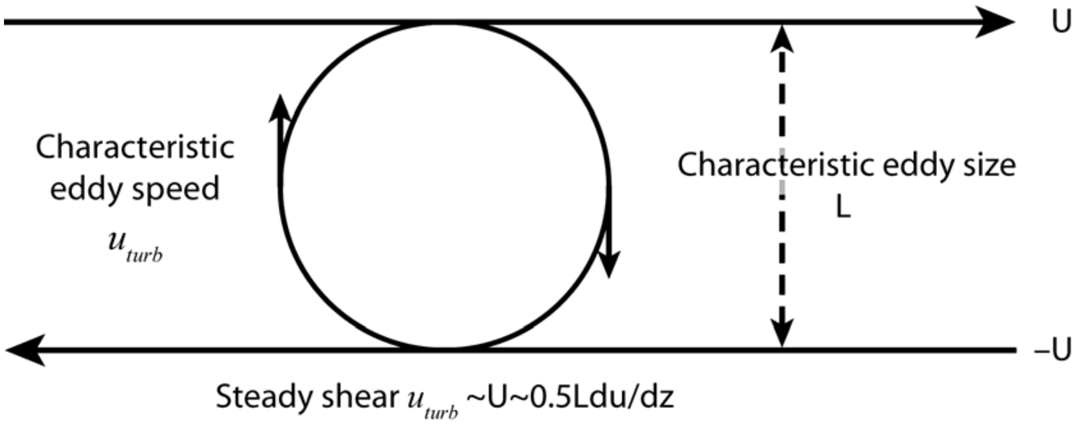

Figure 15 Schematic of Prandtl's mixing length hypothesis.

As discussed above, specification of the eddy length scale completes this approach to closure. A very common form of LES closure ('Smagorinsky-Lilly' – see Lilly, 1967) combines a specification of L as proportional to the grid length, L=CsΔx, with the homogeneous steady state result above. These assumptions are (only) consistent with the length scale being (well) within the inertial subrange of turbulence. Reynolds averaging requires knowledge of the characteristic outer length-scale of the turbulent flow. This may be argued from characteristic scales of the boundary conditions; for example, in the 'surface layer' the only length scale available is the height above the ground, while in the convective boundary layer the boundary-layer depth is an important length scale. Some higher order schemes derive equations for this length scale, but considerable success has also been achieved by combining the 'non-local' derivation of mixing length (e.g. boundary layer depth) with a non-local derivation of turbulent velocity scale, based upon boundary fluxes (Lock, *et al*, 2000). Such schemes can also include fluxes unrelated to (and often the opposite sign of) the gradient.

In considering these simple ideas of turbulent length scales leaves us with a very obvious problem. Given that a model with ~1 km horizontal resolution only properly represents scales several times this, it is clear that the Reynolds averaged approach is generally appropriate for the boundary layer. This is certainly true for a dry boundary layer, though some doubts may be raised when moisture becomes involved (see, for example De Roode *et al*, 2004). However, DCPMs are certainly LES when it comes to the deep convective clouds themselves, if one considers the clouds to be the energy containing turbulent eddies. However, the inertial subrange is (probably) at scales no larger than 100 m (Bryan et al, 2003). Around 1 km is precisely around the scale of the most energy containing eddies, so the parametrization of (volume-averaged) turbulence is anything but straightforward. We shall now consider some of the approaches taken in practice.

6.2. Approaches to turbulence and mixing

The above considerations suggest that parametrized treatments of vertical mixing very similar to treatments in larger-scale models, with the same choices of local, non-local and higher-order local schemes in the boundary layer, are still appropriate for DCPMs. Non-local schemes are especially used in the convective BL (CBL), based on bulk diagnosis of BL depth, surface and entrainment fluxes, plus possible counter gradient terms. Higher-order local schemes emphasise vertical transport of turbulence e.g. from middle of the CBL into the stable entrainment layer. Their success in the CBL probably relies on their ability to continue mixing in the stable entrainment layer where a first order local scheme would tend to reduce mixing. Non-local schemes succeed here by dint of separately parametrizing the entrainment. Local schemes (or nothing!) may also be appropriate outside the BL (though there is scope to develop non-local schemes for problems such as alto-cumulus layers). Local higher order schemes are popular; this may be because they can more easily be extended to sub-grid condensation (via $<w'^2>$ or more complex approaches). They may be more numerically smooth than 1st order schemes (see Diamantakis, 2007 analyses of non-linear diffusion)

As discussed above, the true horizontal resolution of practical DCPMs is (much) larger than the largest energy containing eddies in the convective boundary layer. In general, $\Delta x >> \Delta z$ (even outside the PBL) which means that $\partial_2 / \partial x^2 << \partial_2 / \partial z^2$. Thus, in principle, horizontal mixing is negligible and no change is required to the BL scheme. Even outside the BL, horizontal mixing is usually relatively small-scale compared with a few horizontal gridlengths, so horizontal 'diffusion' is slow. In principle, 'do nothing' for horizontal mixing and 'business as usual' for vertical is a valid starting point.

Following this approach, however, can lead to practical problems. 'Do nothing' is not really a meaningful option, as 'nothing' does something! Stable dynamical cores generally dissipate at the grid scale – the model behaves roughly as a high viscosity fluid; solutions may develop which may (or may not) resemble true atmosphere but with the wrong scales. For example convective cells, and BL roll vortices develop with some realistic structure but typically too large in scale and modified intensity. The nature of deep convective cells which develop is very dependent on effective horizontal mixing.

It is also not clear what to do with vertical mixing. Shallow convection is a new 'grey zone'. In principle, we definitely need to parametrize shallow convection as most of the motion is unresolved but how then do we handle transition from shallow to deep? This involves interaction with near-gridscale dynamics and so is very dependent on model.

No theoretically well founded solutions to this problem exist in the literature (as far as the author knows!) The approach adopted in practice often depends on history of model. Approaches include:

- Fixed diffusion (or hyper-diffusion such as ∇^4, which is more scale-selective).
- 'Horizontal Smagorinsky' with or without stability corrections, with or without vertical components of shear, coupled to standard vertical schemes.
- '3D turbulence closure'.
- 'LES' 3D Filter (Smagorinsky-Lilly)

However, whichever we apply, we must recognise that 1-2 km is **not** a very good resolution for deep convection. Clouds are very energetic at these scales. We should not expect faithful or converged solutions. Instead we are looking for solutions which retain some characteristics of the true solution and may have to accept some systematic errors.

6.3. Practical examples

Figure 16 shows power spectra of vertical velocity from the MetUM using various constant horizontal diffusion schemes. Here, diffusivity is expressed in terms of the number of timesteps to reduce waves of wavelength 2dx by a factor *e*. As expected, 'more' diffusion damps the tail of the spectra more. See also Skamarok, 2004. However, figures such as this ignore the phase information in the flow, and give the false impression that the flow is merely smoothed. Instead, when deep convection occurs, the scale of cells *and inter-cell spacing* can be affected. This is illustrated in Figure 17, which shows a small area of the data leading to the spectra in Figure 16. The red areas are strong convective updraughts.

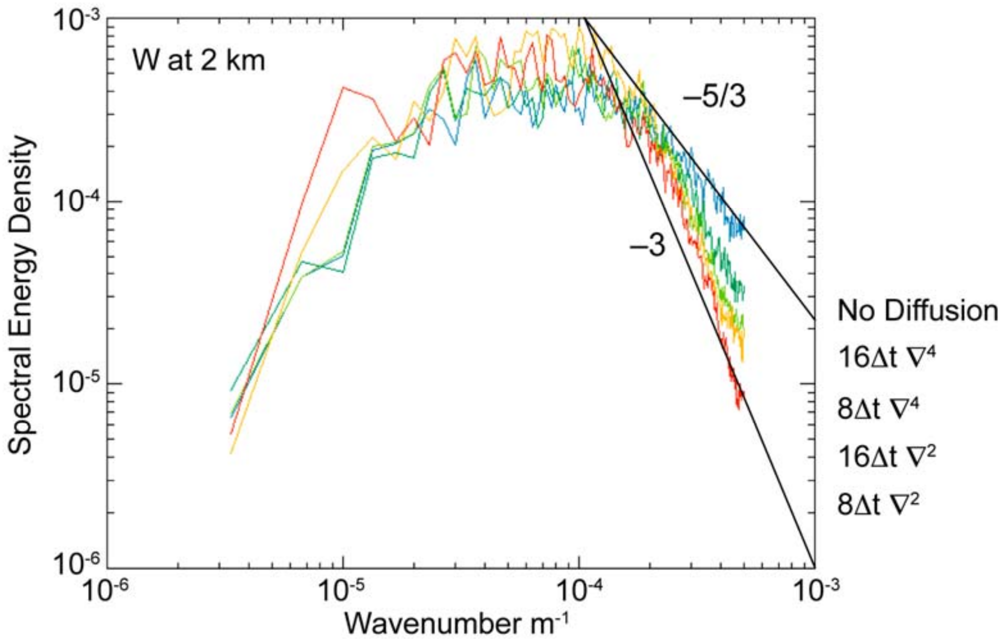

Figure 16 Vertical velocity power spectra from MetUM simulations of deep convection using different approaches to horizontal mixing.

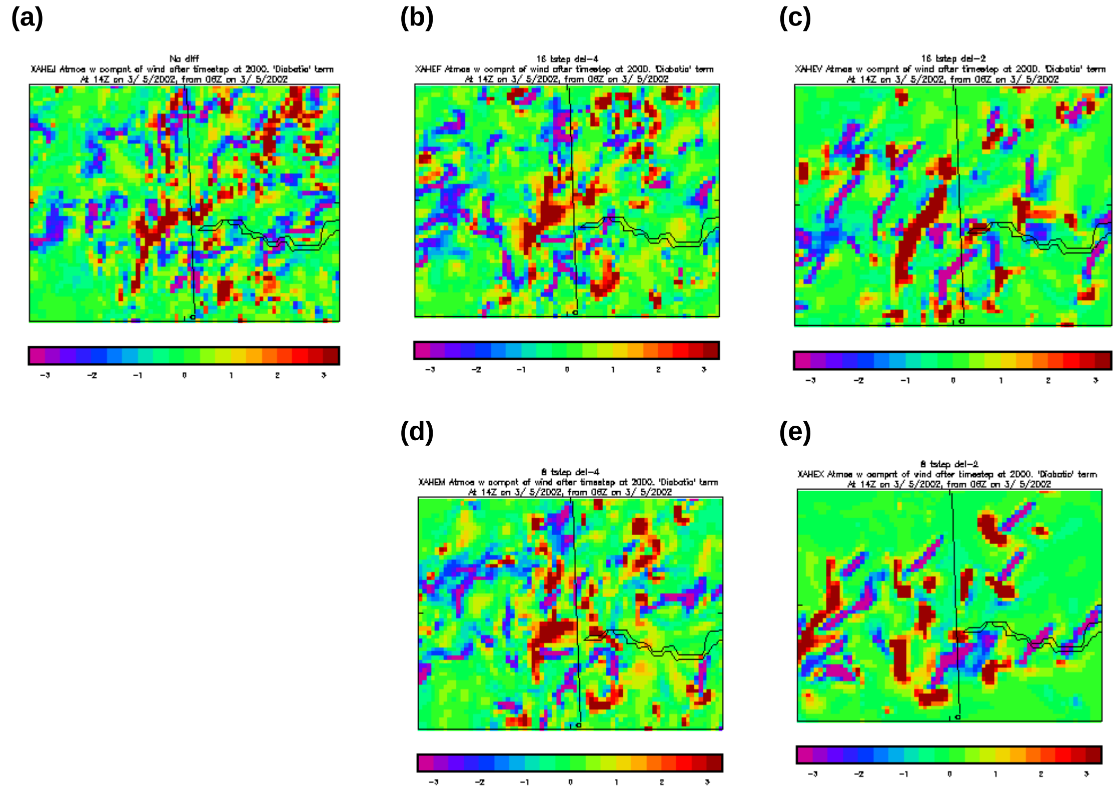

Figure 17 Vertical velocity at 2 km from a part of a set of MetUM simulations of a convection case at 1 km resolution. (a) No horizontal diffusion. (b) $16\Delta t\,\nabla^4$ (c) $16\Delta t\,\nabla^2$ (d) $8\Delta t\,\nabla^4$ (e) $8\Delta t\,\nabla^2$.

The larger effective horizontal dissipation in lower resolution models also has an impact on the timing and character of convective initiation. Most models show a tendency to delay and 'overshoot'; the lack of sufficient variability at the gridscale prevents growth of instabilities, often allowing more Convectively Available Potential Energy (CAPE) to build up and generating an initial pulse of rain. Figure 18 shows this effect in the MetUM. Smagorinsky-Lilly mixing helps to control the overshoot at the expense of further delay in initiation. Further analysis shows that the horizontal and vertical mixing have very different roles – efficient vertical mixing promotes earlier triggering by eroding inhibition, while more active horizontal mixing delays the initiation as well as controlling cloud intensity once initiated. Very similar results are shown by Petch (2006).

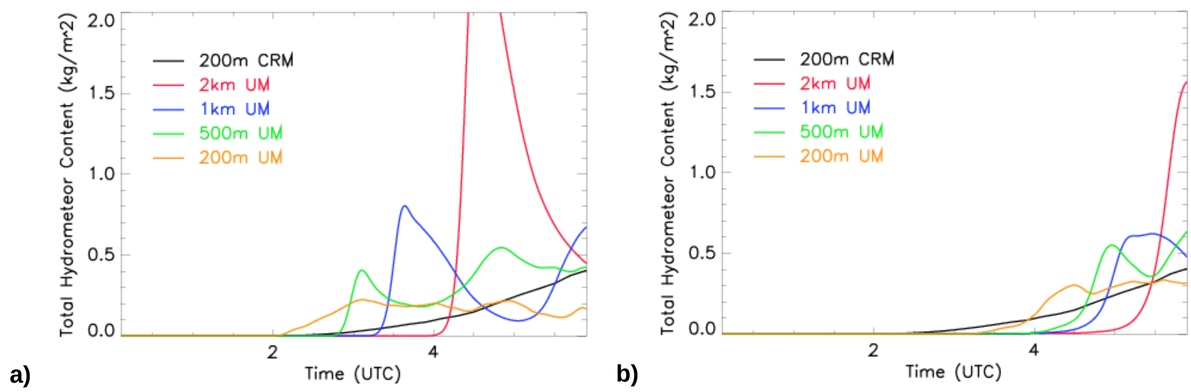

Figure 18 Area-averaged total hydrometeor content for the MetUM simulating the LBA diurnal cycle at various resolutions compared with the Met Office CRM at 200 m resolution. (a) 'Reference' setup, with standard non-local 1D BL scheme and horizontal ∇^4 diffusion compared with (b) 3D Smagorinsky-Lilly with Cs=0.23.

This leaves a slight quandary. As discussed by Petch, 2006, running with large amounts of horizontal diffusion results in fields which may more accurately represent the average over the small clouds which form in practice, but running with no sub-grid model at all produces cumulus-like clouds much sooner (albeit on too gross a scale). Figure 19 shows a sensitivity study with the MetUM, showing the impact of horizontal diffusion on the characteristic scale of precipitating clouds. This may be quantified in terms of number and size of precipitation areas as compared to radar estimates (Figure 20). This illustrates that the horizontal mixing has a systematic impact and may be tuneable for a given model (though there is no guarantee that the tuning will be optimal for every case).

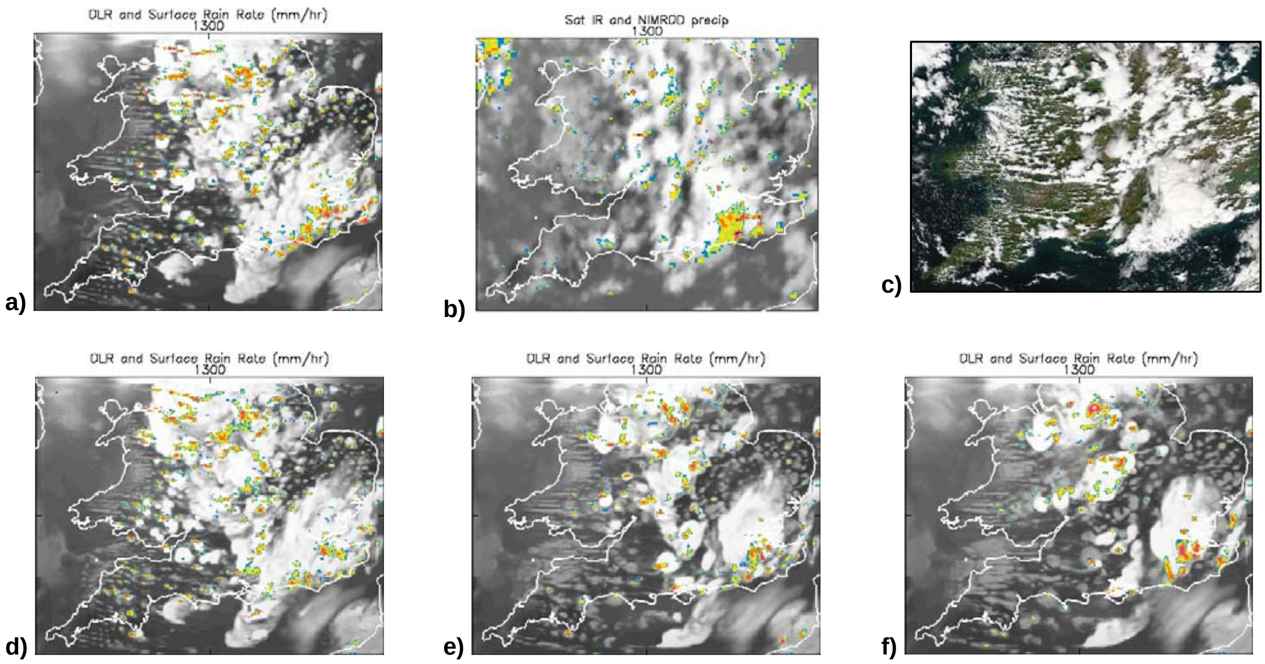

Figure 19 Simulations at 1.5 km horizontal resolution of CSIP IOP 18 (25th August, 2005) with the MetUM ((a),(c),(d),(e)) compared with imagery. Model simulations show a broad band IR radiance temperature (grey scale) and rainfall rates (colour) at 1300 UTC. (b) Meteosat IR radiance (grey scale) and network radar rainfall (colour) at 1300 UTC. (c) MODIS visible at 1325 UTC. (a) Reference run with ∇^4 horizontal diffusion. Bottom row Smagorinsky-Lilly horizontal diffusion with (d) Cs=0.075 (e) Cs=0.1 (f) cs=0.15. All model runs use standard non-local BL scheme.

Problems with transition from convective boundary layer, through shallow to deep moist convection arise from the consequences of the discussion introducing this section; our boundary-layer scheme (and probably our shallow convection scheme) is meant to represent Reynolds averaged flow, while deep convection is LES. The parametrization schemes have been designed on the basis that there is no variability at the interface between resolved and parametrized flow. This situation occurs at two other well-known interfaces; between surface-layer and boundary-layer in LES simulations of the boundary layer and between parametrized convection and resolved flow in large-scale models. In both cases, some success has been achieved by simulating the upscale transport missing from the parametrized flow at the interface using a form of ‘stochastic backscatter’ – see, for example Mason and Thomson, 1992 in the former case and Shutts, 2005 in the latter case (and Shutts’ contribution to this Seminar series). Weinbrecht and Mason, 2008, have applied similar backscatter techniques to the cumulus transition problem with some success, and it seems likely that ultimately a stochastic treatment of shallow cumulus will be required to break the symmetry imposed by the Reynolds averaged scheme.

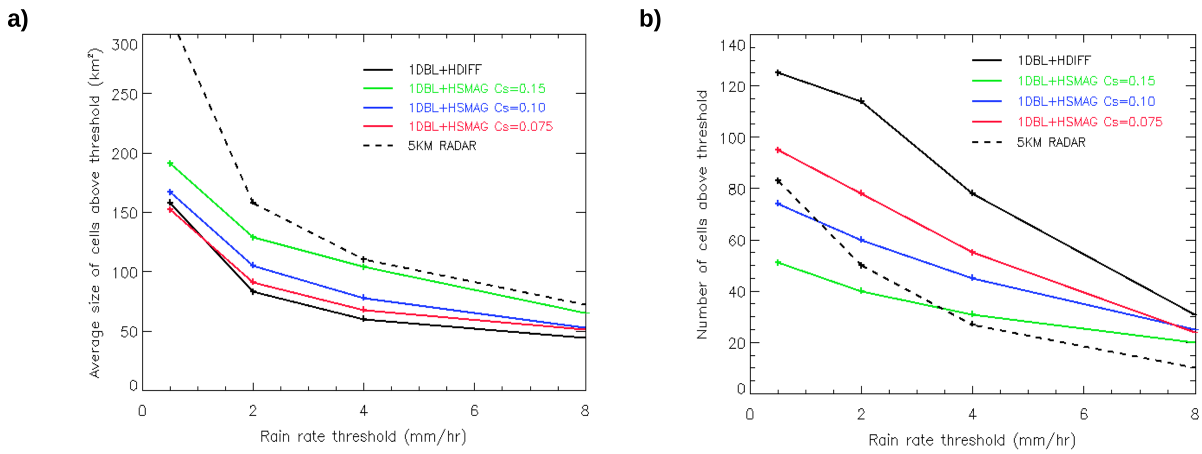

Figure 20 Precipitation cell statistics for 1.5 km MetUM simulations of CSIP IOP 18. Model precipitation has been averaged to a 5 km grid and compared with 5km resolution radar estimates. (a) Average convective cell size as a function of rainrate threshold. (b) Average number of convective cells as a function of rainrate threshold.

7. Concluding remarks: priorities for DCPMs

Above we have discussed at least some of the issues facing the 'convective-scale' modeller, and some of the enhancements required to physical parametrizations. While there are undoubtedly many improvements that can and need to be made, it is perhaps worth highlighting a few issues:

1) Surface schemes can undoubtedly be improved and we probably know how to make some basic technical improvements. In practice, the main limitation may be the availability of input data describing surface and sub-surface characteristics with sufficient accuracy and detail to make improvements worthwhile.

2) What are the prospects of affordable 3D radiation schemes? It is likely that we can continue to 'get away' with using 1D schemes, but some adjustments to the treatment of direct incoming solar radiation for cloud shadowing may be worth pursuing.

3) There remains plenty of scope to improve microphysics, especially interaction with aerosols and ice nucleation but proving the robust value of improvements will be hard. New verification techniques show promise in helping with this.

4) Treatment of turbulence is crucially important for controlling the initiation and nature of convective clouds. Currently this is something of a black art and should be a focus of research.

8. Acknowledgements

Many of the examples using the MetUM in this paper were generated using versions of the model developed within the JCMM as part of the Met Office Public Weather Service Research Programme. The author would like to acknowledge the work of all who contributed, and especially Richard Forbes, Humphrey Lean and Nigel Roberts who contributed many of the figures. Thanks also to Prof. Brian Golding for his detailed review of the paper.

9. References

Barlow, J.F. and Belcher, S.E., 2002, A windtunnel model for quantifying fluxes in the urban boundary layer, *Boundary-Layer Meteorology* **104**, 131-150

Best, M.J., 2005, Representing urban areas within operational numerical weather prediction models. *Boundary-Layer Meteorol.*, **114**, 91-109

Best, M.J., Grimmond, C.S.B., Villani, Maria Gabrialla, 2006, *Boundary-Layer Meteorol*., **118**, 503-525

Blackadar, A.K., 1979, High resolution models of the planetary boundary layer. In: Advances in Environmental Science and Engineering, Editors J. Pfaflin and E. Ziegler, 1, N 1, Gordon and Breach.

Bryan, B.H., Wyngaard, J.C. and Fritsch, J.M., 2003, Resolution Requirements for the Simulation of Deep Moist Convection, *Mon. Weather. Rev.*, **131**, 2394-2416.

Claussen, M., 1990, Area-averaging of surface fluxes in a neutrally stratified, horizontally inhomogeneous atmospheric boundary layer. *Atmos. Environ.*, **24A**, 1349-1360.

Claussen, M., 1995, Flux aggregation at large scales: on the limits of validity of the concept of blending height, *Journal of Hydrology*, **166**, 371-382.

Cohard, J.M. and Pinty,J.C., 2000, A comprehensive two-moment warm microphysical bulk scheme. I : Description and tests, *Quart. J. Roy. Meteorol. Soc*., **126**, 1815-1842

Cuxart', J., Bougeault, P. and Redelsperger, J.L., 2000, A turbulence scheme allowing for mesoscale and large-eddy simulations, *Quart. J. R. Meteorol. Soc*., **126**, 1-30

De Roode, S.R., Duynkerke, P.G. and Jonker, H.J.J., 2004, Large-Eddy Simulation: How Large is Large Enough? *J. Atmos. Sci*., **61**, 403-421.

Diamantakis, M., Wood, N., Davies, T., 2007, An improved implicit predictor-corrector scheme for boundary layer vertical diffusion, *Quart. J. R. Met. Soc.*, **132** , 959 - 978

Ebert, E.E. – Verification web page: http://www.bom.gov.au/bmrc/wefor/staff/eee/verif/verif_web_page.html

Essery, R. , Best, M. an Cox, P. 2001, MOSES 2.2 Technical Documentation, Hadley Centre Technical Report No. 30, Met Office Hadley Centre

Essery, R. L. H. , Best, M. J., Betts, R. A., Cox, P. M., 2003, Explicit Representation of Subgrid Heterogeneity in a GCM Land Surface Scheme, *Journal of Hydrometeorology*, **4**, 530-543

Gerard, L. and Geleyn, J.-F., 2007, Evolution of a subgrid deep convection parametrization in a limited-area model with increasing resolution, *Quart. J. R. Met. Soc.*, **131**, 2293-2312.

Grabowski, W. W., P. Bechtold, A. Cheng, R. Forbes, C. Halliwell, M. Khairoutdinov, S. Lang, T. Nasuno, J. Petch, W.-K. Tao, R. Wong, X. Wu, K.-M. Xu, 2006: Daytime convective development over land: A model intercomparison project based on LBA observations. *Quart. J. Roy. Meteorol. Soc*., **132**, 317-344.

Grimmond, C.S.B. & Oke, T.R., 1999, Aerodynamic properties of urban areas derived from analysis of surface form, *J. Applied Meteorol.*, **38**, 1262-1292.

(See also Grimmond, C.S.B. & Oke, T.R., Corrigendum, 2000, *J. Applied Meteorol.*, **39**, 2494.)

Golding, B.W., Clark, P.A. and May, B., 2005, The Boscastle Flood: Meteorological Analysis of the Conditions Leading to Flooding on 16 August 2004, *Weather*, **60**, 230-235.

Lean, H.W., Clark, P.A., Dixon, M., Fitch, A., Forbes, R.M., Halliwell, C. and Roberts, N.M, Characteristics of High Resolution NWP Models for Forecasting Convection over the UK, 2008, Mon. Wea. Rev., (In Press).

Harman, I.N., 2003, The energy balance of urban areas, Reading University PhD thesis.

Lilly, D.K., 1967, The representation of small-scale turbulence in numerical simulation experiments. Pp. 195-210 in *Proceedings of the Tenth IBM scientific computing symposium on environmental sciences*. Thomas J. Watson Research Center, Yorktown Heights

Lock, A.P., Brown, A.R., Bush, M.R., Martin, G.M. and Smith, R.N.B. 2000 A New Boundary Layer Mixing Scheme. Part 1: Scheme Description and Single-Column Model Tests. *Mon. Wea. Rev.,* **128** 3187-3199.

Lorenz, E.N., 1969: The predictability of a flow which possesses many scales of motion. Tellus, **21**, 289-307

Macdonald, R.W., Griffiths, R.F. & Hall, D.J., 1998, An improved method for the estimation of surface roughness of obstacle arrays, *Atmos. Environment*, **32**, 1857-1864.

Mason, P.J. and Thomson, D.J., 1992, Stochastic backscatter in large-eddy simulations of boundary layers. *J. Fluid Mech.*, **242**, 51-78

Marsham, J.H., Morcrette, C. J. Browning, K. A., Blyth, A. M., Parker, D. J., Corsmeier, U., Kalthoff, N. and Kohler, M., 2007, Variable cirrus shading during CSIP IOP 5. I: Effects on the initiation of convection, *Q. J. R. Meteorol. Soc.,* **133**: 1643–1660 (2007)

Martilli, Alberto, Clappier, Alain and Rotach, Mathias W., 2002, An Urban Surface Exchange Parameterisation for Mesoscale Models, *Boundary-Layer Meteorol.*, **104**, 261-2002.

Mason, P. J., 1988, The formation of areally-averaged roughness lengths, *Quart. J. Roy. Meteorol. Soc.*, **114**, 399-420.

Mass, C.F., Ovens, D., Westrick, K. and Colle, B.A., 2004, Does Increasing Horizontal Resolution Produce More Skillful Forecasts? *B.A.M.S.,* **83**, 407-430.

Masson, V., 2000, A Physically-Based Scheme For The Urban Energy Budget In Atmospheric models, *Boundary-Layer Meteorol.,* **94**, 357-397.

Mellor and Yamada, T., 1974, A Hierarchy of Turbulence Closure Models for Planetary Boundary Layers, *J. Atmos. Sci*., **31**, 1791-1806.

Möller, M.D. and Scherer, D., 2005, A Grid- and Subgrid-Scale Radiation Parameterization of Topographic Effects for Mesoscale Weather Forecast Models, *Mon. Weather. Rev.*, **133**, 1431-1442.

Oliphant, A.J., Spronken-Smith, R. A., Sturman, A. P. and Owens, I. F., 2003, Spatial Variability of Surface Radiation Fluxes in Mountainous Terrain, *J. Appl. Met.*, **42**, 113–128.

Petch, J. C., 2006, Sensitivity studies of developing convection in a cloud-resolving model, *Quart. J. R. Met. Soc.*, **132**, 345-358

Porson A., Harman I.N., Belcher S.E., 2008, How many facets are needed to represent the energy balance of urban areas? submitted to *Boundary-Layer Meteorol.*.

Raupach, M. R., 1994, Simplified expressions for vegetation roughness length and zero-plane displacement as functions of canopy height and area index, *Boundary-Layer Meteorol.*, **71**, 211-216.

(See also Raupach, M. R., 1995, Corrigenda, *Boundary-Layer Meteorol.*, **76**, 303-304.)

Roberts, N.M., 2008, Assessing the spatial and temporal variation in the skill of precipitation forecasts from an NWP model, *Meteorol. Apps.*, **15,** 163-169

Roberts, N.M. and Lean, H., 2008, Scale-selective verification of rainfall accumulations from high-resolution forecasts of convective events. *Mon. Weather. Rev.*, **136,** 78-97

Rutledge, S.A. and Hobbs, P.V., 1983, The Mesoscale and Microscale Structure and Organization of Clouds and Precipitation in Midlatitude Cyclones. VIII: A Model for the "Seeder-Feeder" Process in Warm-Frontal Rainbands, *J. Atmos. Sci*., **40**, 1185–1206.

Shutts, G. 2005, A kinetic energy backscatter algorithm for use in ensemble prediction systems, *Quart. J. R. Met. Soc.*, **131**, 3079-3102.

Skamarock, W.C., 2004, Evaluating Mesoscale NWP Models Using Kinetic Energy Spectra, *Mon. Weather. Rev.*, **132**, 3019-3032

Weinbrecht, S. and Mason, P.J., 2008, Stochastic Backscatter for Cloud-Resolving Models. Part I: Implementation and Testing in a Dry Convective Boundary Layer, *J. Atmos. Sci*., **65**, 123–139.

Wieringa, J., 1986, Roughness-dependent geographical interpolation of surface wind speed averages, *Quart. J. R. Met. Soc.*, **112**, 867-889

Yee, S.Y.K., 1988: The force-restore method revisited. *Boundary-Layer Meteor.*, **43**, 85-90.