The numerics of physical parametrization

Anton Beljaars, Peter Bechtold, Martin Köhler, Jean-Jacques Morcrette, Adrian Tompkins, Pedro Viterbo and Nils Wedi

ECMWF, Shinfield Park, Reading

ABSTRACT

The numerical aspects of the physical parametrization are discussed mainly in the context of the ECMWF Integrated Forecasting System. Two time integration techniques are discussed. With parallel splitting the tendencies of all the parametrized processes are computed independently of each other. With sequential splitting, tendencies of the explicit processes are computed first and are used as input to the subsequent implicit fast process. It is argued that sequential splitting is better than parallel splitting for problems with multiple time scales, because a balance between processes is obtained during the time integration. It is shown that sequential splitting applied to boundary layer diffusion in the ECMWF model leads to much smaller time truncation errors than does parallel splitting.

The attempt in the ECMWF model to obtain 2nd order accuracy with the so called Semi-Lagrangian Averaging of Physical Parametrizations (SLAVEPP) is explained. The scheme reduces time truncation errors, although a few implementation questions remain. In the scheme fast and slow processes are handled differently and it remains a research topic to find the optimal way of handling convection and clouds.

Process specific numerical issues are discussed using the ECMWF parametrization package as an example. Examples are the non-linear stability problems in the vertical diffusion scheme, the stability related mass flux limit in the convection scheme and the fast processes in the cloud microphysics. Vertical resolution in the land surface scheme is inspired by the requirement to represent diurnal to annual time scales. Finally, a new coupling strategy between atmospheric models and land surface schemes is discussed. It allows for fully implicit coupling also for tiled land surface schemes.

1   Introduction

Sub-grid processes play an important role in numerical weather prediction and climate models and parametrization development has been a major research activity for many years (see ECMWF, 2001). Numerics of parametrization is perhaps less developed as it is neither pure numerics nor parametrization. Traditionally the numerics of parametrization is handled by parametrization experts and they focus on the formulation of the equations. Solving them is often considered to be a secondary issue. Recently, the topic has received more attention, as it is realized that parametrization assumptions can be completely overwhelmed by errors in the numerical approximation. In other words the physics of the numerical solution may be different from the parametrized equations due to numerical errors. In that case the choice of numerical scheme and its optimization has become part of the parametrization assumptions which is undesirable. From the model development point of view, a more attractive approach is to have a numerical scheme that solves the parametrized equations with an accuracy that is better than the uncertainty of the parametrization. In this way, parametrization questions can be separated from numerical issues.

In designing a parametrization code for a large scale model, there are a number of considerations. Parametrization packages have separate modules for different processes. Such modularity is desirable from the model development and code maintenance point of view, but may be in conflict with numerical issues. A highly modular system in which e.g. the radiation, cloud and turbulence schemes are completely independent during

a single time step, is very practical because the two schemes can be modified and improved independently and the code does not need to interact (except through the main model variables). On the other hand, e.g. for stratocumulus clouds, radiation and turbulent diffusion have to be in equilibrium and with long time steps it may be necessary to enforce a balance between processes during the time integration. Therefore, from the parametrization and modularity point of view, explicit schemes are the preferred option. However, fast processes with stiff equations, in which fast and slow time scales co-exist, require implicit solutions for stability and accuracy.

While the use of sufficiently small time steps may be acceptable in the context of idealized process studies, comparably long time steps are essential in operational large scale production models for efficiency. It is obvious that the numerics of the parametrized processes has to be compatible with the dynamics of the model. The ECMWF model uses a 2 time level scheme for time integration and a non-staggered grid in the vertical. Time steps are used ranging from 15 minutes for the deterministic T511 forecasts ($\approx 40\ km$ in grid point space) to 60 minutes for the seasonal forecasts at T95 ($\approx 200\ km$). The accuracy of the numerical approximation of the parametrized equations is often ignored and parametrizations are sometimes optimized for a given resolution and time step. This is also undesirable in an operational context, because it will result in a suboptimal model for other time steps and resolutions. Vertical resolution is often not sufficient to resolve the relevant physical processes. An example is the long wave radiative flux divergence near the top of a thick cloud. In reality this flux divergence is concentrated in a layer of the order of 10 m, whereas models with good vertical resolution have 100 m at best.

It is by no means simple to find an optimal solution with conflicting requirements. In this paper a number of numerical aspects related to parametrization are presented. Time stepping is a major topic and will be discussed in the following section. In the later sections process specific issues will be discussed using the ECMWF model as an example. It is argued that it is important to understand the nature of the physical process in order to make a proper numerical approximation.

2   Time stepping

Time stepping is a rather critical issue, because the computational cost of parametrized physics tends to be high ($\approx 50\%$ of the overall computational effort) and time steps have to be long for economy. In order to be suitable for long time steps, the time stepping scheme has to fulfill a number of requirements. First of all stability is an obvious and absolute requirement. The most common strategy is to use explicit time integration for the slow processes (e.g. radiation) and implicit schemes for rapid processes if necessary for stability (e.g. vertical diffusion). As will be demonstrated, for long time steps (compared to the time scale of the involved processes) it is important to assure a good balance between these processes. It requires coupling between processes and can be in conflict with the wish to keep the model code modular. However, for code maintenance and to facilitate parametrization development it is highly desirable to keep the code as modular as possible. Therefore traditionally, the parametrization in most cases is split from the dynamical part of the model and parametrized tendencies from different processes are computed independently and then added together to obtain a first order increment from all subgrid processes. However, this is not always possible as processes may require interaction (e.g. convection and clouds; Tiedtke 1993) and it may be necessary to enforce a balance (e.g. between dynamics and boundary layer diffusion; Beljaars 1991).

Accuracy of the numerics is of course important although it is difficult to specify explicitly what level of accuracy is needed. It does not make sense to insist on an accuracy that is a lot better than the uncertainty in the parametrization if such an accuracy comes at a high price. On the other hand it is undesirable to have numerical errors resulting from individual schemes or splitting errors that overwhelm the effect of parametrization. Although often difficult to quantify, from the parametrization development point of view, it is best to have a

a

b
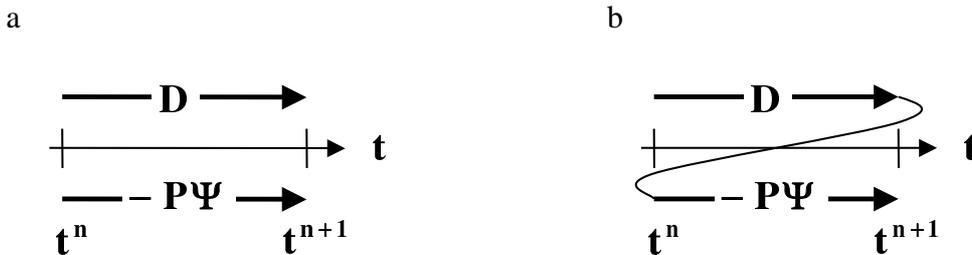
Figure 1: Parallel time splitting (a) and sequential time splitting (b). D represents the dynamics tendency and $-P\Psi$ represents the physics tendency.

time stepping and an overall coupling to the dynamical part of the model that is consistent with other right hand side terms and that does not "change" the parametrization significantly. However, unlike time integration in the dynamics, most time stepping procedures for parametrized processes are first order only. Recently, steps have been made towards a more consistent incorporation of the physical parameterizations to the dynamical part of the model and towards 2nd order accuracy (e.g. Wedi 1999; Cullen and Salmond 2002). A major issue is how to combine the processes and how to couple them to the dynamics. The accurate time integration of multiple time scale systems is a continuing challenge in computational physics (Knoll et al. 2003). A prototype of a stiff system is given by reaction-diffusion equations, the numerical accuracy of which has been assessed in Sportisse (2000) and more recently in Ropp et al. (2004), comparing the accuracy of different implicit or operator-split methods. The problem of multiple time scales is inherent in meteorology and has been investigated by Beljaars (1991), Browning and Kreiss (1994), Caya et al. (1998), McDonald (1998), Wedi (1999), Williamson (2002) and Dubal et al. (2004). In Knoll et al. (2003) it is demonstrated that splitting can result in accuracy degradation when a computational time step is larger than the competing (fast) time scales employed. Therefore, implicitly balanced methods, e.g. preconditioned Jacobian-free Newton-Krylov methods (e.g. Smolarkiewicz and Margolin 1994) are advocated for multiple time scale problems. While promising, this approach has been avoided so far because of its mathematical and algorithmic complexity and in the past the lack of efficient implicit solvers (Knoll et al. 2003). Also because of the desired modularity of physics packages, splitting has been the method of choice.

Here we will discuss two basic splitting techniques: parallel splitting and sequential splitting. Splitting is a standard technique in large scale models as it allows a modular code design (e.g. Caya et al. 1998; Williamson 2002). Here, a simple example for multiple time scales in meterology (Browning and Kreiss 1994; Dubal et al. 2004) is used, where a rate equation is given for Ψ with a dynamics tendency D and a physics tendency $-P\Psi$.

$$\frac{d\Psi}{dt} = D(\Psi) - P(\Psi)\Psi \tag{1}$$

The idea is to illustrate the situation of a slow dynamic forcing with a fast responding physics term. The steady state solution is $\Psi = D/P$. Ideally this should be the solution for time steps that are long compared to the time scale of the fast process.

2.1 Parallel splitting

Parallel splitting, also called process splitting, is the technique where all processes are integrated separately forward in time without communication of tendencies between the processes (see figure 1a). The time stepping

of equation (1) is represented by a tendency from the dynamics and a tendency from the physics:

$$\frac{\Psi_D^{n+1} - \Psi^n}{\Delta t} \quad = \quad D(\Psi^n) \,, \tag{2}$$

$$\frac{\Psi_P^{n+1} - \Psi^n}{\Delta t} \quad = \quad -P(\Psi^n)\Psi_P^{n+1} \quad \Rightarrow \quad \frac{\Psi_P^{n+1} - \Psi^n}{\Delta t} = \frac{-P(\Psi^n)\Psi^n}{1 + \Delta t P(\Psi^n)} \,. \tag{3}$$

The total tendency is:

$$\frac{\Psi^{n+1} - \Psi^n}{\Delta t} = \frac{\Psi_D^{n+1} - \Psi^n}{\Delta t} + \frac{\Psi_P^{n+1} - \Psi^n}{\Delta t} = D(\Psi^n) - \frac{P(\Psi^n)\Psi^n}{1 + \Delta t P(\Psi^n)} \,, \tag{4}$$

with a steady state solution:

$$\Psi^n = \Psi^{n+1} = \frac{D(\Psi^n)[1 + \Delta t P(\Psi^n)]}{P(\Psi^n)} \,. \tag{5}$$

This steady state solution is only independent of the time step if $\Delta t P(\Psi^n) << 1$, in other words if the time step is short compared to the time scale of the physical process. However, there is a clear advantage with parallel splitting because it allows for a modular code design with minimum interaction between the code for different processes.

2.2 Sequential splitting

Sequential splitting, also called time splitting or fractional stepping, is the method where processes are integrated one after the other, but the tendencies from the previous processes are used in the next process (see Fig. 1b). For the simple equation (1), sequential splitting leads to the following finite difference equations:

$$\frac{\Psi^* - \Psi^n}{\Delta t} \quad = \quad D(\Psi^n) \,, \tag{6}$$

$$\frac{\Psi^{n+1} - \Psi^*}{\Delta t} \quad = \quad -P(\Psi^n)\Psi^{n+1} \quad or \quad \frac{\Psi^{n+1} - \Psi^n}{\Delta t} = D(\Psi^n) - P(\Psi^n)\Psi^{n+1} \,. \tag{7}$$

The total tendency is:

$$\frac{\Psi^{n+1} - \Psi^n}{\Delta t} = \frac{D(\Psi^n) - P(\Psi^n)\Psi^n}{1 + \Delta t P(\Psi^n)} \,, \tag{8}$$

with the steady state solution:

$$\Psi^{n+1} = \Psi^n = \frac{D(\Psi^n)}{P(\Psi^n)} \,. \tag{9}$$

This steady state solution is correct and independent of the time step. The reason is that the implicit process is called last and therefore can achieve a balance with the other process because it "knows" about the tendency of the other process.

Equation (7) can be interpreted in two different ways: (i) Variable Ψ^*, updated by the dynamics $D(\Psi^n)$ is used as starting point of the integration with the physics $-P(\Psi^n)\Psi^n$, or (ii) $D(\Psi^n)$ is used as source term in the time integration of the physics. The latter interpretation has the preference, because the first suggests that the physics takes Ψ^* as input, which is not entirely correct. If P at the right hand side of equation (7) is evaluated with Ψ^*, the steady state solution contains Ψ^* rather than Ψ^n, which leads to a time step dependence. In other words, P has to be evaluated at a full time level and not at the "in between" time level between processes. In

a)

Wind speed diff 24-fcts from 20020115; ej4k(m60R1psV1F)-ej4l(m05R1psV1F); Mean=0.76; RMS=1.39

b)

Wind speed diff 24-fcts from 20020115; ej4n(m60R1tsV1F)-ej4m(m05R1tsV1F); Mean=0.11; RMS=0.74

c)

Wind speed diff 24-fcts from 20020115; ej4x(m60R1btsV1F)-ej4y(m05R1btsV1F); Mean=-0.11; RMS=0.76
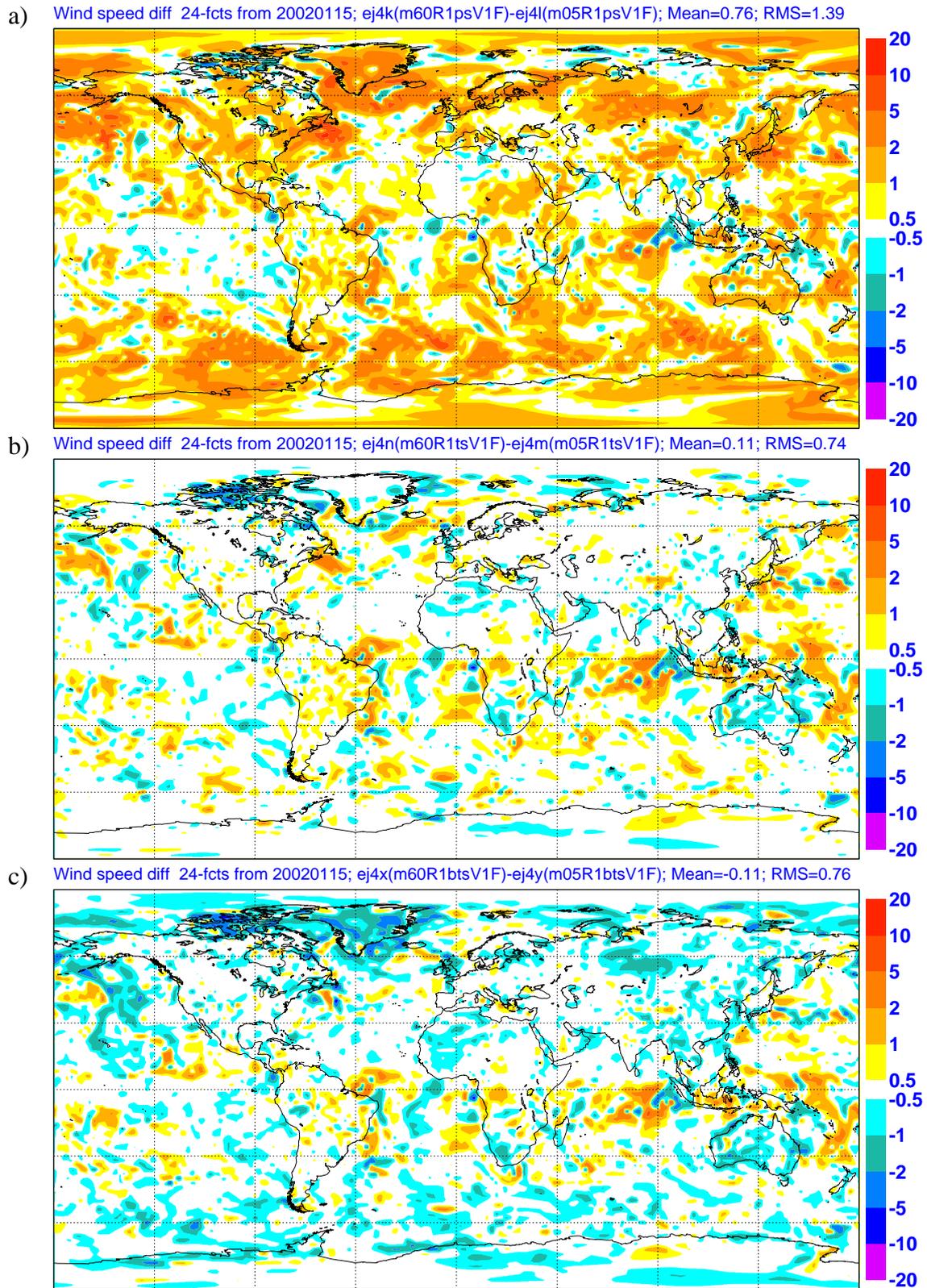
Figure 2: Wind speed (10 m) time truncation errors of a 24 hour forecasts with a 60 minute time step with different time stepping procedures for the vertical diffusion scheme: (a) parallel splitting (b) sequential splitting and (c) sequential splitting but the diffusion coefficients are computed after the dynamics increments have been added. An integration with a 5 minute time step is used as a reference (differences between schemes with 5 minute time steps are small).

117

the case of vertical diffusion $P(\Psi^n)$ represents the computation of the diffusion coefficients and they need to be evaluated with full time level profiles.

An interesting and well know example in meteorology is the process of condensation. Supersaturation with respect to water does not exist in the atmosphere because condensation is so fast that any excess water vapor is instantly converted into water. Therefore condensation is seldom included as a short time scale process; instead an iterative procedure is applied to remove the supersaturation. In fact, this is compatible with sequential splitting. Say dynamics lifts a volume of air above saturation, then an adjustment is made to temperature and moisture in order to reach saturation. The latter is implicit because it is an iterative procedure in which the saturation is evaluated at the new time level, but the tendency of the dynamics has been included in order to have the correct balance.

Time truncation errors can be evaluated by using a short time step integration as a reference. Here we consider the 24 hour forecast of wind speed at the lowest model level (10 m) of the ECMWF model with a 60 minute time step and compare with a 5 minute time step integration. At the lowest model level (in this example) the dominant terms are the pressure gradient (dynamics), the Coriolis term (dynamics) and the turbulent stress divergence. The latter is the fast process that is handled by an implicit computation. Three different schemes are used for the vertical diffusion scheme: (i) parallel splitting, (ii) sequential splitting in which the dynamics tendency is used as source term during the implicit integration of turbulent diffusion, and (iii) sequential splitting as in (ii) but the diffusion coefficients are not computed from time level n, but after the profiles have been incremented with the dynamics. It has been verified that the short time step integrations with all schemes are very similar so they can be used as "truth". The time truncation errors of the 3 schemes are illustrated in Fig. 2 for a T159 integration with a 60 minute time step. The errors with parallel splitting (Fig. 2a) are systematically positive and large (0.76 m/s). The reason is that the diffusion scheme decelerates the flow less than it should because through its implicit nature it "sees" a lower velocity at the end of the diffusion time step. Parallel splitting (Fig. 2b) is much better and has mixed errors, because the implicit vertical diffusion computation includes the dynamics tendency and therefore achieves a better balance. Parallel splitting with the diffusion coefficient computed after the dynamics increment has been added, has systematic negative errors although the global mean error is not really smaller than with standard parallel splitting (-0.11 m/s versus 0.11 m/s). The reason for the reduction in wind speed is that the diffusion coefficients and particularly the transfer coefficients at the lowest model level see a higher wind speed due to dynamics increment (dynamics accelerates the flow and diffusion decelerates it) and therefore the vertical diffusion applies more deceleration.

The conclusion to be drawn from these simple examples is that sequential splitting is the preferred option in multiple time scale problems, which is consistent with the more general literature on this topic (e.g. Sportisse 2000; Ropp et al. 2004; Williamson 2002; Dubal et al. 2004). Sequential splitting is effective and straightforward in a problem with a single fast (implicit) process. In problems with more than one fast process it would be necessary to combine all fast processes in a single implicit solver to obtain balance within the time step which is in agreement with the findings of Knoll et al. (2003).

2.3   Towards 2nd order accuracy in the ECMWF model

Most models with parametrized physics do time integration of the physics with 1st order splitting methods e.g. explicit forward or implicit backward, and process by process. It is possible to achieve higher order accuracy (e.g. Wicker and Skamarock 1998; Cullen and Salmond 2002) but it may be difficult to justify the additional computational costs given the uncertainty in the parametrization. Recently a new time stepping scheme has been implemented in the ECMWF model in which an attempt is made to improve the consistency with the dynamical part of the model while enhancing the accuracy of the time stepping, but without increasing the computational burden (Wedi 1999). The method is called Semi-Lagrangian Averaging of Physical Parametrizations (SLAVEPP). To illustrate the basic principles of SLAVEPP, the following simple equation
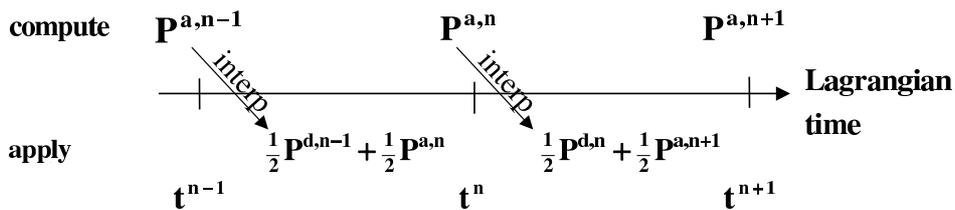
Figure 3: Schematic of Semi-Lagrangian Averaging of Physical Parametrizations (SLAVEPP). P represents the physics tendency.

can be formulated in Lagrangian form

$$\frac{d\Psi}{dt} = D + P,$$ (10)

where $d\Psi/dt$ denotes the total derivative, D denotes the right hand side of the equations attributed to the dynamical part of the model and P is physics. In the physics we distinguish 5 processes, namely *radiation*, *convection*, *clouds*, *vertical diffusion*, and *subgrid orography*. On a discrete mesh, the transported quantity Ψ, which arrives at a gridpoint, and part of D are consistently re-mapped to the departure points of the flow trajectory using a semi-Lagrangian advection algorithm. Here the main focus is now on the evaluation of P which itself depends on Ψ.

The basic idea of SLAVEPP is to approximate P in accord with the dynamical part D, using a second-order trapezoidal rule approximation, where P is evaluated in part at the departure point at time level n and in part at the end of the time step ($n + 1$) at the arrival point, hence averaging the two tendencies "along" the flow trajectory (see Fig. 3). Superscripts d and a are used to indicate departure and arrival points; superscripts n and $n + 1$ refer to the old and the new time level, and in the semi-Lagrangian sense they should correspond to departure and arrival point, respectively. Because the physics tendency at the arrival point can be interpolated to the departure point for the next time step, the physics needs to be evaluated only once per time step, so there is no extra computational cost and only a moderate increase in storage requirements compared to a simple first-order forward time stepping. The time integration can be represented in the following way:

$$\frac{\Psi^{a,n+1} - \Psi^{d,n}}{\Delta t} = D + \tfrac{1}{2}P^{d,n}_{rad+cnv+cld} + \tfrac{1}{2}P^{a,n+1}_{rad+cnv+cld} + P^{a,n+1}_{vdf+sgo},$$ (11)

where the subscripts *rad*, *cnv*, *cld*, *vdf*, and *sgo* refer to the processes radiation, convection, clouds, vertical diffusion and subgrid orography, respectively. Vertical diffusion and subgrid orography are fast processes, so they can not be averaged and are evaluated at the new time level in line with the implicit time integration. Wedi (1999) experimented with averaging of *vdf+sgo* but found large time truncation errors in surface wind over mountainous areas as a result. This suggests that near the surface at locations with high surface stress, the balance between dynamic processes and turbulent friction is more important than the time integration aspect.

The accuracy of the resulting scheme depends critically on the ability to evaluate physics tendencies at time level $n + 1$. Equation (11) suggests that the time integration is implicit in all physical processes which is not feasible with the existing physics codes. Instead, an estimate Ψ^* is made of $\Psi^{a,n+1}$ at the new time level $n + 1$. The estimate is made by using all the tendencies as far as they have been computed already, and to use the previous time level tendencies for the processes that have not yet been computed. One possible scenario would lead to the following sequence of computations (for simplicity the explicit processes *rad*/*cnv*/*cld* and

the implicit processes vdf/sgo have been grouped). First $P_{rad+cnv+cld}^{a,n+1}(\Psi^*)$ is evaluated using

$$\Psi^* = \Psi^{a,n} + \Delta t(D + P_{rad+cnv+cld}^{d,n} + P_{vdf+sgo}^{d,n}) \,, \tag{12}$$

and then equation (11) is applied to compute implicitly the tendencies from vdf+sgo.

For technical and scientific reasons, this is not the way, SLAVEPP could be implemented in the ECMWF model. Firstly, the vertical diffusion (vdf) has to be called before convection (cnv) because surface fluxes are needed by the convection scheme for closure; secondly, it is technically difficult to provide previous time level tendencies to the radiation scheme because it runs on a coarse grid to save computer time, and thirdly the basic scheme reduced convective activity in an unacceptable way. The configuration of SLAVEPP that has been implemented, computes tendencies at time level $n+1$ sequentially for different processes

$$P_{rad}^{a,n+1} = P_{rad}(\Psi^*) \; with \; \Psi^* = \Psi^{a,n} \,, \tag{13}$$

$$P_{vdf+sgo}^{a,n+1} = P_{vdf+sgo}(\Psi^*) \; with \; \Psi^* \; from \; \frac{\Psi^* - \Psi^{a,n}}{\Delta t} = D + P_{rad}^{a,n+1} + P_{vdf+sgo}(\Psi^*) \,, \tag{14}$$

$$P_{cnv+cld}^{a,n+1} = P_{cnv+cld}(\Psi^*) \; with \; \Psi^* = \Psi^{a,n} + \Delta t(D + \tfrac{1}{2}P_{rad}^{d,n} + \tfrac{1}{2}P_{rad}^{a,n+1} + P_{vdf+sgo}^{a,n+1} + \tfrac{1}{2}P_{cnv+cld}^{d,n}) \,. \tag{15}$$

In spite of the compromises in the implementation of SLAVEPP, the scheme is beneficial in the ECMWF model. Wedi (1999) reports a typical reduction of time truncation errors in physics tendencies of up to 25% by comparing long and short time step integrations. This improvement is also seen in the wind speed at the lowest model level (10 m) as shown in Fig. 4. The global mean bias is reduced from 0.17 m/s to 0.10 m/s, and the global RMS error is reduced from 0.74 m/s to 0.65 m/s. As will be seen later, a key appears to be the inclusion of $P_{cnv+cld}^{d,n}$ into Ψ^* which, if it was replaced with $P_{cnv+cld}^{a,n}$, acquires a sense of a second order Runge-Kutta method with respect to the evaluation of the cloud and convection tendencies.

2.4 Towards 2nd order accuracy in the ECMWF model: revised version

Detailed diagnostics in the EPS (Ensemble Prediction System) which runs at T255 with a 45 minute time step showed a reduction of activity in the convection scheme and a deterioration of scores for long time steps. In order to improve, various options in SLAVEPP were explored. As a result, the following configuration of SLAVEPP has been implemented as part of CY28R3:

$$P_{rad}^{a,n+1} = P_{rad}(\Psi^*) \; with \; \Psi^* = \Psi^{a,n} \,, \tag{16}$$

$$P_{vdf+sgo}^{a,n+1} = P_{vdf+sgo}(\Psi^*) \; with \; \Psi^* \; from \; \frac{\Psi^* - \Psi^{a,n}}{\Delta t} = D + P_{rad}^{a,n+1} + P_{vdf+sgo}(\Psi^*) \,, \tag{17}$$

$$P_{cnv}^{a,n+1} = P_{cnv}(\Psi^*) \; with \; \Psi^* = \Psi^{a,n} + \Delta t(D + P_{rad}^{a,n+1} + P_{vdf+sgo}^{a,n+1} + \tfrac{1}{2}P_{cldguess}^{a,n+1}) \,, \tag{18}$$

$$P_{cld}^{a,n+1} = P_{cld}(\Psi^*) \; with \; \Psi^* = \Psi^{a,n} + \Delta t(D + P_{rad}^{a,n+1} + P_{vdf+sgo}^{a,n+1} + P_{cnv}^{a,n+1}) \,, \tag{19}$$

The difference with respect to CY28R1 (compare equations 18-19 to equation 15), is in the guess that is made of Ψ at time level $n+1$ for the different processes. The revised version does not use the previous time level convection tendency, but has half of the guess of the cloud scheme tendency before entering convection and then it uses the full new convection tendency before computing the tendency from the cloud scheme. The consequence is that the cloud scheme has to be called twice, once to provide a guess for convection and once after convection. The full cloud scheme can not be called before convection because the cloud scheme needs detrainment from the convection scheme as source term for the cloud variables (Tiedtke 1993).

Again, the specific configuration of SLAVEPP is inspired by the results. This can be seen in a variety of diagnostics. Fig. 4 shows that the time truncation errors in the surface wind have been reduced (compare Figs. 4b and 4c). This is particularly noticeable in convecting areas in the tropics, which suggests that the convective

a)   Wind speed diff 24-fcts from 20020115; ej5d(m60R1tsV3F)-ej5c(m05R1tsV3F); Mean=0.17; RMS=0.74

b)   Wind speed diff 24-fcts from 20020115; ej5f(m60R1tsV3T)-ej5e(m05R1tsV3T); Mean=0.1; RMS=0.65

c)   Wind speed diff 24-fcts from 20020115; ej5n(m60R3tsV3T)-ej5m(m05R3tsV3T); Mean=0.07; RMS=0.6
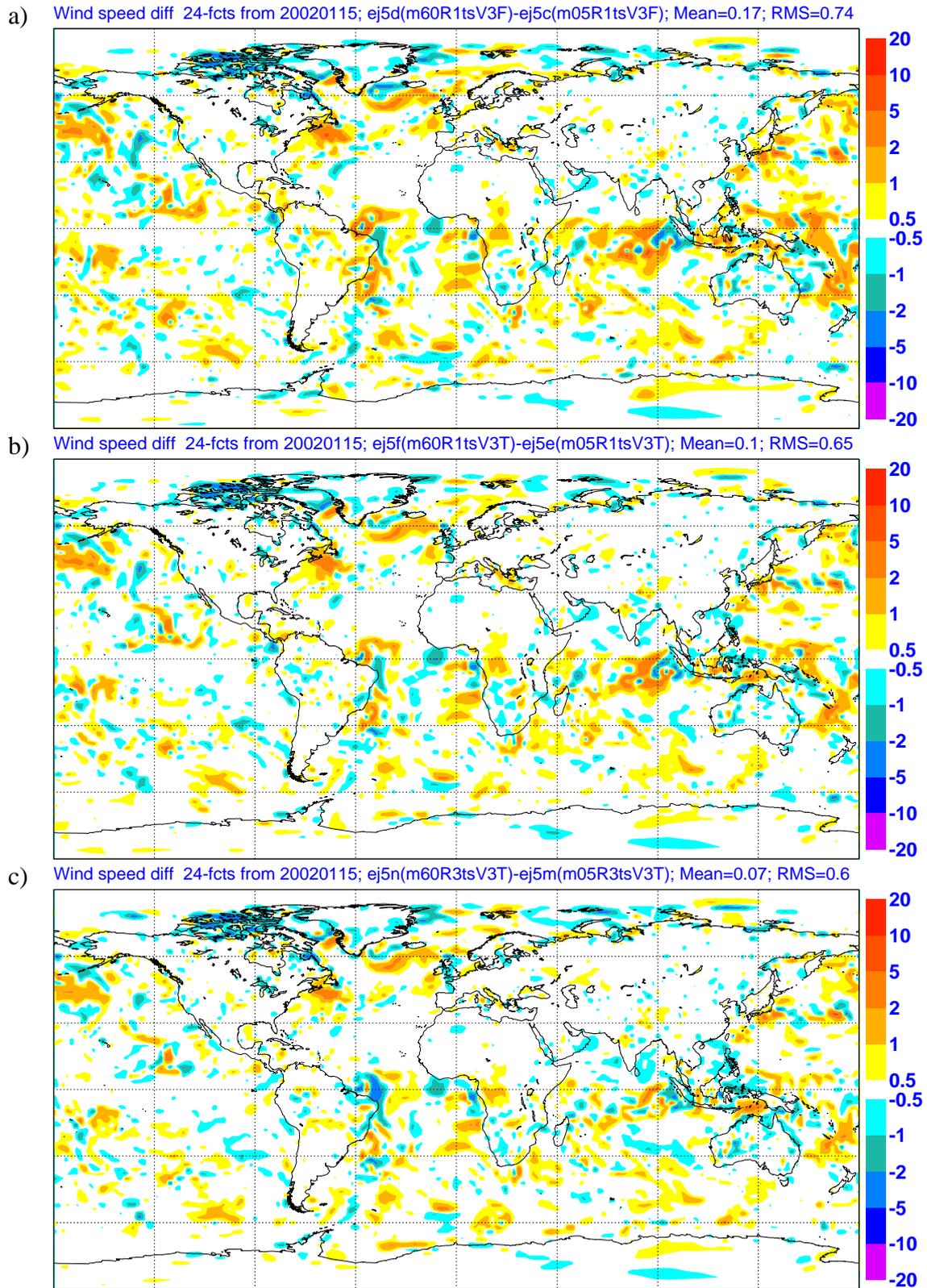
Figure 4: Wind speed (10 m) time truncation errors of 24 hour forecasts with a 60 minute time step with (a) standard time stepping (b) SLAVEPP as in CY28R1 and (c) revised SLAVEPP as in CY28R3 with an additional call to the cloud scheme before convection. An integration with a 5 minute time step is used as the reference.
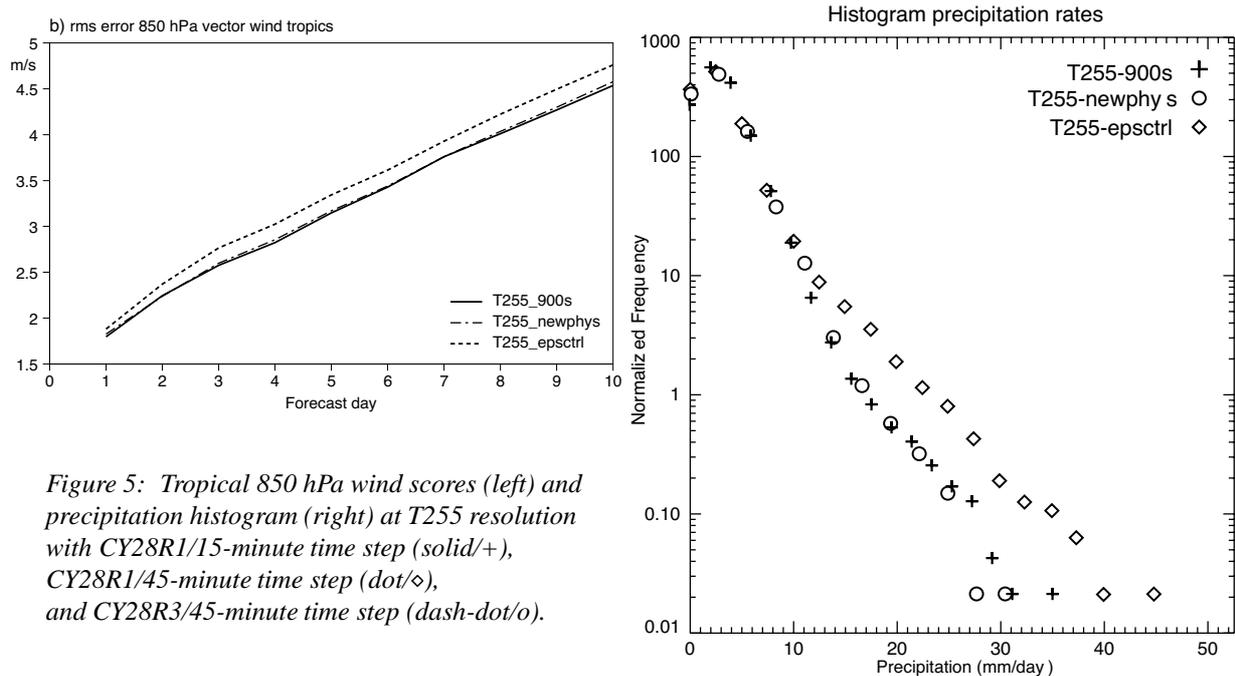
Figure 5: Tropical 850 hPa wind scores (left) and precipitation histogram (right) at T255 resolution with CY28R1/15-minute time step (solid/+), CY28R1/45-minute time step (dot/◇), and CY28R3/45-minute time step (dash-dot/o).

tendencies are handled favorably. It suggests that implicitness is important but that inclusion of tendencies (earlier in time or space) from the same parameterized process that is about to be simulated (as in equation 15) appears less beneficial.

3 Process specific issues

Subgrid processes all have their specific numerical problems. In this section, numerical issues of the various processes will be discussed in the context of the ECMWF model. Although some of the issues are model specific, many numerical problems are rather general.

3.1 Radiation

Radiation is probably the simplest of all processes from the numerics point of view. In the ECMWF model, temperature, moisture, cloud water/ice and cloud cover are the dynamic fields that are used as input; aerosols, absorbing gases etc. are specified as climatological fields or as constants. The process is fairly slow, so explicit time integration is stable. The fluxes F are computed on half levels based on the profiles at time level n and time integration is performed by computing the flux divergence at every model level j (counted from the top of the model)

$$\frac{dT}{dt} = \frac{g}{C_p}\frac{dF}{dp}, \tag{20}$$

$$\left(T_j^{n+1} - T_j^n\right)_{rad} = \Delta t \frac{g}{C_p}\frac{F_{j+1/2}^n - F_{j-1/2}^n}{p_{j+1/2} - p_{j-1/2}}, \tag{21}$$

with C_p for heat capacity at constant pressure, g for the gravitation constant and p for pressure.

However, radiation computations are expensive, so in order to save computer time, the computations are done on a coarse grid and interpolated to the full resolution of the model. Furthermore, the full radiation code is not

122
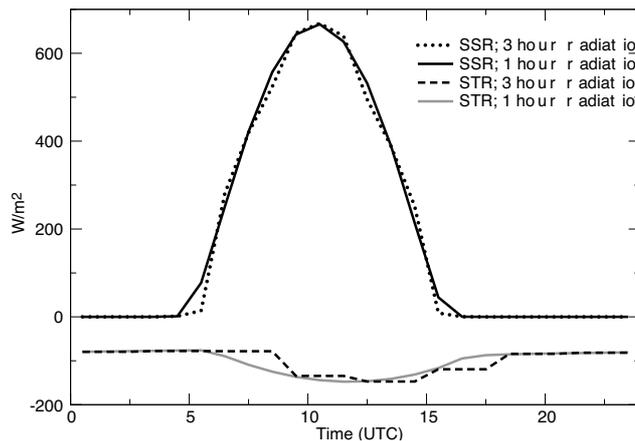
Figure 6: Diurnal cycle of net long wave (lower curves) and net short wave (upper curves) radiation at the surface over Africa (20-30E/10-20S), with 3-hourly (dash) and 1-hourly radiation computations (solid).

called every time step but every 3 hours only up to CY28R2 and every hour with the introduction of CY28R3 on 28 Sep 2004. To obtain tendencies at every time step, the long wave fluxes are kept constant. For the short wave radiation the transmission coefficients are kept from the low frequency full radiation computation, but the solar angle is adjusted every time step to obtain a realistic diurnal cycle. The effect of spatial and temporal sampling on the large scale fields is fairly small (Morcrette 1999) although detailed features are clearly affected. An example is the diurnal cycle of long wave radiation over land which shows a phase shift due to the low frequency of the radiation computation (Fig. 6). A side effect of the spatial sampling is a resolution reduction in surface albedo. The real resolution of the albedo is not the one that is provided on the full model grid but the one that is seen by the radiation code on the coarse grid. This can lead to inconsistencies with the land surface code which needs radiation as input.

With increasing vertical resolution, the numerics of the radiation code is more and more stretched. A clear example is the flux divergence near a cloud edge (e.g. the top of a stratocumulus layer) which is concentrated in a very thin layer. In a model the divergence is spread over at least one layer and the tendency increases with increasing resolution. The typical flux divergence across a stratocumulus top is $80 \ W/m^2$ leading to a tendency of $35 \ K/day$ for a $20 \ hPa$ layer. Doubling the resolution will double the tendency in the top cloud layer. The physics of this problem is that radiation cools the cloud top and that turbulent diffusion redistributes this cooling over the entire cloud depth and often the subcloud layer. With long time steps and future high vertical resolution it will probably be necessary to enforce balance between radiation and turbulent diffusion during the time integration. Using sequential splitting may be sufficient, but it might even be necessary to introduce some implicitness in the radiation code for stability.

3.2 Vertical diffusion

Vertical diffusion describes the effect of vertical transport including the coupling to the surface through turbulence. A popular way of representing turbulent transport is by prescribing eddy diffusion coefficients e.g. as a function of shear and stability (as in the ECMWF model) or with help of a turbulent kinetic energy equation. The equation for model variable Ψ (wind components, dry static energy and specific humidity) reads:

$$\frac{d\Psi}{dt} \quad = \quad g\frac{dF}{dp} \ , \ F_\Psi = K_\Psi \rho \frac{d\Psi}{dz} \ , \tag{22}$$

123

a) RMS(U-tend); 24-hr-av; L60; 20020115; ej4k(m60R1psV1F); Mean=12.68; RMS=17.29

b) RMS(U-tend); 24-hr-av; L60; 20020115; ej4n(m60R1tsV1F); Mean=8.88; RMS=10.76

c)

(90 W, 60 S) T159 forecasts 2002011512, dt=60 min

- parallel split (ej4k)
- sequential split (ej4n)
- bad sequential split (ej4x)
- sequential split, dt=5 min (ej4m)

U (m/s)

Forecast step (hours)
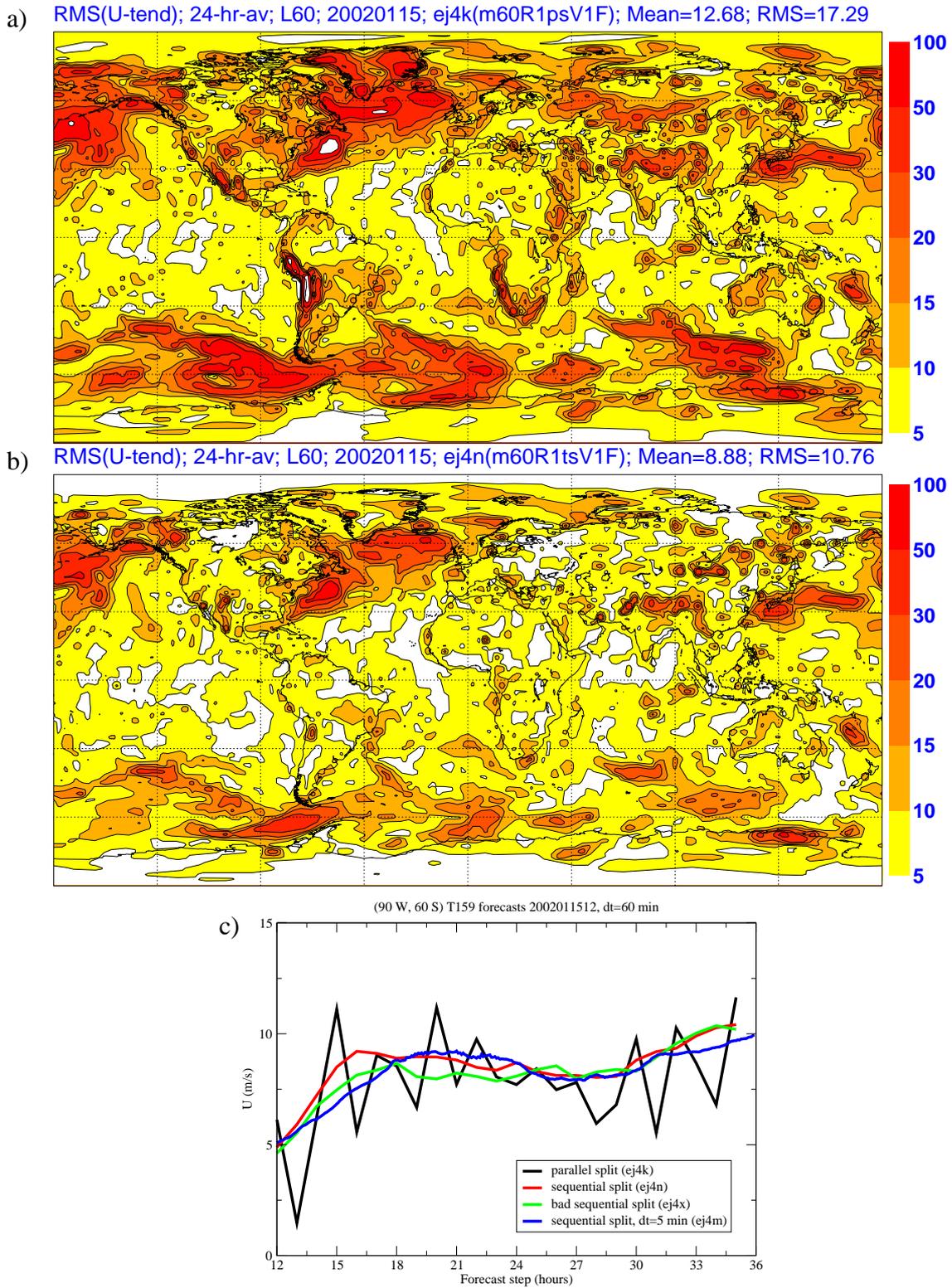
Figure 7: RMS of the zonal wind speed tendency (RMS(dU/dt)) at the 10 m level with a) parallel splitting for vertical diffusion, and b) sequential splitting. The units are $ms^{-1}day^{-1}$. The RMS is obtained by averaging over the first 24 hour of a T159 forecast with a 60 minute time step. Figure c represents time series of zonal wind (every time step) at $60^oS/90^oW$ with parallel splitting (black) sequential splitting (red), sequential splitting with diffusion coefficient after dynamics has been added (green), and sequential splitting with a 5 min. time step (blue).

with K_{Ψ} for the eddy diffusion coefficient which typically depends on Ψ, ρ for density and z for height above the surface. Since the time scale of the diffusion process is small near the surface (of the order of 10 s at a height of 10 m), it is necessary to have implicit time stepping for stability. The discretized equations are implicit in the linear part of the equation and explicit in the diffusion coefficient:

$$\Psi_j^* - \Psi_j^n = \frac{\Delta t g}{\Delta p_j} \left(K_{j+1/2}^n \frac{\hat{\Psi}_{j+1} - \hat{\Psi}_j}{z_{j+1} - z_j} - K_{j-1/2}^n \frac{\hat{\Psi}_j - \hat{\Psi}_{j-1}}{z_j - z_{j-1}} \right) + (\Delta\Psi)_{dyn} + (\Delta\Psi)_{rad} , \qquad (23)$$

$$with \ \hat{\Psi} = \alpha\Psi^* + (1 - \alpha)\Psi^n , \ with \ \alpha = 1.5 . \qquad (24)$$

The dynamics and radiation tendencies are used as source term in the time integration of vertical diffusion. As shown in the sections on time stepping, this is essential to achieve balance between processes. The balance between pressure gradient, Coriolis term and vertical diffusion ensures bias free surface winds for long time steps and the balance between radiative cooling and turbulent transport in cloud layers is essential to have time step independent solutions near cloud tops.

Another major issue in relation to vertical diffusion is numerical stability. Although fully implicit time integration (i.e. $\alpha = 1$ in equation 24), ensures absolute stability in a linear system, non-linear instabilities can still occur in case the diffusion coefficients (which are computed explicitly) depend strongly on the model variables. This is indeed the case in the atmosphere; the diffusion coefficients are a strong function of shear and stability. Particularly the transition from stable to unstable is very pronounced, with diffusion being weak in stable situations and strong in unstable situations. The classic paper by Kalnay and Kanamitsu (1988) discusses the problem extensively using a simple ordinary non-linear differential equation as an example. They analyse stability of a number of schemes e.g. predictor corrector, over-implicit and fully implicit based on linearized diffusion coefficients. The latter is absolutely stable but difficult to implement. Predictor-corrector appears attractive, but does not always give the correct result (see also McDonald 1998). An over-implicit scheme (e.g. with $\alpha = 1.5$) seems to be a good compromise between simplicity and effectiveness. Girard and Delage (1990) use the implicitness factor in a dynamic way by selecting a value that depends on a local stability criterion. More complicated schemes can be beneficial, but usually add to the costs (Hammerstrand 1997).

The problem of non-linear instability does not only manifest inself as fatal blow-ups, but can also be seen sometimes as finite amplitude noise. Different authors report noise in atmospheric models and claim advantages of one scheme over the other (e.g. Beljaars 1991 and Janssen and Doyle 1997). However, noise due to the non-linear vertical diffusion equation is quite common but erratic. Improvements may be seen in one situation whereas more noise may occur in another case.

In order to get a more global perspective in the ECMWF system, the RMS of dU/dt (zonal wind) is computed as a global field over the first 24 hours of a forecast. In Figs. 7ab parallel and sequential splitting are compared and it is clear that parallel splitting is very noisy, particularly in the storm tracks of the Northern and Southern Hemisphere. Fig. 7c clearly shows the finite amplitude $2\Delta t$ noise, which is typical for the non-linear instability. The point where the diffusion coefficient is computed (e.g. after the dynamics tendency has been added) has little impact in contrast with earlier suggestions (Hammerstrand 1997). Whether these results are universal is difficult to say, because the interaction of the non-linear instability with the splitting scheme is not well understood and may depend on various details of numerics and dynamics.

Another numerical issue related to vertical diffusion is vertical resolution. One would expect strong sensitivity to resolution in the stable boundary layer because the stable boundary layer has a lot of structure in a shallow layer (typically 200 m deep). Surprisingly, resolution has little impact on the simulation of the stable boundary. The numerical handling of the surface layer is probably an important aspect, because the finite differencing in the surface layer uses similarity profiles (instead of the linear finite differencing above the surface layer) which is exact in the constant flux layer. The lack of sensitivity of the stable boundary to resolution was found by
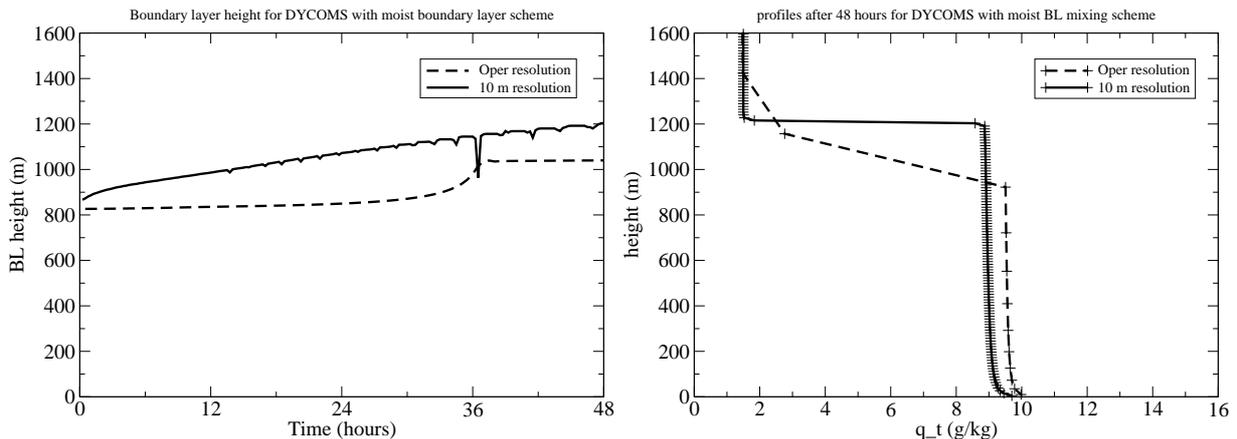
Figure 8: Time evolution of inversion height (left figure) and vertical profile of specific humidity (right figure) from single column simulations with the ECMWF model of a stratocumulus case with operational resolution (typically 200 m at inversion height, dashed) and 10 m resolution (solid).

Beljaars (1991) and confirmed by Cuxart et al. (2004) in an inter-comparison study.

On the other hand, resolution does have impact on inversion structure. Fig. 8 shows a simulation with the ECMWF single column model at operational resolution (typically 200 m at inversion level) and at 10 m resolution. This is a stratocumulus case with a very steep inversion; steps of 10 K in temperature and 10 g/kg in specific humidity are typical. As expected, resolution affects the sharpness of the inversion and the details of the time evolution of its height. As explained in the section on radiation, a good balance between radiative cooling at cloud top and vertical diffusion is necessary. This is achieved by having the radiative tendency at the right hand side of equation (23) as a source term.

Resolution is not the only cause for numerical errors at sharp inversions. With the inversion resolved as a transition between two levels, numerical diffusion can overwhelm the physics in the parametrization as e.g. reported by Lenderink and Holtslag (2000). These inversions tend to be rather stationary and are the result of a balance between downward large scale motion (typical subsidence rates are 0.5 cm/s) and entrainment by turbulence moving the inversion up with the same speed as the large scale subsidence. Since the inversion is strong, entrainment corresponds to a small diffusion coefficient, so numerical diffusion can easily dominate the entrainment. In order to obtain a good balance, Lock (2001) and Grenier and Bretherton (2001) developed special methods to represent the inversion dynamics. Lock (2001) makes an estimate of the numerical errors in vertical advection and adjusts the parametrized entrainment accordingly. Alternatively, Suarez et al (1983) use a mixed-layer approach with the BL top being a coordinate surface. This has the advantage of a simple treatment of the BL top entrainment as well as the interaction of the BL and convection. It has been recently extended to relax the mixed-layer assumption while introducing multiple layers within the BL.

3.3 Subgrid orography

The subgrid orography scheme in the ECMWF model developed by Lott and Miller (1997), consists of two parts: (i) the low level blocking part and (ii) the gravity wave part. The low level blocking part provides strong surface drag with momentum extracted from the low level flow. The gravity wave part also leads to surface drag but the momentum exchange is with higher levels through gravity wave breaking. The surface stress associated with this scheme can be very large in mountainous areas (up to 10 N/m^2). These very high numbers typically occur at isolated points which may lead to unrealistic effects in adjacent points through numerical smoothing of the fields.

Because of the strong tendencies at the lowest model levels, the scheme has to be implicit to ensure numerical stability. Currently, the subgrid orography scheme is called after the vertical diffusion scheme in the ECMWF model, so a proper balance is not obtained during the time stepping. In future, the subgrid orography scheme will be made part of the vertical diffusion scheme in such a way that both processes are solved simultaneously within the same implicit solver.

3.4   Convection

The convection scheme in the ECMWF model uses a bulk mass flux approach (derived from Tiedtke 1989) for up- and down-draughts. Here we consider the updraught only in the convection tendency for variable Ψ (momentum, dry static energy or specific humidity):

$$\frac{d\Psi}{dt} \quad = \quad g\frac{d}{dp}[M_u(\Psi_u - \Psi)] + S \,, \tag{25}$$

where M_u is the updraught mass flux, Ψ_u the updraught property, and S a possible source term for e.g. latent heat release due to condensation. The properties of the updraught Ψ_u are computed by an additional ordinary differential equation. A parcel is followed on its way upwards by considering buoyancy, entrainment, and condensation. Entrainment and detrainment determine the mass flux profile and cloud base mass flux is scaled with the amount of instability for deep convection (CAPE) and with subcloud moist static energy convergence for shallow convection. Moisture convergence is used for the entrainment formulation. These parametrization assumptions obviously require a link with the dynamics for the convergence terms and with the vertical diffusion scheme for the surface fluxes. The latter is the motivation to call the vertical diffusion scheme before the convection scheme. The mass flux closure for deep convection is designed in such a way that it decreases CAPE over a prescribed time scale. The time scale is 1 hour at low resolution and 15 minutes at high resolution. This time scale is of the order of the time step which makes it difficult from the numerical point of view to classify the process as fast or slow. Although equation (25) is valid below cloud base, it is not used there, but instead the fluxes are prescribed by linear interpolation between cloud base and the surface.

With prescribed Ψ_u and M_u profiles, equation (25) is solved with an upwind differencing scheme:

$$(\Psi_j^{n+1} - \Psi_j^n)_{cnv} \quad = \quad \frac{\Delta t g}{\Delta p_j} \left((M_u\Psi_u)_{j+1/2} - (M_u\Psi_u)_{j-1/2} - (M_u)_{j+1/2}\Psi_j + (M_u)_{j-1/2}\Psi_{j-1} \right) + S_j \,. \tag{26}$$

The numerical solution of this advection equation is only stable if the mass flux (equivalent to an advection velocity) is sufficiently small in order not to violate the CFL criterion ($M_u < \Delta p/g\Delta t$). Such a numerical limit alters the parametrization (convection closure) at high vertical resolution and long time steps, which is undesirable. In Fig. 9 zonal mean mass flux profiles are shown for time steps of 15 and 45 minutes. The mass fluxes in the lower troposphere tend to be high due to shallow convection and are limited by the CFL criterion. Since the reduced mass fluxes had a negative impact on the T255 EPS forecasts ($\Delta t = 45$ $min.$), with the introduction of CY26R3 (7 Oct 2003) the mass flux limit has been relaxed by a factor 3 for the temperature and moisture equations of shallow convection. In spite of violating the CFL limit, stability is not compromised in this particular configuration as for these quantities the source term S in (25) appears to dominate.

With increasing vertical resolution and more emphasis on the parametrization of boundary layer clouds it might not be possible in future to have explicit time integration for convection. Attempts have been made at ECMWF to replace the explicit formulation of equation (26) by an implicit formulation. Updraught properties and mass fluxes are still precomputed and prescribed explicitly. Although rather simple in principle, complications do arise from (i) the fact that Ψ_u is itself a non-linear function of Ψ, (ii) from the condensation terms and (iii) from the handling of the subcloud layer. Imposing a linear flux profile in the subcloud layer is not possible any more and it turned out that this assumption was rather crucial to the scheme. On the other hand, an implicit

a) Average of p104/t128 20030115 1200 step 24 Expver ecm6 (180.0W-180.0E)

b) Average of p104/t128 20030115 1200 step 24 Expver ecma (180.0W-180.0E)

c) Average of p104/t128 20030115 1200 step 24 Expver ecqx (180.0W-180.0E)
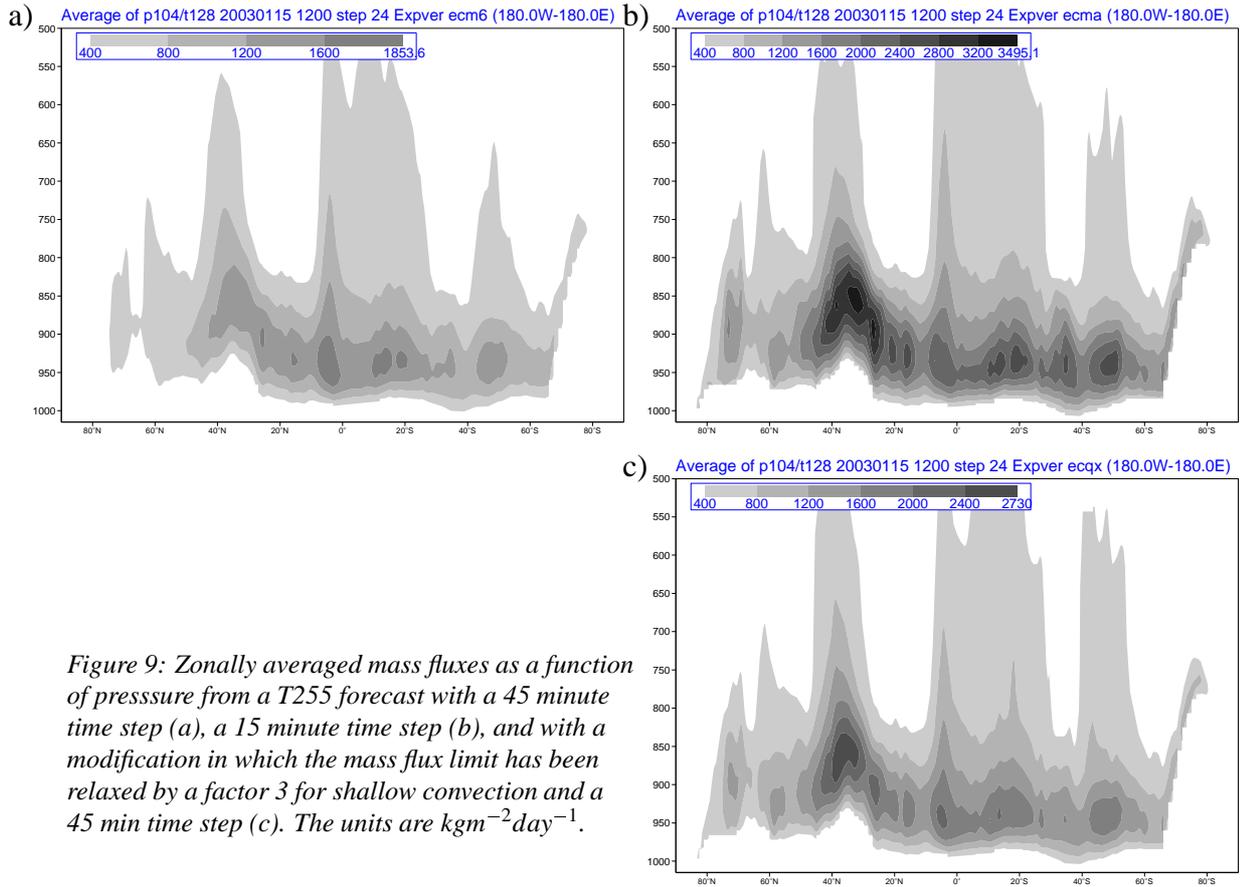
Figure 9: Zonally averaged mass fluxes as a function of presssure from a T255 forecast with a 45 minute time step (a), a 15 minute time step (b), and with a modification in which the mass flux limit has been relaxed by a factor 3 for shallow convection and a 45 min time step (c). The units are $kgm^{-2}day^{-1}$.

mass flux term has been successfully implemented in a new boundary layer scheme for dry boundary layers and stratocumulus mixing processes (Tompkins et al. 2004). This numerical framework may be used in future to include the shallow convection parametrization. Solving for all fast boundary layer processes together in a single implicit solver is obviously an advantage and beneficial to a good balance for long time steps.

3.5 Clouds

The cloud scheme in the ECMWF model has two prognostic variables, namely the combined cloud water/ice variable ℓ and cloud cover a (Tiedtke 1993). The scheme uses source and sink terms from dynamics, radiation, convection, vertical diffusion and adds processes like precipitation and cloud erosion. Cloud processes are notoriously difficult to integrate in time because of the short time scales of some of the processes e.g. the microphysics. To obtain a stable scheme, the prognostic equations are written in the following form

$$\frac{d\ell}{dt} = C - D\ell \,, \tag{27}$$

$$\frac{da}{dt} = A - Ba \,, \tag{28}$$

where the different processes contribute to the coefficients A, B, C, and D. These equation are integrated analytically over a time step which leads to an exponential solution. Tompkins et al. (2004) discuss in more detail the numerics of the cloud scheme and its sensitivity to some upgrades. The ice fallout formulation is particularly sensitive and therefore we will discuss it here. A typical situation is an ice source term from convective detrainment and a sink term due to ice settling. The balance between the two processes controls the
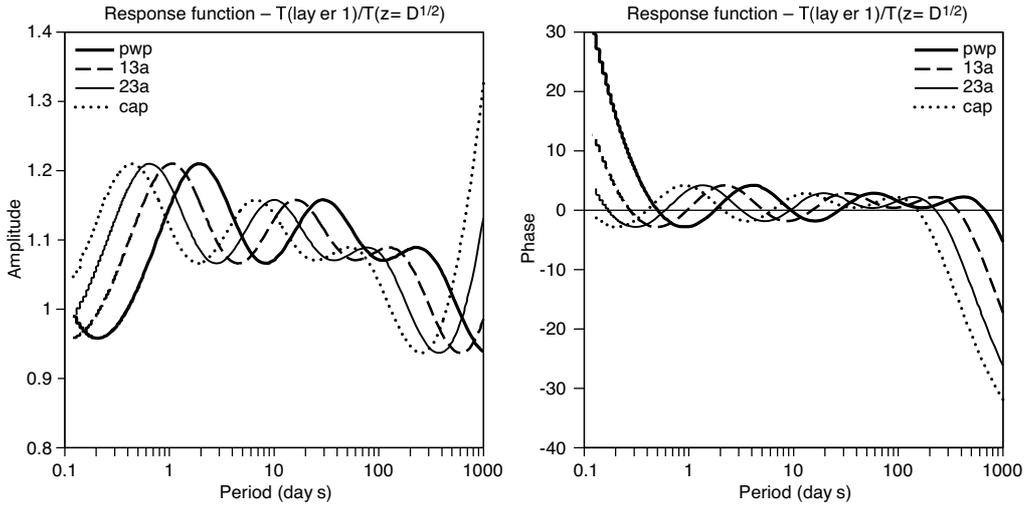
Figure 10: Error in amplitude and phase of the discretized soil temperature equations with layer depths of 7, 21, 72 and 189 cm. Four different constant diffusion coefficients are used corresponding to the soil moisture range from permanent wilting point to field capacity.

ice water content which is relevant for the radiation computation. The rate equation for ℓ in this situation reads

$$\frac{d\ell}{dt} = \frac{1}{\rho}D_{up}\ell_u - \rho g w_{ice}\frac{d\ell}{dp}, \quad w_{ice} = c_1\ell^{c_2}. \tag{29}$$

The first term is the source term from the detrainment of the convection updraught with ℓ_u for updraught ice water, and D_{up} the mass detrainment rate. The second term is the ice settling term, with the ice vertical velocity w_{ice} parametrized as a non-linear function of ℓ. The coefficients C and D for level j are defined as follows

$$C = \frac{1}{\rho}D_{up,j}(\ell_{u,j} - \ell_j) - \rho_{j-1}g w_{ice,j-1}\frac{\ell_{j-1}}{\Delta p_{j-1}}, \tag{30}$$

$$D = -\rho_j g w_{ice,j}\frac{1}{\Delta p_j}. \tag{31}$$

This numerical formulation is by no means obvious and is inspired by practical considerations and the requirement of conservation. The convection detrainment term has been evaluated already in the convection scheme so it is taken as an explicit source. The ice settling velocity is a non-linear function of ℓ and is evaluated from ℓ at the old time level. The ice settling term is evaluated from top to bottom, i.e. what is falling out of a layer is computed implicitly and this amount is used as an explicit source term in the layer below. This has the advantage that conservation is always guaranteed. To obtain a proper equilibrium between the convective source term, the ice falling into the layer from above and the ice settling towards lower layers, it is essential to have all the source terms in the implicit calculation. Whether the current layer by layer formulation is optimal is subject of further research.

3.6 Land surface scheme

Land surface schemes are used to provide a boundary condition for temperature and moisture over land. In contrast to the ocean where the sea surface temperature is quasi-constant (i.e. it varies on a much longer time scale than the boundary layer processes), over land the boundary condition is a combination between fluxes

and temperature. During day time the boundary condition even has the character of a flux condition controlled by radiation and Bowen ratio, the latter depending on soil moisture.

Typically land surface schemes, calculate the energy and water balance in the soil by solving for diffusion type equations for temperature and soil moisture. These equations are non-linear because the diffusion coefficients and hydraulic properties depend on soil moisture. An example is the TESSEL scheme (Tiled ECMWF Scheme for Surface Exchanges over Land; Van den Hurk et al. 2000), which has 4 soil layers, tiles for a simple description of surface heterogeneity, a snow module, an interception reservoir for fast evaporation of water on top of the vegetation and bare soil and a skin layer without heat capacity for instantaneous temperature response to the forcing by radiation and other components of the surface energy balance.

For the numerical aspects it is important to realize that soil and vegetation properties are highly uncertain and variable even within a grid box. It is therefore difficult to justify very accurate and potentially expensive numerical methods. The emphasis is on stability, conservation of water and energy, a representation of time scales from hours to a year, and on a crude representation of the soil profiles. Therefore the vertical structure is simulated with a few layers only. Fig. (10) illustrates the phase and amplitude errors due to the finite vertical resolution in the TESSEL scheme with 4 layers (7, 21, 72 and 189 *cm*) as a function of frequency for sinusoidal forcing (Viterbo 1996). The layer depths have been optimized to obtain flat response functions within the constraint of a very limited number of layers.. Further features of the TESSEL numerics are the implicit solver for the diffusion equations for soil moisture and temperature and an implicit coupling of the skin temperature with the vertical diffusion tridiagonal solver.

3.7 Atmosphere land coupling

Coupling between the vertical diffusion and land surface schemes deserves particular attention, because processes in the boundary layer as well as near the surface in the land surface scheme have short times scales and therefore require implicit numerics. In the TESSEL scheme, the surface skin temperature has fully implicit coupling with the vertical diffusion scheme. This is achieved by solving for the surface energy balance between the downward elimination sweep of the vertical diffusion tri-diagonal matrix and the upward back substitution. The tile scheme raises additional difficulties which are solved in TESSEL by computing a skin temperature for each tile from its surface energy balance as if the tile occupies the entire grid box. Then, the skin temperatures are kept fixed and tile averaged fluxes are computed. This procedure is only implicit for the dominant tiles and not for the tiles that occupy small fractions.

More recently, a new coupler has been introduced as proposed by Best et al. (2004). It has the advantage of being fully implicit on all tiles and it allows for a "universal" way of coupling land surface schemes with atmospheric models. The latter aspect is important, because it provides a framework in which land surface schemes can be exchanged between models without too much difficulty. Because of its practical importance, we give a brief description here (see Best et al. 2004 for more details).

We start with the description of the surface energy balance for each tile. The turbulent fluxes are expressed in terms of differences between the lowest model level dry static energy $\hat{S}_l$ ($S = C_p T + gz$) and specific humidity $\hat{q}_l$ and their corresponding skin values

$$H = \rho C_H^n |\vec{U}_l^n|(\hat{S}_l - \hat{S}_{sk}) , \tag{32}$$

$$E = \rho C_Q^n |\vec{U}_l^n|\beta_s(\hat{q}_l - \alpha_s \hat{q}_{sat}(\hat{T}_{sk}) , \tag{33}$$

where C_H^n and C_Q^n are the transfer coefficients for heat and moisture, $|\vec{U}_l^n|$ is the absolute wind speed at the lowest model level and α_s, β_s are the coefficients that control the moisture availability at the surface through parametric relations with e.g. soil moisture. Note that the implicit variables of the vertical diffusion scheme
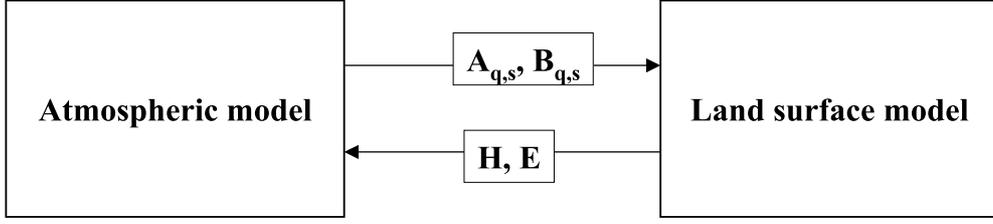
Figure 11: Exchange of parameters between atmospheric models and land surface schemes for universal coupling (Best et al. 2004).

have been used . They can be fully implicit or extrapolated in time (over-implicit) dependent on coefficient α (see equation 24). If the variables are extrapolated in time, the consequence is that the surface skin temperature will also be extrapolated in time. The surface energy balance reads:

$$H + LE + R_{SW} + R_{LW} = \Lambda(\hat{T}_{sk} - T_s^n) . \tag{34}$$

Parameter L is the latent heat of evaporation, R_{SW} and R_{LW} are the net short wave and long wave radiative fluxes, and the right hand side of equation (34) represents the ground heat flux through the skin layer conductivity Λ and the difference of the skin temperature with the first soil layer T_s^n (taken from the old time level). Although this formulation looks rather specific for TESSEL, it can be generalized quite easily by e.g. including an inertia term dT_{sk}/dt.

The next step is to linearize $q_{sat}(\hat{T}_{sk})$ with respect to the previous time level skin temperature T_{sk}^n and to eliminate $\hat{T}_{sk}$ from equations (32) and (33) using (34). This results in two equations expressing the relation between turbulent fluxes and lowest model level variables in the following form:

$$H = D_{H1}\hat{S}_l + D_{H2}\hat{q}_l + D_{H3} , \tag{35}$$
$$E = D_{E1}\hat{S}_l + D_{E2}\hat{q}_l + D_{E3} , \tag{36}$$

with coefficients D that can be computed explicitly from the old time level variables. The averaging over tiles reduces to an averaging over coefficients because the lowest model level variables are assumed to be independent of the tiles. The grid box averaged fluxes read

$$\bar{H} = \hat{S}_l \sum_i v^i D_{H1}^i + \hat{q}_l \sum_i v^i D_{H2}^i + \sum_i v^i D_{H3} , \tag{37}$$
$$\bar{E} = \hat{S}_l \sum_i v^i D_{E1}^i + \hat{q}_l \sum_i v^i D_{E2}^i + \sum_i v^i D_{E3} , \tag{38}$$

with superscript i indicating the tile index and v^i representing the tile fraction assuming that $\Sigma_i v^i = 1$. Equations (37) and (38) can be combined with the result of the downward elimination sweep of the vertical diffusion tridiagonal matrix which is also a relation between lowest model level variables and surface fluxes:

$$\hat{S}_l = A_s \bar{H} + B_s , \tag{39}$$
$$\hat{q}_l = A_q \bar{E} + B_q . \tag{40}$$

Equations (37-38) and (39-40) can now be solved for $\bar{H}$, $\bar{E}$, $\hat{S}_l$, and $\hat{q}_l$, after which the solution of the tridiagonal matrix can be completed by an upward back substitution. In this procedure the vertical diffusion solver straddles the surface scheme.

In spite of the complicated coupling, the procedure outlined above, leads to a clean coupling strategy between atmospheric and land surface models. As suggested by Best et al. (2004), the atmospheric model provides
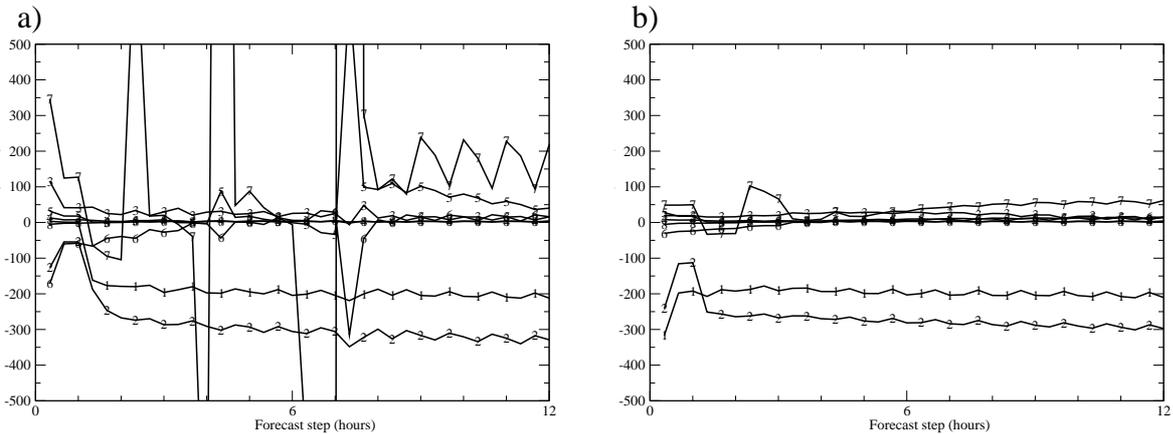
Figure 12: Time series of sensible heat flux (Wm^{-2}) for the 8 TESSEL tiles from a 12 hour forecast in NE Siberia with the old coupling (a) and the new fully implicit coupling (b). The fractional cover of the tiles for this point is: 1. water:0.0; 2. ice:0.0; 3. wet interception:0.53; 4. low vegetation:0.04; 5. exposed snow:0.00; 6. high vegetation:0.37; 7. snow under high vegetation:0.00; 8. bare oil:0.06.

the *A* and *B* coefficients from (39-40) to the land surface model and the land surface model returns fluxes (see Fig. 11). In this coupling, the atmosphere and land are separated at the lowest model level and the atmospheric surface layer is considered to be part of the land surface scheme. Also the complications of surface heterogeneity as reflected by the tile structure are only seen by the land surface model. The implicit tile coupling as described above has been implemented in the ECMWF model with model cycle 28R3 (introduced 28 Sep 2004). The scheme reduces noise for tiles with small fractions as illustrated in Fig. 12.

4   Concluding remarks

The numerics of physical parametrization has been discussed mainly in the context of the ECMWF Integrated Forecasting System. Parametrization schemes have a modular design and therefore splitting techniques are essential for the time stepping. Two major splitting techniques have been discussed: parallel splitting and sequential splitting. It is argued that sequential splitting is best with the explicit processes first and the implicit processes last. It is essential for the implicit process to have the tendencies of the other processes as input, otherwise it is not possible to obtain a proper balance for long time steps. Sequential splitting in multiple time scale problems can only work well with a single implicit process which is used last. With more than one fast process, it is desirable to combine them in a single solver.

Numerical integration of parametrized processes is typically first order accurate only. In the ECMWF model an attempt has been made to move towards 2nd order accuracy without increasing the costs. The idea is to average the parametrized tendencies "along" the semi-lagrangian trajectory. The method is successful as it decreases time truncation errors. However, some questions remain. Details of the formulation needed "optimization" to get good results, which is not satisfactory. An essential distinction has to be made between slow and fast processes in the design of the scheme. At this stage it is not clear in what category convection and clouds should be, because they have time scales of the order of the time step. Further research is necessary to find optimal handling of these processes in the context of the 2nd order physics.

A review of the numerical aspects of parametrized processes in the ECMWF model is given. Different processes have different numerical problems. However, in order to design the numerics it is important to have a good insight into the physics of the problem and to have an understanding of how the different processes

interact. Future increase in horizontal and vertical resolution will be even more demanding on the numerics and may require substantial effort to ensure stability and accuracy for long time steps.

Acknowledgment

Literature

Beljaars, A.C.M. (1991): Numerical schemes for parametrizations, ECMWF Seminar proceedings 9-13 September 1991, *Numerical methods in atmospheric models*, Vol II, 1-42.

Best, M.J., Beljaars, A.C.M., Polcher, J. and Viterbo, P. (2004): A proposed structure for coupling tiled surfaces with the planetary boundary layer, *J. Hydr. Meteor.*, in press.

Browning, G.L. and Kreiss, H.-O. (1994): Splitting methods for problems with different time scales, *Month. Weath. Rev.*, **122**, 2614-2622.

Caya, A., Laprise, R. and Zwack, P. (1998): Consequences of using the splitting method for implementing physical forcings in a semi-implicit semi-Lagrangian model, *Month. Weath. Rev.*, **126**, 1707-1713.

Cullen, M.J.P. and Salmond, D.J. (2002): *On the use of a predictor-corrector scheme to couple the dynamics with the physical parametrizations in the ECMWF model*, ECMWF Tech. Memo no 357, 22p.

Cuxart, J., Holtslag, A.A.M., Beare, R.J., Bazile, E., Beljaars, A., Cheng, A., Conagla, L., Ek, M., Freedman, F., Hamdi, R., Kerstein, A., Kitagawa, H., Lenderink, G., Lewellen, D., Mailhot, J., Mauritsen, T., Perov, V., Schayes, G., Steeneveld, G.-J., Svensson, G., Taylor, P., Wunsch, S., Weng, W. and Xu, K.-M. (2004): Single-column model intercomparison for a stably stratified atmospheric boundary layer, *Bound.-Layer Meteor.*, in press.

Dubal, M., Wood, N. and Staniforth, A. (2004): Analysis of parallel versus sequential splittings for time-stepping physical parametrizations, *Month. Weath. Rev.*, **132**, 121-132.

ECMWF (2001): *Key issues in the parametrization of subgrid physical processes*, ECMWF Seminar proceedings 3-7 September 2001, 349p.

Girard, C. and Delage, Y. (1990): Stable schemes for the vertical diffusion in atmospheric circulation models, *Month. Weath. Rev.*, **118**, 737-746.

Grenier, H. and Bretherton, C.S. (2001): A moist PBL parametrization for large-scale models and its application to subtropical cloud-topped marine boundary layers, *Month. Weath. Rev.*, **129**, 357-377.

Hammarstrand, U. (1997): Two time-step oscillations in numerical weather prediction models, *Month. Weath. Rev.*, **125**, 3368-3372.

Janssen, P.A.E.M. and Doyle, J.D. (1997): *On spurious chaotic behaviour in the discretized ECMWF physics scheme*, ECMWF Tech. Memo no 234.

Kalnay, E. and Kanamitsu, M. (1988): Time schemes for strongly nonlinear damping equations, *Month. Weath. Rev.*, **116**, 1945-1958.

Knoll, D.A., Chacon L., Margolin L.G. and Mousseau V.A. (2003): On balanced approximations for time integration of multiple time scale systems, *J. Comp. Phys.*, **185**, 583-611.

Lenderink, G. and Holtslag, A.A.M. (2000): Evaluation of the kinetic energy approach for modeling turbulent fluxes in stratocumulus, *Month. Weath. Rev.*, **128**, 244-258.

Lock, A. (2001): The numerical representation of entrainment in parametrizations of boundary layer turbulent mixing, *Month. Weath. Rev.*, **129**, 1148-1163.

Lott, F. and Miller, M.J. (1997): A new subgrid-scale orographic drag parametrization: Its formulation and testing, *Quart. J. Roy. Meteor. Soc.*, **123**, 101-127.

McDonald, A. (1998): The origin of noise in semi-Lagrangian integrations, ECMWF Seminar proceedings 7-11 September 1998, *Recent developmenmts in numerical methods for atmospheric modelling*, 308-334.

Morcrette, J.-J. (1999): *On the effects of the temporal and spatial sampling of radiation fields on the ECMWF forecasts and analyses*, ECMWF Tech. Memo 277.

Ropp, D.L., Shadid, J.N. and Ober, C.C. (2004): Studies of the accuracy of time integration methods for reaction-diffusion equations, *J. Comp. Phys.*, **194**, 544-574.

Smolarkiewicz, P.K. and Margolin L.G. (1994): Variational solver for elliptic problems in atmospheric flows, *Appl. Math. and Comp. Sci.*, **4**(4), 527-551.

Sportisse, B. (2000): An analysis of operator splitting techniques in the stiff case, *J. Comp. Phys.*, **161**, 140-168.

Suarez, M.J., Arakawa, A. and Randall, D.A. (1983): The parameterization of the planetary boundary layer in the UCLA general circulation model: formulation and results, *Month. Weath. Rev.*, **111**, 2224-2243.

Tiedtke, M. (1989): A comprehensive mass flux scheme for cumulus parametrization in large-scale models, *Month. Weath. Rev.*, **117**, 1779-1800.

Tiedtke, M. (1993): Representation of clouds in large-scale models, *Month. Weath. Rev.*, **121**, 3040-3061.

Tompkins, A.M., Bechtold, P., Beljaars, A.C.M., Benedetti, A., Janiskova, M., Köhler, M., Lopez, P. and Morcrette, J.-J. (2004): *Moist physical processes in the IFS: progress and plans*, ECMWF Tech. Memo, in press.

Van den Hurk, B.J.J.M., Viterbo, P., Beljaars, A.C.M. and Betts, A.K. (2000): *Offline validation of the ERA40 surface scheme*, ECMWF Tech. Memo no 295, 42p.

Viterbo, P. (1996): *The representation of surface processes in general circulation models*, PhD thesis, University of Lisbon, Available from ECMWF, 201p.

Wedi, N.P. (1999): *The numerical coupling of the physical parametrization to the "dynamical" equations in a forecast model*, ECMWF Tech. Memo no 274, 37p.

Wicker, L.J. and Skamarock, W.C. (1998): A time-splitting scheme for the elastic equations incorporating second-order Runga-Kutta time differencing, *Month. Weath. Rev.*, **126**, 1992-2097.

Williamson, D.L. (2002): Time-split versus process-split coupling of parameterizations and dynamical core, *Month. Weath. Rev.*, **130**, 2024-2041.