TECHNICAL MEMORANDUM

ECRAD: A new radiation scheme for the IFS

Robin J. Hogan and Alessio Bozzo

Research Department

Series: ECMWF Technical Memoranda

A full list of ECMWF Publications can be found on our web site under:
http://www.ecmwf.int/en/research/publications

Contact: library@ecmwf.int

Abstract

A new radiation scheme for the IFS is described. It is flexible, allowing the spectral resolution, the description of cloud and aerosol optical properties, and the solver, to be changed independently of one another. In its default configuration, the new 'McICA' solver produces less noise in heating rates, which has been found to lead to a small but significant reduction in forecast temperature errors. Moreover, the new scheme is around two-thirds the computational cost of the old (with the same spectral resolution), so additional forecast skill is possible in model configurations that could only afford to call the radiation scheme only every 3 h by reducing this interval to 2 h. A correction to the way the longwave equations are solved has led to an improvement in stratospheric mean temperatures. The new scheme has been used to test the impact of various changes to the description of clouds, including improvement of the ice and liquid optical properties, representation of longwave scattering, use of cloud overlap assumptions more consistent with observations, changing the width and shape of the sub-grid cloud water distribution, and for the first time representing 3D radiative effects.

1   Introduction and motivation

The radiation scheme currently operational in the ECMWF Integrated Forecasting System (IFS) was described by Morcrette et al. (2008). It incorporates the Rapid Radiative Transfer Model for GCMs (RRTM-G) of Mlawer et al. (1997), a correlated-k scheme to compute gaseous absorption in a total of 112 spectral intervals in the shortwave and 140 intervals in the longwave. This is combined with the Monte Carlo Independent Column Approximation (McICA; Pincus et al., 2003) for representing cloud sub-grid structure. Morcrette et al. (2008) showed that compared to the previous scheme, this led to an improvement in most aspects of the model, including agreement with satellite observations of top-of-atmosphere flux distributions, better placement of tropical convection, and a modest improvement in forecast skill. The older radiation scheme described by Morcrette (1991) is still available in the IFS code, and due to its speed it is used in the inner loop of data assimilation, while the shortwave component of is used for operational forecasts at Météo-France. An additional highly parameterized radiation scheme 'ACRANEB' (Mašek et al., 2015) is available in the Météo-France part of the code and is used in the HARMONIE limited-area model. In this context, one may ask why introduce yet another radiation scheme?

The first motivation is that, in order to strive towards an ever more accurate representation of atmospheric radiative transfer, we need a scheme that is *flexible*, so that it is easy to test new ideas. There is a particular need to be able to switch between different solvers to represent processes not captured in the current scheme such as longwave scattering (e.g. Costa and Shine, 2006) and the 3D transport of radiation through cloud sides (Hogan et al., 2016). Unfortunately, the calculations required for the current McICA solver are scattered through the existing code so it is not possible to cleanly switch to a different solver without a major rewrite.

The second, related motivation is the need to make the radiation scheme, and hence the whole IFS, more *computationally efficient*. The introduction of RRTM-G slowed the scheme down by a factor of 3.5 compared to what went before (Morcrette et al., 2008), and now we can only afford to run it on a reduced spatial grid (with around 6.25 times fewer points as the full grid) and infrequently in time (every 1 h in the HRES configuration and 3 h in all other operational model configurations). Bozzo et al. (2014) tested a spectral sampling technique built on top of RRTM-G; this enabled more frequent calls to the radiation scheme in time and space that mitigated some of the errors at coastlines due to the reduced spatial grid, but at the cost of noise in the radiative fluxes. Hogan and Bozzo (2015) found a solution to problems of coastal errors that did not introduce noise, and this is currently operational at ECMWF. Their method

of approximate radiation updates also improved the temperature evolution over land in configurations when the radiation scheme is called only every 3 h, but it did not allow the radiative fluxes to adjust to changing clouds within these 3 h intervals. For this reason, reducing the radiation timestep from 3 h to 1 h leads to a statistically significant improvement in forecast skill, but this is currently unaffordable. More sophisticated solvers should improve the climate of the model, but their likely additional computational expense would have to be offset by an equivalent reduction in the expense of other aspects of the radiation scheme.

The most promising source of extra efficiency is likely to be a more efficient gas absorption scheme to RRTM-G; other correlated-*k* schemes are used operationally elsewhere that have fewer than half as many spectral intervals as RRTM-G (e.g. Fu and Liou, 1992; Cusack et al., 1999), and the cost of all parts of the radiation scheme scales linearly with the number of intervals. Some are flexible enough that the spectral resolution can be chosen at run-time (e.g. Edwards and Slingo, 1996). At ECMWF this would enable the trade-off between computational accuracy to be different for different forecast horizons, and in particular we could then use the same radiation scheme for data assimilation as for forecasts, but with different spectral resolution. Unfortunately, the RRTM-G gas optics module is completely intertwined with the other parts of the radiation scheme, and it would be impossible to modify it cleanly to enable switching between different gas-optics modules. Additional opportunities for optimization involve array and loop reordering, and therefore would also involve substantial modification of many routines.

This memorandum describes a new radiation scheme designed to overcome these shortcomings and provide a maintainable framework for radiation developments into the future. The scheme is modular; specifically, the four main components of the scheme (gas optics, aerosol optics, cloud optics and solver) are almost completely independent of each other, so that any component can be modified or replaced without affecting the functioning of any of the other three components. This is essential to facilitate the development and testing of new features, and indeed this memorandum describes the results of testing a number of new features in the representation of clouds. The new scheme can also be run in a configuration that closely matches the old. Bit-reproducibility is not possible due to scientific improvements in the new scheme, and a much more efficient McICA cloud generator that uses the random numbers in a different way. During the process of rewriting, a number of opportunities for optimization in terms of both computational cost and memory usage presented themselves and are described in this document.

The opportunity to test the sensitivity to different ways of representing clouds in the new radiation scheme has enabled us to begin to understand a puzzling aspect of the behaviour of the current model. Three factors ought to lead to too much warming of the surface. Firstly, the overlap scheme correlates vertically separated cloud layers when observations show they should be randomly overlapped (e.g. Hogan and Illingworth, 2000); this would be expected to lead to an underestimate of total cloud cover. Secondly, cloud heterogeneity is overestimated compared to the values reported in the review of Shonk et al. (2010), which would also be expected to lead to too much solar radiation reaching the surface. Thirdly, a bug in the current longwave ice optics scheme makes clouds effectively too optically thick, which would be expected to cause too much downwelling longwave at the surface. Despite these warming biases, 2-m temperature forecasts tend to be too cold, by around 0.5 K, on average. This cold bias extends through most of the troposphere. We investigate the extent to which this behaviour can be explained by the lack of longwave scattering and 3D effects in the radiation scheme.

In section 2, the internal structure and overall computational performance of the new scheme is described. Section 3 describes the computation of gas optical properties and section 4 the aerosol optical properties. Section 5 describes differences in the way the two-stream equations are solved in the new scheme, and their meteorological impact. Section 6 describes the cloud optical properties available. Section 7 describes the new McICA implementation of McICA scheme that reduces noise, and section 8 outlines

the other solvers available. Then in section 9 we describe the computation of radiation diagnostics, such as photosynthetically active radiation at the surface. Note that the primary purpose of this memorandum is to explain the differences between the new scheme and the old; for a full description of the underlying physics of the scheme (most of which is common to the old scheme), the reader is referred to the IFS documentation.

2   Technical overview of new scheme

The radiation scheme takes as input surface properties and profiles of atmospheric properties, and returns profiles of broadband longwave and shortwave fluxes. It is the job of the host model to prepare the inputs by perform tasks such as computing solar zenith angle, aggregating surface tiles to compute average surface albedo in user-specified spectral intervals, computing cloud effective radius and interpolating aerosol and trace-gas concentrations from a climatology. It is also the host model's responsibility to compute atmospheric heating rates from the flux profiles and, if required, to use approximate methods to optimally interpolate the radiative fluxes in time and space (Hogan and Bozzo, 2015). This functionality is already provided by the IFS and is largely unchanged.

The radiation scheme has been coded in Fortran using modern programming practices to maximize readability, extensibility and modularity. All except the most trivial data required by the scheme are read at run-time from self-describing NetCDF data files (e.g. cloud and aerosol optical properties in each spectral interval). Object orientation is used where it improves clarity to do so. No global variables are used, and configuration information is carried in a single configuration object that is passed through to the routines that need it. Long argument lists are avoided by use of derived types (equivalent to classes in C++) for input and output data, as listed Fig. 1. This approach also ensures that the model is easily extensible in future: if, for example, the radiative impact of rain was to be accounted for, then rain mixing ratio could be added to the `cloud` derived type and no routine would need its argument list to be modified.

`single_level`
Cosine of solar zenith angle (col)
Skin temperature (col)
SW albedo to diffuse radiation (col, albedo band)
SW albedo to direct radiation (col, albedo band)
LW emissivity (col, emissivity band)
Seed for random number generator (col)
Broadband solar irradiance

`thermodynamics`
Temperature at half-levels (col, half-level)
Pressure at half-levels (col, half-level)
Specific humidity at liquid saturation (col, level)

`gas`
Mixing ratio (col, level, gas ID)
Flags indicating units of gases
Flags indicating which gases are well mixed

`aerosol`
Mass mixing ratio (col, level, aerosol ID)

`cloud`
Liquid water mixing ratio (col, level)
Ice water mixing ratio (col, level)
Liquid effective radius (col, level)
Ice effective radius (col, level)
Cloud fraction (col, level)
Cloud water fractional standard deviation (col, level)
Cloud overlap parameter (col, half-level)

`flux`
SW/LW up/downwelling clear/total-sky fluxes (col, half-level)
Direct downwelling SW clear/total-sky flux (col, half-level)
Surface downwelling spectral SW flux (col, SW band)
Cloud cover diagnosed by SW/LW solver (col)
LW derivatives of Hogan and Bozzo (2015) (col, half-level)

Figure 1: Contents of the five derived types used to provide input variables to the scheme, and the one derived type (`flux`) used to convey the outputs back to the host model. The dimensions of the arrays are indicated in brackets, where 'col' indicates atmospheric column. An additional derived type containing constant configuration information (`config`) is not shown here.

Table 1: Summary of the main options available in the new radiation scheme.

| Property	Options
Gas optics model	RRTM-G (faster options are planned)
Aerosol optics model	Generalized (includes Tegen and CAMS)
Liquid cloud optics	Slingo (1989) and Lindner and Li (2000); SOCRATES
Ice cloud optics	Fu (1996) and Fu et al. (1998); Baran et al. (2014)
SW and LW solver	McICA; Tripleclouds; SPARTACUS
Cloud overlap scheme	EXP-EXP (only available with McICA); EXP-RAN; MAX-RAN
Cloud overlap parameter	α (Hogan and Illingworth 2000); β (Shonk et al. 2010)
Longwave scattering	Off; cloud only; cloud and aerosols
Cloud water PDF shape	Gamma; Lognormal
δ-Eddington mode	Particles and gas; Particles only
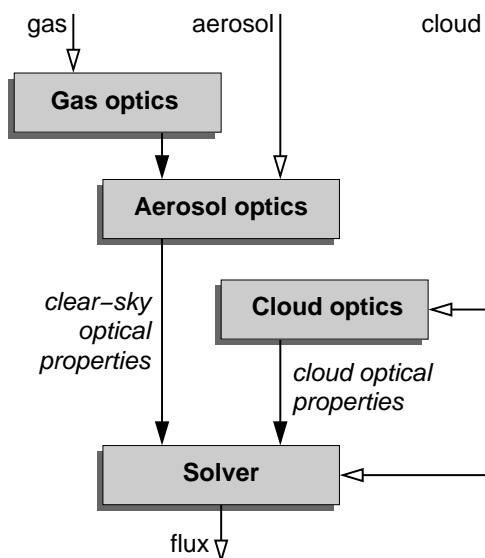
Figure 2: Schematic of the flow of data between the four components of the scheme, where the open-headed arrows denote derived types described in Fig. 1, and the solid-headed arrows denote groups of arrays containing clear-sky and cloud optical properties of the atmosphere. Note that all four components also take as input the `config`, `thermodynamics` *and* `single_level` *derived types (not shown here).*

First, the user partially populates the `config` object to configure the scheme, and this is passed to a setup routine that loads and stores the necessary look-up tables and other data, referenced from the `config` object. The main non-numeric options currently available are listed in Table 1. The user then populates the five input objects shown in Fig. 1, and passes them along with the `config` object to the interface routine. Figure 2 depicts the subsequent flow of data between the four components of the scheme. The gas optics module computes the shortwave and longwave optical properties of gases in each spectral interval (g-point), as well as the spectrally-resolved Planck function and top-of-atmosphere incoming solar radiation. The resulting arrays are passed to the aerosol module, which *adds* the contribution from aerosols with the assumption that aerosols are horizontally well mixed within each model gridbox. The cloud optics module computes the in-cloud-mean optical properties of clouds in each spectral band, where a band typically contains between 2 and 16 g-points. It is then the solver's job to combine the clear-sky and cloudy optical properties according to the cloud fraction and appropriate assumptions about in-cloud heterogeneity, and from this compute the broadband flux profile.

Figure 3 compares the average computational time spent on each profile by the old and new radiation schemes when implemented in the IFS, along with the costs of the individual components. All activities performed after interpolation on to the lower resolution radiation grid are included. This includes the
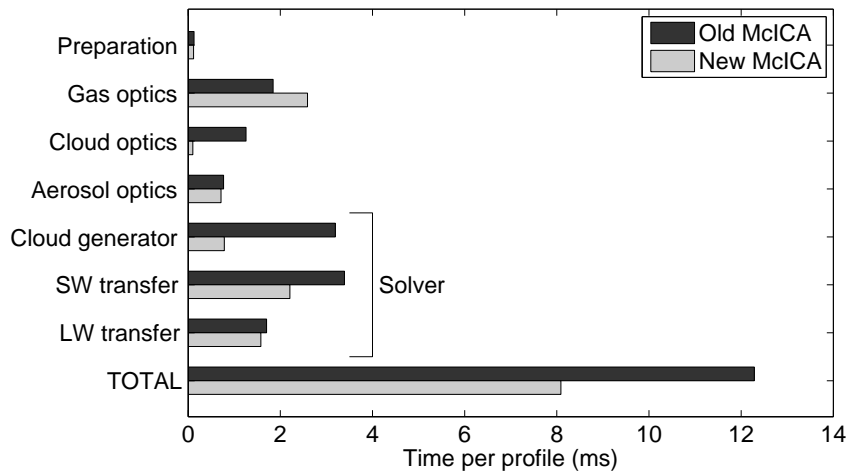
Figure 3: Comparison of the computational cost of various parts of the old and new radiation schemes, determined using the 'Dr Hook' profiling system in a 1-day forecast of the IFS at T639 resolution with 137 levels. All calculations performed after interpolation to the radiation grid are included. 'Preparation' consists of preparing the inputs to the radiation scheme. This is followed by the four components of the new scheme (shown in Fig. 2) but with the 'Solver' component divided into the McICA cloud generator and then the radiative transfer calculations in the shortwave (SW) and longwave (LW). The modularity of the new scheme made it straightforward to time the individual components, but obtaining the equivalent timing for the old scheme required careful placement of Dr Hook calls at numerous places within larger routines. In both runs the model was configured to process atmospheric profiles in batches of eight (i.e. NRPROMA=8).

small cost of 'preparation' of inputs to the radiation scheme, including interpolating the trace gas and aerosol concentrations from climatological values, computing cloud effective radius from water content and temperature, and overlap decorrelation length from latitude. It also includes allocating memory for the arrays in the derived types each time the radiation scheme is called, but this accounts for only 0.2% of the total cost of the radiation scheme. The new scheme was configured similarly to the old, using a McICA solver that neglects longwave scattering. We see that the new scheme is 34% faster than the old, although this reduces to 31% when the computational cost of interpolating to the reduced grid is included. Nonetheless, this suggests that an immediate route to improving forecast skill for the same computer time would be to increase the frequency of radiation calls from every 3 h (used in all operational configurations except HRES) to every 2 h, explored further in section 7.3. The subsequent sections explain the very different speeds of the individual components of the scheme.

3 Gas optics

The particular gas optics model is selected at run-time, and determines the spectral discretization used by the whole of the radiation scheme. The shortwave and longwave regions are split into a number of bands, within which the rapid spectral variation of gaseous absorption is treated by the correlated *k*-distribution method: the gas absorptions have been *reordered* within each band from the least to the most absorbing, as a function of the new coordinate 'g' that varies from 0 (least absorbing) to 1 (most absorbing). The smooth absorption distribution versus g has then been discretized by a handful of 'g-points'. Each g-point is treated as monochromatic by the other parts of the radiation scheme. A small dependence therefore arises between the gas, aerosol and cloud optics modules since they must all use

the same spectral discretization. This is simply a case of ensuring that the aerosol and cloud optics modules load scattering data files with a matching spectral discretization; the code in these modules is unchanged. Likewise, merging aerosol and cloud optical properties (in bands) with gas optical properties (in g-points) is done via configuration arrays that map g-points to bands, rather than differences in the code.

The user provides the mixing ratios of a selection of gases in either volume- or mass-mixing-ratio units. Gas optics models differ in the trace gases they represent, so any gases not provided by the user are assumed to be absent, and any gases provided that are not represented by the model are ignored. This will make it easy in future to switch between models that represent different gases. The only model currently available in the new radiation scheme is RRTM-G, using the implementation currently in the IFS. It should be clarified that 'RRTM-G' is actually an entire radiation scheme including solver and cloud optics components, but since the new radiation scheme uses only its gas optics component and spectral discretization, references to 'RRTM-G' in this memorandum are usually concerned only with these aspects.

The following arrays are output by the gas optics module:

- Shortwave optical depth and single-scattering albedo of each layer of the clear-sky atmosphere (g-point, level, column);

- Top-of-atmosphere solar flux for each shortwave g-point (W m^{-2}) (g-point, column);

- Longwave optical depth of each layer of the clear-sky atmosphere (g-point, level, column);

- Planck function of the atmosphere integrated across each g-point (W m^{-2}) (g-point, half-level, column);

- Planck function of the surface integrated across each g-point (W m^{-2}) (g-point, column).

The array dimensions are listed in brackets with the fastest-varying dimension first. We have found that optimal performance through most of the new radiation code is achieved using spectral interval (g-point) as the innermost array dimension and loop index. This is because conditional operations (which inhibit vectorization) depend on the presence of cloud or whether the sun is above the horizon, factors that are functions of level or column, but not g-point. By computing the Planck function and top-of-atmosphere solar flux in the gas optics module, the subsequent parts of the code can treat all g-points equally. Thus, to compute broadband fluxes the solver simply sums the fluxes for each g-point.

Figure 3 shows that gas optics is the only component that is slower in the new scheme. This is because the IFS implementation of RRTM-G is currently coded to produce arrays with g-point as the slowest varying dimension, and these arrays need to be permuted before they can be passed on. This slowing is more than compensated by speed-ups in other components, but there is nonetheless scope for optimization of RRTM-G in the context of the new radiation scheme.

4   Aerosol optics

Up to IFS Cycle 43R1, the radiation scheme has two quite different ways to specify aerosols. The older Tegen et al. (1997) climatology is specified in terms of the 550-nm optical depth of six (hard-coded) aerosol types in each layer, and then a simple rule is used to extend this to all other parts of the spectrum,

including the longwave. None of the aerosol types are hydrophilic. The newer 'CAMS' (Copernicus Atmosphere Monitoring Service) aerosol description (formerly known as 'MACC') may either be used with prognostic aerosols in the CAMS version of the IFS, or with a climatology of aerosols derived from the CAMS system (Bozzo et al., 2016). CAMS specifies aerosols in terms of the mass mixing ratio of 12 aerosol types, and a hard-coded database of scattering properties for each of the 14 shortwave and 16 longwave RRTM-G bands is used to treat the optical properties rigorously. Three of the CAMS types are hydrophilic (sea salt, hydrophilic organic matter and ammonium sulphate), and their optical properties change according to the local relative humidity. To represent the variation in aerosol size, sea salt and desert dust are represented by mass mixing ratios in three size bins. Both the Tegen and CAMS climatologies are used by default with a low concentration of 'background' aerosol, which are specified via a tropospheric and stratospheric optical depth at 550 nm. In the old radiation scheme, it is fairly straightforward to change the optical properties of individual aerosol types, but to change the number of categories or the interpretation of a category requires much more involved code changes.

The new radiation scheme takes a completely flexible approach to specifying aerosol properties at run time. A NetCDF data file is read that contains the optical properties (mass-extinction coefficient, single scattering albedo and asymmetry factor in each shortwave and longwave band) of any number of hydrophilic and hydrophobic aerosol types. A user-specified mapping then controls how the aerosol mass mixing ratios provided by the host model map on to the hydrophilic and hydrophobic types in the data file. This allows the 'wrapper' routine to convert the Tegen and CAMS aerosol distributions from the model into inputs to the new radiation scheme, such that the scheme does not need to do anything differently for the two sorts of aerosol. In the case of Tegen, where the inputs are in terms of 550-nm optical depth, the NetCDF file does not contain mass-extinction coefficient but rather the ratio of extinction at the wavelength in question to that at 550 nm. To represent background aerosol, a specific aerosol type must be specified from those available in the NetCDF file. The wrapper routine uses the mass-extinction coefficient of that type at the band encompassing 550 nm to convert the requested background-aerosol optical depth into an equivalent mass mixing ratio. We have found that shortwave clear-sky fluxes agree to within a fraction of a Watt-per-square-metre between the old and the new radiation schemes, confirming that the aerosol and gas optics treatments are the same in the two schemes.

Figure 2 shows the aerosol optics module of the scheme in the context of the other modules. The arrays listed in section 3 are passed from the gas optics module, along with the user-provided aerosol mixing ratios in the `aerosol` object, to the aerosol optics module. This computes the scattering properties of aerosols and *adds* them to the clear-sky optical property arrays. We make the assumption that aerosols are horizontally well mixed within each model gridbox. Several new arrays are also produced:

- Shortwave asymmetry factor (g-point, level, column). Note that shortwave gaseous scattering is isotropic so its asymmetry factor is zero until the aerosol contribution is included.

- Longwave single-scattering albedo and asymmetry factor (g-point, level, column). These arrays are only needed if longwave aerosol scattering is enabled.

All arrays describing the optical properties of the gas-aerosol combination are passed to the solver. It can be seen in Fig. 3 that the new aerosol optics component is marginally faster than the old (with both using the Tegen climatology). Using the CAMS aerosol scheme makes the aerosol optics module very slightly slower.

5 Solutions to the two-stream equations

We are confident that the gas and aerosol optics components of the new radiation scheme (discussed in the previous two sections) compute nearly identical optical properties to the equivalent parts of the existing scheme. But the first implementation of the new scheme into the IFS produced significantly different fluxes, and most of the work involved in implementing the scheme has been in tracking down and understanding the reason for these differences, and where necessary changing the behaviour of the new scheme to match the old. The most substantial differences are in cloudy skies, due both to differences in cloud optical properties (discussed in section 6) and in the way cloud structure is described in the solver (discussed in section 7).

The present section describes differences arising from the way the two-stream equations are solved, and their impact on model climate. Underpinning both the new and the old radiation schemes are the two-stream equations: differential equations that describe how radiation from either the direct solar beam or from thermal emission enters the diffuse upwelling and downwelling streams, and how radiation is exchanged between these two streams by scattering. These differential equations are solved in each model layer to obtain layer-wise properties such as the reflectance and transmittance to diffuse radiation, but in the current radiation scheme the analytic formulas for these quantities appear to have been approximated in a way that produces small but significant differences in the model climate, as described in the following subsections for shortwave (section 5.1) and longwave radiation (section 5.2).

5.1 Shortwave conservative-scattering approximation

The McICA solvers of both the old and the new scheme use the Meador and Weaver (1980) analytic solutions to the two-stream equations in the shortwave. These solutions take as input the layer optical depth, single scattering albedo, asymmetry factor and solar zenith angle, and predict the transmission and reflection of the layer to direct and diffuse radiation. They involve several calls to the exponential function for each layer and g-point, so can account for a significant fraction of the cost of the radiation scheme as a whole.

The current radiation scheme performs an optimization: if the shortwave single scattering albedo, ω, is greater than 0.9995 in a spectral interval then conservative scattering is assumed in which case the Meador and Weaver (1980) equations simplify. It turns out that the absorption of such layers, which will almost always be cloudy, cannot strictly be neglected. Later versions of the RRTM-G radiation scheme available from Atmospheric and Environmental Research (AER) only assume conservative scattering when $\omega > 0.999,999,5$. Therefore we have increased the threshold in the old scheme, so that it matches the new scheme that never makes the conservative-scattering approximation. Climate simulations have been performed with the old radiation scheme before and after this change and item 1 in Table 2 shows that the effect was to increase net downward shortwave at top-of-atmosphere (i.e. to *reduce* reflected shortwave radiation) by 0.5 W m^{-2}. Figure 4 shows that the change to net shortwave flux at the surface is smaller, indicating that this reduced reflection is associated with an increase in atmospheric absorption. This modification has been made to the old radiation scheme, which from Cycle 43R3 is still available as an option even though the new scheme is used by default.

Table 2: Impact of various modifications to the radiation scheme on global cloud cover, shortwave and longwave top-of-atmosphere (TOA) net flux, and mean land temperature. They are computed from four 1-year free-running coupled simulations at T255 resolution. The control configuration consists of: IFS Cycle 43R1 using the original radiation scheme called every hour. This scheme does not simulate longwave scattering, assumes EXP-EXP overlap and uses a McICA solver with an in-cloud water content described by a gamma distribution with a fractional standard deviation (FSD) of 1. The first line indicates the change to each quantity when a bug is fixed in the shortwave scheme. Subsequent modifications are added cumulatively, so each modification is implemented in the model on top of all the modifications above in the table, but the numbers indicate the incremental change due to one new feature. In some cases alternative modifications are listed, and numbered with an 'a' or 'b', but these are not used by modifications below in the table. Changes that are a direct result of modifications to the radiation scheme are shown in bold; other changes are due to the response of the model. The cumulative impact of these changes is depicted in Fig. 4. The 'typical uncertainties' listed in the final row are the standard error of the mean change, treating the four simulations in each experiment as independent estimates of the mean change.

Modification	Cloud cover (%)	SW TOA (W m^{-2})	LW TOA (W m^{-2})	Land temperature (°C)
1. Fix SW scattering	−0.2	**+0.5**	−0.1	+0.0
2. New radiation scheme	**−0.4**	**−1.8**	**+0.4**	−0.1
3. Fix liquid optics	−0.0	**+1.6**	**−0.4**	+0.1
4. Fix ice optics	−0.5	+0.4	**−1.6**	−0.2
5. Represent LW scattering	−0.2	+0.2	**+1.4**	+0.3
6. δ-scaling applied to particles only	−0.0	**+0.1**	+0.0	−0.0
7. EXP-RAN overlap	**+3.0**	**−1.4**	**+0.5**	−0.1
7a. MAX-RAN overlap	**+1.4**	**−0.2**	**−0.2**	−0.1
8. FSD=0.75	+0.2	**−2.7**	**+1.5**	−0.1
9. 3D effects using SPARTACUS solver	−0.5	**+3.0**	**−0.2**	+0.6
9a. Lognormal PDF	+0.1	**−0.7**	**+0.4**	−0.1
9b. Tripleclouds solver	−0.0	**+0.4**	**+0.1**	+0.1
Typical uncertainty	±0.1	±0.1	±0.1	±0.1

5.2   Longwave layer-wise emission

A difference between the old and new scheme exists in the solution to longwave radiative transfer that has a significant impact on stratospheric and surface temperatures. The new scheme uses temperature at half-levels only, and the Planck function is computed at these points. It then makes the common assumption that Planck function varies linearly with optical depth through the layer. If scattering is ignored then the two-stream equations decouple and in layer *i*, the downwelling flux obeys the following equation:

$$\frac{dF^\downarrow}{d\delta} = D\left(-F^\downarrow + B_{i-1/2} + \delta B'\right), \tag{1}$$

where δ is the optical depth measured downwards from zero at the top of the layer, $D = 1.66$ is the diffusivity factor accounting for the typical range of zenith angles of longwave radiation transport (e.g. Fu et al., 1997), $B_{i-1/2}$ is the value of the Planck function (in flux units) at the top of the layer, and $B' = dB/d\delta$ is the gradient of the Planck function. The solution to this equation provides the flux exiting the base of the layer:

$$F^\downarrow_{i+1/2} = \tau_i F^\downarrow_{i-1/2} + (1 - \tau_i)\left[B_{i-1/2} + B'/D\right] + B'\delta, \tag{2}$$
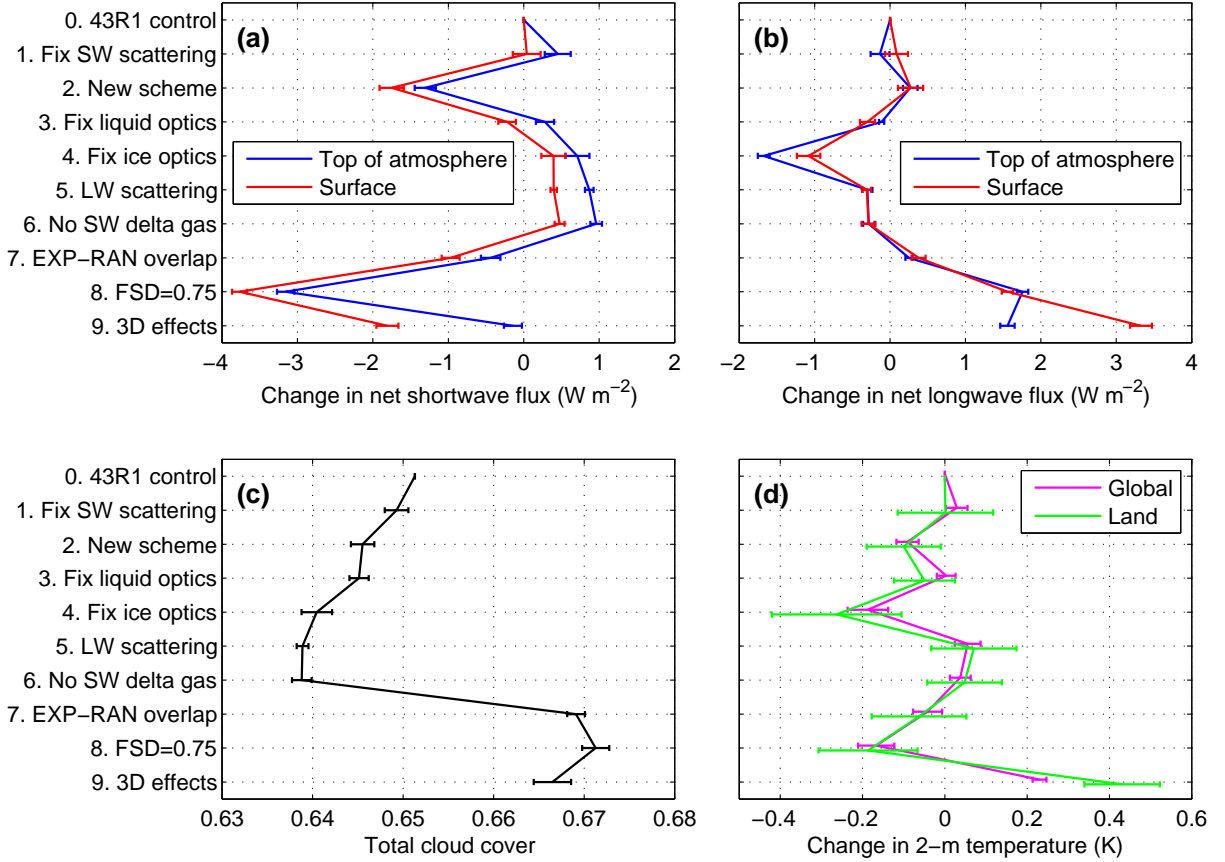
Figure 4: Impact of the cumulative modifications listed in Table 2 on net fluxes (positive downwards) at top-of-atmosphere and the surface in (a) the shortwave and (b) the longwave, and on (c) total cloud cover and (d) global and Northern-Hemisphere-land temperatures (20–70°N). These results are computed from the same simulations as used in Table 2.

where $\tau_i = \exp(-D\delta_i)$ is the transmittance of the layer, δ_i is the optical depth of the layer and $F^{\downarrow}_{i-1/2}$ is the flux entering the top of the layer. This equation is implemented in the new scheme, and if longwave scattering is represented then a more complex formula is used that reduces to this equation in the no-scattering limit.

The old radiation scheme uses the following formula instead:

$$F^{\downarrow}_{i+1/2} = \tau_i F^{\downarrow}_{i-1/2} + (1 - \tau_i) \left[p B_{i+1/2} + (1 - p) \bar{B}_i \right], \tag{3}$$

in which we see that the flux exiting the base of the layer equals the flux transmitted through the layer (as in Eq. 2) plus an emission term, where the emission is expressed as a weighted average of the Planck function at the base of the layer and the mean Planck function of the layer. The linear-in-optical-depth assumption leads to $\bar{B}_i = (B_{i-1/2} + B_{i+1/2})/2$. The weight, p, varies from 0 in the optically thin limit to 1 in the optically thick limit; physically, optically thin layers emit with an effective emission temperature of something like the average temperature of the layer, while optically thick layers emit downwards with an effective emission temperature of the base of the layer. The equation for p consistent with (2) is

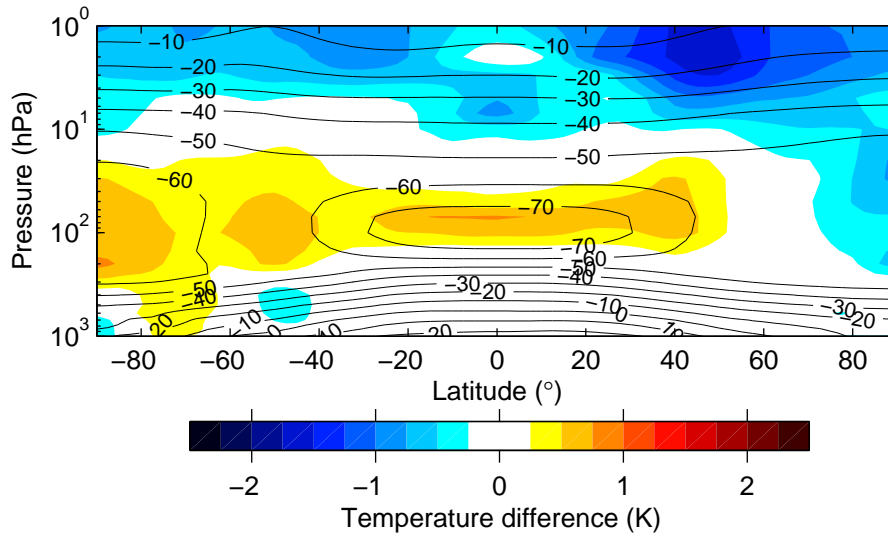$$p = \frac{1 + \tau}{1 - \tau} - \frac{2}{D\delta_i}, \tag{4}$$

Figure 5: Difference in zonal-mean temperature between the new scheme with fixed liquid optics, and the old scheme in the 43R1 Control, for climate simulations used in Table 2.

However, the old scheme uses a simpler Padé approximation of the form

$$p = \frac{\delta_i}{3.59712 + \delta_i}. \tag{5}$$

This formula predicts systematically lower p than (4), with the result that the effective emission temperature will be closer to the mean temperature of the layer than to the temperature at the near edge of the layer. This approximate formula might be appropriate in a band model in which a range of optical depths are treated at once, but for a correlated-k model that is intended to reduce the spectrum to a finite number of monochromatic g-points, it is more appropriate to use the exact solution to (1). Moreover, if the differential equation is not solved exactly, then the solution becomes resolution-dependent. A further difference is that the old scheme computes $\bar{B}_i$ by re-evaluating the Planck function at the temperature of the mid-point of the layer. Since the half-level temperatures were interpolated from the full-level temperatures, for all interior layers this does not bring any extra information to the profile, and since the interpolation is performed linearly in pressure, the difference with a linear-in-optical-depth approximation is likely to be very small. The exception is in the lowest model layer, where the difference has the potential to impact near-surface air temperature. We now examine the impact of these two differences.

The primary atmospheric impact of the difference in the definition of p appears to be on stratospheric temperature since it affects the efficiency of longwave cooling to space. Figure 5 shows that the impact of the new radiation scheme is a warming of up to 0.75 K centred at 100 hPa, and a cooling in the upper stratosphere of up to 1.5 K. Note that this effect is broadly the same for all configurations of the new scheme listed in Table 2. These changes tend to push the free-running model in the direction of better agreement with ERA-Interim in these areas.

The different definitions of both p and $\bar{B}_i$ have a significant impact on near-surface air temperature. In much of the longwave spectrum the atmosphere is completely opaque and the surface receives significant downward emission from the lowest model layer, even though it is only around 20 m thick in the 137-level model. The model solves prognostic equations for the mean temperature of each layer, and we denote the temperature of the lowest layer of an n-layer model as T_n. Half-level temperatures are found by interpolating the pressure-weighted temperatures at the mid-points of the layers in pressure space
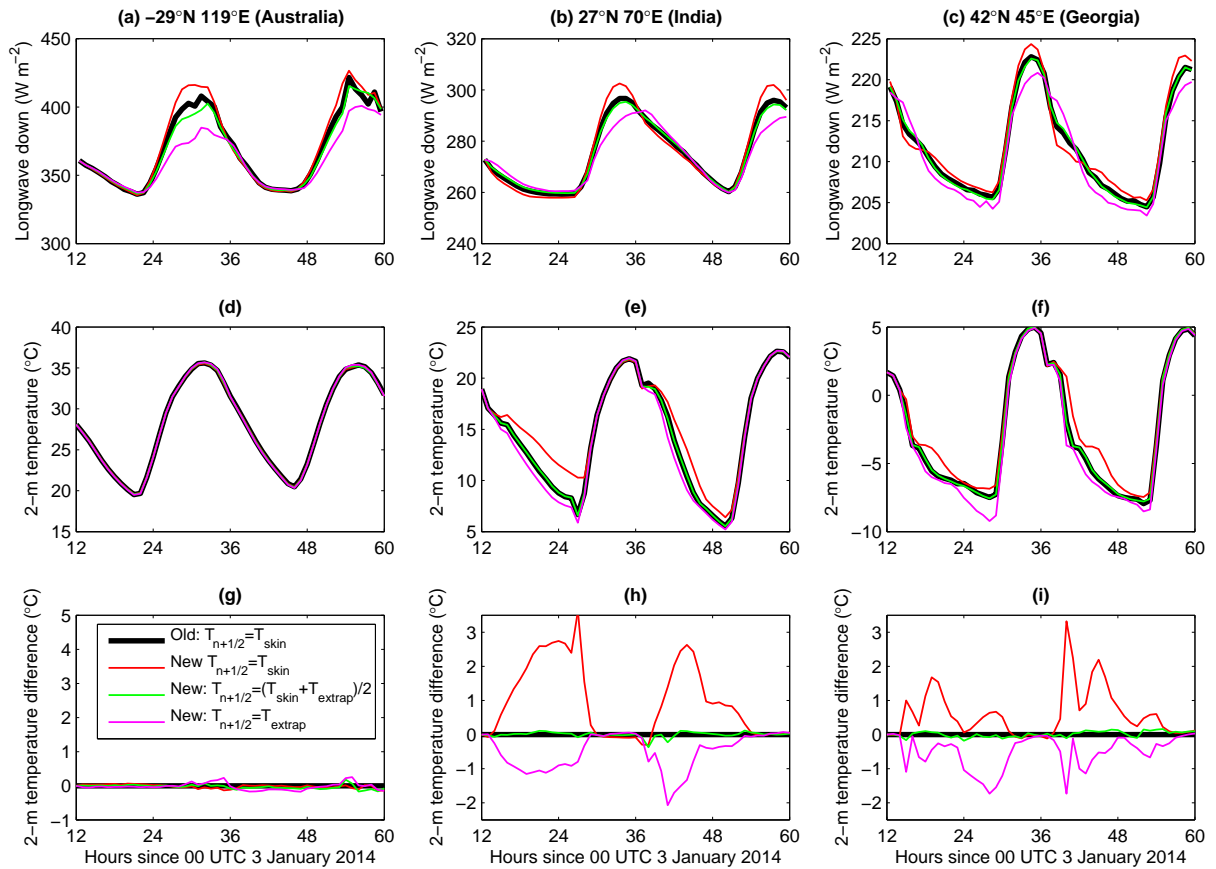
Figure 6: Each column of panels corresponds to contrasting cloud-free land point from T255 forecasts initialized at 12 UTC on 3 January 2014. The properties are (a–c) downwelling longwave radiation, (d–f) 2-m temperature, and (g–i) difference in 2-m temperature from the 43R1 control simulation using the old radiation scheme, shown by the thick black lines. The thin coloured lines show the results of the new scheme with three different assumptions for the air temperature at the surface, $T_{n+1/2}$, as shown in the legend in panel g. These assumptions are depicted schematically in Fig. 7a.

to each half-level. This leaves some ambiguity about what to do for the air temperature at the base of the lowest layer, $T_{n+1/2}$. The current radiation scheme assumes the air temperature to be equal to the skin temperature of the surface: $T_{n+1/2} = T_{\text{skin}}$, and the corresponding evolution of temperature at three cloud-free land points is shown by the black lines in Fig. 6. The red lines show the result of making the same assumption in the new scheme, and we see that this leads to larger surface downwelling longwave flux in the day and typically (but not always) a reduction at night. While 2-m temperatures are virtually unchanged during the day, they are up to 3 K larger at night.

This can be understood further from Fig. 7a, which shows schematically the temperature assumptions that these two schemes are attempting to use within a layer at night. We stress that since the old scheme does not solve the differential equation exactly, it is not providing the exact solution to the black temperature profile shown in Fig. 7a. The red line is everywhere lower than the black, explaining the decreased downwelling longwave flux. Figure 7b depicts the actual temperature profile shortly before dawn in the India case shown in Fig. 6e. We see that the main effect of the reduced downward emission from the lowest model level is to reduce its rate of cooling during the night. The skin temperature is virtually unchanged. Since 2-m temperature is a diagnostic computed from the skin temperature, the lowest model level temperature and the Obukhov length, the warmer temperatures in the lowest model level lead to the
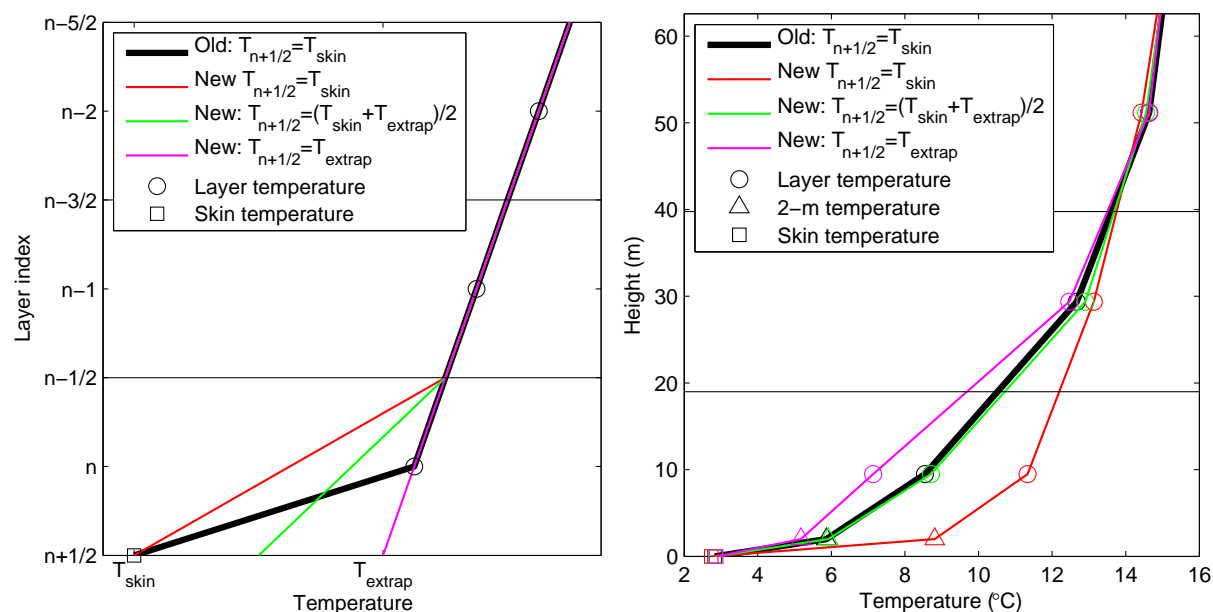
Figure 7: (Left) Schematic of assumed temperature profiles in the lowest model layer (indexed 'n') in the old and new radiation schemes, the latter with three different assumptions about the near-surface air temperature. The impact on 2-m temperature forecasts is shown in Fig. 6. (Right) The corresponding forecast temperature profiles at 00 UTC 4 January 2014 for the point over India shown in Fig. 6, 12 h into the forecast.

elevated 2-m temperature prediction.

In practice the largest temperature gradients occur very close to the surface; a 10 K difference is possible between the 2-m temperature and the skin temperature. Thus, a better approximation than the red line in Fig. 7a might be simply to extrapolate down to the surface in a way that conserves the mean temperature of the layer; this could be written as $T_{\text{extrap}} = 2T_n - T_{n-1/2}$. The magenta line in Fig. 6 shows that this leads to significantly cooler night-time temperatures, by as much as 2 K. Since the IFS already has a cold bias at the surface, this would not be a desirable change from a forecasting perspective. The green line in Fig. 7a shows an assumed temperature profile half-way between these extremes: $T_{n+1/2} = \left(T_{\text{skin}} + T_{\text{extrap}}\right)/2$, which has a layer-mean temperature equal to that of the old scheme. Figure 6 shows that this produces very similar 2-m temperature forecasts to the old scheme while Fig. 7b shows that the temperature profile is very similar, and so it is what we use by default in the IFS implementation of the new scheme.

In summary, the new scheme solves the longwave two-stream equations exactly, which improves stratospheric temperatures, but the assumed temperature profile in the lowest model level must be approximated in a particular way to get similar 2-m temperature forecasts to the current model. The strong sensitivity of 2-m temperature forecasts to the assumed temperature profile in the lowest layer suggests further work is necessary, for example, to replace the linear approximation with an approximation that better matches the nonlinear surface-layer temperature profiles predicted by Monin-Obukhov theory. Observations of the longwave flux divergence across the lowest 10 or 20 m could also provide valuable validation data (Steeneveld et al., 2010).

6 Cloud optics

The cloud-optics module uses the in-cloud liquid and ice mixing ratios (that is, the gridbox-mean mixing ratios from the host model, divided by the cloud fraction), along with the liquid and ice effective radii values, to compute the following arrays:

- Shortwave cloud optical depth, single scattering albedo and asymmetry factor at each height and in each spectral band.

- Longwave cloud optical depth, single scattering albedo and asymmetry factor at each height and in each spectral band. If longwave scattering is not active then the single scattering albedo and asymmetry factor are not used and have zero dimensions.

All arrays are dimensioned (band, level, column), with the first dimension varying fastest. These arrays are passed to the solver. Currently the snow and ice water contents from the model are combined before entering the radiation scheme. A possible improvement in future would be to generalize the treatment of different hydrometeor types, with the number of types being user-configurable. This would give the user the freedom to treat ice and snow separately in the radiation scheme, as well as to include rain which is currently neglected.

Figure 3 shows that the cloud-optics calculation in the new scheme is very efficient and indeed is many times faster than the equivalent parts of the old scheme. This is simply because of a deficiency in the old scheme: it recomputes cloud optical properties for every g-point using the perturbed cloud water contents from the McICA cloud generator. However, the perturbed water contents from the McICA scheme simply have the effect of scaling the cloud optical depth. Therefore a more efficient implementation of the same scheme is to generate the cloud optical properties once for each band (and there are 8–9 times fewer bands than g-points in the shortwave and longwave parts of the spectrum), and then use the cloud generator to provide optical depth scalings for each g-point.

We stress that both the old and the new schemes treat effective radius as constant within a model level. For slightly more realism, it would be possible to use the cloud generator to perturb water content rather than optical depth, and then use the water content to recompute effective radius in each sub-column. However, this would introduce significant extra computational cost but with minimal meteorological impact. The main effect can easily be accounted for by choosing an appropriate value for the fractional standard deviation (FSD) input to the McICA scheme, recognising that this is interpreted within the scheme as the FSD of optical depth, δ. In the case of liquid clouds with constant number concentration, effective radius, $r_e \propto \text{LWC}^{1/3}$ (Martin et al., 1994), where LWC is the liquid water content. Since in the geometric optics approximation $\delta \propto \text{LWC}/r_e$, we expect the FSD for optical depth to be around two-thirds the FSD for water content.

There are significant problems with the optical properties of both liquid droplets and ice particles, discussed in section 6.1 and 6.2, respectively. Section 6.2 also discusses the impact of representing longwave scattering. Then in section 6.3 we investigate a detail of the way delta-Eddington scaling of the scattering phase function is implemented.

6.1 Optical properties of liquid clouds

The default liquid cloud optical properties in IFS Cycle 43R1 are from Slingo (1989) in the shortwave and Lindner and Li (2000) in the longwave. However, previous studies have found that the Slingo (1989)
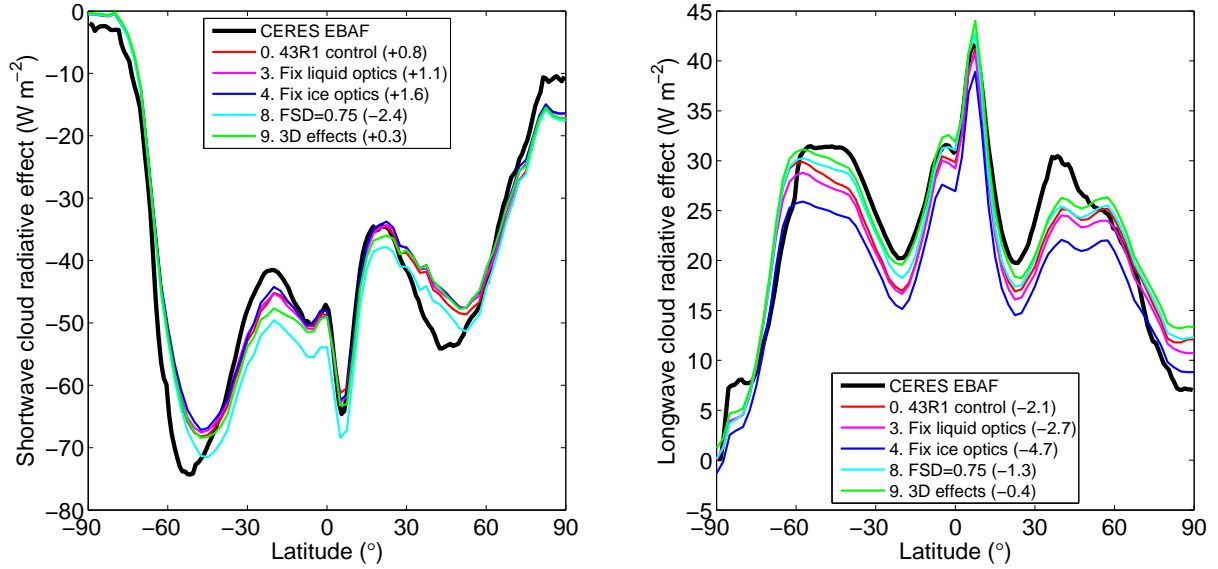
Figure 8: Zonally averaged (a) shortwave and (b) longwave cloud radiative effect from the CERES instrument (the EBAF dataset) and from selected coupled simulations shown in Table 2 and Fig. 4. The values in brackets are the differences in W m^{-2} between the global mean value from the simulation and the value from CERES.

scheme tends to overestimate cloud optical depth (Nielsen et al., 2014) and its simple dependence of optical properties on droplet effective radius, r_e, leads to poor agreement with Mie theory for larger cloud droplets (Dobbie et al., 1999). Schemes that overcome these limitations are available but are typically presented in the literature in terms of coefficients for radiation schemes with considerably fewer spectral bands than RRTM-G, while some are unnecessarily computationally costly via their use of wavelength-dependent non-integer powers of effective radius. We therefore adopt the solution used by the Met Office's SOCRATES radiation scheme (Suite of Community Radiation Codes based on Edwards and Slingo, 1996) in which the optical properties in all shortwave and longwave bands are expressed in terms of a Padé-like approximation of the following form:

$$K_{\text{ext}} = \frac{p_1 + p_2 r_e + p_3 r_e^2}{1 + p_4 r_e + p_5 r_e^2 + p_6 r_e^3}; \tag{6}$$

$$\omega = 1 - \frac{p_7 + p_8 r_e + p_9 r_e^2}{1 + p_{10} r_e + p_{11} r_e^2}; \tag{7}$$

$$g = \frac{p_{12} + p_{13} r_e + p_{14} r_e^2}{1 + p_{15} r_e + p_{16} r_e^2}, \tag{8}$$

where K_{ext} is the mass-extinction coefficient (the extinction cross-section per unit mass of cloud water), ω is the single-scattering albedo and g is the asymmetry factor. We have used the SOCRATES support software to generate the coefficients p_1-p_{16} in each of the 30 RRTM-G shortwave and longwave bands. The size distributions were as described by Rockel et al. (1991).

Item 2 in Table 2 and Fig. 4 indicates the impact of changing from the old to the new radiation scheme while retaining the same cloud optics scheme. The 1.8 W m^{-2} reduction in global-mean net downward shortwave radiation at top-of-atmosphere and the surface can be attributed to differences in the McICA cloud generator described in section 7.2. Item 3 then shows the impact of changing from the Slingo (1989) and Lindner and Li (2000) parameterizations to the new SOCRATES parameterization that provides a much better fit to Mie scattering calculations. It increases global-mean net downward shortwave

radiation by around 1.6 W m^{-2}, thereby almost completely cancelling out the change due to the new McICA cloud generator. This entry represents the configuration of the new scheme that leads to the least meteorological impact compared to the old scheme. This is also indicated in Fig. 8, which compares zonally averaged shortwave and longwave cloud radiative effect to values from the CERES instrument; the magenta line corresponding to item 3 is very close to the red line indicating the control experiment.

6.2 Optical properties of ice clouds and longwave scattering

For computing the optical properties of ice clouds, the current radiation scheme uses by default the parameterizations of Fu (1996) in the shortwave and Fu et al. (1998) in the longwave. In the longwave, for ice clouds only, it uses the Chou et al. (1999) method to modify the optical properties to partially offset the bias due to the neglect of longwave scattering, whereby the effective absorption optical depth used in the scheme in each layer and g-point is $\delta_{\text{abs}} = (1 - \omega f)\delta$, where δ is the true optical depth, ω is the single scattering albedo and f is the factor proposed by Chou et al. (1999). The factor f has a value of around 0.83 for typical ice particles, or a value of 1 if the Chou et al. (1999) parameterization is not used (which makes the assumption that radiation that would have been scattered is neither scattered nor absorbed).

It has been known for a few years that the implementation of the longwave ice optical properties has a bug in that the single scattering albedo (ω) used by the scheme is actually one minus what it ought to be, but that fixing it has a detrimental impact on forecasts. In much of the thermal infrared spectrum, $\omega \sim 0.5$, so this bug has only a small impact, but at shorter wavelengths in the longwave spectrum, ω can be significantly higher than 0.5. Thus the effect of the bug is to increase the effective optical depth of the cloud, and to fix it is to reduce the optical depth of the cloud.

The Fu (1996) and Fu et al. (1998) parameterizations are available in the new scheme. To allow for consistency with the old scheme, the new scheme includes the option to reproduce the bug, and indeed that option has been used in the examples shown from the new scheme until this point, although the Chou et al. (1999) approximation has not been implemented. Item 4 in Table 2 and Fig. 4 indicates that the impact of fixing the ice optics bug is to reduce the downwelling longwave at the surface and to cool the land surface by 0.25 K, which would exacerbate the existing cold bias of the model. Fig. 8 shows that this also leads to a poorer agreement in top-of-atmosphere longwave cloud radiative effect compared to measurements from the CERES instrument.

However, the next entry in Table 2 and Fig. 4 indicates that if longwave scattering by clouds is switched on then the ensuing warming effect at the surface would largely compensate the cooling from fixing the bug. Physically, the introduction of scattering results in optically thick clouds no longer behaving as a black body but rather reflecting a certain fraction of incident radiation. Thus, the introduction of scattering leads to clouds reflecting some of the upwelling longwave radiation from the surface back down, thereby increasing the downwelling radiation at the surface compared to the situation in which the only radiation emanating from the base of the cloud was due to emission. The impact of longwave scattering on the global radiation budget was studied by Costa and Shine (2006).

These results show that the combination of fixing the bug and introducing longwave scattering is a viable configuration for operational use. Representation of longwave scattering introduces an additional computational cost, but Fig. 9 shows that this is quite modest if longwave scattering is neglected for aerosols in clear-sky profiles. However, there have been a number of advances in modelling ice-particle scattering since the Fu parameterizations were proposed, notably the recognition that the facets of ice particles need to be treated as rough in order to match measurements (e.g. Yi et al., 2013). The new
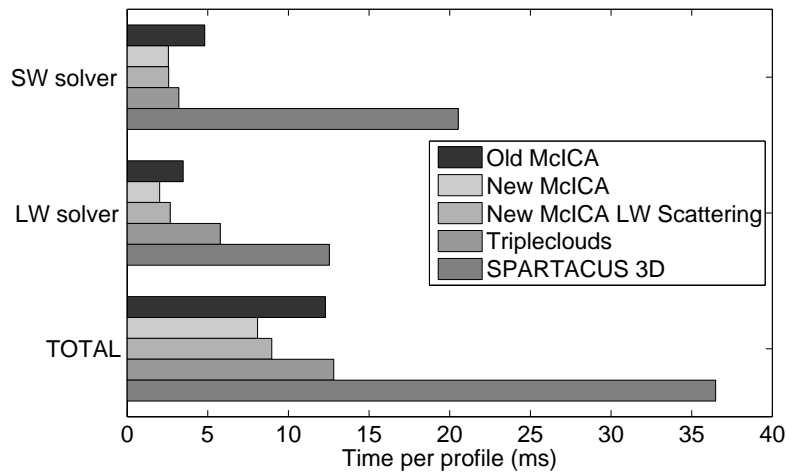
Figure 9: Comparison of the computational cost of the old scheme with the McICA solver to the new scheme with four different solver configurations: McICA without longwave scattering (as in Fig. 3), McICA with longwave scattering, Triplecloude (Shonk and Hogan, 2008) with longwave scattering and SPARTACUS (Hogan et al., 2016) with longwave scattering and 3D effects. Note that the Tegen aerosol climatology does not include the properties needed to represent longwave scattering, which the new McICA scheme exploits by only performing the more expensive calculations including longwave scattering on cloudy profiles..

radiation scheme has the option to use alternative parameterizations for ice optical properties based on the work of Baran et al. (2014), but these are experimental and will be tested more thoroughly in future work. Therefore, the first implementation of the new radiation scheme in the IFS uses the Fu parameterizations, reproducing the longwave bug found in the old scheme and not including longwave scattering.

6.3 Ordering of delta-Eddington scaling

When particles are much larger than the wavelength of the radiation, a significant fraction of the radiation that they intercept is scattered by only a small angle. For two-stream schemes that have only a very crude description of scattering directions, it is better to consider such forward-scattered radiation as if it had not been scattered at all. Joseph et al. (1976) proposed the 'delta-Eddington' method to scale down the optical depth to account for this effect, as well as adjusting the single-scattering albedo and asymmetry factor g. They proposed that a fraction $f = g^2$ of the scattered radiation should be treated as if it had not been scattered at all. Liquid droplets at visible wavelengths have $g \simeq 0.85$, implying a 72% reduction in the effective optical depth of liquid clouds. This is an essential ingredient for a two-stream radiation scheme to be accurate.

When clouds, aerosols and gases are all present in a volume, the asymmetry factor of the mixture is found by averaging the contribution from each constituent weighted by its scattering coefficient. Delta-Eddington scaling has no effect on Rayleigh-scattering gases since they have $g = 0$. However, the non-linear dependence of f on g means that it makes a difference whether delta-Eddington scaling is performed on the clouds and aerosols before merging with gases, or whether the scaling is applied after merging. It seems more physically justified to do the former, since it is only clouds and aerosols that have a forward lobe that needs to be treated, and Joseph et al. (1976) developed their theory primarily without considering mixtures containing Rayleigh-scattering gases. However, the current IFS radiation

scheme performs delta-Eddington scaling on the mixture.

To illustrate the potential effect of this difference, consider a mixture of liquid cloud ($g = 0.85$) and Rayleigh-scattering gas ($g = 0$) of equal optical depth, at a wavelength for which absorption is negligible. If delta-Eddington scaling is applied to the cloud first then its optical depth is reduced by 72%, which means that the optical depth of the mixture is reduced by 36%. If instead the constituents are merged first then the mixture has $g = 0.43$ so the optical depth of the mixture is reduced by only 18%. The delta-Eddington-scaled asymmetry factor is also different depending on the order of scaling and merging, but it is different in such a way as to partially ameliorate the difference in scaled optical depth.

Item 6 in Table 2 and Fig. 4 shows the impact of changing to a scheme where delta-Eddington is performed only on particles. The reduced effective optical depth of the atmosphere leads to an increase in net downwards shortwave flux at the surface and top-of-atmosphere, but only by around 0.1 W m^{-2}. The impact on other variables is not statistically significant. Nonetheless, since this change is believed to be more accurate, subsequent changes in later sections are built on top of this change.

7 New implementation of McICA

Up to now we have considered the impact of clouds on radiation only in terms of particle scattering properties; now we turn to the macroscopic properties of clouds. The host model provides a profile of cloud fraction to its radiation scheme, yet radiation schemes show significant sensitivity to the assumptions they make on both the degree to which clouds overlap in the vertical, and the in-cloud horizontal heterogeneity (e.g. Shonk et al., 2012). Both these properties are represented via the McICA solver, at the heart of which is a cloud generator that stochastically generates cloud profiles intended to be consistent with these assumptions. The McICA solver has been rewritten in the new radiation scheme to improve its efficiency and physical realism as well as reducing noise. Section 7.1 describes some of the shortcomings of the current McICA scheme, with a particular focus on cloud overlap, and section 7.2 describes how they are overcome by the cloud generator in the new scheme. Section 7.3 demonstrates the reduction in noise in heating rates and how this leads to a small improvement in medium-range forecast skill. Section 7.4 then describes the impact of using a more realistic treatment of cloud overlap and cloud horizontal heterogeneity.

7.1 Cloud overlap in the current implementation of McICA

Observations by cloud radar and lidar (Hogan and Illingworth, 2000; Di Giuseppe and Tompkins, 2015) have found that cloud occurrence in different vertical layers is correlated but that the correlation coefficient decreases approximately exponentially as the layer separation is increased with a decorrelation distance z_0. If there is a clear layer between these two cloudy layers then their overlap becomes random. This model for cloud overlap is therefore commonly known as exponential-random (hereafter 'EXP-RAN'). It is a generalization of the older maximum-random (hereafter 'MAX-RAN') assumption of Geleyn and Hollingsworth (1979) obtained in the limit of $z_0 \to \infty$.

The current McICA implementation generates sub-columns that are intended to implement EXP-RAN overlap stochastically using the Räisänen et al. (2004) cloud generator. It generates a sub-column as follows: starting in the first cloudy layer i measured down from the top of the atmosphere, a random number is generated R_i with a uniform probability between 0 and 1 (hereafter all random numbers are taken from the same distribution). If it is less than the cloud fraction at that height, a_i, then cloud
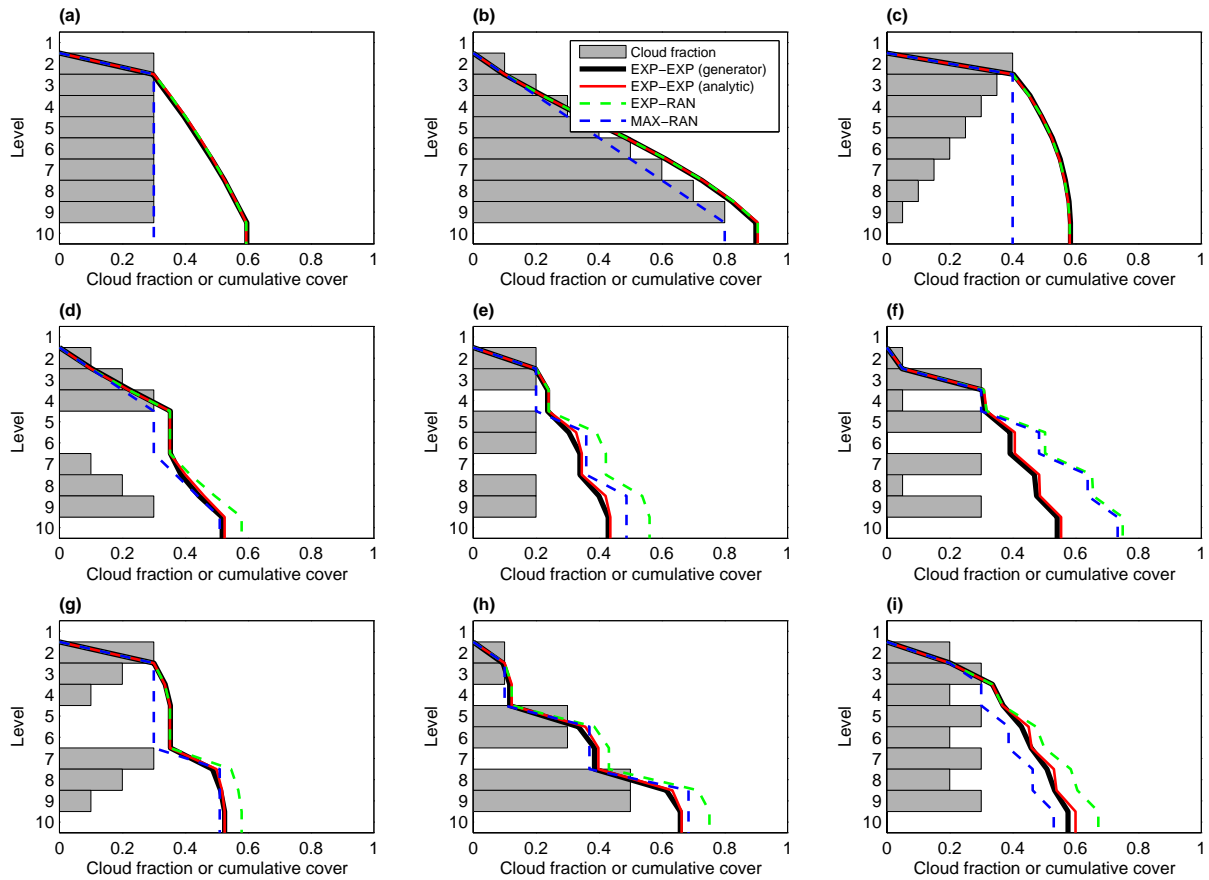
Figure 10: Comparison of different cloud overlap schemes for nine idealized cloud profiles. The cloud fraction is indicated by the grey bars, while the cumulative cloud cover down to each layer interface is shown by the various lines. The black 'EXP-EXP' shows the result of the Räisänen et al. (2004) cloud generator as used in the IFS, in which cloud occurrence is determined by random numbers that are correlated from layer to layer with a correlation coefficient of $\alpha = 0.75$ regardless of whether a clear-sky layer is passed. The lines shown have been averaged over 10,000 sub-columns. The red 'EXP-EXP' line shows the new analytic method that uses the same correlation. 'MAX-RAN' shows the result of the maximum-random formula of Geleyn and Hollingsworth (1979) while 'EXP-RAN' shows the result of the exponential-random formulation (Eqs. 10 and 11), using the same correlation coefficient for vertically continuous clouds.

is deemed to be present. We need to then generate a number R_{i+1} at the next layer that is partially correlated with R_i. In an EXP-RAN scheme, the correlation between layers i and $i + 1$, with a spacing Δz is given by $\alpha = \exp(-\Delta z/z_0)$ (Hogan and Illingworth, 2000). To implement this correlation in the generator, a new random number is generated, S. If $S < \alpha$ then $R_{i+1} = R_i$, otherwise a new random number is used for R_{i+1}. This process is continued down through the atmosphere and then repeated for each of the sub-columns required, one for each spectral g-point.

There are three main shortcomings of this approach. Firstly, contrary to what has been found in observations, the overlap of vertically separated clouds produced by this generator is not random. Randomness could be imposed by setting $\alpha = 0$ between any pair of layers where the cloud fraction in either is zero, but this has not been done. Therefore the scheme does not implement EXP-RAN, but rather 'EXP-EXP': clouds are correlated in the vertical regardless of whether the clouds are vertically separated, essentially the same assumption as made by Bergman and Rasch (2002). This problem was highlighted by Adrian Tompkins (personal communication). It is illustrated in Fig. 10 for nine different cloud-fraction profiles.

The thick black lines depict the total cloud cover measured down from top-of-atmosphere, according to this cloud generator. The value at the lowest level therefore represents the total cloud cover. For comparison the dashed lines show the MAX-RAN and EXP-RAN profiles (the latter with the same value of adjacent-layer α as EXP-EXP). We see that not only does EXP-EXP systematically underestimate cloud cover compared to EXP-RAN, in many situations it predicts a lower cloud cover even than MAX-RAN.

The second shortcoming is that the fluxes and heating-rate profiles can be quite noisy, especially when the total cloud cover is low. This is because if the cloud cover is, say, 0.2 then on average only 20% of the g-points will see any cloud. Since these g-points will be random, they may be associated with preferentially high gas absorption, resulting in the cloud radiative effect being underestimated, or preferentially low gas absorption resulting in the cloud effect being overestimated. Moreover, in this example the (on-average) 80% of the g-points computing clear-sky columns will duplicate the clear-sky calculations that are carried out for diagnostic purposes already. One approach to overcome this problem is to use all the g-points on sub-columns containing cloud, and then then perform a weighted average with the clear-sky calculation according to the total cloud cover to obtain the total-sky flux profile. Räisänen et al. (2004) proposed generating more sub-columns and simply throwing away the ones that do not contain cloud, although this could become expensive: for a cloud cover of 0.2, five times as many sub-columns would need to be generated. While it has been found that unbiased noise in radiative fluxes has no measurable impact on climate models, it has been found to degrade weather forecasts (Hill et al., 2011).

The third shortcoming is computational cost: even with the simple scheme described above, many more random numbers are generated than are actually needed. We next describe an approach that is computationally more efficient, generates less noise, and can implement MAX-RAN, EXP-RAN or EXP-EXP overlap assumptions.

7.2   Description of new cloud generator

In partially cloudy situations we wish to use all the available g-points on cloudy sub-columns, but without generating more columns than are needed and throwing away the clear ones. This is achieved by first calculating $c_{i+1/2}$, the profile of cumulative cloud cover measured from top-of-atmosphere down to a particular half-level $i+1/2$, i.e. the lines shown in Fig. 10. Thus for an n-layer atmosphere, the total cloud cover is $C = c_{n+1/2}$. We also require $p_{i+1/2}$, the combined cloud cover of an adjacent pair of layers i and $i+1$. MAX-RAN overlap is obtained with (Geleyn and Hollingsworth, 1979; Morcrette and Jakob, 2000)

$$p_{i-1/2} = \max(a_{i-1}, a_i); \tag{9}$$

$$(1 - c_{i+1/2}) = (1 - c_{i-1/2})\frac{1 - p_{i-1/2}}{1 - a_{i-1}}. \tag{10}$$

where a_i is the cloud fraction in layer i. These equations are applied recursively down through the atmosphere. They may also be used for cloud-cover diagnostics that are computed on the model grid, outside the radiation scheme.

At this point it is worth clarifying a subtle aspect about this implementation of MAX-RAN. Consider two 3-layer profiles of cloud fraction: (a) 0.5, 0.5, 0.5 and (b) 0.5, 0, 0.5. Any MAX-RAN implementation would diagnose a total cloud cover of 0.5 for the first profile and 0.75 for the second. But what should happen to total cloud cover if we vary the cloud fraction of the middle layer? The definition of MAX-RAN would suggest that if there is even a very small amount of cloud in the middle layer, the layers constitute a vertically continuous cloud profile and so should be treated as maximally overlapped with
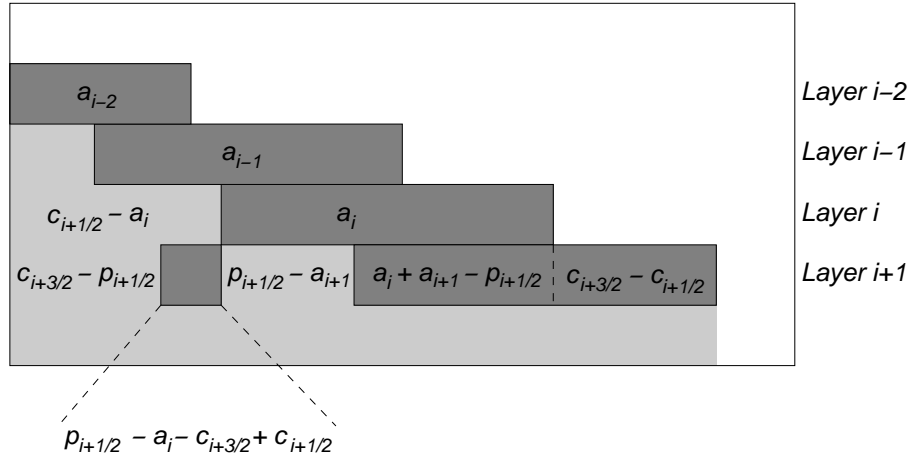
Figure 11: Schematic illustrating how the new McICA cloud generator is implemented, and in particular how to compute the conditional probabilities of cloud in layer $i+1$ dependent on whether cloud is present in layer i. The variables have the following definitions: a_i is the cloud fraction in layer i, $p_{i+1/2}$ is the combined cloud cover of layers i and $i+1$ and $c_{i+1/2}$ is the cloud cover of all the layers from top-of-atmosphere down to half-level $i+1/2$.

a total cloud cover of 0.5. Then if the middle cloud fraction is reduced to exactly zero then the total cloud cover jumps immediately to 0.75. By contrast, the equations above predict a smooth transition between 0.5 and 0.75 as the middle cloud fraction is reduced. Essentially MAX-RAN is being applied at the sub-grid scale: for a cloud fraction profile of 0.5, 0.25, 0.5, the equations treat 0.25 of the entire column as maximally overlapped, with the remaining 0.75 of the gridbox containing cloud only in the top and bottom layers; the clouds in this fraction of the gridbox is therefore treated as randomly overlapped. We regard this behaviour as preferable not only because of the simplicity of the equations, but because the continuous radiative response to a change in cloud fraction at any level is desirable should such an overlap scheme be used in a data assimilation context in future.

EXP-RAN overlap is obtained simply by replacing (9) with

$$p_{i-1/2} = \alpha_{i-1/2} \max(a_{i-1}, a_i) + (1 - \alpha_{i-1/2})(a_{i-1} + a_i - a_{i-1}a_i). \tag{11}$$

This equation also leads to the property that total cloud cover varies continuously in response to a change to cloud fraction at any individual level.

In the case of EXP-EXP, we do not believe that exact formulas exist for p and c that are consistent with the Räisänen et al. (2004) cloud generator, and for this reason, the cloud-cover diagnostic in the model currently uses the same expensive stochastic cloud generator as used in the radiation scheme (see section 9). In the Appendix we describe a new approximate but deterministic method to compute p and c consistent with the EXP-EXP overlap assumption. The cumulative cloud cover profiles it produces (red lines) in Fig. 10 closely match the average values from the cloud generator (black lines).

With the c function describing the cumulative cloud cover for the chosen overlap scheme, it is straight-forward to generate only cloudy sub-columns. We use the first random number in each sub-column, R_0, to work out the highest layer containing cloud: layer i is deemed to be the highest cloudy layer if $c_{i-1/2}/C < R_0 \leq c_{i+1/2}/C$. This way we significantly reduce the number of random numbers generated, since none are needed above the highest cloudy layer in a sub-column. We proceed to layer $i+1$ below and generate a further random number R_{i+1} to decide whether it is cloudy. Figure 11 shows the area fractions of cloudy and clear regions in layers i and $i+1$. If there is a cloud present in layer i then we wish to know the conditional probability of cloud in layer $i+1$ given cloud in layer i. From Fig. 11 this
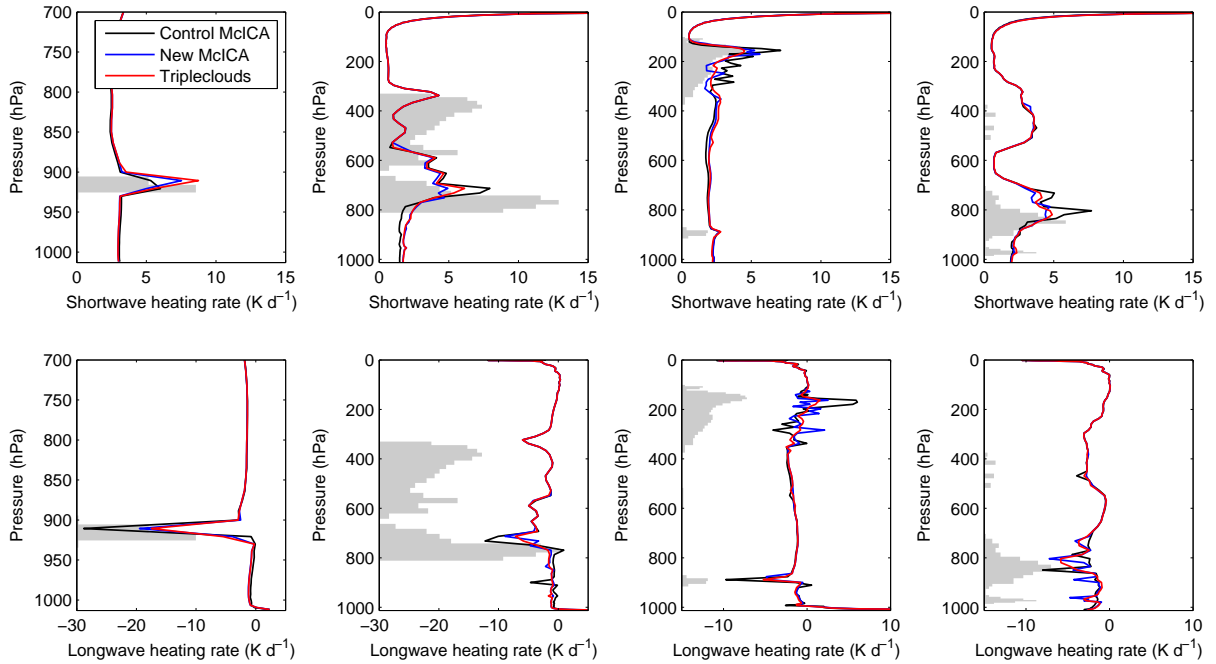
Figure 12: *Comparison of instantaneous shortwave (top row) and longwave (bottom row) heating rates between the old McICA radiation scheme (black line), the new McICA radiation scheme (blue line) and the new scheme using the Tripleclouds solver (red line). Each column is a different profile taken from the first timestep of a T255 model simulation. The grey shading shows the cloud fraction profile ranging from 0 to 1 across the axes.*

can be worked out, leading to cloud being generated in layer $i+1$ if

$$R_{i+1} < \frac{a_i + a_{i+1} - p_{i+1/2}}{a_i}. \tag{12}$$

If cloud is not present in layer i (but there is cloud in a layer above) then cloud should be placed in layer $i+1$ if

$$R_{i+1} < \frac{p_{i+1/2} - a_i - c_{i+3/2} + c_{i+1/2}}{c_{i+1/2} - a_i} \tag{13}$$

This process is continued to the lowest cloudy layer in the sub-column.

7.3   Impact of reduced noise in heating rates

Figure 12 compares shortwave and longwave heating-rate profiles in four partially cloudy columns from the IFS. The three lines show the existing (control) McICA implementation, the new McICA implementation described in section 7.2, and the Tripleclouds solver (see section 8.1) which generates no stochastic noise (although is somewhat more computationally expensive than McICA). It is apparent that the new McICA implementation is significantly less noisy than the old, and the difference between the old McICA implementation and Tripleclouds can be as much as 10 K d^{-1}. This would be expected to lead to some degradation in forecasts, and Fig. 13 shows that the new scheme does indeed lead to a slight reduction in tropospheric temperature errors in medium-range forecasts. These results are for a model configuration in which the radiation scheme is called only every 3 h, so heating-rate errors in the original scheme persist for this time which exacerbates the problem. The new scheme is also found to reduce

root-mean-squared temperature errors in model configurations when the radiation scheme is called every one hour; the reduction is a little less but still significant (not shown).

Figure 3 shows that the computational cost of the new scheme is around two-thirds of the old, so the radiation scheme may be called every 2 h instead of every 3 h for around the same overall cost. The red lines in Fig. 13 shows that this leads to a more significant improvement in forecast skill at all latitudes and heights.

7.4 Impact of cloud structure assumptions

We are now in a position to investigate the impact of changing the various assumptions on the sub-grid distribution of clouds, specifically the cloud overlap assumption, and the assumed width and shape of the horizontal cloud water distribution. As discussed in section 7.1, the current McICA scheme implements an EXP-EXP overlap assumption, while observations support EXP-RAN. Item 7 in Table 2 and Fig. 4 shows the impact of changing to the more realistic EXP-RAN overlap scheme while retaining the existing parameterization of the overlap vertical decorrelation length as a function of latitude, proposed by Shonk et al. (2010). We see that global mean cloud cover increases by around 0.03, which in turn increases the shortwave reflectance by around 1.4 W m^{-2}. If instead the MAX-RAN overlap assumption is used (item 7a) then global mean cloud cover increases by 0.014. This highlights how the current EXP-EXP scheme underestimates cloud cover even with respect to MAX-RAN.

Next we consider the impact of changing the width of the horizontal distribution of sub-grid water content. This is described in terms of the fractional standard deviation (FSD) of in-cloud water content, and the existing scheme assumes a gamma distribution with FSD = 1. Shonk et al. (2010) reviewed a number of observational papers and reported that, while there is a wide spread for different cloud types, a typical global value was FSD = 0.75. Item 8 in Table 2 and Fig. 4 shows that reducing FSD to 0.75 globally increases reflected shortwave radiation by 2.7 W m^{-2}. Figure 8 suggests that this leads to poorer agreement with the CERES instrument. Figure 4 shows that the combination of using EXP-RAN overlap and a lower value of FSD leads to a mean cooling of global land areas by 0.2 K, which if implemented would exacerbate the existing cold bias even though these changes would bring assumptions on cloud structure closer to observed values, on average. In section 8.2 we present evidence that the inclusion of 3D radiative effects could help resolve this issue. There is also a need to implement more sophisticated treatements of cloud overlap and heterogeneity, for example cloud overlap as a function of wind shear (Di Giuseppe and Tompkins, 2015) and regime-dependent FSD (Ahlgrimm and Forbes, 2016).

Lastly we investigate the impact of switching from a gamma distribution to a lognormal distribution of cloud water content, while keeping FSD at 0.75. Entry 9a of Table 2 shows that this leads to a modest increase in shortwave reflectance and an associated cooling. We are not aware of observational evidence supporting one distribution shape over another; Hogan and Illingworth (2003) reported that horizontal distributions of ice water content were equally well described by a gamma as a lognormal distribution.

8 Alternative solvers

The new radiation scheme has three alternative solvers to McICA that we now outline. The first is the 'Homogeneous' solver: it ignores cloud fraction and treats cloud water to be homogeneously mixed horizontally within each gridbox. This solver is not suited to NWP or climate models, but is useful for offline calculations on high-resolution cloud fields using the Independent Column Approximation. The
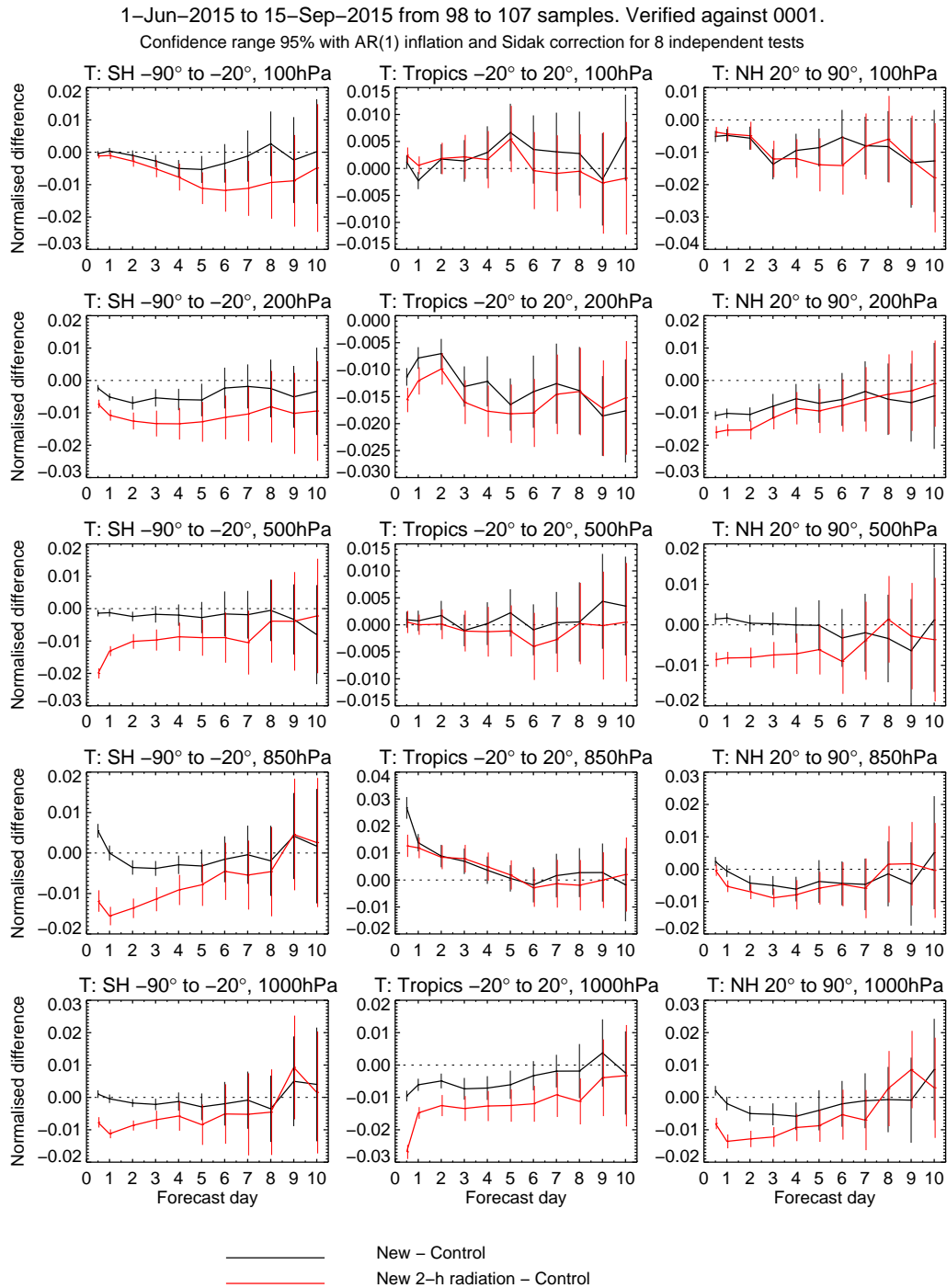
Figure 13: *Difference in root-mean-squared tropospheric temperature errors (such that* −0.01 *implies a 1% improvement) between the new radiation scheme scheme and the old, computed from 3.5 months of 10-day T399 forecasts verified against the ECMWF operational analyses. The black lines show the results when both the old and new radiation schemes are called every 3 h, while the red lines show the results when the new scheme is called every 2 h (exploiting the fact that the new scheme is only around two-thirds the computational cost of the old).*

other two solvers are described in the following subsections.

8.1 The Tripleclouds solver

The Tripleclouds solver was devised by Shonk and Hogan (2008): it divides each cloudy model layer into three regions, a clear-sky region and two cloudy regions. The two cloudy regions can be thought of as a two-point approximation to the sub-grid cloud optical depth distribution, with one region representing the optically thin half of the cloud and the other the optically thick half. It was found by Shonk and Hogan (2008) that the Tripleclouds approach provided a good approximation to the radiative impact of cloud heterogeneity if the optically thin region had an optical depth equal to the 16th percentile of the distribution, and the optically thick region then had the optical depth needed for the in-cloud mean to be correct. The two-stream equations are solved by using EXP-RAN overlap rules to map the fluxes exiting the three regions of one layer into fluxes entering the three regions of an adjacent layer. MAX-RAN overlap is therefore possible as it is a limiting case of EXP-RAN. EXP-EXP overlap cannot be represented by Tripleclouds because clear-sky layers are treated as a single region that fills the gridbox horizontally, and therefore fluxes passing through this region have no memory of the configuration of clouds in layers they originated from. In effect, this enforces random overlap for clouds separated by a clear-sky layer.

In terms of performance, Fig. 9 shows that Tripleclouds is slower than the new implementation of McICA although it is around the same speed as the old McICA implementation. It can be seen in Fig. 12 that unlike McICA it does not produce any noise in heating rates. This advantage would become more significant in future if an alternative gas-optics model were developed with fewer spectral intervals than RRTM-G, since this would increase the noise in McICA but not affect Tripleclouds. Item 9b in Table 2 shows that the impact on fluxes compared to the McICA solver with the same assumptions on cloud overlap and heterogeneity is to reduce the Earth's reflectance by around 0.4 W m^{-2}, resulting in a small warming at the surface.

8.2 The SPARTACUS solver

The SPARTACUS solver (SPeedy Algorithm for Radiative TrAnsfer through CloUd Sides) is based on the Tripleclouds division of a model layer into one clear-sky and two cloudy regions, but adds additional terms to the equations to represent lateral fluxes between regions (Hogan et al., 2016). It therefore makes this the only radiation scheme currently available in an NWP or climate model that can represent shortwave and longwave 3D radiative effects associated with clouds. Three-dimensional effects have been studied for a long time in the shortwave, but have been largely neglected in the longwave. The development of the longwave capability of SPARTACUS was described by Schäfer et al. (2016).

Figure 14 shows results from Hogan et al. (2016): a comparison between the new radiation scheme and benchmark shortwave calculations from the libRadtran package, for the 'I3RC' high-resolution cumulus test case. The dark blue lines show the results of applying the Independent Column Approximation (ICA), i.e. running the models separately on each column of the scene and therefore neglecting 3D transport. The agreement is excellent. The cyan dashed lines show that the much faster McICA scheme (fed with the cloud fraction profile from the scene) gives an unbiased approximation to ICA. The red lines compare fully 3D calculations from the libRadtran's Monte Carlo model to the much faster calculations using the SPARTACUS solver. We see that SPARTACUS agrees well with fully 3D calculations, and also that there is a significant difference between ICA and 3D, as shown in the bottom two panels.
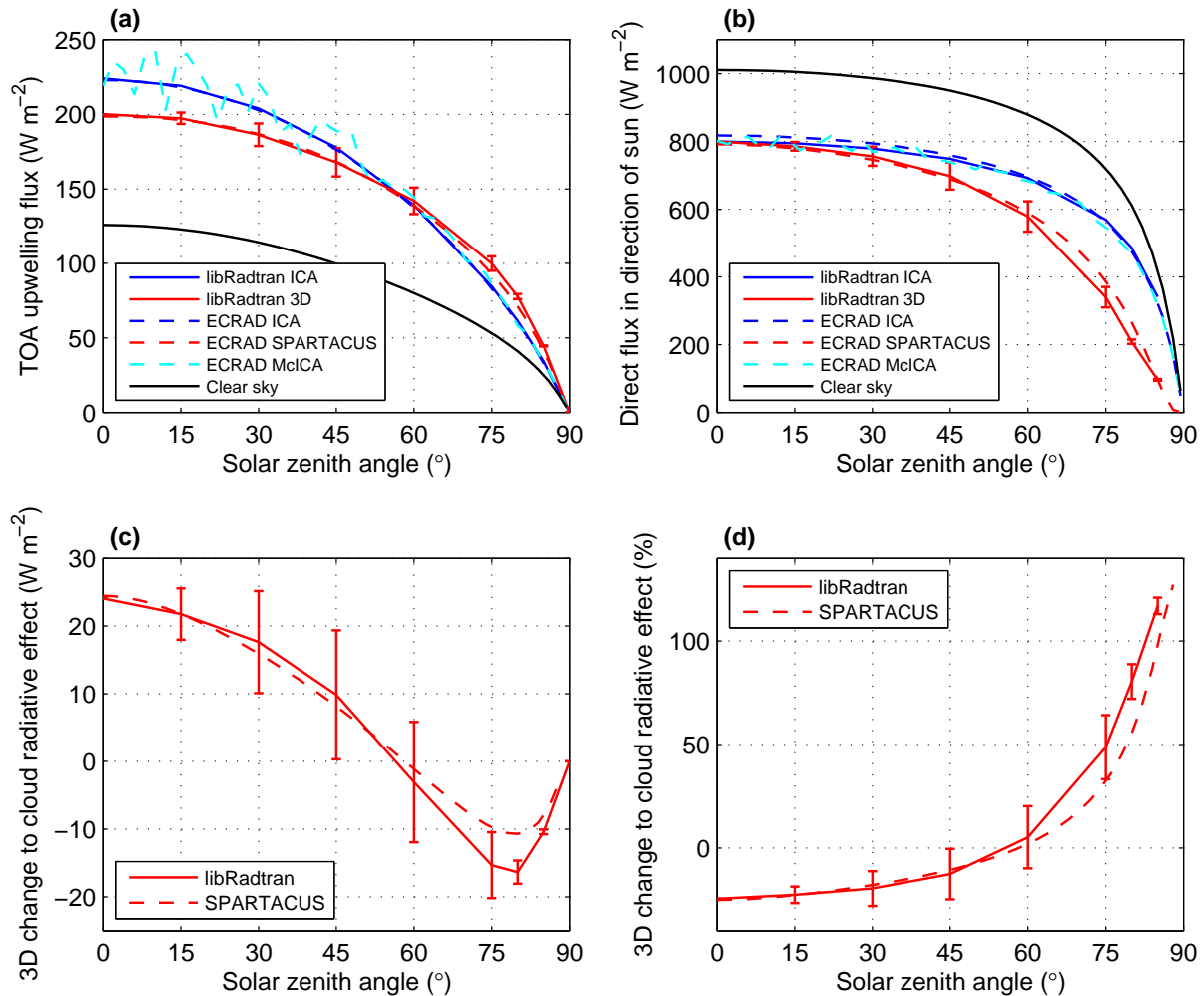
Figure 14: Comparison of simulated broadband shortwave fluxes from libRadtran (using the MYSTIC Monte Carlo solver for 3D calculations) and the new radiation scheme 'ECRAD' (using SPARTACUS for 3D calculations) for the I3RC cumulus test case, versus solar zenith angle. (a) Top-of-atmosphere (TOA) upwelling fluxes both including lateral transport (3D) and excluding it (ICA), along with the ECRAD McICA solver and the values for clear skies. (b) As panel a but comparing the direct component of surface solar flux into a plane perpendicular to the sun. (d) The change to TOA cloud radiative effect due to the inclusion of 3D effects. (d) As panel c but as a percentage. The MYSTIC 3D results are the average over four calculations with the solar azimuth angle 90° apart, and the error bars in each panel represent the standard deviation of these four simulations. Figure adapted from Hogan et al. (2016)..

The computational cost of the scheme is shown in Fig. 9; using the SPARTACUS solver, the new radiation scheme is 4.5 times slower, which is 3 times slower than the old radiation scheme. This is too slow to consider for operational use at the moment, but it is fast enough for detailed experimentation of the impact of a more sophisticated and hopefully more accurate representation of radiative transfer on the model climate and potentially on aspects of forecast performance.

SPARTACUS requires one additional piece of information on cloud geometry in each model layer: the effective length of cloud edge per unit area of gridbox. In Fig. 14, this was extracted from the high-resolution cloud scene, but in an atmospheric model we need a parameterization in terms of variables within the model. Schäfer (2016) showed that this may be conveniently described by an *effective cloud*

scale that is independent of cloud fraction. She analyzed a number of high-resolution observational and modelling datasets of three-dimensional cloud fields, finding that for both cumulus and stratocumulus, the effective cloud scale was around 1 km, but with a typical range of 0.7–1.4 km. For clouds above the boundary layer the effective cloud scale was around 10 km, but with typical values in the range 5–20 km. She then used these values in off-line radiation calculations on ERA-Interim cloud fields to estimate the impact of 3D radiative transfer on the global radiation budget. The change to fluxes was found to be in the sense of a warming of the Earth system in both the shortwave and the longwave, with a change at top-of-atmosphere of 3.0 W m^{-2} in the shortwave and 0.9 W m^{-2} in the longwave. The change at the surface was found to be 2.3 W m^{-2} in the shortwave and 2.1 W m^{-2} in the longwave.

Item 9 in Table 2 and Fig. 4 shows the impact on the climate of the coupled IFS model when SPARTACUS is run on-line with an effective cloud scale of 1 km for low clouds and 10 km for middle and high clouds. The changes to shortwave fluxes are very similar to those found by Schäfer (2016), while changes to net longwave fluxes are somewhat different because of the temperature response of the coupled model. Nonetheless, we see that the net downward flux at the surface is increased by around 2 W m^{-2} in both the longwave and the shortwave. This leads to a significant warming, and an increase in the mean 2-m temperature over land of around 0.6 K. (Note that this is on top of all previous changes to the representation of clouds, in particular the cloud structure changes described in section 7.4 that led to a cooling of mean 2-m temperature of 0.2 K compared to the control experiment using the old radiation scheme.) This change could help explain why the current operational model appears to be too cold at the surface, and yet the current treatments of cloud overlap and heterogeneity ought to lead to it letting too much solar radiation reach the surface. However, considerable further work needs to be done to further validate the SPARTACUS model in a wide range of cloud types, and to refine the appropriate values of effective cloud scale to use in a large-scale model. If this confirms that 3D effects are an important piece of missing physics then further work will be needed to optimize the SPARTACUS solver and wider aspects of the radiation scheme.

9 Diagnostics

The primary outputs from the radiation scheme are the shortwave and longwave upwelling and downwelling all-sky flux profiles at half-levels, which are used by the host model to compute the evolution of surface and atmospheric temperature. In addition, a number of diagnostics are required that are also stored in the output `flux` object shown in Fig. 1. The clear-sky flux profiles may be used to compute cloud radiative effect (shown in Fig. 8), an essential diagnostic of the model climate. In the new implementation of McICA, clear-sky fluxes come 'for free' since they are used to reduce noise in partially cloudy situations (see section 7.2).

The 'direct' downwelling shortwave flux at the surface is a diagnostic that enables the model to be evaluated against global/diffuse pyranometer measurements, and is used by some in the solar energy industry. The model also uses it to compute sunshine duration, which according to the World Meteorological Organization definition is the time for which the surface direct shortwave flux measured into a plane perpendicular to the solar beam exceeds 120 W m^{-2}. It should be stressed that, in common with most radiation schemes in NWP models, the direct-flux diagnostic is the same as used to solve the two-stream equations, and due to the use of delta-Eddington scaling (see section 6.3), it includes some forward scattered radiation.

The remaining surface diagnostics consist of solar fluxes from a particular spectral interval, specifically the Photosynthetically Active Radiation (clear-sky and all-sky) from the wavelength range 0.4–0.7 μm,

and ultraviolet radiation currently computed from the RRTM-G bands in the range 0.2–0.44 μm. In practice, the radiation scheme returns the clear-sky and all-sky surface downwelling fluxes in each short-wave band, and the interface routine in the IFS performs the weighted average of the available bands to get the wavelength interval required. In this way, additional spectral flux diagnostics could be computed (for example, in the passband of a particular photovoltaic cell) without needing to modify the radiation scheme itself.

Two further diagnostics are computed outside the radiation scheme but need to be consistent with it: cloud cover and visibility. In order to be consistent with EXP-EXP overlap, the current cloud-cover diagnostic uses a stochastic cloud generator with the same overlap rules as used in the radiation scheme. As well as being noisy, this is also expensive as it is computed on the full model grid rather than on the reduced radiation grid: in the current IFS it accounts for around 1.6% of the total model time. With the introduction of the new radiation scheme, the cloud-cover diagnostic now uses the faster deterministic formulation of EXP-EXP described in the Appendix, reducing this cost to 0.6% of total model time. Cloud cover assuming EXP-RAN overlap is even faster to compute. The visibility diagnostic is computed by interpolating the various species from the aerosol climatology on to the full model grid. Previously the Tegen optical properties were used for the visibility calculation, but now a routine from the new radiation scheme is called to compute the aerosol extinction coefficient taking account also of aerosol humidification. This ensures consistency between the aerosol optical properties used in the radiation scheme and in the visibility diagnostic.

10   Conclusions and future work

In this memorandum we have described a flexible new radiation scheme for the IFS. Its modular design makes it easily extensible, thereby providing a framework for radical changes to sub-components of the scheme in future. While the default configuration of the new scheme in its first operational implementation will be scientifically quite similar to the old, it will nonetheless bring several immediate benefits to the accuracy and efficiency of the model:

- The new scheme is two-thirds the computational cost of the old. This enables the numerous operational IFS configurations that currently can only afford to call the radiation scheme every 3 h to reduce the radiation timestep to 2 h, with a measurable increase in forecast skill.

- The new implementation of the cloud generator in the stochastic 'McICA' solver generates less noise in heating rates, leading to a further modest improvement in forecast skill.

- The longwave two-stream equations are now solved exactly, leading to a warming of the lower stratosphere and a cooling of the upper stratosphere, both of which reduce existing and longstanding biases in these regions.

- A fast, deterministic method has been developed to compute cumulative cloud cover consistent with the model's cloud overlap assumption. As well as being integral to the speed improvements and noise reduction mentioned above, this also significantly speeds up the model's cloud-cover diagnostic, which was previously computed using the same slow stochastic generator from the old McICA implementation.

- It fixes two bugs in the old scheme that led to an overestimate of the shortwave absorption and optical depth of liquid clouds (items 1 and 3 in Table 2).

A bug in the longwave optical properties of ice particles in the old scheme has been investigated and found to be largely compensated by the lack of longwave scattering, which the new scheme is optionally able to represent at modest additional computational cost. In the near future we plan to substantially revise the ice optics model and activate longwave scattering, but pending these improvements the new scheme has been configured to reproduce the old (including reimplementing the bug).

One of the solvers available in the new scheme is SPARTACUS, making the scheme the first in an NWP or climate model able to represent 3D cloud-radiation effects in both the shortwave and the longwave. While this solver is currently too slow to consider for operational use, and work to verify its accuracy in different cloud conditions is still ongoing, it does have the potential to provide a very useful benchmark that should help to prioritize radiation developments in the future. Initial IFS experiments with SPARTACUS reported in this memorandum suggest that clouds in McICA schemes are too reflective. This could help to explain the apparent cold bias in surface temperature in the IFS, despite the assumptions on cloud overlap and cloud heterogeneity being such that they ought to let too much solar radiation reach the surface. However, other factors need investigation in parallel, such as introducing a regime dependence to the width of the sub-grid cloud water distribution.

The separation of the gas-optics model (governing spectral discretization) from the other parts of the scheme will facilitate the possible future development of an alternative model with a configurable number of spectral intervals. This would allow exploration of the trade-offs between more accurate configurations with many spectral intervals (like RRTM-G), and faster configurations that can be called more frequently in time and on a higher resolution radiation grid. Moreover, if a fast enough configuration is available then the new radiation scheme could be used within the data assimilation system, requiring its tangent-linear and adjoint to be coded. The use of the same scheme in the minimization as in the nonlinear model should improve the quality of the analysis.

Acknowledgements

Appendix: Deterministic calculation of cumulative cloud cover for the EXP-EXP overlap assumption

This appendix describes a deterministic algorithm for calculating the profile of cumulative cloud cover $c_{i+1/2}$ (i.e. the cloud cover between the top-of-atmosphere and half-level $i+1/2$) that is approximately consistent with the EXP-EXP overlap assumption. Thus, for an n-level model the total cloud cover is $C = c_{n+1/2}$. This algorithm is used by both the new cloud generator within the radiation scheme (section 7.2) and the cloud cover diagnostic computed outside the radiation scheme (section 9), and consequently both are now significantly faster. The fidelity with which the algorithm reproduces the cloud cover produced by a stochastic EXP-EXP model in the limit of a very large number of sub-columns is indicated by the agreement between the red and black lines in Fig. 10.

The algorithm takes as input a profile of cloud fraction and first divides it into a number of cloud *objects:* a contiguous set of cloudy layers for which cloud fraction monotonically increases with height up to a certain layer and then monotonically decreases. Therefore there are as many cloud objects as there are maxima in the cloud-fraction profile. This task is straightforward.

The cumulative cloud cover profile is then computed separately for each object j, denoted as $c_{i+1/2}^j$. This is achieved by applying the EXP-RAN formulas (10) and (11) separately for each object, yielding also the total cloud cover of each object, C^j.

The next part of the algorithm is more involved: a pair of adjacent objects is combined into a compound object and its cloud cover is computed. This process is repeated until there is one object left, whose cloud cover is the total cloud cover of the column. We first compute the overlap parameter between each adjacent pair of objects. If objects j and $j+1$ have maximum cloud fractions in layers p and q, respectively, then the overlap parameter $\alpha^{j+1/2}$ describing the correlation between the two objects we assume to be given by

$$\alpha^{j+1/2} = \prod_{i=p}^{q-1} \alpha_{i+1/2}, \tag{14}$$

where $\alpha_{i+1/2}$ is the overlap parameter between layers i and $i+1$, and is computed from the decorrelation length and the layer separation as discussed in section 7.1. At any given step in the process, the next pair to be combined is the pair that is most correlated, i.e. have the largest value of $\alpha^{j+1/2}$.

The first part of combining two objects is to compute their combined cloud cover, C^{new}, by applying (11) but to adjacent objects rather than adjacent layers:

$$C^{\text{new}} = \alpha^{j+1/2} \max(C^j, C^{j+1}) + (1 - \alpha^{j+1/2})(C^j + C^{j+1} - C^j C^{j+1}). \tag{15}$$

The second part of the combination process is to compute the profile of cumulative cloud cover for the combined object, $c_{i+1/2}^{\text{new}}$. For the layers of the upper of the two objects (j), the values are the same:

$$c_{i+1/2}^{\text{new}} = c_{i+1/2}^j. \tag{16}$$

For the lower of the two objects ($j+1$), we need to account for the fact that it is partially overlapped from above by the upper object. This involves the following scaling:

$$c_{i+1/2}^{\text{new}} = C^j + c_{i+1/2}^{j+1} \frac{C^{\text{new}} - C^j}{C^{j+1}}. \tag{17}$$

Once all the objects have been merged, the final task is to compute the combined cloud cover of each pair of layers $p_{i-1/2}$, using (11). However, in some situations, EXP-EXP produces a profile of cumulative

cloud cover that is not consistent with this equation, so the following modification is required:

$$p_{i-1/2} \leftarrow \max\left(p_{i-1/2}, a_{i-1} + c_{i+1/2} - c_{i-1/2}\right), \tag{18}$$

which basically says that the combined cloud cover of the two layers must be at least as big as the cloud fraction of the upper layer (a_{i-1}) plus the amount of cloud in the lower layer that is exposed to space ($c_{i+1/2} - c_{i-1/2}$).

References

Ahlgrimm, M., and R. M. Forbes, 2016: Regime dependence of cloud condensate variability observed at the Atmospheric Radiation Measurement sites. *Q. J. R. Meteorol. Soc.,* **142,** 1605–1617.

Baran, A. J., P. Hill, K. Furtado, P. Field and J. Manners, 2014: A coupled cloud physics–radiation parameterization of the bulk optical properties of cirrus and its impact on the Met Office Unified Model Global Atmosphere 5.0 configuration. *J. Clim.,* **27,** 7725–7752.

Bergman, J. W., and P. J. Rasch, 2002: Parameterizing vertically coherent cloud distributions. *J. Atmos. Sci.,* **59,** 2165–2182.

Bozzo, A., R. Pincus, I. Sandu, and J.-J. Morcrette, 2015: Impact of a spectral sampling technique for radiation on ECMWF weather forecasts. *J. Adv. Model. Earth Syst.,* **6,** 1288–1300.

Bozzo, A., A. Benedetti, S. Remy, P. Bechtold, M. J. Rodwell and J.-J. Morcrette, 2016: Implementation of a CAMS-based aerosol climatology in the IFS. ECMWF Technical Memorandum *in preparation.*

Chou, M.-D., K.-T. Lee, S.-C. Tsay and Q. Fu, 1999: Parameterization for cloud longwave scattering for use in atmosheric models. *J. Clim.,* **12,** 159–169.

Costa, S. M. S., and K. P. Shine, 2006: An estimate of the global impact of multiple scattering by clouds on outgoing long-wave radiation. *Q. J. R. Meteorol. Soc.,* **132,** 885–895.

Cusack, S., J. M. Edwards and J. M. Crowther, 1999: Investigating k distribution methods for parameterizing gaseous absorption in the Hadley Centre Climate Model. *J. Geophys. Res.,* **104,** 2051–2057.

Dobbie, J. S., J. Li and P. Chýlek, 1999: Two and four stream optical properties for water clouds and solar wavelengths. *J. Geophys. Res.,* **104,** 2067–2079.

Edwards, J. M., and A. Slingo, 1996: Studies with a flexible new radiation code: 1. Choosing a configuration for a large-scale model. *Q. J. R. Meteorol. Soc.,* **122,** 689–719.

Di Giuseppe, F., and A. M. Tompkins, 2015: Generalizing cloud overlap treatment to include the effect of wind shear. *J. Atmos. Sci.,* **72,** 2865–2876.

Fu, Q., 1996: An accurate parameterization of the solar radiative properties of cirrus clouds. *J. Clim.,* **9,** 2058–2082.

Fu, Q., and K.-N. Liou, 1992: On the correlated k-distribution method for radiative transfer in nonhomogeneous atmospheres. *J. Atmos. Sci.,* **49,** 2139–2156.

Fu, Q., K. N. Liou, M. C. Cribb, T. P. Charlock and A. Grossman, 1997: Multiple scattering parameterization in thermal infrared radiative transfer. *J. Atmos. Sci.,* **54,** 2799–2812.

Fu, Q., P. Yang and W. B. Sun, 1998: An accurate parameterization of the infrared radiative properties of cirrus clouds of climate models. *J. Clim.,* **11,** 2223–2237.

Geleyn, J.-F., and A. Hollingsworth, 1979: An economical analytical method for the computation of the interaction between scattering and line absorption of radiation. *Beitr. Phys. Atmos.,* **52,** 1–16.

Hill, P. G., J. Manners and J. Petch, 2011: Reducing noise associated with the Monte Carlo Independent Column Approximation for weather forecasting models. *Q. J. R. Meteorol. Soc.,* **137,** 219–228.

Hogan, R. J., and A. J. Illingworth, 2003: Parameterizing ice cloud inhomogeneity and the overlap of inhomogeneities using cloud radar data. *J. Atmos. Sci.,* **60,** 756–767.

Hogan, R. J., and A. Bozzo, 2015: Mitigating errors in surface temperature forecasts using approximate radiation updates. *J. Adv. Modeling. Earth. Sys.,* **7,** 836–853 (see also ECMWF Technical Memorandum 746).

Hogan, R. J. and A. J. Illingworth, 2000: Deriving cloud overlap statistics from radar. *Q. J. R. Meteorol. Soc.,* **126,** 2903–2909.

Hogan, R. J., S. A. K. Schäfer, C. Klinger, J.-C. Chiu and B. Mayer, 2016: Representing 3D cloud-radiation effects in two-stream schemes: 2. Matrix formulation and broadband evaluation. *J. Geophys. Res.,* **121,** 8583–8599.

Joseph, J. H., W. J. Wiscombe and J. A. Weinman, 1976: The delta-Eddington approximation for radiative flux transfer. *J. Atmos. Sci.,* **33,** 2452–2459.

Lindner, T. H., and J. Li, 2000: Parameterization of the optical properties for water clouds in the infrared. *J. Clim.,* **13,** 1797–1805.

Martin, G. M., D. W. Johnson and A. Spice, 1994: The measurement and parameterization of effective radius of droplets in warm stratocumulus clouds. *J. Atmos. Sci.,* **51,** 1823–1842.

Mašek, J., J.-F. Geleyn, R. Brožková, O. Giot, H. O. Achom and P. Kuma, 2016: Single interval shortwave radiation scheme with parameterized optical saturation and spectral overlaps. *Q. J. R. Meteorol. Soc.,* **142,** 304–326.

Meador, W. E., and W. R. Weaver, 1980: Two-stream approximations to radiative transefer in planetary atmospheres: a unified description of existing methods and a new improvement. *J. Atmos. Sci.,* **37,** 630–643.

Mlawer, E. J., S. J. Taubman, P. D. Brown, M. J. Iacono, and S. A. Clough, 1997: Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave. *J. Geophys. Res.,* **102,** 16 663–16 682.

Morcrette, J.-J., 1991: Radiation and cloud radiative properties in the ECMWF operational weather forecast model. *J. Geophys. Res.,* **96,** 9121–9132.

Morcrette, J.-J., 2000: On the effects of the temporal and spatial sampling of radiation fields on the ECMWF forecasts and analyses. *Mon. Weath. Rev.,* **128,** 876–887.

Morcrette, J.-J., and C. Jakob, 2000: The response of the ECMWF model to changes in cloud overlap assumption. *Mon. Weath. Rev.,* **128,** 1707–1732.

Morcrette, J.-J., H. W. Barker, J. N. S. Cole, M. J. Iacono, R. Pincus, 2008: Impact of a new radiation package, McRad, in the ECMWF Integrated Forecasting System. *Mon. Weath. Rev.,* **136,** 4773–4798.

Nielsen, K. P., E. Gleeson and L. Rontu, 2014: Radiation sensitivity tests of the HARMONIE 37h1 NWP model. *Geosci. Model Dev.,* **7,** 1433–1449.

Pincus, R., H. W. Barker, and J.-J. Morcrette, 2003: A fast, flexible, approximate technique for computing radiative transfer in inhomogeneous clouds. *J. Geophys. Res.,* **108,** 4376, doi:10.1029/2002JD003322.

Räisänen, P., H. W. Barker, M. F. Khairoutdinov, J. Li and D. A. Randall, 2004: Stochastic generation of subgrid-scale cloudy columns for large-scale models. *Q. J. R. Meteorol. Soc.,* **130,** 2047–2067.

Rockel, B., E. Raschke and B. Weyres, 1991: A parametrization of broad band radiative transfer properties of water, ice and mixed clouds. *Beiträge Phys. Atmosph.,* **64,** 1–12.

Schäfer, S. A. K., 2016: *What is the global impact of 3D cloud-radiation interactions?* PhD thesis (submitted), University of Reading, UK.

Schäfer, S. A. K., R. J. Hogan, C. Klinger, J.-C. Chiu and B. Mayer, 2016: Representing 3D cloud-radiation effects in two-stream schemes: 1. Longwave considerations and effective cloud edge length. *J. Geophys. Res.,* **121,** 8567–8582.

Slingo, A., 1989: A GCM parameterization for the shortwave radiative properties of water clouds. *J. Atmos. Sci.,* **46,** 1419–1427.

Shonk, J. K. P., and R. J. Hogan, 2008: Tripleclouds: an efficient method for representing horizontal cloud inhomogeneity in 1D radiation schemes by using three regions at each height. *J. Climate,* **21,** 2352–2370.

Shonk, J. K. P., R. J. Hogan, J. M. Edwards and G. G. Mace, 2010: Effect of improving representation of horizontal and vertical cloud structure on the Earth's radiation budget: 1. Review and parameterisation. *Q. J. R. Meteorol. Soc.,* **136,** 1191–1204.

Shonk, J. K. P., R. J. Hogan and J. Manners, 2012: Impact of improved representation of horizontal and vertical cloud structure in a climate model. *Clim. Dyn.,* **38,** 2365–2376.

Steeneveld, G. J., M. J. J. Wokke, C. D. Groot Zwaaftink, S. Pijlman, B. G. Heusinkveld, A. F. G. Jacobs and A. A. M. Holtslag, 2010: Observations of the radiation divergence in the surface layer and its implication for its parameterization in numerical weather prediction models. *J. Geophys. Res.,* **115,** D06107.

Tegen, I., P. Hollrig, M. Chin, I. Fung, D. Jacob and J. Penner, 1997: Contribution of different aerosol species to the global aerosol extinction optical thickness: Estimates from model results. *J. Geophys. Res.,* **102,** 23895–23915.

Yi, B., P. Yang, B. A. Baum, T. LEcuyer, L. Oreopoulos, E. J. Mlawer, A. J. Heymsfield and K.-N. Liou, 2013: Influence of ice particle surface roughness on the global cloud radiative effect. *J. Atmos. Sci.,* **70,** 2794–2807.