A flexible and efficient radiation scheme for the ECMWF model

Article

www.reading.ac.uk/centaur

CentAUR

Key Points:
- A new radiation scheme for the ECMWF model is described that is 41% faster than the original scheme
- We describe how longwave scattering by clouds can be represented with only a 4% increase in computational cost, improving forecast skill
- A sequence of changes have reduced the long-standing warm bias in the middle to upper stratosphere of the ECMWF model

Correspondence to:
R. J. Hogan,
r.j.hogan@ecmwf.int

A Flexible and Efficient Radiation Scheme for the ECMWF Model

Robin J. Hogan[1] (ID) **and Alessio Bozzo**[1] (ID)

[1] European Centre for Medium-Range Weather Forecasts, Reading, UK

Abstract This paper describes a new radiation scheme *ecRad* for use both in the model of the European Centre for Medium-Range Weather Forecasts (ECMWF), and off-line for noncommercial research. Its modular structure allows the spectral resolution, the description of cloud and aerosol optical properties, and the solver, to be changed independently. The available solvers include the Monte Carlo Independent Column Approximation (McICA), *Tripleclouds*, and the Speedy Algorithm for Radiative Transfer through Cloud Sides (SPARTACUS), the latter which makes ECMWF the first global model capable of representing the 3-D radiative effects of clouds. The new implementation of the operational McICA solver produces less noise in atmospheric heating rates, and is 41% faster, which can yield indirect forecast skill improvements via calling the radiation scheme more frequently. We demonstrate how longwave scattering may be implemented for clouds but not aerosols, which is only 4% more computationally costly overall than neglecting longwave scattering and yields further modest forecast improvements. It is also shown how a sequence of radiation changes in the last few years has led to a substantial reduction in stratospheric temperature biases.

Plain Language Summary Solar and thermal infrared radiation provide the energy that drives weather systems and ultimately controls the Earth's climate. Accurately simulating these energy flows is therefore a crucial part of the computer models used for weather and climate prediction. This paper describes a flexible and efficient new software package, *ecRad*, for computing radiation exchange. It became operational in the forecast model of the European Centre for Medium-Range Weather Forecasts (ECMWF) in July 2017, and is 41% computationally faster than the previous package. This offers the possibility to update the radiation fields in the model simulations more frequently for the same overall computational cost, which we show in turn can improve the skill of weather forecasts. A unique feature for a radiation package of this kind is the ability to simulate radiation flows through the sides of clouds, not just through the base and top, making it well suited as a tool for research into atmospheric radiation exchange.

1. Introduction

Radiation provides the energy that drives both the large-scale circulation and smaller-scale processes affecting the weather experienced at the ground. In a weather forecast model, such as European Centre for Medium-Range Weather Forecasts' (ECMWF's) Integrated Forecast System (IFS), a good treatment of radiative transfer is a prerequisite for accurate near-surface temperature forecasts up to a week ahead. But a key challenge for weather prediction is to push the boundaries of predictability at monthly and seasonal time scales, and indeed ECMWF makes operational forecasts up to a year ahead. This puts additional stringent demands on the accuracy of the radiation scheme, requiring not only that the model has an excellent climate but also that the role of radiation in the evolution and predictability of atmospheric weather systems and regimes is well captured.

The radiation scheme used at ECMWF has undergone steady improvement over the years. The Morcrette (1991) scheme was used in the 1990s, and the developments made in that decade were outlined by Morcrette, Barker, et al. (2008). Table 1 shows the changes made since 2000, a year that saw a major upgrade to the longwave scheme with the introduction of the Rapid Radiative Transfer Model for GCMs (RRTM-G; Mlawer et al., 1997). RRTM-G is a correlated-k model of gas absorption that reproduces clear-sky irradiances to within a few Watts per square meter.

Table 1
Timeline of the Main Developments in the ECMWF Radiation Scheme Since 2000, and the Date They Became Operational

|	22R3	23R4	25R1	26R3	28R3	32R2	35R3	41R1	41R2	43R1	43R3
IFS cycle	22R3	23R4	25R1	26R3	28R3	32R2	35R3	41R1	41R2	43R1	43R3
Month/Year	6/2000	6/2001	4/2002	10/2003	9/2004	6/2007	9/2009	5/2015	3/2016	11/2016	7/2017
Package name	Morcrette					McRad					ecRad
SW spectral bands	4		6			14 (RRTM-G)					
LW spectral bands	16 (RRTM-G)							Updated to latest RRTM-G			
Liquid cloud optics	Fouquart (1987), Smith and Shi (1992)					Slingo (1989), Lindner and Li (2000)					SOCRATES
Ice cloud optics	Ebert and Curry (1992)					Fu (1996), Fu et al. (1998)					
Aerosol climatology	Tanré (1984)			Tegen et al. (1997)							CAMS
Ozone climatology	Fortuin and Langematz (1994)					GEMS	MACC			CAMS	
Ozone climatology	Constant mixing ratio					GEMS	MACC				
Solver	Clear/cloudy regions					McICA					Less noise
Surface albedo	ERBE					60-km MODIS	5-km MODIS				
Frequency	3 hr	1 hr in data assimilation			1 hr in HRES				Approx. updates, better sun angle		
Total solar irradiance	1,366 W/m^2										1,361 W/m^2 ($\pm$ solar cycle)

Note. The meanings of the acronyms not defined elsewhere in the text are as follows: GEMS = Global and regional Earth-system Monitoring using Satellite and in situ data; ERBE = Earth Radiation Budget Experiment; and MODIS = Moderate Resolution Imaging Spectroradiometer. Other Acronyms: SW = shortwave; LW = longwave; SPARTACUS = Speedy Algorithm for Radiative Transfer through Cloud Sides; MACC = Monitoring Atmospheric Composition and Climate; CAMS = Copernicus Atmospheric Monitoring Service; RRTM-G = Rapid Radiative Transfer Model for GCMs; See section 1 for further explanation.

This was followed in 2007 by a major upgrade of the radiation package to *McRad*, which involved not only incorporation of the shortwave RRTM-G model but also the McICA scheme of Pincus et al. (2003) to represent realistic cloud heterogeneity and overlap via a stochastic cloud generator. Many weather and climate models have since incorporated one or both of these advances. Morcrette, Barker, et al. (2008) and Ahlgrimm et al. (2016) provided a detailed description of the scheme and the resulting improvements in forecast scores and model climate.

The treatment of aerosols and trace gases in the radiation scheme has also improved over this period. Introduction of the Tegen et al. (1997) aerosol climatology in 2003 significantly reduced the optical depth of aerosol over the Sahara, the direct radiative effect of which improved representation not only of the African Easterly Jet (Tompkins et al., 2005) but also global weather forecasts via the impact on Rossby wave sources (Rodwell & Jung, 2008). Between 2009 and 2017, improved trace gas climatologies were introduced that exploited ECMWF activities in assimilating measurements of atmospheric composition into a configuration of the IFS containing prognostic aerosol and gas chemistry. This included the *Monitoring Atmospheric Composition and Climate* (MACC) project and most recently the Copernicus Atmospheric Monitoring Service (CAMS). The positive impact of the recent implementation of an aerosol climatology based on CAMS, especially for the Indian Summer Monsoon, was reported by Bozzo et al. (2017).

The larger total number of spectral intervals (252) used by RRTM-G made it 3.5 times slower than the Morcrette scheme, necessitating the use of a lower-resolution radiation grid (Morcrette, Mozdzynski, & Leutbecher, 2008). In 2016, two methods were introduced to remedy problems associated with calling the radiation scheme infrequently in time and space. The Hogan and Bozzo (2015) scheme performs approximate updates to the irradiances at every time step and model grid point to correct for errors due to sharp temperature and albedo transitions at coastlines. The improved treatment of solar zenith angle described by Hogan and Hirahara (2016) addressed problems in IFS configurations that call the radiation scheme only every 3 hr, reducing the associated stratospheric warm bias, and largely eliminating spurious longitudinal variations in tropical solar irradiances and heating rates.

Two remaining shortcomings of McRad have motivated the recent development of a new ECMWF radiation scheme, *ecRad*. First flexibility: to facilitate current and future scientific developments (such as the representation of 3-D radiative effects, or alternative treatments of gas optics), we need the ability to swap individual components of the radiation scheme for faster and/or more accurate ones, but the nonmodular design of McRad makes this very difficult. Second efficiency: in all operational model configurations except the highest-resolution deterministic forecast (HRES), we can afford to call the radiation scheme only every 3 hr.
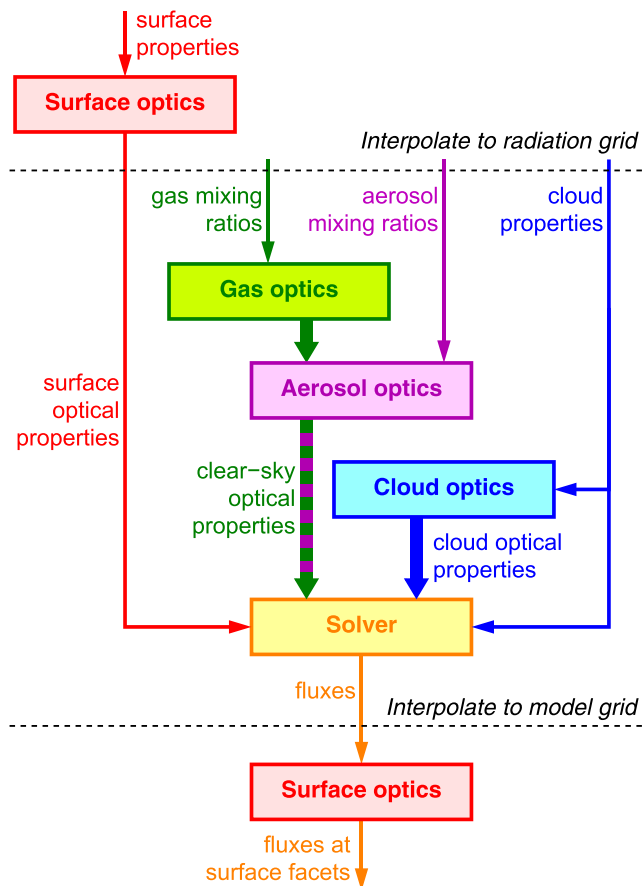
Figure 1. Schematic illustrating the five main components of ecRad (boxes) and the flow of data between them (arrows). Each thin arrow denotes a data structure (*derived type* in Fortran) containing a number of arrays, while the thick arrows denote arrays of optical properties in each spectral interval used by the radiation scheme. Not shown is an additional *thermodynamics* structure containing atmospheric temperature and pressure at layer interfaces that is passed to each component.

Despite the use of approximate updates at intervening model time steps, this is known to degrade forecasts with respect to more frequent calls to the full radiation scheme.

The ecRad scheme became operational in July 2017. In this article we summarize its capabilities and meteorological impact. In section 2, its modular structure is described, followed in section 3 by a description of its revised McICA implementation, which is both much more efficient and generates less noise in atmospheric heating rates than the McRad scheme. In section 4 it is shown how longwave scattering by clouds can be introduced with only a 4% increase in the overall cost of the scheme. In section 5, we describe the cumulative improvement to stratospheric climate due to several recent radiation and ozone changes. Then in section 6, the impact on weather forecast skill is presented.

2. Overview of ecRad

The new scheme has been coded in Fortran 2003 and amounts to around 16,000 lines of code, in addition to the existing RRTM-G implementation that is still used to compute gas optical properties. Figure 1 depicts the five main components of ecRad and the flow of data between them. The modular structure ensures that, in most cases, the internals of each component can be altered without needing to change any other component. The main options available are listed in Table 2. Before discussing ecRad specifically, section 2.1 summarizes the configurations of the IFS currently used to perform operational global medium-range forecasts at ECMWF, and what variables are passed to and from the radiation scheme, whether McRad or ecRad. Sections 2.2–2.6 then describe the components of ecRad in detail, and section 2.7 outlines the status of the ecRad implementation in the IFS.

Table 2
Summary of the Main Options Available in ecRad, Each of Which Can Be Selected at Run-Time

Property	Options
Gas optics model	Monochromatic; **RRTM-G** (other options are planned)
Aerosol optics model	**Generalized** (supporting Tegen and CAMS types); optics computed externally
Liquid cloud optics	Slingo (1989) and Lindner and Li (2000); **SOCRATES**
Ice cloud optics	**Fu (1996) and Fu et al. (1998)**; Yi et al. (2013); Baran et al. (2014)
δ-Eddington mode	Particles and gases; **Particles only**
SW and LW solver	Homogeneous; **McICA**; Tripleclouds; SPARTACUS
Longwave scattering	Off; **cloud only**; cloud and aerosols
Cloud overlap scheme	**EXP-EXP** (only available with McICA); MAX-RAN; EXP-RAN
Cloud overlap parameter	**alpha (Hogan & Illingworth, 2000)**; beta (Shonk et al., 2010)
Cloud water PDF shape	**Gamma**; Lognormal

Note. The settings expected in the next operational cycle (46R1) are shown in bold. The earlier IFS implementation of ecRad in Cycle 43R3 were the same except for longwave scattering being switched off and δ-Eddington scaling being applied to particles and gases. RRTM-G = Rapid Radiative Transfer Model for GCMs; SPARTACUS = Speedy Algorithm for Radiative Transfer through Cloud Sides; SW = shortwave; LW = longwave.

2.1. Radiation in the Context of the ECMWF Model

Since IFS Cycle 41R2, the high-resolution deterministic (HRES) forecasts use $T_{Co}1279$ resolution (around 9 km) with 137 terrain-following levels and a 7.5-min time step. The radiation scheme is called every hour (which we refer to as the *radiation time step*) on a grid with 10.24 times fewer columns than the rest of the model (3.2 times coarser in each horizontal direction). The ensemble (ENS) forecasts use $T_{Co}639$ resolution (around 18 km) with 91 levels and a 12-min time step. The radiation scheme is called every 3 hr on a grid with 6.25 times fewer columns than the rest of the model (2.5 times coarser in each horizontal direction).

In a standard medium-range forecast, the prognostic atmospheric variables fed to the radiation scheme are temperature, cloud fraction, and the mixing ratios of water vapor, liquid water, ice and snow. These are predicted on full model levels and are interpolated horizontally to the radiation grid using cubic interpolation, but retaining the same vertical grid. This is depicted by the upper dashed black line in Figure 1. Pressure, both at full levels and at layer interfaces (*half levels*), is also horizontally interpolated, but since the operational version of the model is hydrostatic, the pressure profile is diagnosed from the prognostic surface pressure. Temperature is interpolated onto half levels, and while McRad requires temperature and pressure at both full and half levels, ecRad only takes these variables at half levels. Cloud effective radius is computed diagnostically at full levels on the radiation grid using the parameterization of Martin et al. (1994) for liquid clouds and Sun and Rikus (1999), with corrections by Sun (2001), for ice clouds.

In chemistry and air quality applications for CAMS, the prognostic mixing ratios for ozone, carbon dioxide, and an arbitrary number of aerosol species may also be interpolated to the radiation grid. Otherwise, these mixing ratios are computed from a climatology on the radiation grid. The cosine of the solar zenith angle is also passed to the radiation scheme, having been computed as the average value over the sunlit part of the radiation time step as described by Hogan and Hirahara (2016). The top-of-atmosphere solar irradiance provided to the radiation scheme accounts for the seasonal variation in Sun-Earth distance, the equation of time and an approximate representation of the solar cycle.

The radiation scheme, whether McRad or ecRad, computes profiles of shortwave and longwave irradiances at half levels, and these are interpolated horizontally back onto the model grid using cubic interpolation, as depicted by the lower dashed black line in Figure 1. A correction is made to the shortwave irradiance profiles to make them consistent with the high-resolution albedo information using the method of Hogan and Bozzo (2015), and then they are normalized by the incoming solar irradiance at top of atmosphere (TOA). These normalized irradiances are then available to be used by the model at the intervening model time steps.

At each model time step, the irradiance profiles from the last call to the radiation scheme are used to compute atmospheric heating rate profiles, but only after some corrections are applied. In the shortwave, we account for the change in solar zenith angle between calls to the radiation scheme: the normalized irradiance profiles are multiplied by the incoming solar irradiance, and an additional approximate correction is applied to

account for the change in path length of the solar beam through the atmosphere (Manners et al., 2009). In the longwave, the irradiance profiles are adjusted to account for the local value of skin temperature using the technique of Hogan and Bozzo (2015). Atmospheric heating rates, surface and TOA irradiances are then computed in a way that ensures energy conservation.

2.2. Surface Optics

In the current operational version of the IFS, which includes the ecRad radiation scheme, the interaction of radiation with the surface is rather simple. The surface scheme computes grid box mean skin temperature, longwave emissivity in two spectral intervals (corresponding to the atmospheric infrared window, and everything else), and shortwave direct and diffuse albedo in six spectral intervals. These variables are interpolated to the radiation grid and passed to the solver, except for skin temperature which is passed to the gas optics component where the Planck emission to the atmosphere is computed in each spectral interval.

A more sophisticated coupling between radiation and the surface is in preparation, aiming to improve the interactions with complex surfaces such as forests and urban areas. Hogan et al. (2018) described the method that will be used to represent 3-D radiative interactions in open forest canopies. Radiative transfer in urban areas will be treated by an analogous approach and will make use of parameters provided by the urban surface scheme currently under development. A consistent spectral treatment between radiative transfer in the surface and atmosphere should then enable, for example, more reliable prediction of photosynthesis rates and street-level temperatures.

To support these developments, ecRad includes an optional surface optics component whose planned deployment in the IFS is depicted by the red box in Figure 1, where it will perform computations on the model grid both before and after the atmospheric calculation. The first part takes as input a physical description of the tiles that make up the land surface in one model grid box and computes the surface properties seen by the atmosphere as the weighted average of each tile in the grid box. The properties are longwave emissivity, the longwave emitted upwelling radiation, and the shortwave direct and diffuse albedo. They may be computed using either the full spectral resolution of the atmospheric model or a reduced resolution according to user preference. These properties are then interpolated onto the radiation grid and passed directly to the solver. After the atmospheric radiative transfer calculation is complete, the second part of the surface component takes as input the downwelling shortwave and longwave irradiances exiting the base of the lowest atmospheric layer, interpolated back onto the model grid, but at the spectral resolution of the first surface computation. It then works out the net irradiances at each facet of the underlying surface. A detailed description of this two-stage process for open forest canopies was provided by Hogan et al. (2018).

2.3. Gas Optics

The gas optics model is selected at runtime, and determines the spectral discretization used by the whole of the radiation scheme. 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. A number of variables are computed in each spectral interval, which we hereafter refer to as *g-points* (e.g., Morcrette, Barker, et al., 2008) to contrast with the fewer number of *bands* used for aerosol and clouds. In the shortwave, these variables are the optical depth and single-scattering albedo of each layer of the clear-sky atmosphere, and the TOA incoming solar irradiance. In the longwave they are the optical depth of each layer and the Planck function integrated across each g-point at layer interfaces. If a simple representation of the surface is used, then the corresponding Planck function of the surface is computed from the skin temperature.

A key part of this design is that subsequent parts of the radiation scheme can treat each g-point equally, and the solver can compute broadband irradiances simply by summing over the irradiances in each g-point. Moreover, the subsequent parts use g-point as the fastest-varying array dimension and this has been found to lead to faster code overall than in McRad, which used atmospheric column as the fastest-varying dimension. 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 and column, but not g-point. Aside from the monochromatic scheme used for testing, the only gas optics model currently available is RRTM-G, which employs 112 g-points in the shortwave and 140 in the longwave. We currently use the same RRTM-G implementation of gas optics as in McRad, but need to permute the arrays it produces to make g-point the fastest-varying dimension. Figure 2b shows that this increases the cost of the gas optics component, although this is more than compensated by the speedup in the subsequent components, particularly
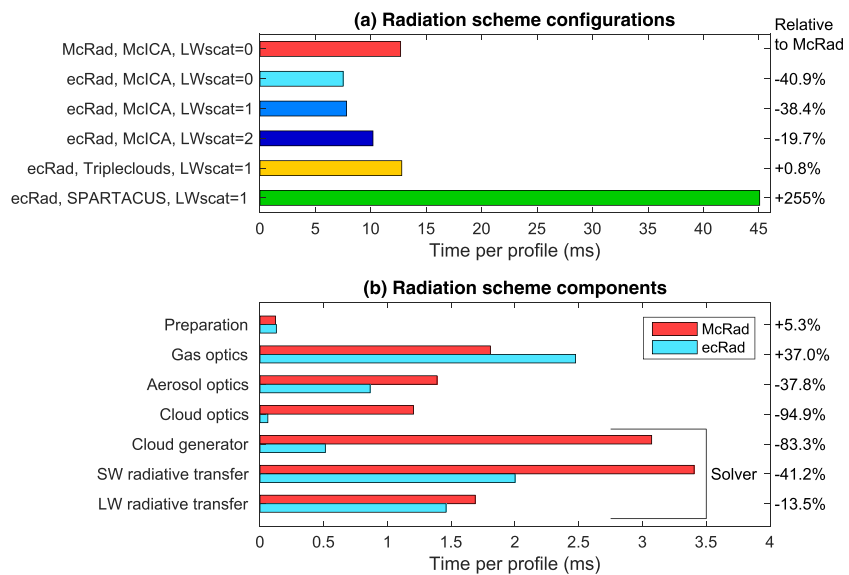
Figure 2. (a) Computational cost of various configurations of the radiation scheme per atmospheric profile, where the labels on the left indicate the scheme name, solver name, and longwave scattering configuration (0 = no longwave scattering, 1 = scattering by clouds only, and 2 = scattering by clouds and aerosols). (b) Breakdown of the cost for the first two configurations: the previous McRad scheme (operational until IFS Cycle 43R1) and the equivalent ecRad scheme (operational from 43R3), both using a McICA solver with no longwave scattering. The percentages on the right of each panel indicate the fractional change to the computational cost compared to McRad. The values have been computed from 24-hr $T_{Co}1279$ forecasts using IFS Cycle 45R1 incorporating ecRad version 1.0.13.

the shortwave solver. In the future we plan to implement a more efficient version of RRTM-G, as well as to consider alternative gas optics models that would offer the user more flexibility in choosing how to make the trade-off between accuracy and speed.

2.4. Aerosol Optics

The aerosol component adds the optical properties of aerosols to those of gases with the assumption that aerosols are horizontally well mixed within each model grid box. Although the clear-sky optical properties are defined at the full resolution of the gas optics scheme, aerosol properties are computed in a coarser set of bands; in the case of RRTM-G this is 14 in the shortwave and 16 in the longwave. Aerosol properties may be provided in two ways. In the first, the mass-mixing ratios of an arbitrary number of run-time-configurable aerosol species are provided. The optical properties of each type are preloaded from a NetCDF data file in terms of the mass extinction coefficient, single-scattering albedo and asymmetry factor in each band. These properties are computed off-line using an assumed size distribution and Mie theory, although a method that can represent nonspherical particles could be used instead. Aerosols may be hydrophilic or hydrophobic, with the optical properties of the hydrophilic types also a function of relative humidity, and the mass-mixing ratio from the host model assumed to be either the dry mass or the mass at a fixed relative humidity. Within the IFS, this *generalized* aerosol formulation is flexible enough to accommodate both the older Tegen et al. (1997) climatology of six hydrophobic aerosol species, and the newer CAMS specification with 11 hydrophilic and hydrophobic species (Flemming et al., 2017). Note that the CAMS specification is used both with prognostic aerosols, and with the climatology derived from the CAMS reanalysis (Bozzo et al., 2017).

This aerosol formulation is not well suited for more advanced applications where the aerosol properties are required to vary smoothly with additional input variables such as aerosol mean size, age, nonsphericity, and so on. Therefore, the option is available for the user to compute the aerosol optical properties in each band externally to ecRad, and to pass them in instead of mass-mixing ratios. This capability is being used in the implementation of the *GLOMAP-mode* aerosol scheme (Mann et al., 2010) in the IFS.

2.5. Cloud Optics

The cloud optics component takes as input the mass-mixing ratio and effective radius of cloud liquid and ice (but note that in the IFS the ice mixing ratio fed to the radiation scheme is actually the sum of the prognostic

ice and snow variables). It divides the grid box mean mixing ratios by cloud fraction to obtain the in-cloud mean mixing ratios and uses them to compute in-cloud optical properties in each spectral band.

As shown in Table 2, by default ice optical properties are computed using the Fu scheme as in McRad, but two additional schemes are available for testing. The default liquid-optics model has been changed to the one from the Met Office's SOCRATES radiation scheme (Suite of Community Radiation Codes based on Edwards & Slingo, 1996), in which the optical properties in all bands are expressed in terms of a Padé approximation. This overcomes the tendency of the previous Slingo (1989) scheme to overestimate cloud optical depth (Nielsen et al., 2014). Coupled 1-year IFS simulations show that this change increases global mean net downward shortwave at TOA by 1.6 W/m^2, but this is almost completely canceled by a 1.8 W/m^2 reduction with the new implementation of McICA. The corresponding changes to surface net shortwave irradiances are the same as those at TOA, so there is no detectable change to atmospheric absorption.

Figure 2b shows that the cloud optics calculation is 20 times faster than in McRad. This is mainly because McRad recomputes cloud optical properties for every g-point using the perturbed cloud water contents from the McICA cloud generator. However, the perturbed water contents 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), and then use the cloud generator to provide optical-depth scalings for each g-point.

An additional efficiency saving arises from the implementation of delta-Eddington scaling (Joseph et al., 1976), the scheme that accounts for strong forward scattering by cloud and aerosol particles by treating a fraction $f = g^2$ of the scattered radiation as if it had not been scattered at all (where g is the asymmetry factor). In McRad, this scaling is applied at every g-point to the gas-particle mixture, and in the solver distributed with RRTM-G it is applied at every g-point to the gas-aerosol mixture. However, since Rayleigh scattering by gases in the shortwave does not preferentially scatter in the forward direction, it is more physically justified to apply delta-Eddington scaling to cloud and aerosol separately, before merging the scaled optical properties with those of the gases. The nonlinear dependence of f on g means that the resulting merged optical properties are different. To illustrate this, 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 subsequent scaling reduces the optical depth of the mixture 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. In practice we find from coupled IFS simulations that this only changes globally averaged net shortwave irradiances by 0.1 W/m^2, presumably because it is quite rare for the Rayleigh and particle optical depths in a given atmospheric layer and g-point to be of a similar magnitude. But because delta-Eddington scaling is now only applied on bands rather than g-points, it reduces the computational cost of the shortwave solver (excluding the cloud generator) by 12%, and of ecRad overall by 3%.

2.6. Solver

The job of the solver is to combine the clear-sky and cloudy optical properties according to the cloud fraction and appropriate assumptions about in-cloud heterogeneity and overlap, and to use them to compute irradiance profiles. Four solvers are currently available. The 'homogeneous' solver is suitable for off-line calculations on high-resolution scenes using the independent column approximation. It ignores cloud fraction, assuming that clouds are homogenized horizontally within each grid box. The McICA solver has been completely reformulated and uses an improved cloud generator that is more efficient and produces less stochastic noise in atmospheric heating rates than McRad (see section 3). The Tripleclouds solver (Shonk & Hogan, 2008) uses an alternative approach to represent in-cloud heterogeneity, which has the benefit of generating no stochastic noise, but the drawback of making the radiation scheme 61% slower overall (for the same treatment of longwave scattering).

All three solvers discussed so far are based on a classical two-stream formulation in both the shortwave and longwave, but with different ways to treat cloud structure. The final solver, SPARTACUS (Speedy Algorithm for Radiative Transfer through Cloud Sides; Hogan et al., 2016; Schäfer et al., 2016) is somewhat more complex, and makes the IFS the first global model capable of representing the 3-D radiative effects of clouds.

Figure 2a shows that with this solver, ecRad is 5.8 times slower than with McICA, and 3.6 times slower than the original McRad scheme. This is too slow to be justified for operational weather forecasting at ECMWF, but is fast enough to use for research.

Table 2 lists some of the more detailed options available to the solvers, including different treatments of cloud overlap and the shape of the cloud-water probability distribution function (see section 3), as well as efficient methods to represent longwave scattering (see section 4.1). Generally, all these options are available for the McICA, Tripleclouds and SPARTACUS solvers. Moreover, each solver has the capability of computing the derivative of upwelling longwave irradiance at each height with respect to the surface value, needed as input to the longwave part of the approximate update scheme of Hogan and Bozzo (2015).

In addition to outputting the full irradiance profile, ecRad outputs the downwelling direct and diffuse surface shortwave irradiances in each spectral band. This facilitates the calculation of diagnostics such as photosynthetically active radiation and ultraviolet index.

2.7. Status in the IFS

Version 1.0.1 of ecRad was implemented in IFS Cycle 43R3, which became operational in July 2017. This version had longwave scattering switched off and was therefore scientifically rather similar to McRad. It was 31% faster than McRad (including the cost of interpolating to and from the radiation grid), reducing the cost of radiation from 5% to 3.5% of total model time in the HRES model configuration. Version 1.0.13 of ecRad, on which this paper is based, includes further optimizations and as shown by the first two bars in Figure 2a, when longwave scattering is off, it is 41% faster than McRad excluding the cost of interpolation, or 33% faster when the interpolation cost is included. This version will be used in the next operational cycle (46R1), but with longwave scattering by clouds switched on (the third bar in Figure 2a).

Cycle 46R1 also fixes a long-standing bug in the treatment of the longwave optical properties of ice clouds, which was present in McRad. Fixing the bug reduces global mean downwelling longwave irradiance at the surface by 0.7 W/m^2 and cools the land surface by 0.2 K (in coupled free-running IFS simulations). Unfortunately, this fix on its own degrades medium-range forecasts, so the bug had to be reproduced in the 43R3 implementation of ecRad. However, the introduction of longwave scattering increases downwelling surface longwave by 0.7 W/m^2 and warms the land surface by 0.3 K (in free-running simulations), so when accompanied by the bug-fix leads to a small surface warming, which is positive for forecasts given that the IFS is generally a little too cold.

3. New Implementation of McICA

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 ecRad to improve its efficiency, to reduce noise, and to add flexibility via the ability to represent three different overlap scenarios and to take the fractional standard deviation of cloud water content as input. We first describe some of the shortcomings of the current McICA scheme, with a particular focus on its representation of cloud overlap. We then describe how they are overcome by the cloud generator in the new scheme and demonstrate the reduction in heating rate noise.

3.1. The McRad Implementation of McICA

Observations by cloud radar and lidar (e.g., Di Giuseppe & Tompkins, 2015; Hogan & Illingworth, 2000) have found that cloud occurrence in different vertical layers is correlated but that the *overlap parameter α*, which quantifies the correlation, decreases approximately exponentially with layer separation according to a decorrelation distance z_0. If there is a clear layer between these two cloudy layers, then their overlap becomes random. This model is therefore commonly known as *exponential-random* overlap (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 $z_0 \rightarrow \infty$. In the IFS, z_0 is parameterized as a function of latitude following Shonk et al. (2010).

The McRad implementation of McICA generates subcolumns that are intended to implement EXP-RAN overlap stochastically using the Räisänen et al. (2004) cloud generator. It generates a subcolumn 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 R_i is less than the cloud fraction at that height, a_i, then cloud is deemed to be present. It then generates 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 & 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 subcolumns required, one for each g-point.

There are three shortcomings of this approach. First, 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).

The second shortcoming is that the irradiances 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 subcolumns containing cloud, and then perform a weighted average with the clear-sky calculation according to the total cloud cover to obtain the total sky irradiance profile. Räisänen et al. (2004) proposed generating more subcolumns and simply throwing away the ones that do not contain cloud, although this could become expensive for low cloud covers. While it has been found that unbiased noise in irradiances has no measurable impact on seasonal forecasts (Pincus et al., 2003), it can add random noise to short-range forecasts of near-surface temperature (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 the MAX-RAN, EXP-RAN, or EXP-EXP overlap assumptions. The key to reducing noise is in two parts. First, total cloud cover is computed deterministically from the cloud fraction profile and the overlap rules, so removing stochastic noise in cloud cover (even though the cloud profiles within the cloudy part of the grid box are still stochastically generated). This is described in section 3.2. Second, we make use of the fact that clear-sky radiation calculations (i.e., fed by the same atmospheric profiles except with the clouds removed) are already performed for diagnostic purposes, so we can use the cloud generator to produce only cloudy profiles (described in section 3.3), and then compute the total-sky irradiance profile as a weighted average of the clear and cloudy profiles. Since all 252 RRTM-G g-points see a cloud, this leads to much better sampling of the cloudy part of the column in partially cloudy conditions.

3.2. Deterministic Cloud Cover

The deterministic part of the algorithm consists of calculating $c_{i+1/2}$, the profile of cumulative cloud cover measured from TOA down to a particular half-level $i + 1/2$, that is, the small circles shown in Figure 3. 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 $i + 1$. This deterministic part is also used by the Tripleclouds and SPARTACUS solvers and ensures they are exactly consistent with McICA in terms of overlap assumption.

MAX-RAN overlap is obtained with (Geleyn & Hollingsworth, 1979; Morcrette & Jakob, 2000)

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

$$(1 - c_{i+1/2}) = (1 - c_{i-1/2})\frac{1 - p_{i-1/2}}{1 - a_{i-1}}, \tag{2}$$
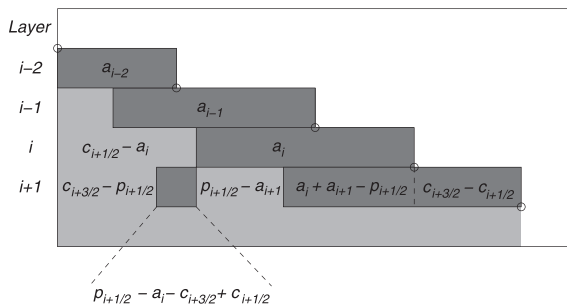
Figure 3. Schematic illustrating how the new McICA cloud generator is implemented and, in particular, how to compute the conditional probabilities of cloud in layer $i + 1$ dependent on whether cloud is present in layer i. The variables have the following definitions: a_i is the cloud fraction in layer i, $p_{i+1/2}$ is the combined cloud cover of layers i and $i + 1$, and $c_{i+1/2}$ (indicated by the small circles) is the cloud cover of all the layers from top of atmosphere down to half-level $i + 1/2$.

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 *MAX* in MAX-RAN says that vertically contiguous cloud layers should be maximally overlapped. This implies that if there is even a very small amount of cloud in the middle layer, the cloudy layers are vertically contiguous and so should be treated as maximally overlapped with a total cloud cover of 0.5. But if the middle cloud fraction is reduced to exactly zero, the total cloud cover would jump immediately to 0.75, making such an overlap implementation *discontinuous*. By contrast, (1) and (2) are a *continuous* implementation of cloud overlap, since they 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 subgrid 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 grid box containing cloud only in the top and bottom layers; the clouds in this fraction of the grid box are therefore treated as randomly overlapped. We regard this behavior as preferable to the discontinuous implementation, 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 a differentiable overlap scheme be used in a variational data assimilation context in future (although then a differentiable solver such as Tripleclouds would also be required).

EXP-RAN overlap is obtained by substituting equations 1 and 2 of Hogan and Illingworth (2000) into their equation 4 to obtain

$$
\begin{aligned}
p_{i-1/2} = {} & \alpha_{i-1/2} \max(a_{i-1}, a_i) \\
& + (1 - \alpha_{i-1/2}) \left(a_{i-1} + a_i - a_{i-1} a_i \right),
\end{aligned}
\tag{3}
$$

where α is the overlap parameter described in section 3.1. This replaces (1) in the algorithm, and 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 formulas exist for p and c that are exactly consistent with the Räisänen et al. (2004) cloud generator, but Hogan and Bozzo (2016) described an approximate but deterministic method to compute p and c consistent with the EXP-EXP overlap assumption, which they found to closely match the average total cloud cover from the EXP-EXP cloud generator. It identifies concave features in the cloud fraction profile and clusters them hierarchically, but is too involved to repeat here.

The availability of fast, deterministic algorithms for cloud cover has also improved the cloud cover diagnostic, which is computed on the model grid for high, midlevel, low, and total cloud cover. Prior to IFS Cycle 43R3, the Räisänen et al. (2004) cloud generator was rerun on the model grid to provide a stochastic estimate of these quantities consistent with the assumptions used in the radiation scheme, but it was unnecessarily costly and also subject to stochastic noise. Since Cycle 43R3, the ecRad deterministic algorithm is used, reducing the cost of the cloud cover diagnostic by a factor of 4 for the operational EXP-EXP overlap assumption.

3.3. Efficient Use of Random Numbers

With the c function describing the cumulative cloud cover for the chosen overlap scheme, we now describe the algorithm to generate only cloudy subcolumns. We use the first random number in each subcolumn, 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 subcolumn. This leads to the ecRad cloud generator being much more efficient than in McRad (see Figure 2b). We proceed to layer $i + 1$ below and generate a further random

number R_{i+1} to decide whether it is cloudy. Figure 3 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 Figure 3 this can be worked out, leading to cloud being generated in layer $i+1$ if

$$R_{i+1} < \frac{a_i + a_{i+1} - p_{i+1/2}}{a_i}. \tag{4}$$

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

$$R_{i+1} < \frac{p_{i+1/2} - a_i - c_{i+3/2} + c_{i+1/2}}{c_{i+1/2} - a_i}. \tag{5}$$

This process is continued down to the lowest cloudy layer in the subcolumn.

The generation of in-cloud heterogeneity follows more closely the approach of Räisänen et al. (2004): a profile of uniformly distributed random numbers between 0 and 1 is generated in the cloudy pixels of a subcolumn that is consistent with the user-specified vertical decorrelation length of cloud heterogeneities. If EXP-EXP overlap is requested then in-cloud heterogeneities are correlated even across cloud-free layers, whereas for MAX-RAN and EXP-RAN overlap, the random numbers are decorrelated whenever passing through a cloud-free layer. The profile of random numbers is then used to index a look-up table to obtain the corresponding value from a gamma or lognormal distribution with the user-specified fractional standard deviation. This in turn is used to scale the in-cloud optical depth.

### 3.4. Impact on Heating Rates

Figures 4a and 4b compare shortwave and longwave heating rate profiles in a partially cloudy column from the IFS. The three lines show the McRad McICA implementation, the ecRad McICA implementation described above, and the Tripleclouds solver which generates no stochastic noise (although is more computationally expensive than McICA). In this example it is apparent that the new McICA implementation is significantly less noisy than the old, and the difference between the McRad McICA implementation and Triplecloud can be as much as 5 K/day.

It should be noted that Triplecloud and SPARTACUS cannot represent EXP-EXP overlap, because the assumption that vertically separated cloud layers are randomly overlapped is built into their formulation. Therefore, in Figure 4, Triplecloud used EXP-RAN overlap, while the two McICA implementations used EXP-EXP overlap, but the impact of overlap assumption on instantaneous heating rates is much smaller than the stochastic noise in McICA. With the assumption that Triplecloud represents an appropriate reference calculation against which to quantify McICA noise, Figures 4c and 4d compare the root-mean-square heating rate errors of the two McICA implementations, considering all profiles at the start of an IFS model forecast. In the boundary layer, it can be seen that the McRad implementation of McICA produces around 30% more noise than the ecRad implementation in both the shortwave and the longwave.

4. Improved Longwave Radiative Transfer

4.1. Fast Implementation of Longwave Scattering by Clouds

For reasons of computational cost, longwave scattering is frequently omitted in the radiation schemes of atmospheric models, yet can increase longwave cloud radiative effect by around 10% globally (Costa & Shine, 2006). A fast approximation to longwave scattering has been proposed by Chou et al. (1999), but it only approximates the effects of scattering for clouds of low and moderate optical depth. Comparing the second and fourth bars in Figure 2, we see that the overall cost of ecRad is increased by 36% when longwave scattering by aerosols and clouds is represented, which is due to the longwave solver itself being twice as costly. In this section we demonstrate that this cost can be reduced dramatically if we represent the far more important scattering by clouds, but not scattering by aerosols (although longwave absorption and emission by aerosols are still represented). We note that Dufresne et al. (2002) estimated that the neglect of longwave scattering for dust with an optical depth of 1 would lead to errors in surface and TOA net longwave irradiances of $3 - 5$ W/m^2. However, our free-running uncoupled IFS simulations suggest that turning on longwave scattering for aerosols changes global mean surface and TOA net longwave irradiances by no more than 0.1 W/m^2, which can be safely neglected in the context of weather forecasts.
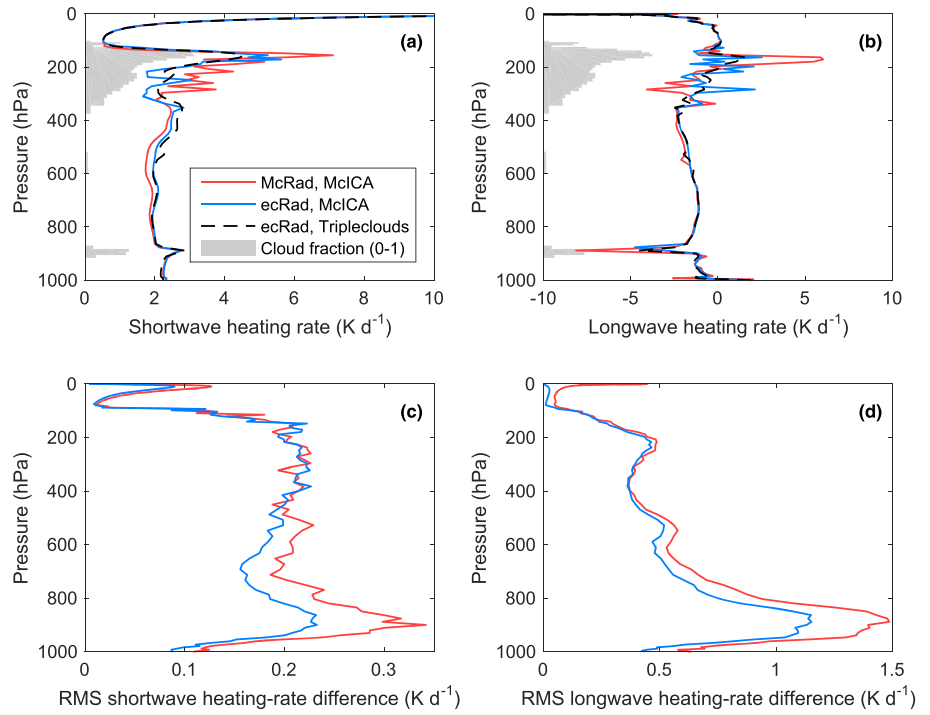
Figure 4. Comparison of instantaneous (a) shortwave and (b) longwave heating rates between the McRad radiation scheme with its McICA solver, and the ecRad scheme with both its McICA and Tripleclouds solvers, for a profile taken from a 137-level T_L255 IFS model simulation. The gray shading shows the cloud fraction profile ranging from 0 to 1 across the horizontal axis. Panels (c) and (d) show the corresponding root-mean-square heating rate differences between each McICA calculation and the Tripleclouds calculations from the same model simulation (the shortwave results in panel c considering only sunlit profiles). All panels consider only radiation calculations from the first call of the radiation calculations when the input cloud profiles are identical.

To explain how longwave scattering by clouds alone can be represented efficiently, we first outline how nonscattering and fully scattering longwave solvers are formulated in the context of the two-stream approximation. First, four variables are computed for each layer i: reflectance R_i, transmittance T_i, and layer sources $S_i^{\uparrow}$ and $S_i^{\downarrow}$ representing the upward emission from the top of the layer and the downward emission from the base of the layer, respectively. Each variable is a function of g-point, and we assume the Planck function to vary linearly with optical depth in the layer (Wiscombe, 1976). In the scattering case the first two variables are provided by Meador and Weaver (1980) and the second two by Stackhouse and Stephens (1991). In the nonscattering case they reduce to simpler forms (e.g., Clough et al., 1992):

$$R_i = 0; \tag{6}$$

$$T_i = \exp(-D\delta_i); \tag{7}$$

$$S_i^{\uparrow} = (1 - T_i)\left(B_{i+1/2} + \frac{\Delta B_i}{D\delta_i}\right) - \Delta B_i; \tag{8}$$

$$S_i^{\downarrow} = (1 - T_i)\left(B_{i-1/2} - \frac{\Delta B_i}{D\delta_i}\right) + \Delta B_i, \tag{9}$$

where δ_i is the absorption optical depth of the layer, $D = 1.66$ is the diffusivity factor (Fu et al., 1997) and $\Delta B_i = B_{i+1/2} - B_{i-1/2}$ is the difference in the Planck function at the base and top of the layer.

A nonscattering longwave solver then proceeds down from the top of atmosphere (with layer index i increasing downward) by repeated application of the following to compute the profile of downwelling irradiances at layer boundaries:

$$F_{i+1/2}^{\downarrow} = T_i F_{i-1/2}^{\downarrow} + S_i^{\downarrow}. \tag{10}$$

The upwelling irradiance at the surface is a combination of emission and reflection:

$$F_s^\uparrow = \epsilon_s B_s + (1 - \epsilon_s) F_s^\downarrow, \tag{11}$$

where subscript s denotes surface quantities and ϵ_s is the surface emissivity. The following is then applied up through the atmosphere to compute the profile of upwelling irradiances at layer boundaries:

$$F_{i-1/2}^\uparrow = T_i F_{i+1/2}^\uparrow + S_i^\uparrow. \tag{12}$$

A full scattering computation is more complex. It works up from the surface computing the albedo $A_{i-1/2}$ of the entire atmosphere (including the surface) below half-level $i - 1/2$, and similarly the total radiation $G_{i-1/2}$ upwelling at the same half level that originated from emission below it (e.g., Shonk & Hogan, 2008):

$$A_{i-1/2} = R_i + \frac{T_i^2 A_{i+1/2}}{1 - A_{i+1/2} R_i}; \tag{13}$$

$$G_{i-1/2} = S_i^\uparrow + \frac{T_i \left(G_{i-1/2} + A_{i+1/2} S_i^\downarrow \right)}{1 - A_{i+1/2} R_i}. \tag{14}$$

It then proceeds down through the atmosphere computing the profile of upwelling and downwelling irradiances:

$$F_{i+1/2}^\downarrow = \frac{T_i F_{i-1/2}^\downarrow + R_i G_{i+1/2} + S_i^\downarrow}{1 - A_{i+1/2} R_i}; \tag{15}$$

$$F_{i+1/2}^\uparrow = A_{i+1/2} F_{i+1/2}^\downarrow + G_{i+1/2}. \tag{16}$$

To represent longwave scattering by cloud particles only, we first note that ecRad is almost always configured to compute the equivalent clear-sky radiation profile at the same time as each all-sky calculation. The clear-sky profile is needed for diagnostic purposes to calculate cloud radiative effect, but is also used within the McICA solver (described in section 3) as an ingredient in the method to reduce noise in all-sky irradiances. Since the clear-sky profile contains only gases and aerosol, we may use the much cheaper nonscattering computation described above. But the all-sky calculation may be accelerated as well. First, the more expensive formulas to compute R, T, and $S^{\uparrow\downarrow}$ are only used in layers containing cloud; otherwise the precomputed values for clear-sky are copied over. Second, we exploit the fact that there are always a large number of clear layers above the highest cloudy layer. The highest cloudy layer is identified, and the profile of clear-sky downwelling irradiances above it is copied over. Then (13)–(16) are applied but only from the surface up to the highest cloud layer. Finally, (11) is applied above the highest cloud to complete the profile of upwelling irradiances to TOA.

Comparing the second and third bars in Figure 2, we see that ecRad with this cloud-scattering-only longwave solver is only 4% more expensive than ecRad without any longwave scattering. The longwave solver itself (excluding the cloud generator) is only 16% more expensive.

4.2. Longwave Layer-Wise Emission

A difference between McRad and ecRad exists in the solution to longwave radiative transfer equation, which has a noticeable impact on stratospheric temperatures. In the absence of scattering, (8) and (9) provide the irradiances due to emission by a single layer. They are exact solutions to the two-stream equations in which scattering has been neglected; for example, (9) is a solution to
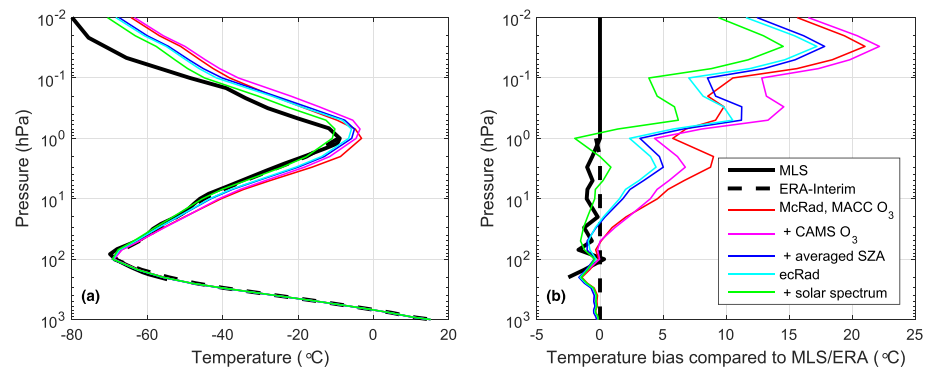
$$dF^\downarrow / d\delta = D \left(-F^\downarrow + B \right), \tag{17}$$

in the case that the Planck function B varies linearly with optical depth, and absorption optical depth δ increases in a downward direction.

McRad, and indeed the longwave solver distributed with RRTM-G, use the following formula instead:

$$S_i^\downarrow = (1 - T_i) \left[p B_{i+1/2} + (1 - p) \bar{B}_i \right], \tag{18}$$

Figure 5. Annual mean temperature from four 1-year uncoupled $T_L 255$ 137-level simulations of the IFS with a 3-hr radiation time step for the years 2000–2004 (colored lines), the Microwave Limb Sounder for the years 2004–2008 (MLS; black solid line) and the ERA-Interim reanalysis for the years 2000–2006 (black dashed line). ERA-Interim has a large warm bias in the mesosphere and so is not shown above 1 hPa. Each model improvement is described in section 5 and is made on top of the previous one in the legend.

in which 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, $\bar{B}_i$, with the weight expressed as a Padé approximation of the form

$$p = \frac{\delta_i}{3.59712 + \delta_i}. \tag{19}$$

This formula results in the effective emission temperature being closer to the mean temperature of the layer than to the temperature at the near edge of the layer. Also, since it is not an exact solution to the differential equation, the result will be resolution dependent as well as being inconsistent with the full scattering formulation. Clough et al. (1992) originally proposed the use of a Padé approximant (although of a different form to that shown above) as a means of avoiding division by zero in (9) in the case of $\delta_i = 0$. On modern hardware this is unnecessary as denominators can be constrained to be greater than or equal to some small positive number without inhibiting vectorization.

The primary atmospheric impact of the difference in the definition of layer-wise emission appears to be on middle-atmosphere temperature since it affects the efficiency of longwave cooling to space. The effect is examined in the next section.

5. Stratospheric Climate

Table 1 lists several recent changes to the radiation scheme and radiatively active trace gases that have an impact on stratospheric temperature. The colored lines in Figure 5a depict the annual and global mean temperature profile of the free-running IFS model using prescribed sea surface temperatures and calling the radiation scheme every 3 hr. The red line is approximately equivalent to IFS Cycle 41R1, which used the McRad radiation scheme and the MACC ozone climatology. The corresponding bias in Figure 5b versus the ERA-Interim reanalysis in the troposphere and stratosphere and the Microwave Limb Sounder (MLS) in the mesosphere reveals that the model has an increasing warm bias with height in the stratosphere, reaching +9 K at 2 hPa, and then increasing further to +20 K in the middle mesosphere. It should be stressed that between 0.1 hPa and the model top at 0.01 hPa, the temperature is always likely to be unrealistic due to a side effect of the *sponge* that modifies the dynamical equations here to prevent reflection of upwardly propagating waves. Figure 5b shows that the global mean climate of the troposphere with McRad is already very good, and this is little changed with the introduction of ecRad.

The magenta line in Figure 5b shows that the upgrade to the CAMS ozone climatology reduced the bias by up to 3 K in the upper stratosphere but increased it in the lower mesosphere. The dark blue line shows the additional effect of introducing the Hogan and Hirahara (2016) method in which the solar zenith angle seen by the radiation scheme is the average over the sunlit part of the 3-hr radiation time step, rather than the

instantaneous value for the time corresponding to the middle of the radiation time step. This leads to a further 1–4 K cooling between the midstratosphere and the top of the model, bringing it much closer to the temperature of the model when the radiation scheme is called every model time step.

The light blue solid line shows the temperature bias after the introduction of ecRad, in a model configuration close to operational IFS Cycle 43R3. The temperature reduction of up to 1 K at the stratopause and above, and temperature increase by 0.2 K at 70 hPa, is a result of the more accurate solution to the longwave equations described in section 4.2. Finally, the green line shows the effect (not yet introduced into an operational cycle) of replacing the default Kurucz (1995) solar spectrum in RRTM-G by the more recent spectrum of Coddington et al. (2016) based on satellite observations of the Sun. The latter has 7 – 8% less ultraviolet than Kurucz, leading to less solar heating by ozone and consequently a cooler middle atmosphere. The greatest reduction is 4 K at the stratopause, in agreement with the findings of Zhong et al. (2008).

The cumulative effect of all these developments since Cycle 41R1 is to have completely removed the large warm bias in the upper stratosphere and substantially improved the mesosphere. The cause of the remaining cold bias in the 10 – 100 hPa range after changing the solar spectrum is being investigated. The cold bias at 200 hPa is untouched by any of the changes described here; it peaks at −5 K at the poles, is common to many climate models, and is believed to be due to excessive transport of water vapor across the tropopause that then cools by longwave emission (Stenke et al., 2008). Work is ongoing to improve the remaining warm bias in the mesosphere, for example, by replacing the diurnal-average ozone climatology with one that represents the reduction of ozone during daylight, and consequent reduction in solar heating. The ecRad radiation scheme also currently assumes local thermodynamic equilibrium (LTE), and inclusion of non-LTE effects may be needed to fully address mesosphere biases.

6. Impact on Forecast Skill

To quantify the impact of the introduction of ecRad on weather forecast skill, 8 months of daily 10-day uncoupled forecasts have been run for a number of model configurations. The control configuration uses IFS Cycle 45R1 at $T_{Co}399$ resolution (around 28 km) with the McRad radiation scheme and a radiation time step of 3 hr. Three further experiments employ the ecRad radiation scheme with a radiation time step of 1, 2, and 3 hr, using the configuration envisaged for the next operational IFS cycle (46R1) shown in bold in Table 2. Each forecast has been evaluated against the operational ECMWF analysis, computing the root-mean-square error in selected model variables.

The black lines in Figure 6 depict the fractional change in root-mean-square error when replacing McRad by ecRad but retaining the 3-hr radiation time step. The first three rows show temperature at three heights, while the last row shows low cloud cover. In general, the change is negative, indicating a reduction of error, although the improvements are modest, lying in the range 0 – 1.5%. The main reason for the improvement appears to be the introduction of longwave scattering by clouds and the bug-fix mentioned in section 2.7; similar experiments with ecRad configured without these changes as in Cycle 43R3 (not shown) suggest no significant difference in skill compared to McRad.

The fact that ecRad is around one-third faster than McRad means that ecRad can be called every 2 hr for the same computational cost as McRad every 3 hr. The red lines in Figure 6 show that this leads to an error reduction of up to 2%, most clearly for 2-m temperature in the tropics at all forecast lead times. For completeness we also show the results for ecRad with a 1-hr radiation time step, which is under consideration as an upgrade for the ECMWF ensemble prediction system and would match the radiation time step used in the HRES forecast. This reduces errors in tropical 2-m temperatures by 3%. The improvement is most pronounced over the Amazon and Congolese rainforests, where low cloud cover forecasts are also significantly improved. This implies that the higher frequency of radiation calls leads to a better interaction of surface heating with shallow convection over land.

These forecasts all used a radiation grid with 6.25 times fewer columns than the rest of the model (2.5 times coarser in each horizontal direction), the same ratio as used in ECMWF's operational ensemble prediction system. An alternative strategy for reinvesting the computational savings would be to use a finer radiation grid, but it was found by Hogan et al. (2017) that even calling the radiation scheme on the same grid as the rest of the model led to very small improvements (less than 1% in 2-m temperature and only in the first 2 days). Reducing the radiation time step appears to be much more effective at improving forecast skill.
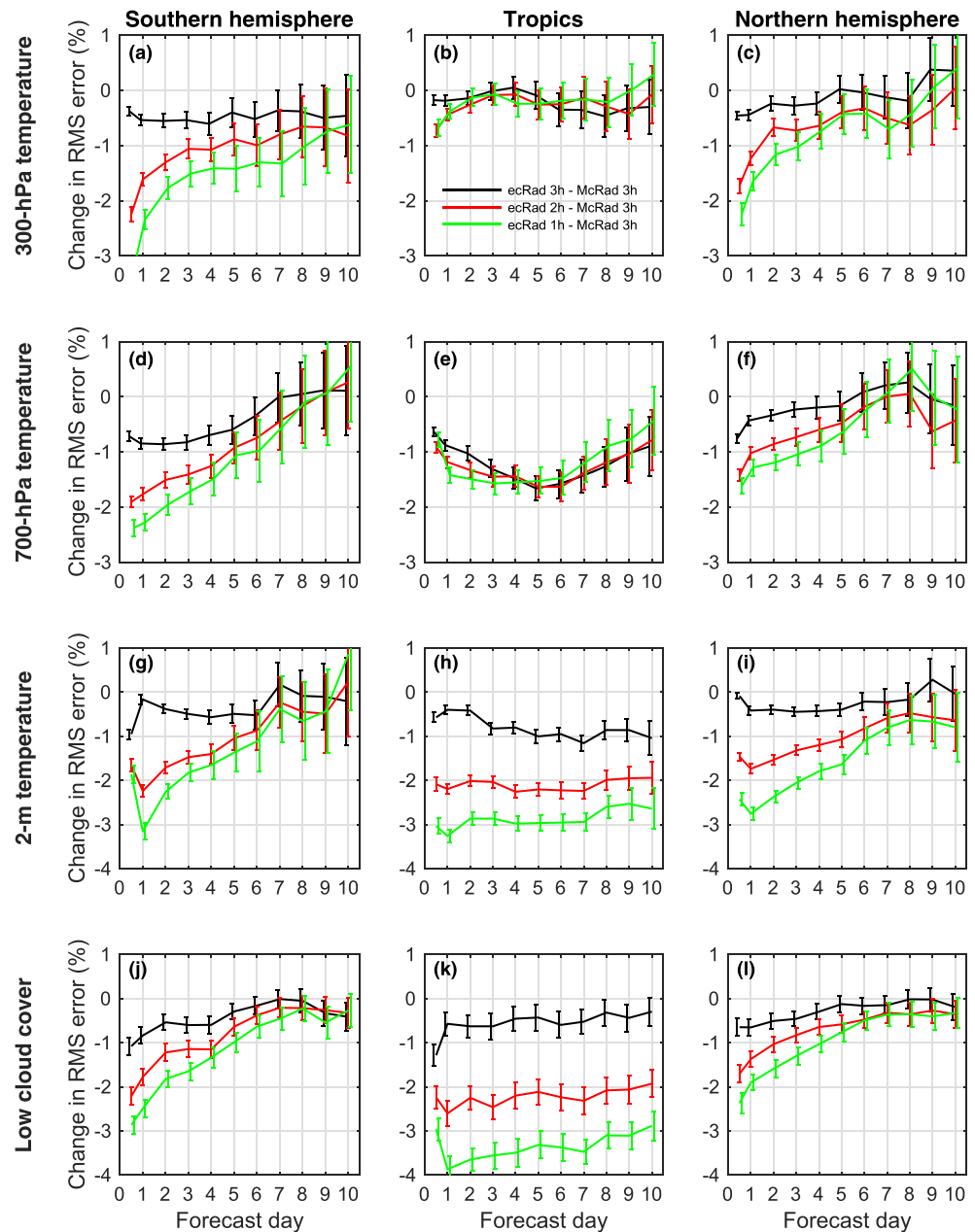
Figure 6. Percentage change to root-mean-square forecast error when switching from the McRad radiation scheme called every 3 hr to ecRad called every (black lines) 3 hr, (red lines) 2 hr, and (green lines) 1 hr. The variables shown are (a–c) 300-hPa temperature, (d–f) 700-hPa temperature, (g–i) 2-m temperature, and (j–l) low cloud cover for the latitude ranges (left column) 90°S to 20°S, (middle column) 20°S to 20°N and (right column) 20°N to 90°N. Eight months of daily atmosphere-only forecasts (June to September 2017 and December 2017 to March 2018) were performed using IFS Cycle 45R1 at $T_{Co}399$ resolution and evaluated against the operational ECMWF analysis. The error bars indicate the 95% confidence intervals and incorporate a 1.1 inflation factor to account for autocorrelated forecasts (Geer, 2016).

7. Conclusions and Future Work

In this paper we have described ecRad, a flexible new radiation scheme for the ECMWF model. It brings several immediate benefits in terms of operational weather forecasting:

1. It is around 41% faster than McRad, a saving that can be reinvested by calling the radiation scheme more frequently, with a measurable increase in forecast skill. It is now planned to reduce the radiation time step in the operational ensemble configuration of the model from 3 hr to 1 hr.

2. It incorporates a method to represent longwave scattering of clouds with an additional computational cost of only 4%, also leading to a modest improvement in forecast skill.

3. The new implementation of the cloud generator in the stochastic McICA solver generates less noise in heating rates and is much faster, also significantly speeding up the cloud cover diagnostic in the IFS.

4. The longwave two-stream equations are now solved exactly, leading to a 1-K cooling of the upper stratosphere and mesosphere, one of a number of recent changes that have reduced the long-standing warm bias in the stratosphere of the IFS.

The last three items above could also be useful in the development of other radiation schemes.

In the longer term, a key advantage of ecRad is its modular design, which makes it straightforward to explore alternative formulations to individual components of the scheme. One area requiring attention is the radiation scheme used at ECMWF for minimizing the cost function in the variational data assimilation system, which must be differentiable and computationally fast. For this reason, the data assimilation system uses a version of the Morcrette scheme that was operational before the introduction of RRTM-G in either the longwave or the shortwave, that is, before Cycle 22R3 in Table 1. The inconsistency between the Morcrette scheme and the scheme used in the full nonlinear model (whether McRad or ecRad) is certainly suboptimal. Ideally an alternative gas optics model would be developed for ecRad with a flexible number of spectral intervals, to allow the user control over the trade-off between speed and accuracy, and hence to provide a configuration of ecRad fast enough to be used in variational data assimilation.

Another challenge is to refine many of the assumptions made in the radiation scheme that determine the radiative effect of clouds, such that they agree as well as possible with the latest observations, and are consistent with the IFS cloud scheme and the observational forward operators in the data assimilation system. This could include switching from EXP-EXP to EXP-RAN overlap, changing the fractional standard deviation of in-cloud water content from 1 to a value closer to observations (e.g., Ahlgrimm & Forbes, 2017; Shonk et al., 2010), representing convective clouds in the radiation scheme, and making use of an ice optics parameterization that represents more recent findings that ice particles are optically *rough* (Baran et al., 2014; Yi et al., 2013). Although not affordable operationally, ecRad also has the option to represent 3-D effects for the first time in a global model. This will enable us to determine the importance of 3-D radiative effects on the Earth's radiation budget, and if it is significant, to explore cheaper ways of approximating the impact in the operational model.

In addition to its use in the IFS, ecRad is being tested in the Meso-NH and ICON (Icosahedral Nonhydrostatic) models. It will be used in the next version of the EC-Earth climate model. An off-line version of ecRad is available for educational and noncommercial research.

References

Ahlgrimm, M., & Forbes, R. M. (2017). Regime dependence of ice cloud heterogeneity — A convective life-cycle effect. *Quarterly Journal of the Royal Meteorological Society*, *143*, 3259–3268.

Ahlgrimm, M., Forbes, R. M., Morcrette, J.-J., & Neggers, R. A. (2016). ARM's impact on numerical weather prediction at ECMWF. *Meteorological Monographs*, *57*, 28.1–28.13.

Baran, A. J., Hill, P., Furtado, K., Field, P., & Manners, J. (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. *Journal of Climate*, *27*, 7725–7752.

Bergman, J. W., & Rasch, P. J. (2002). Parameterizing vertically coherent cloud distributions. *Journal of the Atmospheric Sciences*, *59*, 2165–2182.

Bozzo, A., Remy, S., Benedetti, A., Flemming, J., Bechtold, P., Rodwell, M., & Morcrette, J.-J. (2017). Implementation of a CAMS-based aerosol climatology in the IFS (Tech. Memo. No. 801): ECMWF. 33 pp.

Chou, M.-D., Lee, K.-T., Tsay, S.-C., & Fu, Q. (1999). Parameterization for cloud longwave scattering for use in atmospheric models. *Journal of Climate*, *12*, 159–169.

Clough, S. A., Iacono, M. J., & Moncet, J. L. (1992). Line-by-line calculations of atmospheric fluxes and cooling rates: Application to water vapor. *Journal of Geophysical Research*, *97*, 15,761–15,785.

Coddington, O., Lean, J. L., Pilewskie, P., Snow, M., & Lindholm, D. (2016). A solar irradiance climate data record. *Bulletin of the American Meteorological Society*, *97*, 1265–1282.

Costa, S. M. S., & Shine, K. P. (2006). An estimate of the global impact of multiple scattering by clouds on outgoing long-wave radiation. *Quarterly Journal of the Royal Meteorological Society*, *132*, 885–895.

Di Giuseppe, F., & Tompkins, A. M. (2015). Generalizing cloud overlap treatment to include the effect of wind shear. *Journal of the Atmospheric Sciences*, *72*, 2865–2876.

Dufresne, J.-L., Gautier, C., Ricchiazzi, P., & Fouquart, Y. (2002). Longwave scattering effects of mineral aerosols. *Journal of the Atmospheric Sciences*, *59*, 1959–1966.

Ebert, E. E., & Curry, J. A. (1992). A parameterization of ice cloud optical properties for climate models. *Journal of Geophysical Research*, *97*, 3831–3836.

Edwards, J. M., & Slingo, A. (1996). Studies with a flexible new radiation code: 1. Choosing a configuration for a large-scale model. *Quarterly Journal of the Royal Meteorological Society*, *122*, 689–719.

Flemming, J., Benedetti, A., Inness, A., Engelen, R. J., Jones, L., Huijnen, V., et al. (2017). The CAMS interim reanalysis of carbon monoxide, ozone and aerosol for 2003–2015. *Atmospheric Chemistry and Physics*, *17*, 1945–1983.

Fortuin, J. P. F., & Langematz, U. (1994). An update on the global ozone climatology and on concurrent ozone and temperature trends. *Proceedings of the SPIE Sensing and Modeling*, *2311*, 207–216.

Fouquart, Y. (1987). Radiative transfer in climate modeling, *NATO Adv. Study Inst. on physically based modeling and simulation of climate and climate changes* (pp. 223–283). Erice, Sicily: Springer, Dordrecht. 11–23 May 1986.

Fu, Q. (1996). An accurate parameterization of the solar radiative properties of cirrus clouds. *Journal of Climate*, *9*, 2058–2082.

Fu, Q., Liou, K. N., Cribb, M. C., Charlock, T. P., & Grossman, A. (1997). Multiple scattering parameterization in thermal infrared radiative transfer. *Journal of the Atmospheric Sciences*, *54*, 2799–2812.

Fu, Q., Yang, P., & Sun, W. B. (1998). An accurate parametrization of the infrared radiative properties of cirrus clouds of climate models. *Journal of Climate*, *11*, 2223–2237.

Geer, A. J. (2016). Significance of changes in medium-range forecast scores. *Tellus A*, *68*, 1–21. https://doi.org/10.3402/tellusa.v68.30229

Geleyn, J.-F., & Hollingsworth, A. (1979). An economical analytical method for the computation of the interaction between scattering and line absorption of radiation. *Beitraege zur Physik der Atmosphaere*, *52*, 1–16.

Hill, P. G., Manners, J., & Petch, J. (2011). Reducing noise associated with the Monte Carlo independent column approximation for weather forecasting models. *Quarterly Journal of the Royal Meteorological Society*, *137*, 219–228.

Hogan, R. J., Ahlgrimm, M., Balsamo, G., Beljaars, A. C. M., Berrisford, P., Bozzo, A., et al. (2017). Radiation in numerical weather prediction (Tech. Memo. No. 816): ECMWF. 48 pp.

Hogan, R. J., & Bozzo, A. (2015). Mitigating errors in surface temperature forecasts using approximate radiation updates. *Journal of Advances in Modeling Earth Systems*, *7*, 836–853.

Hogan, R. J., & Bozzo, A. (2016). ECRAD: A new radiation scheme for the IFS (Tech. Memo. No. 787): ECMWF. 33 pp.

Hogan, R. J., & Hirahara, S. (2016). Effect of solar zenith angle specification in models on mean shortwave fluxes and stratospheric temperatures. *Geophysical Research Letters*, *43*, 482–488. https://doi.org/10.1002/2015GL066868

Hogan, R. J., & Illingworth, A. J. (2000). Deriving cloud overlap statistics from radar. *Quarterly Journal of the Royal Meteorological Society*, *126*, 2903–2909.

Hogan, R. J., Quaife, T., & Braghiere, R. (2018). Fast matrix treatment of 3-D radiative transfer in vegetation canopies: SPARTACUS-Vegetation 1.1. *Geoscientific Model Development*, *11*, 339–350.

Hogan, R. J., Schäfer, S. A. K., Klinger, C., Chiu, J.-C., & Mayer, B. (2016). Representing 3-D cloud-radiation effects in two-stream schemes: 2. Matrix formulation and broadband evaluation. *Journal of Geophysical Research: Atmospheres*, *121*, 8583–8599. https://doi.org/10.1002/2016JD024875

Joseph, J. H., Wiscombe, W. J., & Weinman, J. A. (1976). The delta-Eddington approximation for radiative flux transfer. *Journal of the Atmospheric Sciences*, *33*, 2452–2459.

Kurucz, R. L. (1995). The solar spectrum: atlases and line identifications. *In Laboratory and Astronomical High Resolution Spectra*, *81*, 17–31.

Lindner, T. H., & Li, J. (2000). Parameterization of the optical properties for water clouds in the infrared. *Journal of Climate*, *13*, 1797–1805.

Mann, G. W., Carslaw, K. S., Spracklen, D. V., Ridley, D. A., Manktelow, P. T., Chipperfield, M. P., et al. (2010). Description and evaluation of GLOMAP-mode: A modal global aerosol microphysics model for the UKCA composition-climate model. *Geoscientific Model Development*, *3*, 519–551.

Manners, J., Thelen, J.-C., Petch, J., Hill, P., & Edwards, J. M. (2009). Two fast radiative transfer methods to improve the temporal sampling of clouds in numerical weather prediction and climate models. *Quarterly Journal of the Royal Meteorological Society*, *135*, 457–468.

Martin, G. M., Johnson, D. W., & Spice, A. (1994). The measurement and parameterization of effective radius of droplets in warm stratocumulus. *Journal of the Atmospheric Sciences*, *51*, 1823–1842.

Meador, W. E., & Weaver, W. R. (1980). Two-stream approximations to radiative transfer in planetary atmospheres: A unified description of existing methods and a new improvement. *Journal of the Atmospheric Sciences*, *37*, 630–643.

Mlawer, E. J., Taubman, S. J., Brown, P. D., Iacono, M. J., & Clough, S. A. (1997). Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave. *Journal of Geophysical Research*, *102*, 16,663–16,682.

Morcrette, J.-J. (1991). Radiation and cloud radiative properties in the ECMWF operational weather forecast model. *Journal of Geophysical Research*, *96*, 9121–9132.

Morcrette, J.-J., Barker, H. W., Cole, J. N. S., Iacono, M. J., & Pincus, R. (2008). Impact of a new radiation package, McRad, in the ECMWF integrated forecasting system. *Monthly Weather Review*, *136*, 4773–4798.

Morcrette, J.-J., & Jakob, C. (2000). The response of the ECMWF model to changes in cloud overlap assumption. *Monthly Weather Review*, *128*, 1707–1732.

Morcrette, J.-J., Mozdzynski, G., & Leutbecher, M. (2008). A reduced radiation grid for the ECMWF integrated forecasting system. *Monthly Weather Review*, *136*, 4760–4772.

Nielsen, K. P., Gleeson, E., & Rontu, L. (2014). Radiation sensitivity tests of the HARMONIE 37h1 NWP model. *Geoscientific Model Development*, *7*, 1433–1449.

Pincus, R., Barker, H. W., & Morcrette, J.-J. (2003). A fast, flexible, approximate technique for computing radiative transfer in inhomogeneous clouds. *Journal of Geophysical Research*, *108*, 4376. https://doi.org/10.1029/2002JD003322

Räisänen, P., Barker, H. W., Khairoutdinov, M. F., Li, J., & Randall, D. A. (2004). Stochastic generation of subgrid-scale cloudy columns for large-scale models. *Quarterly Journal of the Royal Meteorological Society*, *130*, 2047–2067.

Rodwell, M. J., & Jung, T. (2008). Understanding the local and global impacts of model physics changes: An aerosol example. *Quarterly Journal of the Royal Meteorological Society*, *134*, 1479–1497.

Schäfer, S. A. K., Hogan, R. J., Klinger, C., Chiu, J.-C., & Mayer, B. (2016). Representing 3D cloud-radiation effects in two-stream schemes: 1. Longwave considerations and effective cloud edge length. *Journal of Geophysical Research: Atmospheres*, *121*, 8567–8582. https://doi.org/10.1002/2016JD024876

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

Shonk, J. K. P., Hogan, R. J., Edwards, J. M., & Mace, G. G. (2010). Effect of improving representation of horizontal and vertical cloud structure on the Earth's radiation budget: 1. Review and parameterisation. *Quarterly Journal of the Royal Meteorological Society*, *136*, 1191–1204.

Shonk, J. K. P., Hogan, R. J., & Manners, J. (2012). Impact of improved representation of horizontal and vertical cloud structure in a climate model. *Climate Dynamics*, *38*, 2365–2376.

Slingo, A. (1989). A GCM parametrization for the shortwave radiative properties of water clouds. *Journal of the Atmospheric Sciences*, *46*, 1419–1427.

Smith, E. A., & Shi, L. (1992). Surface forcing of the infrared cooling profile over the Tibetan plateau—1: Influence of relative longwave radiative heating at high altitude. *Journal of the Atmospheric Sciences*, *49*, 805–822.

Stackhouse, P. W., & Stephens, G. L. (1991). A theoretical and observational study of the radiative properties of cirrus: Results from FIRE 1986. *Journal of the Atmospheric Sciences*, *48*, 2044–2059.

Stenke, A., Grewe, V., & Ponater, M. (2008). Lagrangian transport of water vapor and cloud water in the ECHAM4 GCM and its impact on the cold bias. *Climate Dynamics*, *31*, 491–506.

Sun, Z. (2001). Reply to comments by Greg M. McFarquhar on 'Parameterization of effective sizes of cirrus-cloud particles and its verification against observations'. *Quarterly Journal of the Royal Meteorological Society*, *127*, 267–271.

Sun, Z., & Rikus, L. (1999). Parameterization of effective sizes of cirrus-cloud particles and its verification against observations. *Quarterly Journal of the Royal Meteorological Society*, *125*, 3037–3055.

Tanré, D. (1984). First results of the introduction of an advanced aerosol-radiation interaction in the ECMWF low resolution global model. In H. E. Gerber & A. Deepak (Eds.), *Aerosols and their climatic effects* (pp. 133–177). Hampton, VA: A. Deepak Publ.

Tegen, I., Hollrig, P., Chin, M., Fung, I., Jacob, D., & Penner, J. (1997). Contribution of different aerosol species to the global aerosol extinction optical thickness: Estimates from model results. *Journal of Geophysical Research*, *102*, 23,895–23,915.

Tompkins, A. M., Cardinali, C., Morcrette, J.-J., & Rodwell, M. (2005). Influence of aerosol climatology on forecasts of the African easterly jet. *Geophysical Research Letters*, *32*, L10801. https://doi.org/10.1029/2004GL022189

Wiscombe, W. J. (1976). On initialization, error and flux conservation in the doubling method. *Journal of Quantitative Spectroscopy and Radiation Transfer*, *16*, 637–658.

Yi, B., Yang, P., Baum, B. A., L'Ecuyer, T., Oreopoulos, L., Mlawer, E. J., et al. (2013). Influence of ice particle surface roughening on the global cloud radiative effect. *Journal of the Atmospheric Sciences*, *70*, 2794–2807.

Zhong, W., Osprey, S. M., Gray, L. J., & Haigh, J. D. (2008). Influence of the prescribed solar spectrum on calculations of atmospheric temperature. *Geophysical Research Letters*, *35*, L22813. https://doi.org/10.1029/2008GL035993