Process-based climate model development harnessing machine learning: III. The Representation of Cumulus Geometry and their 3D Radiative Effects

Najda Villefranque[1], Danny Williamson[2], Fleur Couvreux[3], Frédéric Hourdin[4], Jacques Gautrais[5], Richard Fournier[6], Robin J Hogan[7], Stéphane Blanco[6], and Victoria Volodina[8]

[1]Centre National de Recherches Météorologiques
[2]University of Exeter
[3]Université Toulouse, CNRM, Meteo-France, CNRS
[4]LMD
[5]CRCA, CBI
[6]LAPLACE
[7]ECMWF
[8]Turing Institute

November 24, 2022

Abstract

Process-scale development, evaluation and calibration of physically-based parameterizations are key to improve weather and climate models. Cloud–radiation interactions are a central issue because of their major role in global energy balance and climate sensitivity. In a series of papers, we propose papers a strategy for process-based calibration of climate models that uses machine learning techniques. It relies on systematic comparisons of single-column versions of climate models with explicit simulations of boundary-layer clouds (LES). Parts I and II apply this framework to the calibration of boundary layer parameters targeting first boundary layer characteristics and then global radiation balance at the top of the atmosphere. This third part focuses on the calibration of cloud geometry parameters that appear in the parameterization of radiation. The solar component of a radiative transfer scheme (ecRad) is run in offline single-column mode on input cloud profiles synthesized from an ensemble of LES outputs. A recent version of ecRad that includes explicit representation of the effects of cloud geometry and horizontal transport is evaluated and calibrated by comparing radiative metrics to reference values provided by Monte Carlo 3D radiative transfer computations. Errors on TOA, surface and absorbed fluxes estimated by ecRad are computed for an ensemble of cumulus fields. The average root-mean-square error can be less than 5 Wm^{-2} provided that 3D effects are represented and that cloud geometry parameters are well calibrated. A key result is that configurations using calibrated parameters yield better predictions than those using parameter values diagnosed in the LES fields.

1

Process-based climate model development harnessing machine learning: III. The Representation of Cumulus Geometry and their 3D Radiative Effects

Najda Villefranque[1], Stéphane Blanco[2], Fleur Couvreux[1], Richard Fournier[2], Jacques Gautrais[3], Robin J. Hogan[4], Frédéric Hourdin[5], Victoria Volodina[6], Daniel Williamson[6,7]

[1]CNRM, Universit de Toulouse, Meteo-France, CNRS, Toulouse, France
[2]LAPLACE, Universit de Toulouse, CNRS, Toulouse, France
[3]Centre de Recherches sur la Cognition Animale (CRCA), Centre de Biologie Intgrative (CBI), Universit de Toulouse, CNRS, Universit Paul Sabatier, France
[4]ECMWF, Reading, United Kingdom
[5]LMD-IPSL, Sorbonne University, CNRS, 4 pl Jussieu, Paris, France
[6]The Alan Turing Institute, 96 Euston Road, London, United Kingdom
[7]Exeter University, Exeter, United Kingdom

Key Points:

- A state-of-the-art radiative transfer scheme is calibrated using Monte Carlo 3D radiative transfer in LES 3D cumulus fields
- The errors due to various approximations commonly made in radiative transfer schemes are untangled and quantified
- Radiative estimates are more accurate using calibrated than observed 3D-cloud-geometry parameters

Corresponding author: Najda Villefranque, `najda.villefranque@gmail.com`

Abstract

Process-scale development, evaluation and calibration of physically-based parameterizations are key to improve weather and climate models. Cloud–radiation interactions are a central issue because of their major role in global energy balance and climate sensitivity. In a series of papers, we propose papers a strategy for process-based calibration of climate models that uses machine learning techniques. It relies on systematic comparisons of single-column versions of climate models with explicit simulations of boundary-layer clouds (LES). Parts I and II apply this framework to the calibration of boundary layer parameters targeting first boundary layer characteristics and then global radiation balance at the top of the atmosphere. This third part focuses on the calibration of cloud geometry parameters that appear in the parameterization of radiation. The solar component of a radiative transfer scheme (ecRad) is run in offline single-column mode on input cloud profiles synthesized from an ensemble of LES outputs. A recent version of ecRad that includes explicit representation of the effects of cloud geometry and horizontal transport is evaluated and calibrated by comparing radiative metrics to reference values provided by Monte Carlo 3D radiative transfer computations. Errors on TOA, surface and absorbed fluxes estimated by ecRad are computed for an ensemble of cumulus fields. The average root-mean-square error can be less than 5 Wm^{-2} provided that 3D effects are represented and that cloud geometry parameters are well calibrated. A key result is that configurations using calibrated parameters yield better predictions than those using parameter values diagnosed in the LES fields.

1 Introduction

Cloud–radiation interactions, through their strong impact on the Earth's global energy balance (Ramanathan et al., 1989), are key processes in the evolution of the Earth's climate. The radiative effect of cumulus clouds is particularly important due to their permanent presence in large regions of the Earth's troposphere and their large optical thickness (Berg et al., 2011). They are also responsible for a large part of the uncertainties around climate sensitivity (Dufresne & Bony, 2008; Bony et al., 2015). These results motivate the improvement of the representation of cloud–radiation interactions in large-scale models, who still struggle to accurately represent the radiative effects of these small short-lived complex clouds (Dolinar et al., 2015).

Cloud geometry affects radiation in several ways. In particular, poor representation of vertical overlap and horizontal heterogeneity, as well as the neglect of horizontal transport, have been identified as sources of radiative biases in large-scale models for decades (see e.g. McKee and Cox (1974); Várnai and Davies (1999); Barker et al. (2003) among many others, or Marshak and Davis (2005)). Recent developments of radiation parameterizations include realistic representations of these effects, for example, McICA (Pincus et al., 2003) or Tripleclouds (Shonk & Hogan, 2008) for the heterogeneity, the exponential-random model of Hogan and Illingworth (2000) for vertical overlap and SPARTACUS (Hogan & Shonk, 2013; Schäfer et al., 2016; Hogan et al., 2016, 2019) for 3D effects. These recent propositions need more systematic evaluation before they can be routinely used in operational models.

Besides deriving new formulations to represent subgrid-scale physical processes, a crucial aspect of model development is the adjustment of the parameters to calibrate reference configurations of climate models or operational configurations of weather forecast models (Schmidt et al., 2017; Bellprat et al., 2012; Duan et al., 2017; Hourdin et al., 2017). In this calibration process, the objective is to find a configuration consistent with target quantities that measure different aspects of the observed climate or weather. In climate model calibration, the targets are most often radiation observables such as the global solar and thermal top-of-atmosphere (TOA) fluxes, their spatial distribution or their clear sky and cloud radiative effect (CRE) components (Hourdin et al., 2017). Because of the

key role of clouds on the radiation balance, free parameters that appear in cloud related parameterizations are largely used in the calibration process (see e.g. Golaz et al. (2013); Mauritsen et al. (2012); Hourdin et al. (2017)). As highlighted by Hourdin et al. (2017), documentation of existing calibration techniques and propositions of new approaches is bound to accelerate improvement of climate models.

In the first two papers of this series (Couvreux et al., 2020; Hourdin et al., 2020), a strategy to reduce cloud-related uncertainties in large-scale models is proposed. It strongly relies on comparisons between Single Column Model (SCM) and Large-Eddy Simulations (LES) to develop, evaluate and calibrate parameterizations at the process scale, in particular those that model boundary-layer (BL) transport and clouds. Free open-source numerical tools are provided to the community to promote a transparent, rigorous and efficient procedure to calibrate models. The first paper (Couvreux et al., 2020) is focused on the description of High-Tune:explorer (htexplo), a tuning tool based on history matching, developed jointly by physicists in atmospheric science and statisticians in Uncertainty Quantification. Its potential is demonstrated through the calibration of BL parameterizations of an SCM, targeting reference BL metrics computed in the LES fields. It thereby ensures that parameters controling the BL are calibrated to obtain the right BL properties. The second paper (Hourdin et al., 2020) uses the same tool to calibrate the 3D climate model targeting observed radiation at the TOA. In this global calibration, the model is only allowed to explore a restricted parameter space, which is obtained by prior calibration of the parameterization of shallow convection using the SCM/LES framework. It thereby ensures that BL parameters are calibrated to obtain both the right BL properties and the right global radiation, i.e., the right TOA cloud radiative forcing for the right BL clouds.

Preliminary calibration at the process-scale is a way to avoid compensation errors during global calibration. Indeed, a bias in cloud representation might lead to the introduction of a bias in radiation in order to get the right global cloud radiative effect. It is the case of the famous "too few too bright" syndrome found in numerous climate models (Karlsson et al., 2008; Nam et al., 2012), in which deficient representation of BL processes leads to an underestimation of low cloud cover, that is then compensated by an overestimation of cloud optical depth so that reflectance is increased. This overestimation of cloud reflectivity could also come from trying to compensate for too large transmissivity at low sun angles due to the neglect of 3D radiative effects. This same lack of 3D effects could also lead to an artificial increase of cloud cover to mimic the increase of the "effective" cloud cover seen by the sun when it goes down.

To prevent these compensations, this paper adds a step to the multistage calibration procedure: the process-based calibration of an offline parameterisation — the radiation scheme. By targeting radiative metrics computed at the cloud-field scale, we ensure that, given the right clouds, radiation parameters are calibrated to obtain the right local radiation. The set of radiation parameters that are retained at this scale can then be explored in interaction with the other parameterizations during the 3D global calibration exercise; the radiation scheme will not be allowed to unphysically compensate for cloud biases.

For this purpose, the SCM/LES framework used for evaluation and calibration of BL parameterizations is adapted to the offline mode of the ecRad radiative transfer (RT) scheme (Hogan & Bozzo, 2018). Outputs from LES are used as reference cloud fields, in which 3D RT is solved by Monte Carlo (MC) to provide reference radiative metrics. These same cloud fields are reduced to a few vertical profiles to mimic a "perfect" SCM output, in the sense that it perfectly matches the reference LES cloud field statistics. This way, the inaccurate representation of BL processes in large-scale models cannot be blamed for errors on CRE estimates. With this, the errors due to the different approximations used in the RT scheme are quantified. Specific attention is paid to the radiative effect of approximate description of the subgrid clouds 3D structure in the RT scheme. As a

physical interpretation is associated with the parameters that appear in the description of cloud geometry, they can be estimated from the LES fields. Therefore, it is not obvious that the htexplo tool should be used to adjust these parameters. Yet, the results of the calibration process indicate that ecRad can be improved by using tuned parameters instead of measured ones.

The paper is organised as follows: Section 2 describes the ecRad RT scheme, the MC model, the 3D LES and the resulting 1D profiles. In Section 3, ecRad is evaluated against MC simulations, quantifying independently the different sources of errors, first on overcast single-layer homogeneous clouds to exclude errors due to geometrical effects, then on cumulus scenes to test different representations of cloud geometry and 3D effects. In Section 4, the htexplo tool is briefly described before being applied to explore the space of possible values for cloud geometrical parameters in SPARTACUS. Four calibrated configurations are then analysed. Calibrated configurations are found to systematically improve surface and TOA fluxes but not the absorption. The main results are discussed in Section 5.

2 Radiative Transfer Models and Cloudy Atmosphere Data

An atmospheric radiative transfer parameterisation relies on a statistical representation of cloud fields, in the form of 1D profiles of a small number of cloud geometry variables. In order to untangle and estimate the various biases that are inherent to such parameterisations, a 3D RT reference is built using a Monte Carlo (MC) model on fully resolved 3D cloud fields obtained from LES. This section presents both radiative models: the ecRad radiation scheme (Hogan & Bozzo, 2018) and the reference MC model (Villefranque et al., 2019); as well as the LES clouds and the methodology used to translate these 3D fields into the 1D profiles used as inputs to ecRad.

2.1 Radiative transfer models

2.1.1 ecRad: a flexible radiation scheme for large-scale atmospheric models

The radiation scheme ecRad (Hogan & Bozzo, 2018) is operational since 2017 in the Integrated Forecasting System (IFS) at the European Centre for Medium-Range Weather Forecasts (ECMWF). Recent efforts have led to a notable increase in flexibility as well as in efficiency (the authors report a 41% increase in speed for the operational configuration) compared to previous schemes. Another important step was the development of SPARTACUS (Schäfer et al., 2016; Hogan et al., 2016, 2019), a 2-stream based solver that explicitly represents horizontal transport of light.

In this work, the offline version is used, which differs from the coupled version mostly by how inputs and outputs are handled. Most of the configuration will remain the same (see Table 1), but the solver and the geometrical parameters that we intend to calibrate will differ from one run to another.

As far as liquid clouds are concerned, six main sources of errors were identified. They are described in the following, and Section 3 quantifies the errors due to each of these approximations.

1. **Spectral dimension and optical properties.** The RRTMG gas model used in ecRad and other radiation schemes uses the correlated k-distribution method in 14 bands in the solar and 16 bands in the thermal. Here, only solar computations are performed, integrating fluxes on a $0.2 - 12.5$ μm interval. The SOCRATES data that describe the liquid cloud optical properties are the coefficients of a polynomial function fitted to Mie computations performed on a finite number of wave-

Table 1. Configuration of ecRad in the following work.

Property	Option	Reference
Gas model	RRTMG-IFS	Iacono et al. (2008)
Aerosols	None	
Liquid cloud optics	SOCRATES	Manners et al. (2017)
Liquid water content distrib. shape	Gamma	
Cloud overlap scheme	Exp-Ran	Hogan and Illingworth (2000)
Solver	Tripleclouds	Shonk and Hogan (2008)
	or SPARTACUS	Schäfer et al. (2016); Hogan et al. (2016)
Entrapment (SPARTACUS only)	Explicit	Hogan et al. (2019)

lengths and effective radii, spectrally averaged over the RRTM narrow bands. Details can be found in the SOCRATES technical report (Manners et al., 2017) and in the Supporting Information that accompanies this paper.

2. **The two-stream approximation.** RT solvers based on the two-stream approximation are often the most efficient, if not the most accurate ones. They basically consist in summarizing the angular distribution of diffuse fluxes into two "streams": one downward flux and one upward flux, at each interface between two model layers. This reduces the number of unkown variables in the system that couples the directional fluxes at each interface. The "direct" or unscattered flux is treated as an additional stream. This limited amount of represented directions is known to induce biases (see e.g. Barker et al. (2015)).

3. **Angular distribution of scattered light.** Two-stream models do not use detailed angular phase functions. They instead rely on the cosine-weighted average of the phase function, namely the asymmetry parameter g. It appears in the computation of reflected and transmitted fluxes at each layer interface. In addition to this simplification, ecRad and numerous other two-stream based models use the delta-scaling approximation. It corrects for clouds being too reflective due to the incapacity of basic two-stream schemes to account for large amounts of energy scattered in a very small solid angle around the forward direction. The approximation consists in scaling both the optical depth and the asymmetry parameter so as to treat some of the forward scattering as direct transport. The scaling used in ecRad considers a fraction $f = g^2$ of scattered light as unscattered, as per the δ-Eddington model of Joseph et al. (1976).

4. **Vertical overlap of partially cloudy layers.** A unique vertical profile of cloud fractions can be obtained from very different cloud fields, yielding quite different radiative effects. Indeed, the radiative effect of clouds depends on both the total cloud cover and the total cloud optical depth. Given a cloud fraction profile, if cloudy regions are maximally overlapped, then the total cloud cover will be smaller and the total cloud optical depth larger than if cloudy regions are more randomly overlapped. Hogan and Illingworth (2000) proposed to express the cloud cover $C_{true}(i, i+1)$ of two adjacent layers i and $i+1$ of cloud fractions c_i and c_{i+1} as a weighted sum of the two following terms:

- $C_{max}(i, i+1) = \max(c_i, c_{i+1})$ which is the "maximum" cloud cover and
- $C_{rand}(i, i+1) = c_i + c_{i+1} - c_i c_{i+1}$ which is the "random" cloud cover.

Then,

$$C_{true}(i, i+1) = \alpha_{i,i+1} C_{max}(i, i+1) + (1 - \alpha_{i,i+1}) C_{rand}(i, i+1) \qquad (1)$$

and α is called the overlap parameter, with $\alpha = 1$ for maximum overlap and $\alpha = 0$ for random. In the two-stream scheme, it constrains, at the interface between each

pair of layers, the distribution of upward (downward) fluxes into the cloudy and clear regions of the layer above (below). Hogan and Illingworth (2000) observed from radar measurements that α could be modeled as an exponential that decreases with layers separation:

$$\alpha_{i,i+1} = \exp(-\frac{\Delta z(i,i+1)}{z_0})$$ (2)

where $\Delta z(i,i+1)$ is the vertical distance that separates the center of the two layers and z_0 is a decorrelation length: the degree of overlap of two layers separated by half this length is around 60%, and falls down to 5% at three times this length.

5. **Horizontal heterogeneity of in-cloud liquid water.** Since Beer's exponential law is a convex function, Jensen's inequality applies (Jensen, 1906), resulting in systematic bias due to averaging of optical properties (Newman et al., 1995): the mean transmitted flux under a cloud of horizontally varying liquid water content (LWC) is always larger than the transmitted flux under the equivalent homogeneous cloud of average LWC. To represent the effect of horizontal variations of LWC on radiation, the Tripleclouds method has been proposed by Shonk and Hogan (2008). In addition to solving solving radiation independently in cloudy and clear regions of each layer, it further divides the cloudy region into a thin subregion and a thick one. To distribute the LWC into the two sub-regions of a given layer and then infer their respective optical depths, a gamma-shaped distribution of the liquid water is assumed, characterized by a mean and a standard deviation σ. Tripleclouds uses the fractional standard deviation (FSD) of the distribution (ratio of σ to mean in-cloud LWC) to robustly characterize the horizontal variability of LWC in each layer.

6. **Horizontal transport of light.** In Tripleclouds, layers are divided into sub-regions and radiation is solved independently in each sub-region of the layer. Light cannot be transported from clear to cloudy sub-regions of the same layer. However, many studies have shown that horizontal transport of light significantly modifies the distribution of energy in the atmosphere: transport through cloud sides increases transmission at high sun and decreases it at low sun, while multiple reflections combined to horizontal transport leads to entrapment of upward flux, thereby increasing transmission (McKee & Cox, 1974; McKee & Klehr, 1978; Várnai & Davies, 1999; Barker et al., 2003; Hogan & Shonk, 2013; Hogan et al., 2019). SPARTACUS is the first two-stream-based solver that allows these 3D effects to be represented in large-scale models. It explicitly represents entrapment by computing the mean horizontal distance traveled by reflected light. To represent transport through cloud sides, it uses the Tripleclouds approximation for the cloud field then adds exchange terms between regions. This term is proportional to the length of the interface between clear and cloudy regions. For a given cloud fraction, 3D effects will be larger for a large number of small clouds than for a single large cloud. From the total cloud perimeter density p (perimeter length to domain horizontal area, of unit inverse length) and the cloud fraction of the layer c, Schäfer et al. (2016) define the cloud effective scale C_s as:

$$C_s = \frac{4c(1-c)}{p}$$ (3)

Given observed values for c and p, C_s is the size of the cloud that is such that when a virtual layer is filled randomly with instances of this cloud until the cloud fraction of the layer is c, then the total cloud perimeter in the virtual scene is p. Other choices of representation for p have been explored, in particular the recent work of Fielding et al. (2020) has led to a new parameterization for the cloud perimeter in SPARTACUS.

2.1.2 Monte Carlo reference computations of solar 3D RT

A MC code is used to compute solar 3D RT in 3D cloud fields, considered as the "truth" in comparisons to ecRad estimates. The Monte Carlo methods are widely used to accurately compute 3D RT in complex media (see for example Marchuk et al. (1980), Mayer (2009) or Marshak and Davis (2005)). It consists in tracking a large number of virtual photon paths (in this work, ten million per simulation) throughout a virtual medium, explicitly simulating all radiative processes such as emission, absorption, scattering and surface reflection. The model used here is based on the High-Tune library described in Villefranque et al. (2019), and is freely available online [1]. Whenever a path hits the ground or the TOA, its weight is added to a virtual sensor. Paths are terminated upon absorption or escape in space.

Spectral integration over the solar spectrum is performed according to the correlated-k model RRTMG. The k-distribution data were retrieved from ecRad to insure fair comparisons. At each path, a narrow band is sampled as per the in-band ratio of incident solar flux. Then, a quadrature point is sampled in the band as per the quadrature weights provided with the k-distribution data. The gas optical properties are then set accordingly.

Optical properties for liquid droplets (extinction coefficient, single scattering albedo, discretized phase function and/or asymmetry parameter) are provided as an input to the MC code. The droplet size distribution is assumed to be the same everywhere within the clouds, with a homogeneous effective radius of 10 μm (the same hypothesis is used in ecRad runs). Two choices of optical properties models were explored. A first dataset is a 25 nm discretised output of Mishchenko's code (Mishchenko et al., 2002) based on the Lorenz-Mie theory (see Supporting Information for more details). A second dataset corresponds to the SOCRATES parameterization of optical properties used in ecRad (Manners et al., 2017). The underlying gamma-shaped droplet size distributions are the same in the two datasets.

The representation of scattering was also explored. Mie computations output an angularly discretised phase function at each wavelength, which is the most realistic available representation of the angular distribution of scattered light upon each scattering event. To estimate the impact of the delta-scaling approximation, a version of the Monte Carlo was implemented that reproduces the δ-Eddington approximation by scaling the scattering coefficients and using Henyey-Greenstein (HG) analytic phase function with appropriately scaled g, as was done in Barker et al. (2015).

In total, four different MC configurations, combining choices of optical properties and phase function, are tested in Section 3.1, among which one is used as a reference in the rest of Section 3 and in Section 4.

2.2 Cloudy atmosphere data

2.2.1 3D fields from LES

For this study, four idealized cumulus cases have been simulated using the French LES model Meso-NH (Lafore et al., 1997; Lac et al., 2018):

- ARM-Cumulus (ARMCu; Brown et al. (2002)), a case of continental cumulus developing over the Southern Great Plains, with a clear signature of the diurnal cycle of the boundary layer in the cloud characteristics. Cloud cover ranges from 0 to 30%;

---

[1] https://gitlab.com/najdavlf/scart_project

- BOMEX (Siebesma et al., 2003), a case of marine shallow cumulus forced with constant surface fluxes through the simulation. Cloud cover ranges from 10 to 20%;
- RICO (vanZanten et al., 2011), a second case of marine cumulus, forced with constant sea surface temperature through the simulation. Cloud cover ranges from 15 to 25%;
- SCMS (R. Neggers et al., 2003), a case of continental cumulus developing in Florida, with strong moisture advection into the domain caused by the nearby ocean. Cloud cover ranges from 0 to 45%.

All simulations were performed on $(6.4 \text{ km})^2 \times 4$ km domains for 12 hours, with isotropic spatial resolution of 25 m and temporal resolution of 1 s. The horizontal boundary conditions are periodic. The four cases are standards of the literature used in LES intercomparison exercises. Detailed descriptions of the setups, initial conditions and forcings can be found in the reference papers. From these four simulations, thirty-five 3D fields of temperature, pressure, mixing ratio of water vapor and liquid water are used in this study, among which eight will be used in the calibration process of Section 4 (the colored entries in Table 2).

Using an object-identification tool [2] (Brient et al., 2019), individual clouds are labelled in each field. A cloud is defined as an ensemble of contiguous cells where the liquid mixing ratio is greater than 10^{-6} kg/kg. Each scene is then described in terms of cloud characteristics, some of which are presented in Table 2. The cloud cover is the proportion of cloudy columns in the domain. At the very first order, it controls the transmitted and reflected solar fluxes. The number density is the total number of identified clouds in the scene divided by the horizontal surface of the domain. For a given cloud cover, a larger number density indicates a longer interface between clouds and clear sky, hence more 3D radiative effects. The maximum depth is the higher minus lower altitudes where clouds are present. When the sun is not at zenith, the "effective" cloud cover (that is, the cloud cover projected in the sun direction) depends on the cloud layer depth.

2.2.2 1D profiles from 3D fields

From each 3D cloud field output from LES runs, 1D profiles are derived to serve as inputs to ecRad. Temperature, pressure, vapor and liquid mixing ratios are horizontally averaged from the 3D fields on each vertical level and extended above the LES domain top using the I3RC (Cahalan et al., 2005) mid-latitude summer (MLS) cumulus profiles provided in the ecRad package. Gases mixing ratios (other than water vapor) are set as in the I3RC MLS cumulus case. Cloud fraction is computed at each level as the fraction of cells where the liquid mixing ratio is positive in the 3D field. Effective radius for liquid droplets is uniformly set to 10 μm as in the Monte Carlo runs. These are the basic profiles needed by any RT scheme. In order to take into account the geometry of clouds, three more parameters are to be provided to ecRad: the overlap parameter α, the FSD of in-cloud liquid water horizontal distribution and the cloud scale C_s.

The overlap parameter can be computed from a 3D cloud field between each pair of layers by inverting Equation (1). Vertical profiles of overlap diagnosed in the 35 LES scenes are illustrated in Figure 1a. Overlap is most often greater than 0.7, with an average value (over the scenes and the vertical levels) of 0.876. It shows relatively small variations on the vertical as well as between the different scenes. Inverting Equation (2) for the average α yields an average decorrelation length z_0 of around 189 meters, close to the values found by R. A. J. Neggers et al. (2011) in LES cumulus fields yet much smaller

---

[2] https://gitlab.com/tropics/object

Table 2. Cloud characteristics from the 35 scenes issued from four standard cumulus cases simulated by LES. Scenes selected for the calibration process are in bold and colors.

Case	Hour	Cover [%]	Number density [km^{-2}]	Max depth [km]
ARMCu	04	2.722	0.73	0.175
ARMCu	05	13.174	1.59	0.300
ARMCu	06	27.139	1.39	0.525
ARMCu	07	29.416	2.00	0.825
ARMCu	**08**	**26.343**	**1.64**	**1.225**
ARMCu	09	26.180	1.44	1.050
ARMCu	10	23.499	1.61	1.375
ARMCu	11	23.029	1.15	1.275
ARMCu	**12**	**12.663**	**0.81**	**1.450**
BOMEX	**05**	**16.301**	**2.17**	**1.200**
BOMEX	04	13.884	2.71	1.025
BOMEX	05	16.301	2.17	1.200
BOMEX	06	18.001	2.71	1.200
BOMEX	07	18.204	2.69	1.125
BOMEX	08	19.081	2.25	1.375
BOMEX	09	14.175	2.39	1.075
BOMEX	10	16.585	2.05	0.975
BOMEX	**11**	**10.318**	**2.00**	**0.775**
BOMEX	12	14.294	2.15	0.650
RICO	04	13.933	2.27	0.950
RICO	**05**	**13.802**	**2.15**	**0.850**
RICO	06	17.195	2.25	1.025
RICO	**07**	**18.054**	**2.34**	**1.175**
RICO	08	19.252	2.69	1.225
RICO	10	23.451	2.20	1.425
RICO	11	21.048	2.25	1.125
RICO	12	16.768	2.32	1.350
SCMS	04	44.035	4.86	1.050
SCMS	**05**	**37.947**	**3.71**	**1.450**
SCMS	**06**	**32.010**	**2.78**	**1.400**
SCMS	07	29.108	2.51	1.450
SCMS	08	20.961	2.05	1.725
SCMS	09	15.678	1.88	1.600
SCMS	10	18.272	1.81	1.200
SCMS	11	11.980	0.93	1.050
SCMS	12	1.502	0.51	0.325

than the range reported by Hogan and Illingworth (2000) because of our smaller vertical resolution.

The FSD, that is, the ratio of in-cloud LWC horizontal standard deviation to total in-cloud LWC is easily diagnosed in each layer of the LES 3D fields since the LWC horizontal distribution is directly accessible. Computed FSD profiles are illustrated in Figure 1b. Again, relatively small variations are observed as both height and scenes change. The FSD ranges from 0.3 to 1 with an average value of 0.7, in agreement with the literature (see e.g. Shonk et al. (2010)).

In 3D cloud fields, the true (resolution-dependent) cloud perimeter could be diagnosed in each layer. However, Schäfer et al. (2016) have shown that accounting for small-scale fluctuations of cloud edges leads to an overestimation of the radiatively effective perimeter and hence of 3D effects. They therefore advocate the use of a cloud perimeter corresponding to the perimeter of an ellipse fitted to the cloud area. This is done by computing, for each labeled cloud in a given layer, the position of its barycenter and the maximum distance between the barycenter and any position in the cloud. This distance is taken as the length of the ellipse's semi-major axis a. The perimeter of the ellipse is then deduced from a and the cloud area. C_s is computed from the sum of the ellipses perimeters as per Equation (3). Vertical profiles of diagnosed C_s are illustrated in Figure 1c. C_s ranges from 50 to 600 meters with some variability both in height and between the different cloud fields, with an average value of 249 m. They are slightly smaller than those found by Hogan et al. (2016) and Fielding et al. (2020) in the I3RC LES cumulus cloud field of Hinkelman et al. (2005). Their simulation is also based on the ARMCu case but their larger resolution of $(67 \text{ m})^2 \times 40$ m explains the differences.

3 Evaluating ecRad against reference Monte Carlo

The objective of this section is to characterize the different sources of error in ecRad. The two-stream, δ-Eddington and approximate optical properties errors are briefly analyzed in Section 3.1. Section 3.2 focuses on the errors due to the approximate representation of cloud geometry and 3D effects.

3.1 ecRad errors in homogeneous plane parallel clouds

Overcast plane parallel homogeneous clouds are very idealized media that are not representative of a true atmospheric situation. However, they are useful when studying pure radiative transfer, without geometrical or 3D effects.

Nine fields consisting of a single-layer infinite homogeneous cloud of uniform geometrical depth are synthetized with cloud optical depths of 0.1, 0.25, 0.5, 1, 2.5, 5, 25, 50 and 100 (computed at wavelength 800 nm for a 10 μm droplet effective radius). The background atmosphere (temperature, pressure and water vapor) depends on the altitude only and corresponds to the I3RC cumulus case, whose 1D profiles are provided in the ecRad package.

MC SOCRATES and MC δ-Eddington are compared to MC "exact" to quantify independently the errors from approximate optical properties and from approximate representation of scattering. The two-stream error is also estimated by comparing ecRad to MC "as ecRad".

Barker et al. (2015) have documented these same errors as a function of optical depth and solar angle. We find similar results (not shown). Here, we only give values averaged over the fields and solar angles, that quantify the mean relative errors on the CREs at TOA and at the surface, and on cloud absorption. First, relative errors are computed only where the reference CRE is more than 2 Wm^{-2} to avoid dividing by very small values such as the CRE on absorption for low sun angles, or the CRE of very thin clouds.
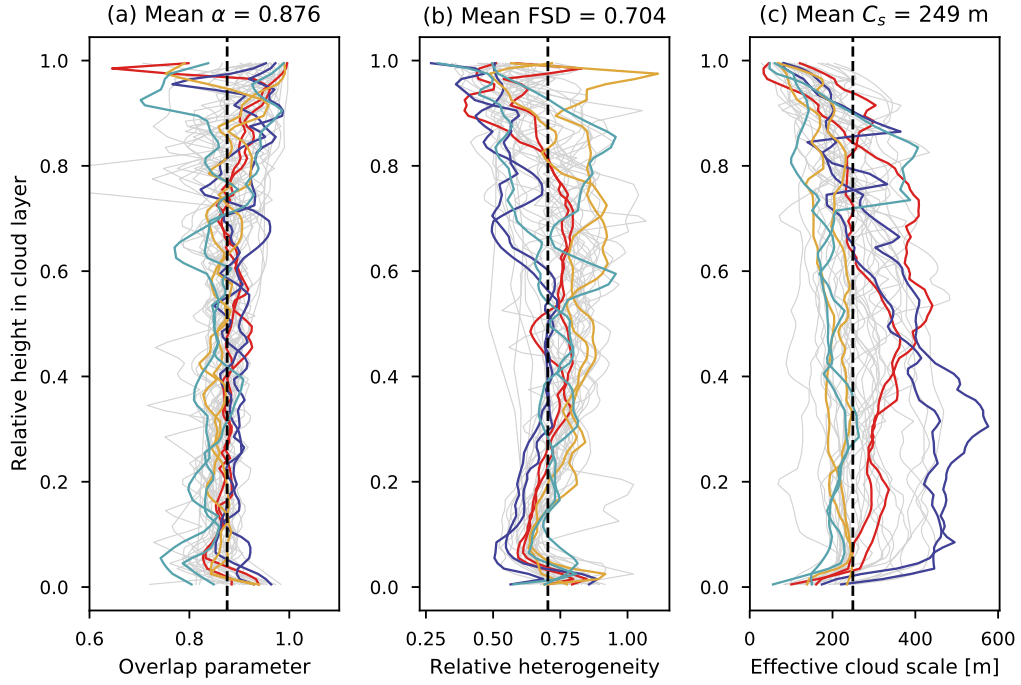
Figure 1. Vertical profiles of the three geometric parameters, scaled on the cloud layer depth (height 0 is bottom of cloud layer, height 1 is top of cloud layer). Gray and colored curves are for individual cloud scenes (colored curves are the fields used for calibration) and dashed black line is the average value over all cloud scenes and heights.

Once RMSEs and mean biases are computed for each solar angle, they are weighted by the cosine of the solar angle so that small relative errors on metrics involving large amounts of energy count more than large relative errors on metrics involving small amounts of energy. They are given in Table 3. Unweighted RMSEs and biases are given for two solar angles (0 and 77) in Supporting Information.

Small relative errors on TOA and surface CREs lead to large relative errors on the absorption, but errors on absorption CREs only represent small amounts of energy. The absorption errors due to the different components are of similar magnitude. The largest source of error on surface and TOA CREs is the approximate representation of scattering, and more generally the representation of transport, while approximate optical properties have small effect. Errors due to transport and optical properties are quasi additive, while sums of errors due to approximate scattering and two-stream formulation are larger than the global transport error because of non-linear effects.

These numbers are given in order to provide some perspective to the rest of the study, where similar errors are computed in cumulus cloud fields for various configurations of ecRad (also given in Table 3, although description of the experiments are later in the text).

3.2 Errors due to cloud geometry and 3D effects

To account for fractional clouds, the Tripleclouds method separates each model layer into cloudy and clear regions, and solves fluxes transmission and reflection independently in each region. This is similar to the Independent Column Approximation (ICA) since fluxes do not travel from one region to its horizontal neighbour. The transmitted (reflected) fluxes from one region are distributed into the regions of the layer below (above) according to the vertical overlap of cloudy regions. In what follows, maximum overlap is used as a "worst case scenario" to estimate the error on fluxes before introducing realistic overlap. In this worst case scenario, each cloudy region is homogeneous and the ICA is maintained.

In Figure 2, differences between ecRad and MC SOCRATES (the choice of MC SOCRATES as a reference to measure ecRad errors is discussed in Section 4.1.2) for different metrics are illustrated for three configurations of ecRad: with maximum overlap, horizontal homogeneity and without 3D effects (PPH max ovp); with realistic overlap and heterogeneity and without 3D effects (Tripleclouds); with realistic overlap and heterogeneity and with 3D effects (SPARTACUS). Profiles of FSD, α and C_s diagnosed in the LES fields are used to constrain the cloud geometry parameters in Tripleclouds and SPARTACUS. Histograms of absolute differences between ecRad and MC estimates of upward TOA, absorbed and downward surface fluxes are plotted for each configuration. The RMS absolute errors are given in the legend of Figure 2 for each metric and configuration, while the weighted averages of biases and RMS relative errors are given in Table 3.

The configuration that corresponds to the basic two-stream scheme with no parameterization of cloud geometry or 3D effects leads to the largest absolute means and largest RMSEs for all the metrics: clouds transmit too much energy to the surface, resulting in a lack of reflectivity and absorption. Introducing realistic overlap and horizontal heterogeneity systematically reduces the mean bias and changes the sign of the tail of the distribution. Realistic overlap increases the total cloud cover, which decreases transmissivity. On the contrary, introducing heterogeneities increases it. The fact that the mean biases in transmissivity and reflectivity changes sign means that the overlap effect dominates the heterogeneity effect: the cloud cover controls radiative transfer at the first order. Introducing 3D effects globally decreases both the distributions width and RMSEs. The strongest biases associated with Tripleclouds estimates of the transmissivity and reflectivity are removed when switching to SPARTACUS, suggesting that these extreme errors are due to the neglect of 3D effects. The biases on absorption are slightly shifted to more positive values: 3D effects increase absorption.

Table 3.   Relative errors [%] for different experiments presented throughout the paper.

(i) Experiment	(ii) Reference	(iii) Model	(iv) TOA up RMS	bias	(v) Absorbed RMS	bias	(vi) Surf. down RMS	bias
(a) In slabs (Section 3.1)								
(1) SOCRATES	MC exact	MC SOCRATES	1.1	0.8	10.8	-10.5	1.9	-1.5
(2) δ-Eddington	MC exact	MC δ-Eddington	11.1	4.6	13.1	4.5	11.8	4.6
(3) two-stream	MC as ecRad	ecRad two-stream	5.4	0.3	15.0	-9.1	4.3	-2.2
Transport (2+3)	MC SOCRATES	ecRad two-stream	8.3	3.2	15.2	-4.4	7.0	1.0
Total (1+2+3)	MC exact	ecRad two-stream	8.6	4.1	17.8	-14.5	5.8	-0.7
(b) In cumulus, MC vs ecRad 1D and 3D solvers, parameters $\lambda = (\alpha, \text{FSD}, C_s)$ (Section 3.2)								
PPH max ovp		1D, $\lambda = (1, 0, \infty)$	23.4	-20.9	54.2	-53.5	28.6	-27.0
Tripleclouds (1D)	MC SOCRATES	1D, $\lambda(z, case)$ LES	29.3	23.0	23.8	-18.9	23.7	15.1
SPARTACUS (3D)		3D, $\lambda(z, case)$ LES	22.7	20.0	20.0	-10.4	18.3	14.4
(c) In cumulus, ecRad SPARTACUS, with LES profiles vs averaged parameters (Section 3.3)								
z-averaged	$\lambda(z, case)$ LES	$\overline{\lambda}(case)$ LES	1.4	-0.1	1.6	-0.4	1.4	-0.1
case-z-averaged	$\lambda(z, case)$ LES	$\overline{\overline{\lambda}}$ LES	3.7	0.6	3.5	-0.4	3.6	0.4
(d) In cumulus, MC vs ecRad SPARTACUS with calibrated parameters (Section 4)								
Best global		$\overline{\overline{\lambda}}$ from htexplo	8.3	-2.7	29.1	-28.1	10.2	-7.2
Best TOA up	MC SOCRATES	(see Table 4)	11.3	-8.5	33.3	-32.6	14.4	-12.8
Best absorption			17.9	12.0	22.1	-18.8	14.9	6.3
Best surface down			9.2	-0.4	28.0	-26.8	9.6	-5.1

For each pair of reference computation (ii) / test approximation (iii), errors on the cloud radiative effects on TOA upward (iv), absorbed (v), and surface downward (vi) fluxes are quantified. For each column, the RMS and mean bias are first computed independently for each solar angle over the different cases, then RMS and mean bias are weighted by the cosine of the solar angle, and averaged over the 8 SZAs. Only data points where reference CRE > 2 Wm^{-2} are used to avoid division by zero. Only solar angles where at least 9 data points were available are used in the cosine-weighted average. The table subsections concern: (a) errors related to non-geometrical effects of clouds, (b) ecRad errors for different solvers, with increasing complexity in the representation of geometrical effects, (c) errors related to the neglect of parameters variations with height and cloud field, (d) ecRad errors for different choices of cloud-geometry parameters, output from the calibration exercise of Section 4.

CRE = total sky - clear sky. Relative error $r = 100 \times (\text{model-ref})/\text{ref}$. RMS $= \sqrt{\langle r^2 \rangle_{fields}}$. bias$= \langle r \rangle_{fields}$
MC exact: detailed Mie optical properties and phase function.
MC SOCRATES: parameterized optical properties and detailed Mie phase function.
MC δ-Eddington: detailed Mie optical properties and HG δ-Eddington phase function.
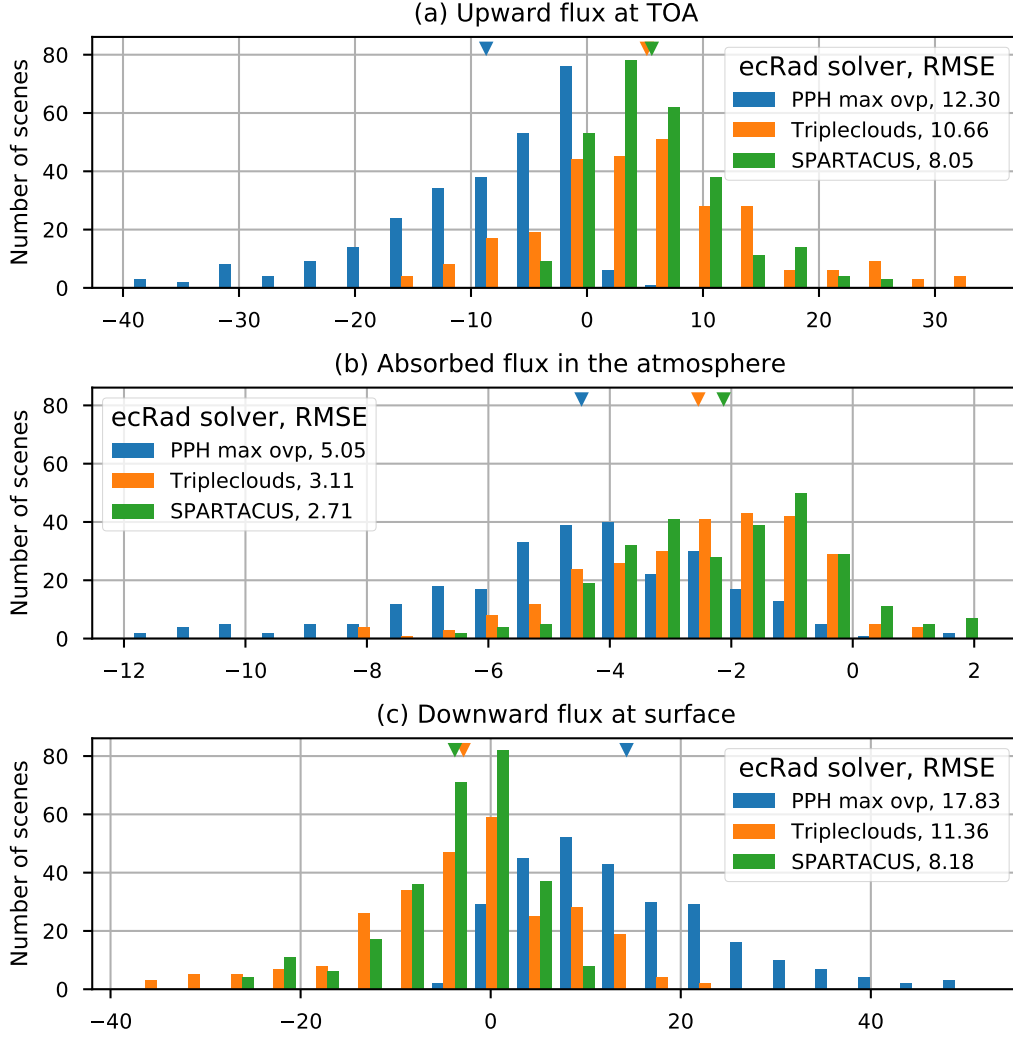MC as ecRad: parameterized optical properties and HG δ-Eddington phase function.

Figure 2. Histograms representing the distributions of differences between ecRad and MC estimates for the three metrics: (a) upward flux at TOA, (b) absorbed flux in the atmosphere and (c) downward flux at the ground. Each histogram represents the distribution of 280 data points: 35 scenes × 8 solar zenith angles (from 0 to 77 with step 11 degrees). Each color corresponds to a different configuration of ecRad. PPH max ovp corresponds to homogeneous clouds with maximum overlap and no 3D effects. Tripleclouds corresponds to heterogeneous clouds with FSD and α as diagnosed in the 3D LES field, without 3D effects. SPARTACUS is as Tripleclouds but with 3D effects, with C_s as diagnosed in the 3D LES fields. The mean error is represented by colored triangles. The RMSEs are given in the legends.

3.3 Errors due to parameter variations with height and cloud scene

By introducing cloud geometry effects and horizontal transport, the parameterization of radiative transfer is globally improved. However, the distance to reference for this new parameterization depends on the choice of values for the newly introduced parameters. In Section 3.2, the parameters were set to the profiles diagnosed in each 3D cloud field.

To evaluate the impact of prescribing a vertically uniform value for each parameter instead of a varying one, ecRad simulations are performed using for each cloud scene the vertically averaged value of the diagnosed parameters. It results in a small change in CRE when compared to the ecRad simulations using the full profiles (see Table 3). This result shows that for a cumulus cloud scene, vertical variations of overlap, relative heterogeneity or cloud size within the cloud layer are not of crucial importance and can be neglected.

In a third step, ecRad is run using the mean value of each parameter, obtained by averaging over height and the 35 cloud scenes, to estimate the impact of inter-scene parameter variability. It results that the impact of inter-scene variability is larger than the effect of inter-layer variability (around 3%, see Table 3), but still inferior to the impact of representing horizontal transport (around 7%).

3.4 Remaining errors

Some discrepancies remain between SPARTACUS and MC SOCRATES. They are in part due to the δ-Eddington and two-stream approximations. Other sources of errors reside in the modeling choices for the treatment of geometry. For example, the overlap model only constrains how pairs of layers interact, but does not provide a vertically-integrated constraint. This leads to a systematic overestimation of the total cloud cover, a quantity that is key to the first order estimation of total transmission and reflexion. The treatment of in-cloud heterogeneity is also idealized in Tripleclouds: it is assumed that two cloudy regions are enough to represent the radiative effect of horizontal heterogeneity, and the distribution of the LWC into these two regions is based on hypothesis such as the LWC distribution shape and the percentile that should be used to partition the condensate into the thin and thick regions. Other modeling assumptions and parameters that have not been discussed here also contribute to the remaining errors, such as free parameters specific to the representation of entrapment, or the degree of vertical overlap of heterogeneities.

Another candidate to explain these remaining errors is of course the choice of the three geometrical parameter values. It was shown that their inter-scene and inter-height variabilities have moderate impacts on the fluxes estimate. Yet, it was all along taken for granted that the best possible choice for these parameter values was to set them to "observed" values, that is, the values deduced from the 3D cloud fields according to their physical interpretation. However, these parameters are used in ecRad as effective properties, and since radiative transfer is highly nonlinear with respect to cloud geometry, there is no fundamental reason for which the mean parameters that are diagnosed in the LES should be the best choice to effectively describe cloud geometry to the radiative transfer parameterization. In the next section, we question this choice by exploring the behaviour of SPARTACUS for different sets of parameter values, randomly sampled in the three-dimensional parameter space.

4 Exploring ecRad behaviour in the cloud geometry parameter space

4.1 The High-Tune:explorer tool

High-Tune:explorer (htexplo) is a statistical tool that helps to efficiently explore the behaviour of a model throughout an arbitrarily large parameter space. It is based on history matching and implements iterative refocusing to reduce the initially submitted parameter space to the set of parameter values (the model configurations) that are "acceptable" in view of a set of predetermined reference targets. The method is thoroughly described and illustrated in a recently published two-part paper (Couvreux et al., 2020; Hourdin et al., 2020).

4.1.1 Overview

The different steps to explore the behaviour of a parameterization are briefly summarized here. Details on the specific ingredients that were used to explore ecRad are provided in the following subsections. For one parameterization or model, one should:

1. Select the target metrics that will serve to evaluate the model, and determine the uncertainty associated with the references.
2. Select the n parameters to calibrate. A default value and a range to explore must be provided for each parameter. The n-dimensional hypercube formed by the cartesian product of parameter ranges is the original parameter space $\mathcal{P}$.
3. Build the "experimental design" by sampling a small number of points (around ten times n) in the parameter space. In htexplo, a maximin Latin Hypercube sampling method is used to ensure the space is efficiently explored, by maximizing the minimum distance between samples (Williamson, 2015). The model is run for the sampled configurations.
4. Compute metrics from the model outputs and use them as a learning basis to build one emulator per metric. In htexplo, each emulator is based on a Gaussian Process. It is a fast surrogate model that provides an estimate (the expectation of the process) for the metric value at any point of the parameter space, along with its statistical uncertainty (the standard deviation of the process).
5. Compute a distance to the reference target for each metric f (f_k is the kth metric) and parameter vector $\check{\lambda}$, using the emulators. In htexplo, this distance, called the *implausibility* $I_f(\check{\lambda})$, is the absolute difference between the emulator estimate $\mathbf{E}[f(\check{\lambda})]$ and the target r_f, divided by the root square of the quadratic sum of three uncertainties:

$$I_f(\check{\lambda}) = \frac{|r_f - \mathbf{E}[f(\check{\lambda})]|}{\sqrt{\sigma_{r,f}^2 + \sigma_{d,f}^2 + \sigma_f(\check{\lambda})^2}} \quad (4)$$

   where $\sigma_{r,f}$ is the uncertainty associated with the reference (or observational error) whose value was set at the first step, $\sigma_{d,f}$ is the model intrinsic error, whose value is unknown (see section 4.1.4), and $\sigma_f(\check{\lambda})$ is the statistical uncertainty associated with the emulator estimate, directly provided by the emulator itself. The points where the implausibility is larger than a threshold for at least one of the N_{met} metrics are removed from the parameter space. This means that points are kept in the parameter space if all the metrics are close enough to their target, or if the local uncertainties are too large to ensure that the configuration is unacceptable. The new parameter space is called the Not-Ruled-Out-Yet (NROY) space.
6. The NROY space is sampled to build a new experimental design and steps 3 to 5 are repeated until the NROY space converges. With each iteration, called "wave", the uncertainties associated with the emulators decrease until convergence, since the sampling of model configurations that serve to build the emulators is denser (the parameter space is smaller and the number of sampled points is unchanged).
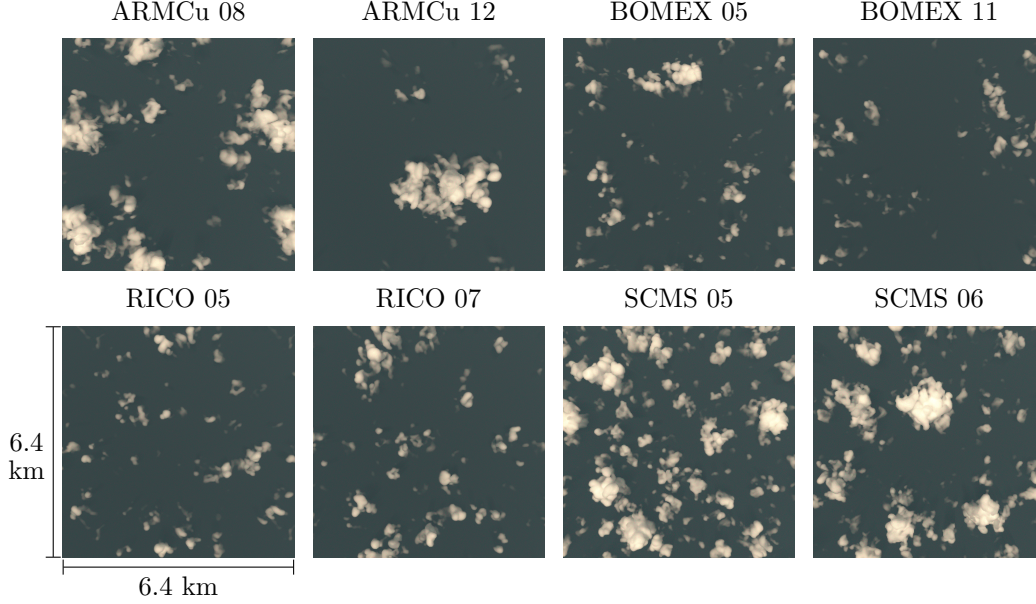
Figure 3. Monte Carlo rendering of the eight selected scenes using htrdr, with nadir view, sun at zenith and a black surface.

4.1.2 Metrics and references to calibrate ecRad

Three metric types were used in the calibration of ecRad, all based on solar fluxes horizontally averaged over the LES domain: the reflected flux at the TOA $F_t^\uparrow$; the total absorbed flux in the atmosphere F_{abs} and the atmospheric radiative effect measured at the surface which is the difference between downward flux at TOA and downward flux at the surface, $F_t^\downarrow - F_s^\downarrow$. These three metrics are not independent since the incoming flux at the TOA is entirely distributed into reflected, absorbed by the atmosphere and absorbed by the ground fluxes:

$$F_t^\downarrow = F_t^\uparrow + F_{abs} + (1-a)F_s^\downarrow \tag{5}$$

where a is the surface albedo. Considering the three metrics instead of two still adds a constraint through their respective uncertainties: each metric's uncertainty is smaller than the sum of the uncertainties associated with the other two metrics. For each of these fluxes, three solar angles are used, at 0, 44 and 77 degrees from zenith.

Each of these nine metrics (three fluxes × three solar angles) are computed in different cloud fields. Eight scenes were selected among the 35 available cumulus fields described in Table 2. These eight scenes were chosen for their contrasted characteristics to explore the distribution of available cumulus fields. They are represented in Figure 3 where the *htrdr* MC path-tracing tool (?, ?) was applied to render nadir views of the different scenes, with the sun at zenith. Image rendering is useful to better visualize the variety of combinations of cloud covers, number of clouds and cloud optical depths of the scenes that will enter the calibration process. The images also show strong similarities, highlighting the relative narrowness of the cloud fields distribution inside the cumulus regime.

The reference values used as targets for these 72 metrics are the MC estimates of the fluxes. The observational error is taken as the standard deviation of the MC estimate, typically smaller than 0.1%. In Section 2.1.2, various options for the MC computations were exposed. Here, we use the SOCRATES data for the optical properties of

cloud droplets, and the detailed Mie phase function to sample the next direction after scattering. By using the same low-spectral-resolution optical properties as in ecRad, the radiative effects of cloud geometry are prevented from compensating for errors that could come from mismatched optical properties. On the other hand, the representation of the phase function is deeply entangled with the two-stream formulation of transport. It might not be the role of geometrical effects to compensate for errors that come from the δ-Eddington approximation, but since these effects were modeled and formulated in direct interaction with the two-stream scheme, they are seen as intrinsic to the parameterization of transport. By targeting fluxes obtained from simulations using detailed Mie phase function, we allow internal compensation errors between cloud geometry and transport as long as at least one geometrical configuration exists that lets the parameterization handle contrasted situations reasonably well.

4.1.3   Parameters to calibrate

The parameters that enter the calibration process are the three geometrical parameters described in Section 2.1.1: the vertical decorrelation length z_0 from the formulation of the overlap parameter; the fractional standard deviation of the horizontal distribution of in-cloud liquid water FSD; and the cloud scale C_s. Since we have shown before that the ecRad estimates were moderately sensitive to the vertical details of the parameter profiles, and to the inter-scene variability, we configure ecRad with a unique value per parameter; the same parameter value is used for all the cloudy layers and all the scenes. Other parameters of ecRad could have entered the process but our work focuses on the modeling of cloud geometry, and these three parameters are of first importance in the representation of geometrical effects. The mean values diagnosed in the LES (given in Figure 1) are taken as default values:

- z_0 ranges in [50, 500] with default value 189 m
- FSD ranges in [0.1, 2] with default value 0.704
- Cs ranges in [50, 1000] with default value 249 m

4.1.4   Tolerance to error

The structural error of the parameterization in Equation (4) stems from inevitable modeling and numerical approximations. It is most often unknown. In a sense, it is the error that would remain after the parameters are well calibrated. However, its characterisation is a prerequisite to the calibration process, as it prevents the tool from rejecting configurations that predict metric values within the structural error around the reference target. Since the structural error is unknown, we rather use a "tolerance to error": an acceptable distance between the parameterization estimate and the reference target, arbitrarily set by the modeler. Here, it is inferred from the relative errors between MC and SPARTACUS runs using the mean LES parameter values, for each type of metric and solar angle. The tolerances to error are set as the third quartile of these distributions:

- for the atmospheric radiative effect at the surface ($F_t^\downarrow$ - $F_s^\downarrow$), the relative tolerances to error are 3% for SZAs 0 and 44, and 4% for SZA 77
- for the absorbed flux in the atmosphere, the relative tolerances to error are 1%, 2% and 4% respectively for SZAs 0, 44 and 77
- for the reflected fluxes at SZAs 0, 44 and 77, the relative tolerances to error are set to 6%, 3% and 4% respectively.
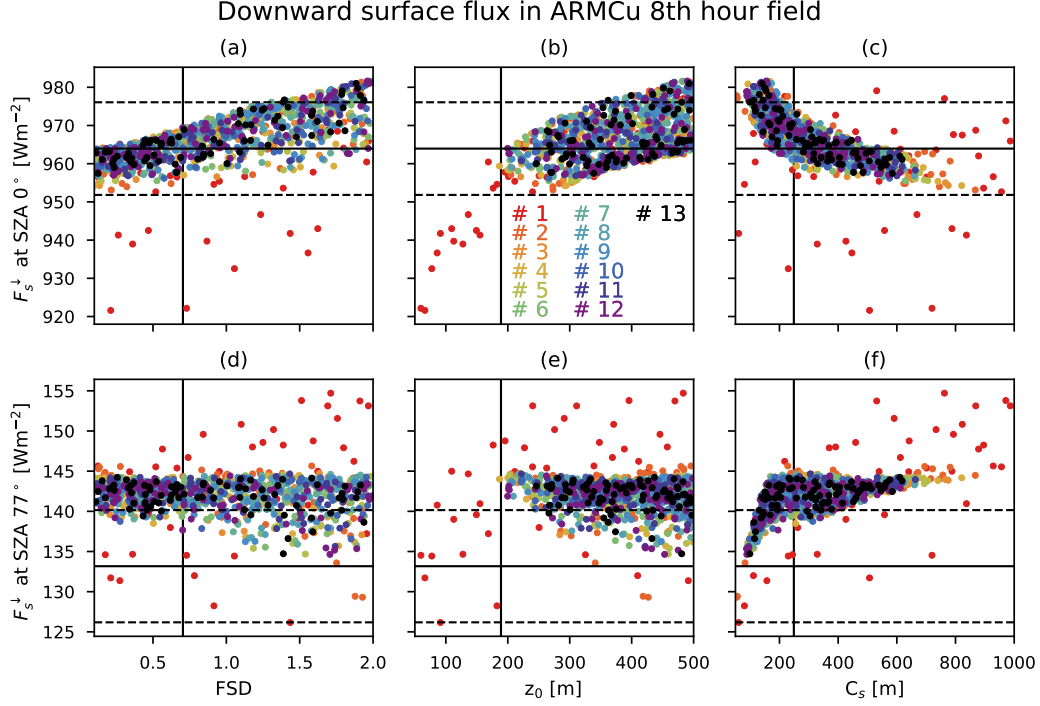
Figure 4. Downward flux at surface for various ecRad runs, as a function of three parameter values: (a,d) fractional standard deviation FSD, (b,e) overlap decorrelation length z_0, (c,f) cloud scale C_s, and of solar zenith angle: (a-c) $0°$ and (d-f) $77°$. Full black horizontal lines represent the Monte Carlo reference value, dashed horizontal lines represent the tolerances to error. Full vertical lines represent the mean parameter value diagnosed in the LES. Different colors represent parameter sets sampled at different waves.

4.2 Exploring ecRad configurations

In this section, htexplo is used to explore the cloud-geometry parameter space and analyze the behaviour of ecRad. Thirteen iterations were applied, with reduction of the NROY space from 11.7% of the original space after the first wave, to 8.4% after the thirteenth wave.

4.2.1 Surface downward flux as a function of parameters

Figure 4 illustrates the dependence of two of the 72 metrics to the geometry parameters. ecRad estimations of the downward flux at the surface under the ARMCu 8th hour clouds at SZA $0°$ and $77°$ are represented for the many ecRad configurations explored during the thirteen waves of history matching.

Figure 4a shows that large surface fluxes at high sun are only obtained when clouds are sufficiently heterogeneous (when FSD is large enough), while the effect of heterogeneity in grazing sun conditions is less obvious (Figure 4d).

Figure 4b shows that the transmitted flux at $0°$ is strongly related to the decorrelation length, but the transmitted flux at $77°$ does not seem driven by this parameter (Figure 4e). Indeed, when the decorrelation length increases, the overlap gets closer to maximum (and further away from random) and the total cloud cover decreases. This

629 leads to more energy reaching the surface, in particular for high sun. As the sun gets closer
630 to the horizon, it is not the total cloud cover that matters but the effective cloud cover,
631 projected in direction of the sun, to which cloud sides contribute largely.

632 In SPARTACUS, 3D effects are inversly proportional to cloud size C_s. At high sun,
633 Figure 4c shows that 3D effects lead to an increase in surface flux, a signature of escape
634 of light from cloud sides and entrapment. At low sun, Figure 4f shows that they lead to
635 a decrease in surface flux, explained by the interception of light by cloud sides. In multi-
636 layered cloud scenes, the entrapment effect would be stronger and the balance between
637 positive and negative 3D effects as a function of SZA could be affected (entrapment leads
638 to an increase of surface flux at all solar angles; Hogan et al. (2019)).

639 *4.2.2 Reduction of NROY space and parameters interdependency*

640 Metrics computed at different iterations in the calibration process are represented
641 in different colors in Figure 4, which evidences that part of the parameter ranges are no
642 longer sampled after a given wave. For instance, after the first wave (red points), decor-
643 relation length values smaller than ∼180 m have been excluded from the parameter space,
644 independently of the values of the other two parameters. This is because for this sub-
645 range of decorrelation length values (in which the cloud cover is large) the $0°$ surface flux
646 emulator predicts values that are too small compared to the MC estimate.

The implausibility matrix presented in Figure 5 reveals the structure of the NROY
space obtained after the thirteenth wave. It is constructed at each wave as follows: with
each point of the original space parameter of dimension n (here, $n = 3$), is associated
the largest metric implausibility computed at the current wave

$$\forall \check{\lambda} \in \mathcal{P}, \ I(\check{\lambda}) = \max_{1 \leq k \leq N_{met}} \left\{ I_{f_k}(\check{\lambda}) \right\}$$

647 The three subplots that form the upper triangle of the matrix show the fraction
648 of points that are still in the NROY space, i.e. that verify $I(\check{\lambda}) \leq 3$. Each projection
649 shows the density of points in the NROY space in the dimension that remains once two
650 parameters are fixed (for each subplot, the horizontal and vertical axes are given by the
651 parameters on the diagonal).

652 The three lower subplots show the minimum implausibility of the points in the di-
653 mension that remains once two parameter values are fixed, as in the three subplots of
654 the upper triangle. These subplots have the same axes as their symmetric in the matrix
655 (hence, axes are not directly given by the diagonals).

656 The upper triangle gives the quantity of acceptable configurations, while the lower
657 triangle informs on the quality of "best" configurations.

658 The gray (red) zones in the upper (lower) triangle subplots represent the regions
659 of the parameter space where no configuration is acceptable given the two parameter val-
660 ues that correspond to the pixel, whatever the value of the third parameter. For instance,
661 the upper-left and lower-right subplots show that small values of the decorrelation length
662 have been rejected, independently of the values of the other two parameters. This was
663 already illustrated in Figure 4. Here, the plots additionally show that the set of param-
664 eters diagnosed from the 3D cloud fields do not belong to the NROY space of the thir-
665 teenth wave, in particular due to too small value of the FSD and/or of z_0.

666 On the upper-right subplot, we see that many (FSD, C_s) pairs have been rejected.
667 The pairs that lead to acceptable configurations of the parameterization are cleary iden-
668 tified: small values of C_s are paired with large values of FSD and conversely (although
669 very large values of C_s were all rejected). This means that an increase in heterogeneity
670 can be compensated by a decrease in cloud size (more intense 3D effects), and that the
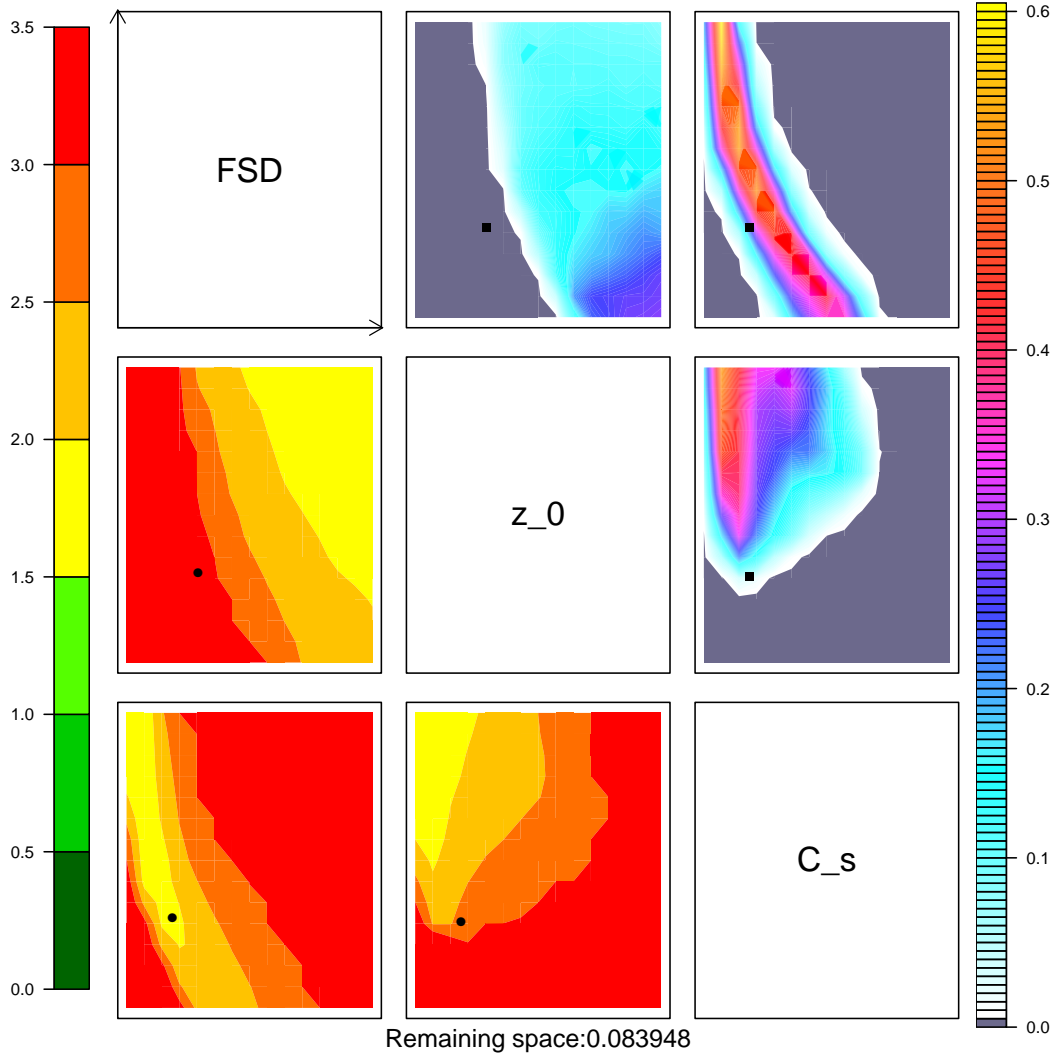
Figure 5. Visualisation of the NROY space density (upper triangle) and of the implausibility hypercube (lower triangle) at wave 13. The implausibility is computed as the maximum over the metrics, and the minimum over the third dimension. Axes of the upper-triangle subplots correspond to the parameters on the diagonal (each x-axis varies as the parameter of the same column and y-axis varies as the parameter of the same line) while the axes of the lower-triangle subplots are the same as the axes of their symmetric subplot in the upper triangle.

Table 4.   Parameter values for the best configurations of ecRad

Parameters	FSD	z_0 [m]	C_s [m]
From LES	0.705	187	247
Best Global	1.079	436	155
Best TOA	1.646	493	119
Best absorbed	0.102	294	821
Best surface	1.469	374	113

uncertainties associated with the target metrics do not allow to determine which mode should be favored between small heterogenous or large homogeneous clouds.

The quantitative information displayed in the lower triangle is strongly dependent on the chosen tolerance to error (since it is in the expression of implausibility, see Equation 4). The variations of implausibility in the parameter space reveal more of the parameterization behaviour than the implausibility absolute values. However, the subplots of the lower triangle show that the implausibility in the parameter space is never less than 1.5. It means that for any configuration, there is always at least one metric that is farther away from its target than 1.5 times the root square sum of its uncertainties, which is dominated by the tolerance to error at wave thirteen. It also shows that the "best" configuration has large decorrelation lengths associated with small heterogeneous clouds, rather than large homogeneous ones.

4.2.3   *Global improvement of fluxes using tuned configuration*

The various configurations that were sampled to construct emulators from true ecRad runs are evaluated using scores associated with each metric and configuration. It is the error between ecRad and the reference MC divided by the tolerance to error. For each simulation of waves three to thirteen, the RMS scores are computed over all metrics, and over reflected fluxes, absorbed fluxes and surface fluxes separately. Then, the configurations with smallest RMS scores of each category are selected as "best" configurations. They are presented in Table 4. Configurations that lead to best upward TOA and best downward surface fluxes are relatively similar, favoring small heterogeneous clouds. The configuration that leads to the better estimates of absorbed fluxes favors the other direction (large homogeneous clouds). The configuration that leads to best global RMS is in between these two modes, but still selects smaller more heterogeneous clouds than in the LES. The overlap decorrelation length parameter is always greater than the one diagnosed in the 3D cloud fields, yielding smaller cloud covers.

These four new configurations, obtained from a calibration process using only eight cloud fields and three solar angles, were tested on the 35 cloud fields and 11 solar zenith angles of Section 3. The distributions of errors are represented in Figure 6. As in Figure 2, the RMSEs are given in the legends for each configuration. These numbers are of different nature from the configuration scores as they are not divided by the tolerance to error. The mean relative errors on surface, TOA and absorbed CREs are given for each configuration in Table 3.

The fluxes at TOA and surface are systematically improved compared to the configuration using the parameter values diagnosed in the LES, but all tuned configurations are slightly worse for the absorption. The configuration corresponding to the best surface score leads to the smallest global RMSE; it is the configuration with smallest and most heterogeneous clouds. With this configuration, as well as for the two configurations corresponding to best TOA score and best global score, the absorption bias is always negative: the absorption is underestimated by ecRad for all scenes and all SZAs when clouds
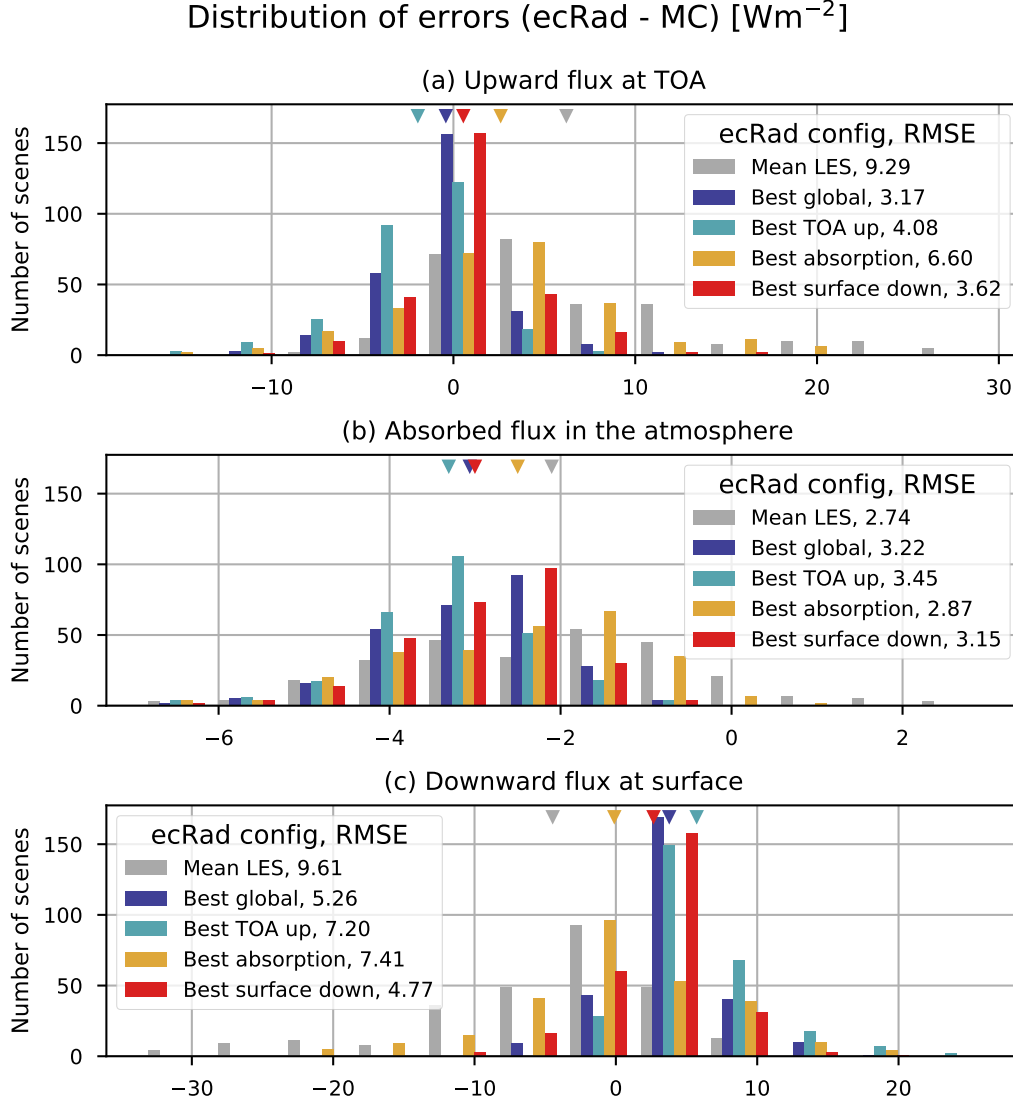
Figure 6. Histograms representing the distributions of differences between ecRad and Monte Carlo estimates for the three metrics: (a) upward flux at TOA, (b) absorbed flux in the atmosphere and (c) downward flux at the ground. Errors for all 35 cumulus scenes and 8 solar angles (from 0 to 77 with step 11 degrees) are distributed together. Each color corresponds to a different configuration of ecRad. The parameters values for each configuration are given in Table 4. Color triangles represent the mean error. The root mean square distances (RMSE) are given in the legends.

are small and heterogeneous. It appears that most of the flux that should have been absorbed reaches the surface, inducing a positive mean bias in the transmitted fluxes.

4.3 Calibration fails to reduce errors on the absorption

To understand what in the "best Surf" configuration leads to wrong estimates of the absorption, the CRE on absorption computed in the SCMS 5th hour cloud field is represented in Figure 7 as a function of SZA and for various sensitivity tests. The "best Surf" is used as the reference ecRad configuration, and one parameter is modified in each of three sensitivity tests: 3D effects are removed by using Tripleclouds instead of SPARTACUS; FSD is set to 0; z_0 is set to 294 m, which leads to a larger cloud cover diagnosed by ecRad (34% instead of 30%; the true cloud cover in the 3D cloud field is 38%).

The reference configuration with intense 3D effects and important heterogeneity accurately reproduces the absorption dependency to solar angle but with a negative shift of 2 to 4 Wm^{-2}, probably due to two-stream errors as hinted by the results of Section 3.1. Homogeneity and 1D radiation both induce errors that are larger than the "best Surf" configuration, but they are of opposit signs. Because increasing 3D effects decreases absorption at low suns, the configuration with small clouds lead to even larger errors for the 77° metric. On the other hand, using error compensations from the two other parameters leads to improvement for the three solar angles, even if the shape of the function is wrong.

This highlights the importance of choosing relevant metrics for the calibration process. In particular, here, additional metrics might need to be introduced to constrain the shape of the fluxes dependency to solar angle, instead of using punctual values only. This is actually an option in htexplo, where functions are decomposed onto a basis of empirical orthogonal functions (EOFs) and the coefficients of the linear combination are tuned (Salter et al., 2019). Further work on exploring ecRad should include tuning functions of solar angles and determine if the absorption metric then agrees that small heterogeneous clouds are more appropriate than large homogeneous ones to *effectively* describe cumulus geometry to the radiation scheme.

5 Discussions and outlooks

A fundamental aspect of the tuning strategy advocated in this series of papers is that a first calibration step should be done at the process scale, using either offline parameterizations on well-mastered cases or the LES/SCM framework. With this approach, the calibration of the 3D climate model that involves all parameterizations is restricted to regions of the parameter space that were judged acceptable on the basis of process-based metrics. This strategy accelerates 3D calibration and prevents errors associated with different parameterizations from compensating each other, for example compensating wrong boundary layer cloud radiative effects by tuning sensitive high clouds parameters. In this third paper, the gap between process-scale tuning of cloud properties and global tuning of radiation was further reduced by providing reference radiative metrics computed at the cloud-field scale. This prevents other possible sources of compensation errors such as compensating the lack of 3D effects by an increase in cloud cover.

A key result of this study is that the parameters diagnosed in the LES fields do not belong to the final NROY space. This result strikes us as important because it questions the conceptual constraints that surround parameterization development and tuning. The main goal of parameterization development should be to derive functional forms that are able to provide accurate source terms for the explicitly resolved variables of the model over a wide range of atmospheric regimes. To achieve this, it is essential to base our developments on our understanding of physical processes. However, we argue that some flexibility should be allowed in the choice of parameter values. Results reported by Bastidas
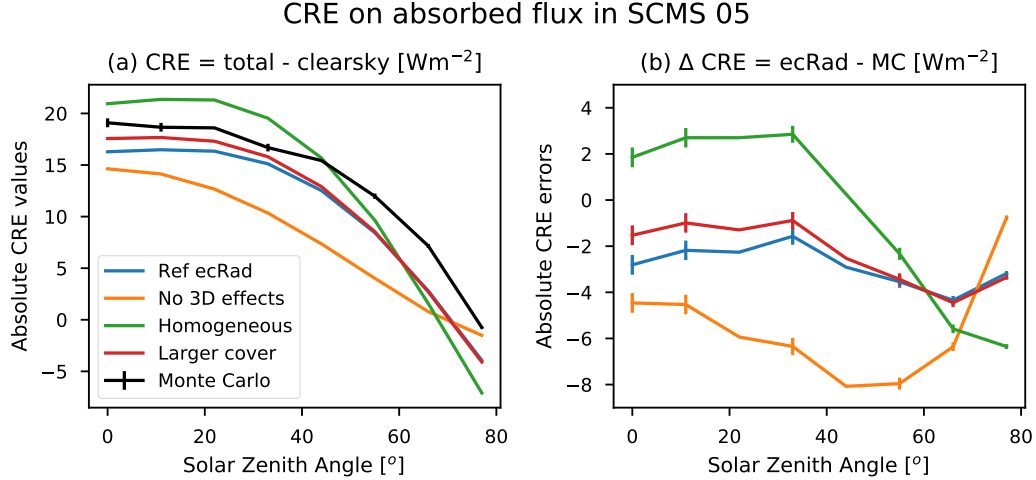
Figure 7. Absorption CRE in the fifth hour of the SCMS case. Subplot (a) shows the absolute values of CRE for various ecRad configurations and the Monte Carlo reference, as a function of solar zenith angle. Subplot (b) shows differences between the ecRad and the MC estimates for each ecRad configuration. The "Ref ecRad" is the configuration that corresponds to the "best Surf" of the previous section, that is: FSD=1.469, z_0=374 m and C_s=113 m

et al. (2006) and Hogue et al. (2006) also support this idea. They show that free parameters should be set to different values through different land surface models even though they are supposed to have the same physical meaning. Their conclusions were limited to so-called "functional" parameters that cannot be associated with physical measurements. We argue that observational constraints on "physical" parameters should also be alleviated. Indeed, it is most often an "effective" value of the parameters rather than a mean observed value that is needed in the models. These can be quite different when parameters impact the metrics in non linear ways. In the context of cloud–radiation interactions, it means that the effective cloud characteristics that are appropriate to derive the average radiative effect of a complex cloud field have no reason to be the detailed characteristics averaged over the cloud population. More generally, since parameterizations are a simplification of reality, there is no fundamental reason to prefer observed parameter values. They should serve as first guesses, but the final retained values should be the result of some model calibration. Nonetheless, a tuned value laying too far from observations could indicate that the physical images that supported the parameterization development are wrong or that important processes are missing.

Another result is that improvement of ecRad was obtained by calibrating a mean parameter, thereby neglecting parameter variations with height and between cloud scenes. This was probably only possible because all cloud fields used here represent cumulus clouds, with relative resemblance between the cases, although both marine and continental clouds were represented. An interesting follow-up would be to repeat this exercise with other cloud types, starting with other boundary-layer clouds such as stratocumulus and transition scenes involving both cloud types. A possible diagnosis of htexplo might then be that a single parameter is not able to represent different clouds. This would mean that a sub-parameterization should be developed to make this parameter depend on atmospheric conditions. Such parameterizations exist for example to predict cloud perimeter length in Fielding et al. (2020), or the degree of overlap in e.g. Sulak et al. (2020). Other parameters appear in these formulations, which can in turn be calibrated using the same procedure as described in this work.

A third result is that our different metrics disagree on the configuration to use. If this disagreement had resulted in an empty NROY space for the predetermined tolerance, it would mean that the parameterization is deficient in parts of its expected operating regime. Here, the final NROY space is not empty. However, the fact that calibration of cloud-geometry parameters fails to significantly reduce the absorption errors suggests that either other parameters control the absorption and might deserve better calibration, or that the formulation of radiative transfer intrisic to the parameterization should be revisited.

References

Barker, H. W., Cole, J. N. S., Li, J., Yi, B., & Yang, P. (2015). Estimation of errors in two-stream approximations of the solar radiative transfer equation for cloudy-sky conditions. *Journal of the Atmospheric Sciences*, *72*(11), 4053-4074. doi: 10.1175/JAS-D-15-0033.1

Barker, H. W., Stephens, G. L., Partain, P. T., Bergman, J. W., Bonnel, B., Campana, K., ... others (2003). Assessing 1D atmospheric solar radiative transfer models: Interpretation and handling of unresolved clouds. *Journal of Climate*, *16*(16), 2676–2699.

Bastidas, L. A., Hogue, T. S., Sorooshian, S., Gupta, H. V., & Shuttleworth, W. J. (2006). Parameter sensitivity analysis for different complexity land surface models using multicriteria methods. *Journal of Geophysical Research: Atmospheres*, *111*(D20). doi: https://doi.org/10.1029/2005JD006377

Bellprat, O., Kotlarski, S., Lthi, D., & Schr, C. (2012). Objective calibration of regional climate models. *Journal of Geophysical Research: Atmospheres*, *117*(D23). doi: https://doi.org/10.1029/2012JD018262

Berg, L. K., Kassianov, E. I., Long, C. N., & Mills, D. L. (2011). Surface summertime radiative forcing by shallow cumuli at the Atmospheric Radiation Measurement Southern Great Plains site. *Journal of Geophysical Research-Atmospheres*, *116*, D01202. doi: 10.1029/2010JD014593

Bony, S., Stevens, B., Frierson, D. M. W., Jakob, C., Kageyama, M., Pincus, R., . . . Webb, M. J. (2015). Clouds, circulation and climate sensitivity. *Nature Geoscience*, *8*(4), 261–268. doi: 10.1038/ngeo2398

Brient, F., Couvreux, F., Villefranque, N., Rio, C., & Honnert, R. (2019). Object-Oriented Identification of Coherent Structures in Large Eddy Simulations: Importance of Downdrafts in Stratocumulus. *Geophysical Research Letters*, *46*(5), 2854-2864. doi: 10.1029/2018GL081499

Brown, A., Cederwall, R., Chlond, A., Duynkerke, P., Golaz, J.-C., Khairoutdinov, M., . . . Stevens, B. (2002). Large-eddy simulation of the diurnal cycle of shallow cumulus convection over land. *Q. J. R. Meteorol. Soc.*, *128*, 1075–1093.

Cahalan, R. F., Oreopoulos, L., Marshak, A., Evans, K. F., Davis, A. B., Pincus, R., . . . Zhuravleva, T. B. (2005). The I3RC: Bringing Together the Most Advanced Radiative Transfer Tools for Cloudy Atmospheres. *Bulletin of the American Meteorological Society*, *86*(9), 1275–1293. doi: 10.1175/BAMS-86-9-1275

Couvreux, F., Hourdin, F., Williamson, D., Roehrig, R., Volodina, V., Villefranque, N., . . . Xu, W. (2020). Process-based climate model development harnessing machine learning: I. A calibration tool for parameterization improvement. *Journal of Advances in Modeling Earth Systems*.

Dolinar, E. K., Dong, X., Xi, B., Jiang, J. H., & Su, H. (2015). Evaluation of CMIP5 simulated clouds and TOA radiation budgets using NASA satellite observations. *Climate Dynamics*, *44*(7), 2229–2247. doi: 10.1007/s00382-014-2158-9

Duan, Q., Di, Z., Quan, J., Wang, C., Gong, W., Gan, Y., . . . Fan, S. (2017). Automatic Model Calibration: A New Way to Improve Numerical Weather Forecasting. *Bulletin of the American Meteorological Society*, *98*(5), 959-970. doi: 10.1175/BAMS-D-15-00104.1

Dufresne, J.-L., & Bony, S. (2008). An assessment of the primary sources of spread of global warming estimates from coupled atmosphereocean models. *Journal of Climate*, *21*(19), 5135-5144. doi: 10.1175/2008JCLI2239.1

Fielding, M. D., Schfer, S. A. K., Hogan, R. J., & Forbes, R. M. (2020). Parametrizing cloud geometry and its application in a subgrid cloud-edge erosion scheme. *Quarterly Journal of the Royal Meteorological Society*, *146*(729), 1651-1667. doi: 10.1002/qj.3758

Golaz, J.-C., Horowitz, L. W., & Levy II, H. (2013). Cloud tuning in a coupled climate model: Impact on 20th century warming. *Geophysical Research Letters*, *40*(10), 2246-2251. doi: https://doi.org/10.1002/grl.50232

Hinkelman, L. M., Stevens, B., & Evans, K. F. (2005). A large-eddy simulation study of anisotropy in fair-weather cumulus cloud fields. *Journal of the Atmospheric Sciences*, *62*(7), 2155-2171. doi: 10.1175/JAS3463.1

Hogan, R. J., & Bozzo, A. (2018). A flexible and efficient radiation scheme for the ECMWF model. *Journal of Advances in Modeling Earth Systems*. doi: 10.1029/2018MS001364

Hogan, R. J., Fielding, M. D., Barker, H. W., Villefranque, N., & Schfer, S. A. K. (2019). Entrapment: An Important Mechanism to Explain the Shortwave 3D Radiative Effect of Clouds. *Journal of the Atmospheric Sciences*, *76*(7), 2123-2141. doi: 10.1175/JAS-D-18-0366.1

Hogan, R. J., & Illingworth, A. J. (2000). Deriving cloud overlap statistics from radar. *Quarterly Journal of the Royal Meteorological Society*, *126*(569), 2903–2909. doi: 10.1002/qj.49712656914

Hogan, R. J., Schäfer, S. A. K., Klinger, C., Chiu, J. C., & Mayer, B. (2016, July). Representing 3-D cloud radiation effects in two-stream schemes: 2. Matrix formulation and broadband evaluation. *Journal of Geophysical Research: Atmospheres*, *121*(14), 2016JD024875. doi: 10.1002/2016JD024875

Hogan, R. J., & Shonk, J. K. P. (2013). Incorporating the Effects of 3d Radia-

tive Transfer in the Presence of Clouds into Two-Stream Multilayer Radiation Schemes. *Journal of the Atmospheric Sciences*, *70*(2), 708–724. doi: 10.1175/JAS-D-12-041.1

Hogue, T. S., Bastidas, L. A., Gupta, H. V., & Sorooshian, S. (2006). Evaluating model performance and parameter behavior for varying levels of land surface model complexity. *Water Resources Research*, *42*(8). doi: https://doi.org/10.1029/2005WR004440

Hourdin, F., Mauritsen, T., Gettelman, A., Golaz, J.-C., Balaji, V., Duan, Q., ... Williamson, D. (2017). The art and science of climate model tuning. *Bulletin of the American Meteorological Society*, *98*(3), 589-602. doi: 10.1175/BAMS-D-15-00135.1

Hourdin, F., Williamson, D., Rio, C., Couvreux, F., Roehrig, R., Villefranque, N., ... Volodina, V. (2020). Process-based climate model development harnessing Machine Learning. Part II: Model Calibration from Single Column to Global. *Journal of Advances in Modeling Earth Systems*.

Iacono, M. J., Delamere, J. S., Mlawer, E. J., Shephard, M. W., Clough, S. A., & Collins, W. D. (2008). Radiative forcing by long-lived greenhouse gases: Calculations with the AER radiative transfer models. *Journal of Geophysical Research: Atmospheres*, *113*(D13). doi: 10.1029/2008JD009944

Jensen, J. L. W. V. (1906). Sur les fonctions convexes et les ingalits entre les valeurs moyennes. *Acta Mathematica*, *30*, 175–193. doi: 10.1007/BF02418571

Joseph, J. H., Wiscombe, W. J., & Weinman, J. A. (1976). The Delta-Eddington Approximation for Radiative Flux Transfer. *Journal of the Atmospheric Sciences*, *33*(12), 2452–2459. doi: 10.1175/1520-0469(1976)033⟨2452:TDEAFR⟩2.0 .CO;2

Karlsson, J., Svensson, G., & Rodhe, H. (2008). Cloud radiative forcing of subtropical low level clouds in global models. *Climate Dynamics*, *30*(7-8), 779–788. doi: 10.1007/s00382-007-0322-1

Lac, C., Chaboureau, J.-P., Masson, V., Pinty, J.-P., Tulet, P., Escobar, J., ... Wautelet, P. (2018). Overview of the meso-nh model version 5.4 and its applications. *Geoscientific Model Development*, *11*(5), 1929–1969. doi: 10.5194/gmd-11-1929-2018

Lafore, J. P., Stein, J., Asencio, N., Bougeault, P., Ducrocq, V., Duron, J., ... de Arellano, J. V.-G. (1997). The Meso-NH Atmospheric Simulation System. Part I: adiabatic formulation and control simulations. *Annales Geophysicae*, *16*(1), 90–109. doi: 10.1007/s00585-997-0090-6

Manners, J., Edwards, J. M., Hill, P., & Thelen, J.-C. (2017). *SOCRATES Technical Guide Suite Of Community RAdiative Transfer codes based on Edwards and Slingo* (Tech. Rep.). FitzRoy Rd, Exeter EX1 3PB: Met Office. Retrieved from `http://homepages.see.leeds.ac.uk/~lecsjed/winscpuse/socrates_techguide.pdf`

Marchuk, G. I., Mikhailov, G. A., Nazareliev, M. A., Darbinjan, R. A., Kargin, B. A., & Elepov, B. S. (1980). *The Monte Carlo Methods in Atmospheric Optics.* Berlin Heidelberg: Springer-Verlag.

Marshak, A., & Davis, A. (Eds.). (2005). *3D Radiative Transfer in Cloudy Atmospheres.* Berlin Heidelberg: Springer-Verlag. (DOI: 10.1007/3-540-28519-9)

Mauritsen, T., Stevens, B., Roeckner, E., Crueger, T., Esch, M., Giorgetta, M., ... Tomassini, L. (2012). Tuning the climate of a global model. *Journal of Advances in Modeling Earth Systems*, *4*(3). doi: https://doi.org/10.1029/2012MS000154

Mayer, B. (2009). Radiative transfer in the cloudy atmosphere. *The European Physical Journal Conferences*, *1*, 75–99. doi: 10.1140/epjconf/e2009-00912-1

McKee, T. B., & Cox, S. K. (1974). Scattering of Visible Radiation by Finite Clouds. *Journal of the Atmospheric Sciences*, *31*(7), 1885–1892. doi: 10.1175/1520-0469(1974)031⟨1885:SOVRBF⟩2.0.CO;2

McKee, T. B., & Klehr, J. T.   (1978).   Effects of Cloud Shape on Scattered Solar Radiation.   *Monthly Weather Review*, *106*(3), 399–404.   doi: 10.1175/1520 -0493(1978)106⟨0399:EOCSOS⟩2.0.CO;2

Mishchenko, M., D. Travis, L., & Lacis, A. (2002). *Scattering, absorption, and emission of light by small particles* (Vol. 4).

Nam, C., Bony, S., Dufresne, J.-L., & Chepfer, H.   (2012).   The 'too few, too bright' tropical low-cloud problem in CMIP5 models.   *Geophysical Research Letters*, *39*, L21801. doi: 10.1029/2012GL053421

Neggers, R., Duynkerke, P., & Rodts, S.   (2003).   Shallow cumulus convection : A validation of large-eddy simulation against aircraft and Landsat observations.   *Quarterly Journal of the Royal Meteorological Society*, *129*, 2671-2696.

Neggers, R. A. J., Heus, T., & Siebesma, A. P.   (2011).   Overlap statistics of cumuliform boundary-layer cloud fields in large-eddy simulations.   *Journal of Geophysical Research: Atmospheres*, *116*(D21). doi: 10.1029/2011JD015650

Newman, W. I., Lew, J. K., Siscoe, G. L., & Fovell, R. G.   (1995, 02).   Systematic Effects of Randomness in Radiative Transfer.   *Journal of the Atmospheric Sciences*, *52*(4), 427-435. doi: 10.1175/1520-0469(1995)052⟨0427:SEORIR⟩2.0.CO; 2

Pincus, R., Barker, H. W., & Morcrette, J.-J.   (2003).   A fast, flexible, approximate technique for computing radiative transfer in inhomogeneous cloud fields.   *Journal of Geophysical Research: Atmospheres*, *108*(D13).   doi: 10.1029/2002JD003322

Ramanathan, V., Cess, R. D., Harrison, E. F., Minnis, P., Barkstrom, B. R., Ahmad, E., & Hartmann, D. (1989). Cloud-radiative forcing and climate: Results from the earth radiation budget experiment.   *Science*, *243*(4887), 57–63.   doi: 10.1126/science.243.4887.57

Salter, J. M., Williamson, D. B., Scinocca, J., & Kharin, V.   (2019).   Uncertainty Quantification for Computer Models With Spatial Output Using Calibration-Optimal Bases.   *Journal of the American Statistical Association*, *114*(528), 1800-1814. doi: 10.1080/01621459.2018.1514306

Schäfer, S. A. K., Hogan, R. J., Klinger, C., Chiu, J. C., & Mayer, B.   (2016, July).   Representing 3-D cloud radiation effects in two-stream schemes: 1. Longwave considerations and effective cloud edge length.   *Journal of Geophysical Research: Atmospheres*, *121*(14), 2016JD024876. doi: 10.1002/2016JD024876

Schmidt, G. A., Bader, D., Donner, L. J., Elsaesser, G. S., Golaz, J.-C., Hannay, C., ... Saha, S.   (2017).   Practice and philosophy of climate model tuning across six us modeling centers.   *Geoscientific Model Development*, *10*(9), 3207–3223. doi: 10.5194/gmd-10-3207-2017

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*(11), 2352–2370.   doi: 10.1175/2007JCLI1940.1

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 global radiation budget. Part I: review and parametrization. *Quarterly Journal of the Royal Meteorological Society*. doi: 10.1002/qj.647

Siebesma, A. P., Bretherton, C. S., Brown, A., Chlond, A., Cuxart, J., Duynkerke, P. G., ... Stevens, D. E.   (2003).   A Large Eddy Simulation Intercomparison Study of Shallow Cumulus Convection.   *Journal of the Atmospheric Sciences*, *60*(10), 1201-1219.

Sulak, A. M., Calabrase, W. J., Ryan, S. D., & Heus, T.   (2020).   The contributions of shear and turbulence to cloud overlap for cumulus clouds.   *Journal of Geophysical Research: Atmospheres*, *125*(10), e2019JD032017. doi: https://doi .org/10.1029/2019JD032017

vanZanten, M. C., Stevens, B., Nuijens, L., Siebesma, A. P., Ackerman, A. S., Bur-

net, F., ... Wyszogrodzki, A. (2011). Controls on precipitation and cloudiness in simulations of trade-wind cumulus as observed during rico. *Journal of Advances in Modeling Earth Systems*, *3*(2).

Várnai, T., & Davies, R. (1999). Effects of cloud heterogeneities on shortwave radiation: Comparison of cloud-top variability and internal heterogeneity. *Journal of the Atmospheric Sciences*, *56*(24), 4206–4224.

Villefranque, N., Fournier, R., Couvreux, F., Blanco, S., Cornet, C., Eymet, V., ... Tregan, J.-M. (2019). A Path-Tracing Monte Carlo Library for 3-D Radiative Transfer in Highly Resolved Cloudy Atmospheres. *Journal of Advances in Modeling Earth Systems*. doi: 10.1029/2018MS001602

Williamson, D. (2015). Exploratory ensemble designs for environmental models using k-extended latin hypercubes. *Environmetrics*, *26*(4), 268-283. doi: 10.1002/env.2335

Supporting Information for "Process-based climate model development harnessing machine learning: III. The Representation of Cumulus Geometry and their 3D Radiative Effects"

Najda Villefranque[1], Stéphane Blanco[2], Fleur Couvreux[1], Richard Fournier[2],
Jacques Gautrais[3], Robin J. Hogan[4], Frédéric Hourdin[5], Victoria Volodina[6],
Daniel Williamson[6,7]

[1]CNRM, Université de Toulouse, Meteo-France, CNRS, Toulouse, France

[2]LAPLACE, Université de Toulouse, CNRS, Toulouse, France

[3]Centre de Recherches sur la Cognition Animale (CRCA), Centre de Biologie Intégrative (CBI), Université de Toulouse, CNRS, Université Paul Sabatier, France

[4]ECMWF, Reading, United Kingdom

[5]LMD-IPSL, Sorbonne University, CNRS, 4 pl Jussieu, Paris, France

[6]The Alan Turing Institute, 96 Euston Road, London, United Kingdom

[7]Exeter University, Exeter, United Kingdom

Contents of this file

Introduction This document details our choices in the production of liquid droplet optical properties. We use Mishchenko's code (Mishchenko et al., 2002) to compute the Mie optical properties that were used to design the SOCRATES (Manners et al., 2017) parameterisation. The routines are available on gitlab (`https://gitlab.com/najdavlf/mie-mdv`). Two figures illustrate the optical properties and solar energy incoming in each spectral band. On another topic, two tables are provided that are as Table 3 from the main text but the errors are given separately for two solar zenith angles (SZAs) instead of being averaged over various SZAs.

Text S1. Droplet size distribution parameters.

Let $n(r)$ be the density number of droplets which radius is in $\mathrm{d}r$, an infinitesimal range around r. The effective radius is defined as the ratio of moments of order 3 and 2 of the droplet size distribution:

$$r_{\mathrm{eff}} = \frac{\int_0^\infty \mathrm{d}r\, n(r) r^3}{\int_0^\infty \mathrm{d}r\, n(r) r^2} \qquad (1)$$

The effective variance represnent a mean quadratic deviation from the effective radius:

$$v_{\mathrm{eff}} = \frac{1}{r_{\mathrm{eff}}^2} \frac{\int_0^\infty \mathrm{d}r\, n(r)(r - r_{\mathrm{eff}})^2 r^2}{\int_0^\infty \mathrm{d}r\, n(r) r^2} \qquad (2)$$

Text S2. Gamma distribution.

$$n(r) = N_0 r^{\frac{1-3b}{b}} \exp(-\frac{r}{ab}) \qquad (3)$$

One can show that:

$$r_{\mathrm{eff}} = a \qquad (4)$$

and that:

$$v_{\mathrm{eff}} = b \qquad (5)$$

Expected parameters in Mishchenko's code are $AA = a$ et $BB = b$, hence directly the effective radius and variance. The modal radius of the distribution (not used here), is:

$$r_m = (1 - 3v_{\mathrm{eff}}) r_{\mathrm{eff}} \qquad (6)$$

and the shape parameter:

$$\alpha = \frac{1 - 3v_{\mathrm{eff}}}{v_{\mathrm{eff}}} \qquad (7)$$

These parameters also appear in the Generalised Gamma distribution.

Text S3. Numerical integration on size distribution.

Mishchenko's code numerically integrates monodisperse optical properties on the selected size distribution. The user must choose an integration interval $[r_{\mathrm{min}}, r_{\mathrm{max}}]$. We set the bounds in order to let:

$$\int_0^{r_{\mathrm{min}}} \mathrm{d}r n(r) \sim \epsilon \qquad (8)$$

and

$$\int_{r_{\mathrm{max}}}^{1000} \mathrm{d}r n(r) \sim \epsilon \qquad (9)$$

That is, the size distribution integral on $[r_{\mathrm{min}}, r_{\mathrm{max}}]$ is $1-2\epsilon$. We set $\epsilon = 10^{-4}$.

Text S4. From cross section to massic cross section.

Mishchenko's code outputs cross sections in m^{-1}. We convert them into massic cross sections, in m^2.kg^{-1}. To do so, we need the effective mass corresponding to the droplets distribution:

$$M_{eff} = \int_{r_{\min}}^{r_{\max}} \mathrm{d}r n(r) \, \rho_w \frac{4}{3}\pi r^3 \qquad (10)$$

again numerically integrated.

Text S5. Phase functions.

Phase functions are computed on 2k angles. They are then cumulated by numerical integration using the trapezoidal rule on the phase function and sinus of the scattering angle. The inverse cumulated is tabulated on 4k points, by linearly interpolating the cumulated.

Text S6. SOCRATES.

To reproduce the data that served as reference to design the SOCRATES parameterisation, four computations were performed using four Gamma distributions for an effective radius of 10 μm, of varying effective variance: 0.01, 0.1, 0.175, 0.25. The spectrum is discretized from 200 nm to 14 μm with one point every 25 nm. The Mie data integrated over these four size distributions were then averaged. The resulting data is plotted in Figure S1. In the Monte Carlo algorithm, a spectral band is sampled according to the incoming sun flux distribution plotted in Figure S2. Then a wavelength is sampled uniformly in the band and the cloud optical properties are interpolated linearly from the 25 nm discretized Mie table.

References

Manners, J., Edwards, J. M., Hill, P., & Thelen, J.-C. (2017). *SOCRATES Technical Guide Suite Of Community RAdiative Transfer codes based on Edwards and Slingo* (Tech. Rep.). FitzRoy Rd, Exeter EX1 3PB: Met Office. Retrieved from `http://homepages.see.leeds.ac.uk/~lecsjed/winscpuse/socrates_techguide.pdf`

Mishchenko, M., D. Travis, L., & Lacis, A. (2002). *Scattering, absorption, and emission of light by small particles* (Vol. 4).
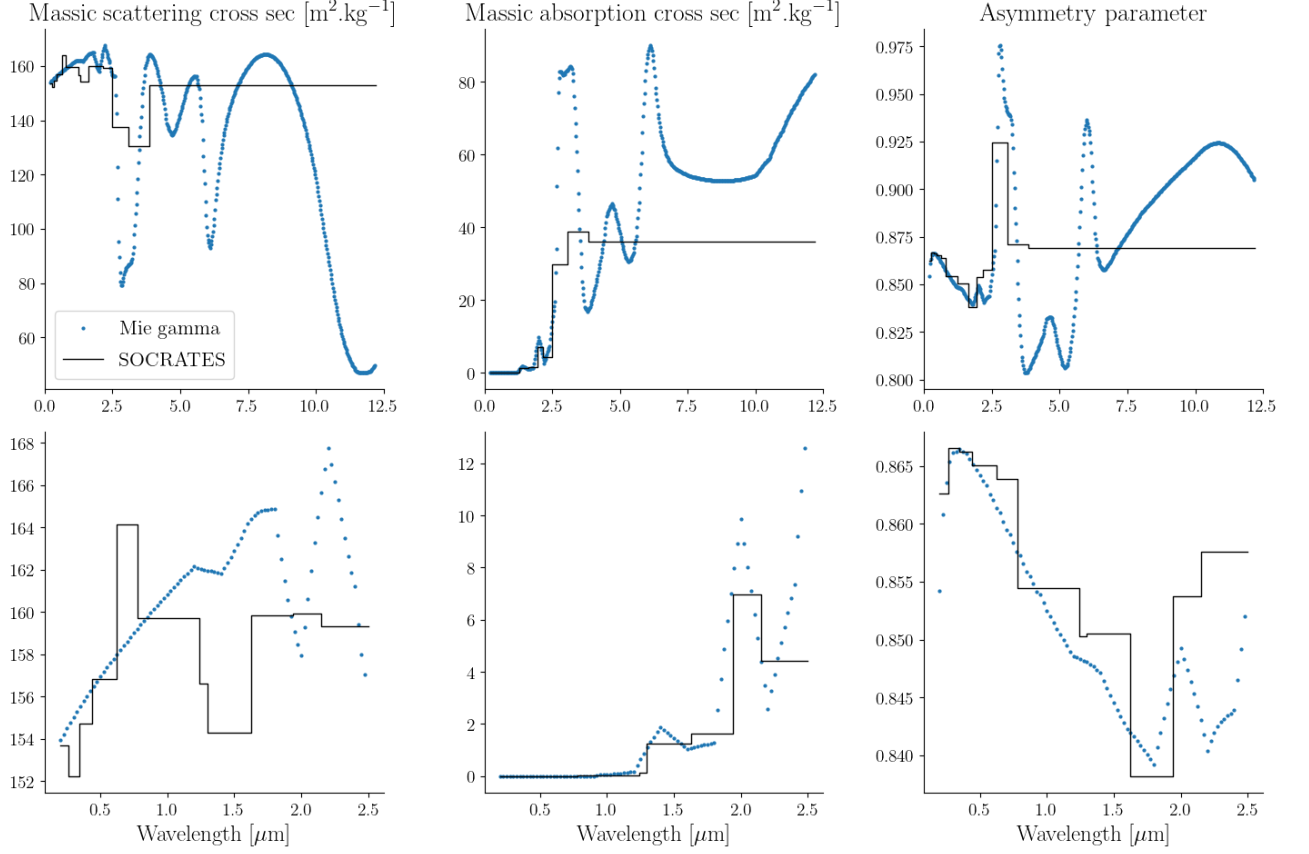
Figure S1. Optical properties for liquid water, following Mie theory (in blue) and the SOCRATES parameterisation (in black). The bottom row is a zoom of the same data on the 0-2.5 μm spectral interval.
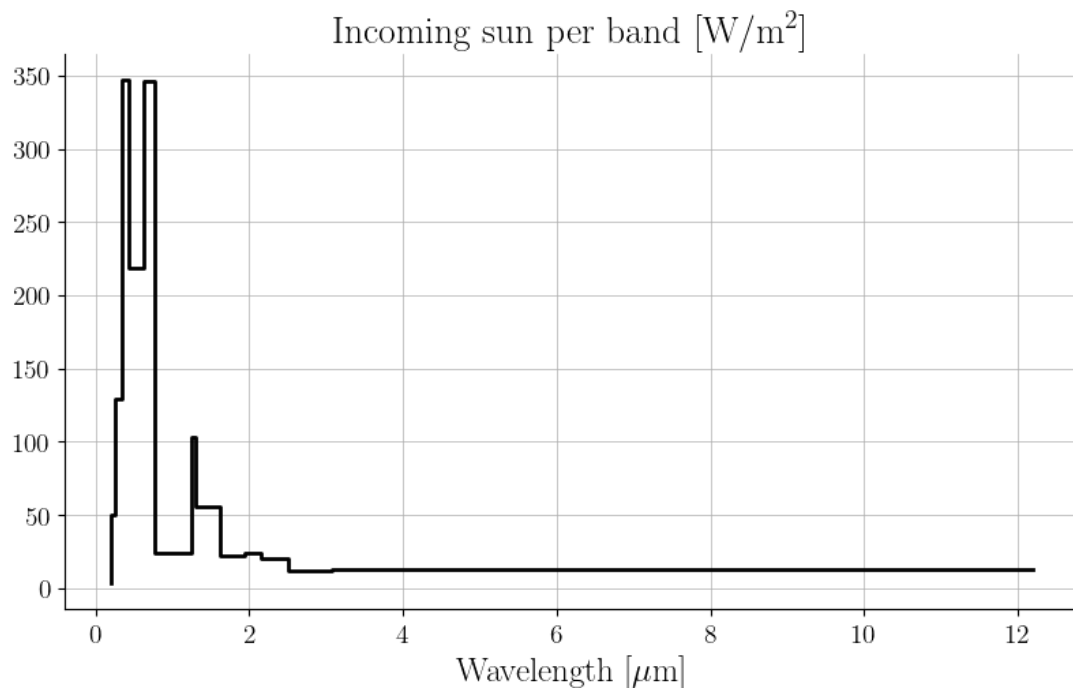
Figure S2.   Incoming solar flux at TOA, integrated in each spectral band of the RRTMG parameterisation. This data is used in ecRad and also in our Monte Carlo computations.

Table S1. Same as Table 3. but relative errors [%] are computed at solar zenith angle 0° instead of averaged

(i) Experiment	(ii) Reference	(iii) Model	(iv) TOA up		(v) Absorbed		(vi) Surf. down	
			RMS	bias	RMS	bias	RMS	bias
(a) In slabs (Section 3.1)								
(1) SOCRATES	MC exact	MC SOCRATES	1.2	0.9	12.0	-11.4	2.8	-2.1
(2) δ-Eddington	MC exact	MC δ-Eddington	14.9	10.3	22.2	14.8	16.3	11.2
(3) two-stream	MC as ecRad	ecRad two-stream	6.3	-4.6	8.4	-3.2	6.9	-4.6
Transport (2+3)	MC SOCRATES	ecRad two-stream	5.6	3.1	17.0	13.4	7.8	5.1
Total (1+2+3)	MC exact	ecRad two-stream	6.0	4.1	7.8	0.3	5.3	2.9
(b) In cumulus, MC vs ecRad 1D and 3D solvers, parameters $\lambda = (\alpha, \mathrm{FSD}, C_s)$ (Section 3.2)								
PPH max ovp		1D, $\lambda = (1, 0, \infty)$	16.5	-4.2	64.6	-28.7	15.4	-12.3
Tripleclouds (1D)	MC SOCRATES	1D, $\lambda(z, case)$ LES	39.6	36.2	51.4	12.3	30.8	28.3
SPARTACUS (3D)		3D, $\lambda(z, case)$ LES	25.5	21.8	51.5	14.5	20.5	17.6
(c) In cumulus, ecRad SPARTACUS, with LES profiles vs averaged parameters (Section 3.3)								
z-averaged	$\lambda(z, case)$ LES	$\overline{\lambda}(case)$ LES	1.5	-0.5	1.9	-0.8	1.6	-0.6
case-z-averaged	$\lambda(z, case)$ LES	$\overline{\overline{\lambda}}$ LES	3.4	1.1	3.3	-0.3	3.3	0.8
(d) In cumulus, MC vs ecRad SPARTACUS with calibrated parameters (Section 4)								
Best global		$\overline{\overline{\lambda}}$ from htexplo	13.2	-5.8	51.7	-12.2	12.1	-9.7
Best TOA up	MC SOCRATES		18.4	-14.5	53.2	-20.5	19.4	-18.2
Best absorption		(see Table 4.)	24.8	20.4	53.0	7.0	17.4	14.7
Best surface down			13.2	-6.3	50.8	-12.5	12.5	-10.1

For each pair of reference computation (ii) / test approximation (iii), errors on the cloud radiative effects on TOA upward (iv), absorbed (v), and surface downward (vi) fluxes are quantified. For each column, the RMS and mean bias are presented for SZA=0°. The table subsections concern: (a) errors related to non-geometrical effects of clouds, (b) ecRad errors for different solvers, with increasing complexity in the representation of geometrical effects, (c) errors related to the neglect of parameters variations with height and cloud field, (d) ecRad errors for different choices of cloud-geometry parameters, output from the calibration exercise of Section 4.

CRE = total sky - clear sky. Relative error $r = 100\times(\text{model-ref})/\text{ref}$. RMS $= \sqrt{\langle r^2 \rangle_{fields}}$. bias=$\langle r \rangle_{fields}$
MC exact: detailed Mie optical properties and phase function.
MC SOCRATES: parameterized optical properties and detailed Mie phase function.
MC δ-Eddington: detailed Mie optical properties and HG δ-Eddington phase function.
MC as ecRad: parameterized optical properties and HG δ-Eddington phase function.

Table S2. Same as Table 3. but relative errors [%] are computed at solar zenith angle 77° instead of averaged

(i) Experiment	(ii) Reference	(iii) Model	(iv) TOA up		(v) Absorbed		(vi) Surf. down	
			RMS	bias	RMS	bias	RMS	bias
(a) In slabs (Section 3.1)								
(1) SOCRATES	MC exact	MC SOCRATES	0.7	0.6	84.3	-32.6	0.7	-0.6
(2) δ-Eddington	MC exact	MC δ-Eddington	15.1	-10.8	32.2	-23.1	16.1	-11.5
(3) two-stream	MC as ecRad	ecRad two-stream	6.4	3.4	1841.1	507.0	2.5	-1.0
Transport (2+3)	MC SOCRATES	ecRad two-stream	11.8	-9.1	696.3	179.1	17.2	-13.2
Total (1+2+3)	MC exact	ecRad two-stream	11.6	-8.6	292.6	-113.3	17.7	-13.7
(b) In cumulus, MC vs ecRad 1D and 3D solvers, parameters $\lambda = (\alpha, \mathrm{FSD}, C_s)$ (Section 3.2)								
PPH max ovp		1D, $\lambda = (1, 0, \infty)$	67.6	-66.9	213.3	-115.9	69.2	-68.6
Tripleclouds (1D)	MC SOCRATES	1D, $\lambda(z, case)$ LES	37.4	-36.8	338.7	-127.2	41.7	-41.1
SPARTACUS (3D)		3D, $\lambda(z, case)$ LES	10.6	-7.0	481.0	-140.7	16.8	-14.8
(c) In cumulus, ecRad SPARTACUS, with LES profiles vs averaged parameters (Section 3.3)								
z-averaged	$\lambda(z, case)$ LES	$\overline{\lambda}(case)$ LES	2.1	1.7	5.1	3.9	2.0	1.5
case-z-averaged	$\lambda(z, case)$ LES	$\overline{\overline{\lambda}}$ LES	5.9	-1.1	11.8	-2.1	5.6	-1.0
(d) In cumulus, MC vs ecRad SPARTACUS with calibrated parameters (Section 4)								
Best global		$\overline{\overline{\lambda}}$ from htexplo	15.5	-11.9	462.5	-139.4	20.8	-18.4
Best TOA up	MC SOCRATES		15.2	-11.8	436.3	-135.1	20.3	-17.8
Best absorption		(see Table 4.)	28.6	-27.5	410.4	-134.9	33.5	-32.5
Best surface down			12.0	-6.0	492.4	-144.9	16.5	-12.9

For each pair of reference computation (ii) / test approximation (iii), errors on the cloud radiative effects on TOA upward (iv), absorbed (v), and surface downward (vi) fluxes are quantified. For each column, the RMS and mean bias are presented for SZA=77°. The table subsections concern: (a) errors related to non-geometrical effects of clouds, (b) ecRad errors for different solvers, with increasing complexity in the representation of geometrical effects, (c) errors related to the neglect of parameters variations with height and cloud field, (d) ecRad errors for different choices of cloud-geometry parameters, output from the calibration exercise of Section 4.

CRE = total sky - clear sky. Relative error $r = 100 \times (\mathrm{model-ref})/\mathrm{ref}$. RMS = $\sqrt{\langle r^2 \rangle_{fields}}$. bias=$\langle r \rangle_{fields}$
MC exact: detailed Mie optical properties and phase function.
MC SOCRATES: parameterized optical properties and detailed Mie phase function.
MC δ-Eddington: detailed Mie optical properties and HG δ-Eddington phase function.
MC as ecRad: parameterized optical properties and HG δ-Eddington phase function.