



**HAL**  
open science

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

Najda Villefranque, Stéphane Blanco, Fleur Couvreur, Richard A Fournier, Jacques Gautrais, Robin James Hogan, Frédéric Hourdin, Victoria Volodina, Daniel Williamson

### ► To cite this version:

Najda Villefranque, Stéphane Blanco, Fleur Couvreur, Richard A Fournier, Jacques Gautrais, et al.. Process-based climate model development harnessing machine learning: III. The Representation of Cumulus Geometry and their 3D Radiative Effects. 2020. hal-03036667v1

**HAL Id: hal-03036667**

**<https://hal.science/hal-03036667v1>**

Preprint submitted on 2 Dec 2020 (v1), last revised 16 Nov 2021 (v2)

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

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

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

7   <sup>1</sup>CNRM, Universit de Toulouse, Meteo-France, CNRS, Toulouse, France

8   <sup>2</sup>LAPLACE, Universit de Toulouse, CNRS, Toulouse, France

9           <sup>3</sup>Centre de Recherches sur la Cognition Animale (CRCA), Centre de Biologie Intgrative (CBI), Universit  
10           de Toulouse, CNRS, Universit Paul Sabatier, France

11   <sup>4</sup>ECMWF, Reading, United Kingdom

12   <sup>5</sup>LMD-IPSL, Sorbonne University, CNRS, 4 pl Jussieu, Paris, France

13   <sup>6</sup>The Alan Turing Institute, 96 Euston Road, London, United Kingdom

14   <sup>7</sup>Exeter University, Exeter, United Kingdom

15           **Key Points:**

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

## 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 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  $5 \text{ Wm}^{-2}$  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 SPAR-TACUS (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

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

78 In the first two papers of this series (Couvreur et al., 2020; Hourdin et al., 2020),  
79 a strategy to reduce cloud-related uncertainties in large-scale models is proposed. It strongly  
80 relies on comparisons between Single Column Model (SCM) and Large-Eddy Simulations  
81 (LES) to develop, evaluate and calibrate parameterizations at the process scale, in par-  
82 ticular those that model boundary-layer (BL) transport and clouds. Free open-source  
83 numerical tools are provided to the community to promote a transparent, rigorous and  
84 efficient procedure to calibrate models. The first paper (Couvreur et al., 2020) is focused  
85 on the description of High-Tune:explorer (htexplo), a tuning tool based on history match-  
86 ing, developed jointly by physicists in atmospheric science and statisticians in Uncer-  
87 tainty Quantification. Its potential is demonstrated through the calibration of BL pa-  
88 rameterizations of an SCM, targeting reference BL metrics computed in the LES fields.  
89 It thereby ensures that parameters controlling the BL are calibrated to obtain the right  
90 BL properties. The second paper (Hourdin et al., 2020) uses the same tool to calibrate  
91 the 3D climate model targeting observed radiation at the TOA. In this global calibra-  
92 tion, the model is only allowed to explore a restricted parameter space, which is obtained  
93 by prior calibration of the parameterization of shallow convection using the SCM/LES  
94 framework. It thereby ensures that BL parameters are calibrated to obtain both the right  
95 BL properties and the right global radiation, i.e., the right TOA cloud radiative forc-  
96 ing for the right BL clouds.

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

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

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

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

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

## 141 2 Radiative Transfer Models and Cloudy Atmosphere Data

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

### 150 2.1 Radiative transfer models

#### 151 2.1.1 *ecRad: a flexible radiation scheme for large-scale atmospheric mod-* 152 *els*

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

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

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

167 **1. Spectral dimension and optical properties.** The RRTMG gas model used  
 168 in ecRad and other radiation schemes uses the correlated k-distribution method  
 169 in 14 bands in the solar and 16 bands in the thermal. Here, only solar computa-  
 170 tions are performed, integrating fluxes on a 0.2 – 12.5  $\mu\text{m}$  interval. The SOCRATES  
 171 data that describe the liquid cloud optical properties are the coefficients of a poly-  
 172 nomial 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<br>or SPARTACUS | Shonk and Hogan (2008)<br>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 unknown 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  $\delta$ -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,

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

and  $\alpha$  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

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

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

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

217 **5. Horizontal heterogeneity of in-cloud liquid water.** Since Beer’s exponen-  
 218 tial law is a convex function, Jensen’s inequality applies (Jensen, 1906), resulting  
 219 in systematic bias due to averaging of optical properties (Newman et al., 1995):  
 220 the mean transmitted flux under a cloud of horizontally varying liquid water con-  
 221 tent (LWC) is always larger than the transmitted flux under the equivalent ho-  
 222 mogeneous cloud of average LWC. To represent the effect of horizontal variations  
 223 of LWC on radiation, the Tripleclouds method has been proposed by Shonk and  
 224 Hogan (2008). In addition to solving solving radiation independently in cloudy  
 225 and clear regions of each layer, it further divides the cloudy region into a thin sub-  
 226 region and a thick one. To distribute the LWC into the two sub-regions of a given  
 227 layer and then infer their respective optical depths, a gamma-shaped distribution  
 228 of the liquid water is assumed, characterized by a mean and a standard deviation  
 229  $\sigma$ . Tripleclouds uses the fractional standard deviation (FSD) of the distribution  
 230 (ratio of  $\sigma$  to mean in-cloud LWC) to robustly characterize the horizontal vari-  
 231 ability of LWC in each layer.

232 **6. Horizontal transport of light.** In Tripleclouds, layers are divided into sub-regions  
 233 and radiation is solved independently in each sub-region of the layer. Light can-  
 234 not be transported from clear to cloudy sub-regions of the same layer. However,  
 235 many studies have shown that horizontal transport of light significantly modifies  
 236 the distribution of energy in the atmosphere: transport through cloud sides in-  
 237 creases transmission at high sun and decreases it at low sun, while multiple re-  
 238 flections combined to horizontal transport leads to entrapment of upward flux, thereby  
 239 increasing transmission (McKee & Cox, 1974; McKee & Klehr, 1978; Várnai & Davies,  
 240 1999; Barker et al., 2003; Hogan & Shonk, 2013; Hogan et al., 2019). SPARTA-  
 241 CUS is the first two-stream-based solver that allows these 3D effects to be rep-  
 242 resented in large-scale models. It explicitly represents entrapment by computing  
 243 the mean horizontal distance traveled by reflected light. To represent transport  
 244 through cloud sides, it uses the Tripleclouds approximation for the cloud field then  
 245 adds exchange terms between regions. This term is proportional to the length of  
 246 the interface between clear and cloudy regions. For a given cloud fraction, 3D ef-  
 247 fects will be larger for a large number of small clouds than for a single large cloud.  
 248 From the total cloud perimeter density  $p$  (perimeter length to domain horizontal  
 249 area, of unit inverse length) and the cloud fraction of the layer  $c$ , Schäfer et al. (2016)  
 250 define the cloud effective scale  $C_s$  as:

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

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

257

### 2.1.2 Monte Carlo reference computations of solar 3D RT

258

259

260

261

262

263

264

265

266

267

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 <sup>1</sup>. 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.

268

269

270

271

272

273

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.

274

275

276

277

278

279

280

281

282

283

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  $\mu\text{m}$  (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.

284

285

286

287

288

289

290

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  $\delta$ -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).

291

292

293

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.

294

## 2.2 Cloudy atmosphere data

295

### 2.2.1 3D fields from LES

296

297

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):

298

299

300

301

- 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%;

<sup>1</sup> [https://gitlab.com/najdavlf/scart\\_project](https://gitlab.com/najdavlf/scart_project)



- 302 • BOMEX (Siebesma et al., 2003), a case of marine shallow cumulus forced with  
303 constant surface fluxes through the simulation. Cloud cover ranges from 10 to 20%;
- 304 • RICO (vanZanten et al., 2011), a second case of marine cumulus, forced with con-  
305 stant sea surface temperature through the simulation. Cloud cover ranges from  
306 15 to 25%;
- 307 • SCMS (R. Neggers et al., 2003), a case of continental cumulus developing in Florida,  
308 with strong moisture advection into the domain caused by the nearby ocean. Cloud  
309 cover ranges from 0 to 45%.

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

318 Using an object-identification tool <sup>2</sup> (Brient et al., 2019), individual clouds are la-  
319 belled in each field. A cloud is defined as an ensemble of contiguous cells where the liq-  
320 uid mixing ratio is greater than  $10^{-6} \text{ kg/kg}$ . Each scene is then described in terms of  
321 cloud characteristics, some of which are presented in Table 2. The cloud cover is the pro-  
322 portion of cloudy columns in the domain. At the very first order, it controls the trans-  
323 mitted and reflected solar fluxes. The number density is the total number of identified  
324 clouds in the scene divided by the horizontal surface of the domain. For a given cloud  
325 cover, a larger number density indicates a longer interface between clouds and clear sky,  
326 hence more 3D radiative effects. The maximum depth is the higher minus lower altitudes  
327 where clouds are present. When the sun is not at zenith, the “effective” cloud cover (that  
328 is, the cloud cover projected in the sun direction) depends on the cloud layer depth.

### 329 **2.2.2 1D profiles from 3D fields**

330 From each 3D cloud field output from LES runs, 1D profiles are derived to serve  
331 as inputs to ecRad. Temperature, pressure, vapor and liquid mixing ratios are horizon-  
332 tally averaged from the 3D fields on each vertical level and extended above the LES do-  
333 main top using the I3RC (Cahalan et al., 2005) mid-latitude summer (MLS) cumulus  
334 profiles provided in the ecRad package. Gases mixing ratios (other than water vapor)  
335 are set as in the I3RC MLS cumulus case. Cloud fraction is computed at each level as  
336 the fraction of cells where the liquid mixing ratio is positive in the 3D field. Effective  
337 radius for liquid droplets is uniformly set to  $10 \mu\text{m}$  as in the Monte Carlo runs. These  
338 are the basic profiles needed by any RT scheme. In order to take into account the ge-  
339 ometry of clouds, three more parameters are to be provided to ecRad: the overlap pa-  
340 rameter  $\alpha$ , the FSD of in-cloud liquid water horizontal distribution and the cloud scale  
341  $C_s$ .

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

<sup>2</sup> <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 [ $\text{km}^{-2}$ ] | 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          |
| <b>ARMCu</b> | <b>08</b> | <b>26.343</b> | <b>1.64</b>                         | <b>1.225</b>   |
| ARMCu        | 09        | 26.180        | 1.44                                | 1.050          |
| ARMCu        | 10        | 23.499        | 1.61                                | 1.375          |
| ARMCu        | 11        | 23.029        | 1.15                                | 1.275          |
| <b>ARMCu</b> | <b>12</b> | <b>12.663</b> | <b>0.81</b>                         | <b>1.450</b>   |
| <b>BOMEX</b> | <b>05</b> | <b>16.301</b> | <b>2.17</b>                         | <b>1.200</b>   |
| 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          |
| <b>BOMEX</b> | <b>11</b> | <b>10.318</b> | <b>2.00</b>                         | <b>0.775</b>   |
| BOMEX        | 12        | 14.294        | 2.15                                | 0.650          |
| RICO         | 04        | 13.933        | 2.27                                | 0.950          |
| <b>RICO</b>  | <b>05</b> | <b>13.802</b> | <b>2.15</b>                         | <b>0.850</b>   |
| RICO         | 06        | 17.195        | 2.25                                | 1.025          |
| <b>RICO</b>  | <b>07</b> | <b>18.054</b> | <b>2.34</b>                         | <b>1.175</b>   |
| 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          |
| <b>SCMS</b>  | <b>05</b> | <b>37.947</b> | <b>3.71</b>                         | <b>1.450</b>   |
| <b>SCMS</b>  | <b>06</b> | <b>32.010</b> | <b>2.78</b>                         | <b>1.400</b>   |
| 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          |

349 than the range reported by Hogan and Illingworth (2000) because of our smaller verti-  
350 cal resolution.

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

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

### 372 **3 Evaluating ecRad against reference Monte Carlo**

373 The objective of this section is to characterize the different sources of error in ecRad.  
374 The two-stream,  $\delta$ -Eddington and approximate optical properties errors are briefly an-  
375 alyzed in Section 3.1. Section 3.2 focuses on the errors due to the approximate repre-  
376 sentation of cloud geometry and 3D effects.

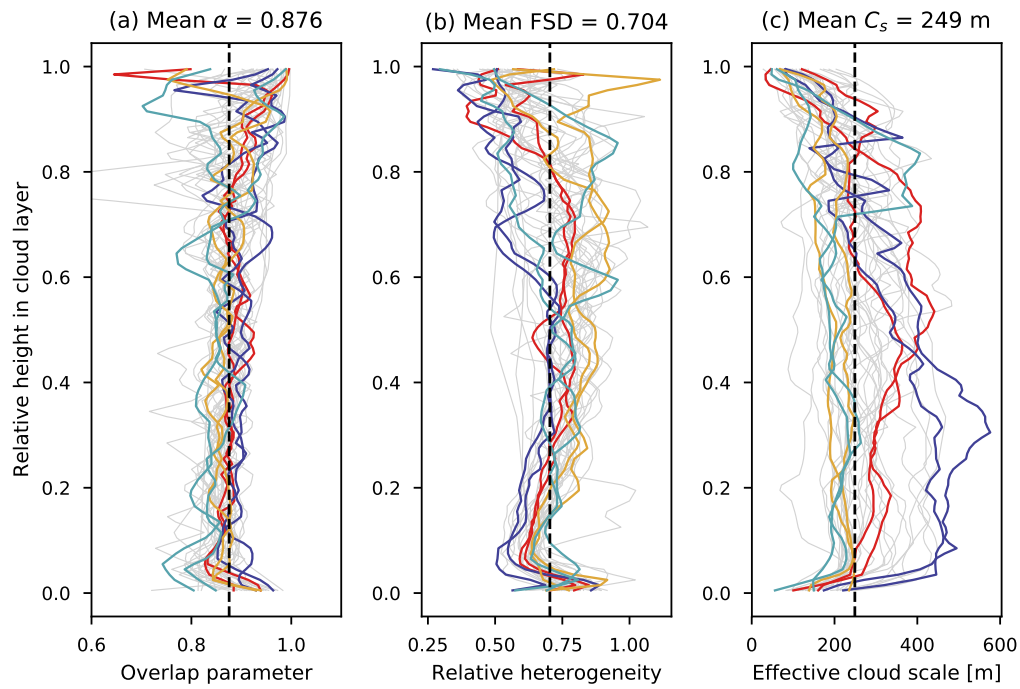
#### 377 **3.1 ecRad errors in homogeneous plane parallel clouds**

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

381 Nine fields consisting of a single-layer infinite homogeneous cloud of uniform ge-  
382 ometrical depth are synthesized with cloud optical depths of 0.1, 0.25, 0.5, 1, 2.5, 5, 25,  
383 50 and 100 (computed at wavelength 800 nm for a  $10\ \mu\text{m}$  droplet effective radius). The  
384 background atmosphere (temperature, pressure and water vapor) depends on the alti-  
385 tude only and corresponds to the I3RC cumulus case, whose 1D profiles are provided in  
386 the ecRad package.

387 MC SOCRATES and MC  $\delta$ -Eddington are compared to MC “exact” to quantify  
388 independently the errors from approximate optical properties and from approximate rep-  
389 resentation of scattering. The two-stream error is also estimated by comparing ecRad  
390 to MC “as ecRad”.

391 Barker et al. (2015) have documented these same errors as a function of optical depth  
392 and solar angle. We find similar results (not shown). Here, we only give values averaged  
393 over the fields and solar angles, that quantify the mean relative errors on the CREs at  
394 TOA and at the surface, and on cloud absorption. First, relative errors are computed  
395 only where the reference CRE is more than  $2\text{ Wm}^{-2}$  to avoid dividing by very small val-  
396 ues 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.

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

402 Small relative errors on TOA and surface CREs lead to large relative errors on the  
 403 absorption, but errors on absorption CREs only represent small amounts of energy. The  
 404 absorption errors due to the different components are of similar magnitude. The largest  
 405 source of error on surface and TOA CREs is the approximate representation of scatter-  
 406 ing, and more generally the representation of transport, while approximate optical prop-  
 407 erties have small effect. Errors due to transport and optical properties are quasi addi-  
 408 tive, while sums of errors due to approximate scattering and two-stream formulation are  
 409 larger than the global transport error because of non-linear effects.

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

### 413 3.2 Errors due to cloud geometry and 3D effects

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

423 In Figure 2, differences between ecRad and MC SOCRATES (the choice of MC SOCRATES  
 424 as a reference to measure ecRad errors is discussed in Section 4.1.2) for different met-  
 425 rics are illustrated for three configurations of ecRad: with maximum overlap, horizon-  
 426 tal homogeneity and without 3D effects (PPH max ovp); with realistic overlap and het-  
 427 erogeneity and without 3D effects (Tripleclouds); with realistic overlap and heterogene-  
 428 ity and with 3D effects (SPARTACUS). Profiles of FSD,  $\alpha$  and  $C_s$  diagnosed in the LES  
 429 fields are used to constrain the cloud geometry parameters in Tripleclouds and SPAR-  
 430 TACUS. Histograms of absolute differences between ecRad and MC estimates of upward  
 431 TOA, absorbed and downward surface fluxes are plotted for each configuration. The RMS  
 432 absolute errors are given in the legend of Figure 2 for each metric and configuration, while  
 433 the weighted averages of biases and RMS relative errors are given in Table 3.

434 The configuration that corresponds to the basic two-stream scheme with no param-  
 435 eterization of cloud geometry or 3D effects leads to the largest absolute means and largest  
 436 RMSEs for all the metrics: clouds transmit too much energy to the surface, resulting in  
 437 a lack of reflectivity and absorption. Introducing realistic overlap and horizontal hetero-  
 438 geneity systematically reduces the mean bias and changes the sign of the tail of the dis-  
 439 tribution. Realistic overlap increases the total cloud cover, which decreases transmissiv-  
 440 ity. On the contrary, introducing heterogeneities increases it. The fact that the mean bi-  
 441 ases in transmissivity and reflectivity changes sign means that the overlap effect domi-  
 442 nates the heterogeneity effect: the cloud cover controls radiative transfer at the first or-  
 443 der. Introducing 3D effects globally decreases both the distributions width and RMSEs.  
 444 The strongest biases associated with Tripleclouds estimates of the transmissivity and re-  
 445 flectivity are removed when switching to SPARTACUS, suggesting that these extreme  
 446 errors are due to the neglect of 3D effects. The biases on absorption are slightly shifted  
 447 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 |       | (v) Absorbed |       | (vi) Surf. down |       |
|---|------------------------|------------------------------------|-------------|-------|--------------|-------|-----------------|-------|
|   |                        |                                    | RMS         | bias  | RMS          | bias  | 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) $\delta$ -Eddington   | MC exact               | MC $\delta$ -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   | MC SOCRATES            | 1D, $\lambda = (1, 0, \infty)$     | 23.4        | -20.9 | 54.2         | -53.5 | 28.6            | -27.0 |
| Tripleclouds (1D)   |                        | 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 | $\bar{\lambda}(case)$ LES          | 1.4         | -0.1  | 1.6          | -0.4  | 1.4             | -0.1  |
| case-z-averaged   | $\lambda(z, case)$ LES | $\bar{\bar{\lambda}}$ 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   | MC SOCRATES            | $\bar{\bar{\lambda}}$ from htexplo | 8.3         | -2.7  | 29.1         | -28.1 | 10.2            | -7.2  |
| Best TOA up   |                        | 11.3                               | -8.5        | 33.3  | -32.6        | 14.4  | -12.8           |       |
| Best absorption   |                        | (see Table 4)                      | 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  $> 2 \text{ Wm}^{-2}$  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.

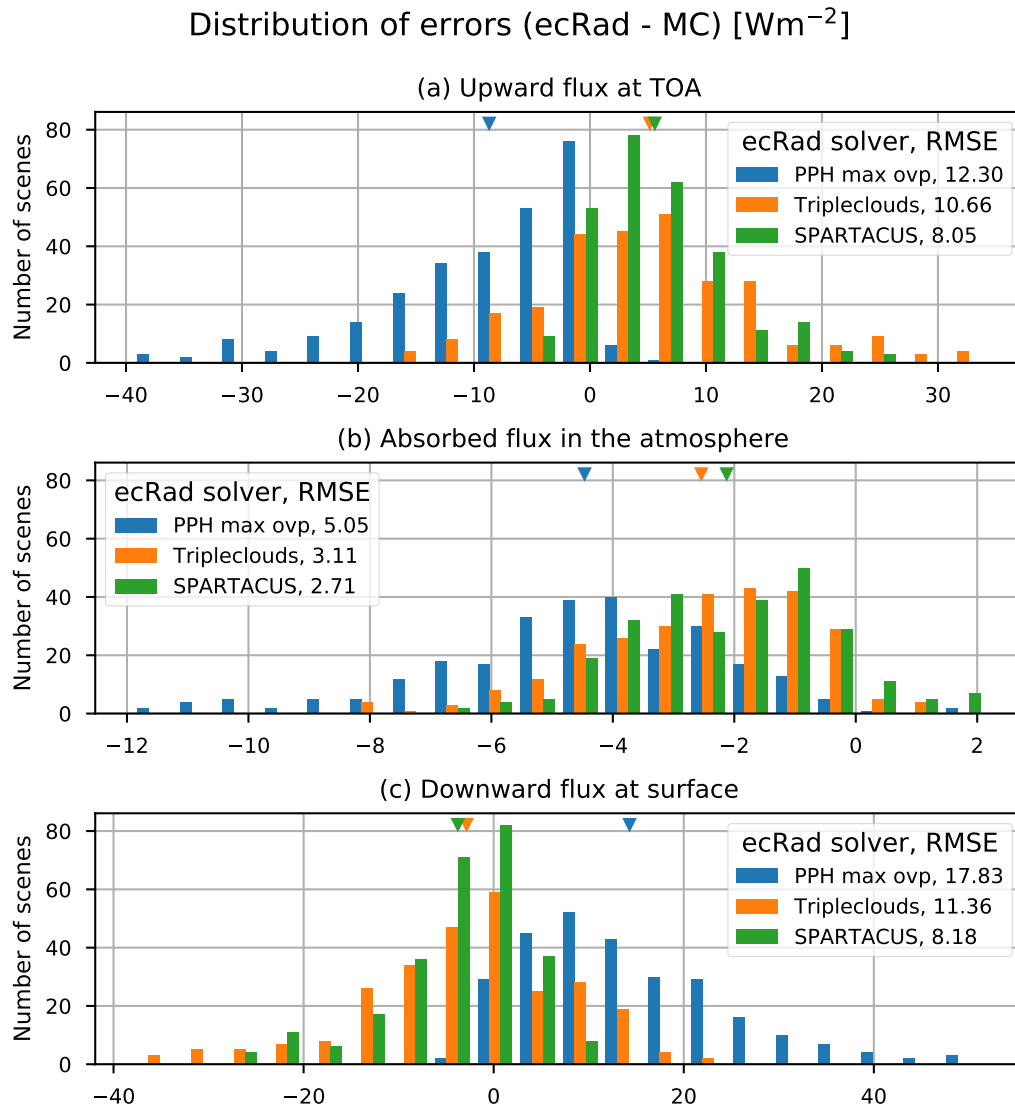
CRE = total sky - clear sky. Relative error  $r = 100 \times (\text{model-ref})/\text{ref}$ .  $\text{RMS} = \sqrt{\langle r^2 \rangle_{fields}}$ .  $\text{bias} = \langle r \rangle_{fields}$

MC exact: detailed Mie optical properties and phase function.

MC SOCRATES: parameterized optical properties and detailed Mie phase function.

MC  $\delta$ -Eddington: detailed Mie optical properties and HG  $\delta$ -Eddington phase function.

MC as ecRad: parameterized optical properties and HG  $\delta$ -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  $\times$  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  $\alpha$  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.

### 448 **3.3 Errors due to parameter variations with height and cloud scene**

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

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

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

### 466 **3.4 Remaining errors**

467 Some discrepancies remain between SPARTACUS and MC SOCRATES. They are  
468 in part due to the  $\delta$ -Eddington and two-stream approximations. Other sources of errors  
469 reside in the modeling choices for the treatment of geometry. For example, the overlap  
470 model only constrains how pairs of layers interact, but does not provide a vertically-integrated  
471 constraint. This leads to a systematic overestimation of the total cloud cover, a quan-  
472 tity that is key to the first order estimation of total transmission and reflexion. The treat-  
473 ment of in-cloud heterogeneity is also idealized in Tripleclouds: it is assumed that two  
474 cloudy regions are enough to represent the radiative effect of horizontal heterogeneity,  
475 and the distribution of the LWC into these two regions is based on hypothesis such as  
476 the LWC distribution shape and the percentile that should be used to partition the con-  
477 densate into the thin and thick regions. Other modeling assumptions and parameters  
478 that have not been discussed here also contribute to the remaining errors, such as free  
479 parameters specific to the representation of entrainment, or the degree of vertical over-  
480 lap of heterogeneities.

481 Another candidate to explain these remaining errors is of course the choice of the  
482 three geometrical parameter values. It was shown that their inter-scene and inter-height  
483 variabilities have moderate impacts on the fluxes estimate. Yet, it was all along taken  
484 for granted that the best possible choice for these parameter values was to set them to  
485 “observed” values, that is, the values deduced from the 3D cloud fields according to their  
486 physical interpretation. However, these parameters are used in ecRad as effective prop-  
487 erties, and since radiative transfer is highly nonlinear with respect to cloud geometry,  
488 there is no fundamental reason for which the mean parameters that are diagnosed in the  
489 LES should be the best choice to effectively describe cloud geometry to the radiative trans-  
490 fer parameterization. In the next section, we question this choice by exploring the be-  
491 haviour of SPARTACUS for different sets of parameter values, randomly sampled in the  
492 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:

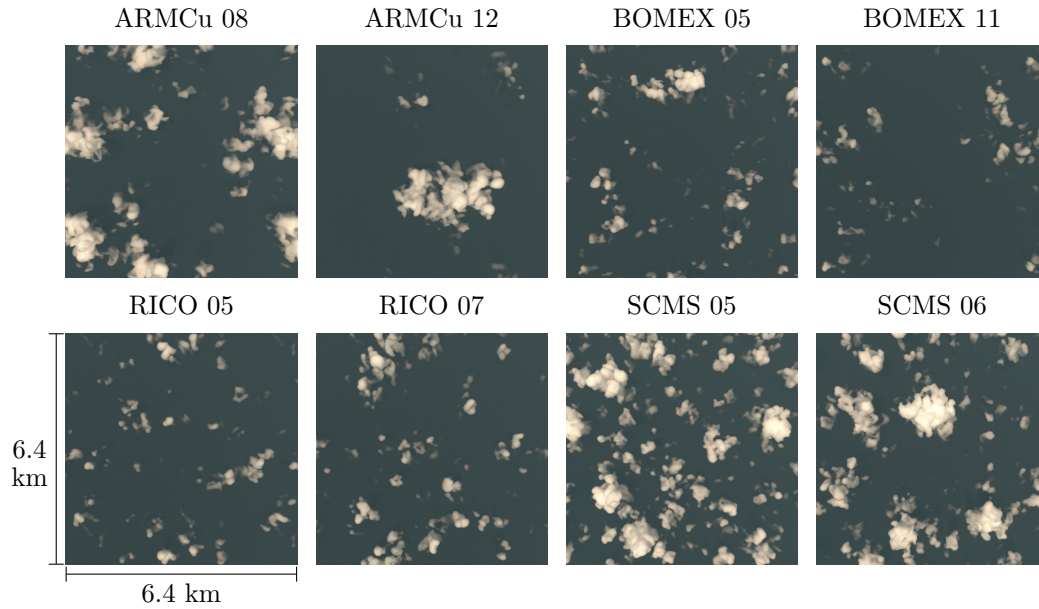
1. Select the target metrics that will serve to evaluate the model, and determine the uncertainty associated with the references.
2. Select the  $n$  parameters to calibrate. A default value and a range to explore must be provided for each parameter. The  $n$ -dimensional hypercube formed by the cartesian 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  $k^{th}$  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.

541 **4.1.2 Metrics and references to calibrate *ecRad***

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

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

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

553 Each of these nine metrics (three fluxes  $\times$  three solar angles) are computed in dif-  
 554 ferent cloud fields. Eight scenes were selected among the 35 available cumulus fields de-  
 555 scribed in Table 2. These eight scenes were chosen for their contrasted characteristics  
 556 to explore the distribution of available cumulus fields. They are represented in Figure 3  
 557 where the *htrdr* MC path-tracing tool ( ?, ? ) was applied to render nadir views of the dif-  
 558 ferent scenes, with the sun at zenith. Image rendering is useful to better visualize the  
 559 variety of combinations of cloud covers, number of clouds and cloud optical depths of  
 560 the scenes that will enter the calibration process. The images also show strong similar-  
 561 ities, highlighting the relative narrowness of the cloud fields distribution inside the cum-  
 562 ulus regime.

563 The reference values used as targets for these 72 metrics are the MC estimates of  
 564 the fluxes. The observational error is taken as the standard deviation of the MC esti-  
 565 mate, typically smaller than 0.1%. In Section 2.1.2, various options for the MC compu-  
 566 tations were exposed. Here, we use the SOCRATES data for the optical properties of

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

#### 579 **4.1.3 Parameters to calibrate**

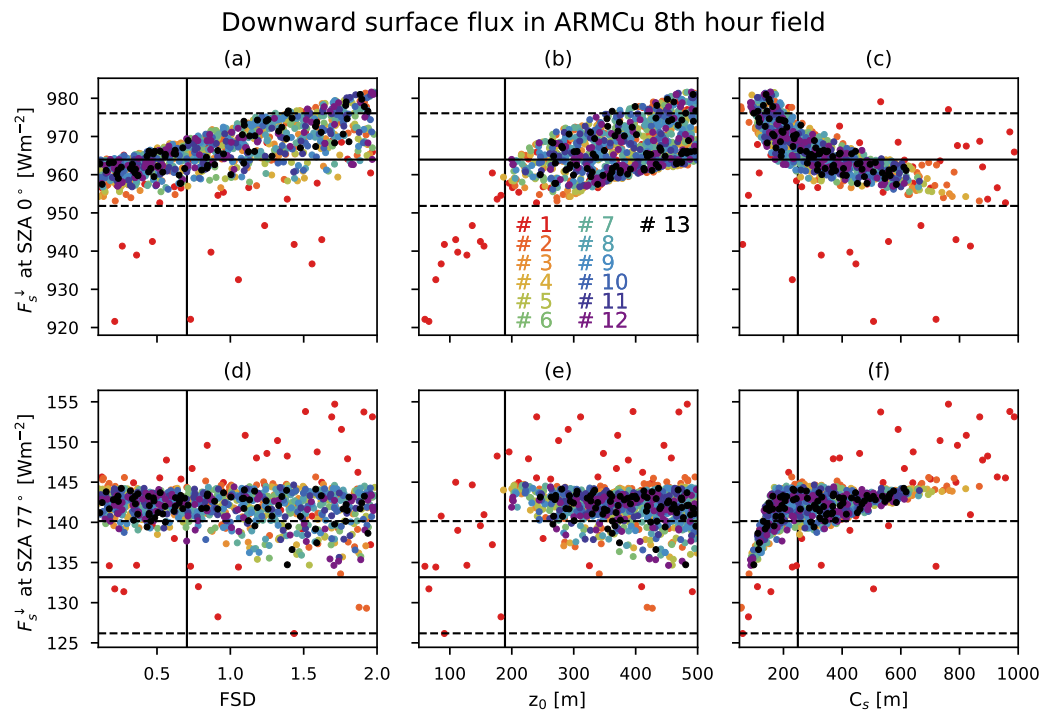
580 The parameters that enter the calibration process are the three geometrical param-  
 581 eters described in Section 2.1.1: the vertical decorrelation length  $z_0$  from the formula-  
 582 tion of the overlap parameter; the fractional standard deviation of the horizontal distri-  
 583 bution of in-cloud liquid water FSD; and the cloud scale  $C_s$ . Since we have shown be-  
 584 fore that the ecRad estimates were moderately sensitive to the vertical details of the pa-  
 585 rameter profiles, and to the inter-scene variability, we configure ecRad with a unique value  
 586 per parameter; the same parameter value is used for all the cloudy layers and all the scenes.  
 587 Other parameters of ecRad could have entered the process but our work focuses on the  
 588 modeling of cloud geometry, and these three parameters are of first importance in the  
 589 representation of geometrical effects. The mean values diagnosed in the LES (given in  
 590 Figure 1) are taken as default values:

- 591 •  $z_0$  ranges in [50, 500] with default value 189 m
- 592 • FSD ranges in [0.1, 2] with default value 0.704
- 593 •  $C_s$  ranges in [50, 1000] with default value 249 m

#### 594 **4.1.4 Tolerance to error**

595 The structural error of the parameterization in Equation (4) stems from inevitable  
 596 modeling and numerical approximations. It is most often unknown. In a sense, it is the  
 597 error that would remain after the parameters are well calibrated. However, its charac-  
 598 terisation is a prerequisite to the calibration process, as it prevents the tool from reject-  
 599 ing configurations that predict metric values within the structural error around the refer-  
 600 ence target. Since the structural error is unknown, we rather use a “tolerance to er-  
 601 ror”: an acceptable distance between the parameterization estimate and the reference  
 602 target, arbitrarily set by the modeler. Here, it is inferred from the relative errors between  
 603 MC and SPARTACUS runs using the mean LES parameter values, for each type of met-  
 604 ric and solar angle. The tolerances to error are set as the third quartile of these distri-  
 605 butions:

- 606 • for the atmospheric radiative effect at the surface ( $F_t^\downarrow - F_s^\downarrow$ ), the relative tolerances  
 607 to error are 3% for SZAs 0 and 44, and 4% for SZA 77
- 608 • for the absorbed flux in the atmosphere, the relative tolerances to error are 1%,  
 609 2% and 4% respectively for SZAs 0, 44 and 77
- 610 • for the reflected fluxes at SZAs 0, 44 and 77, the relative tolerances to error are  
 611 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)  $0^\circ$  and (d-f)  $77^\circ$ . 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.

612

## 4.2 Exploring ecRad configurations

613

In this section, htxplo 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.

616

617

### 4.2.1 Surface downward flux as a function of parameters

618

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  $0^\circ$  and  $77^\circ$  are represented for the many ecRad configurations explored during the thirteen waves of history matching.

620

622

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).

623

625

Figure 4b shows that the transmitted flux at  $0^\circ$  is strongly related to the decorrelation length, but the transmitted flux at  $77^\circ$  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

627

628

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 inversely 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  $\sim 180$  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  $0^\circ$  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

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

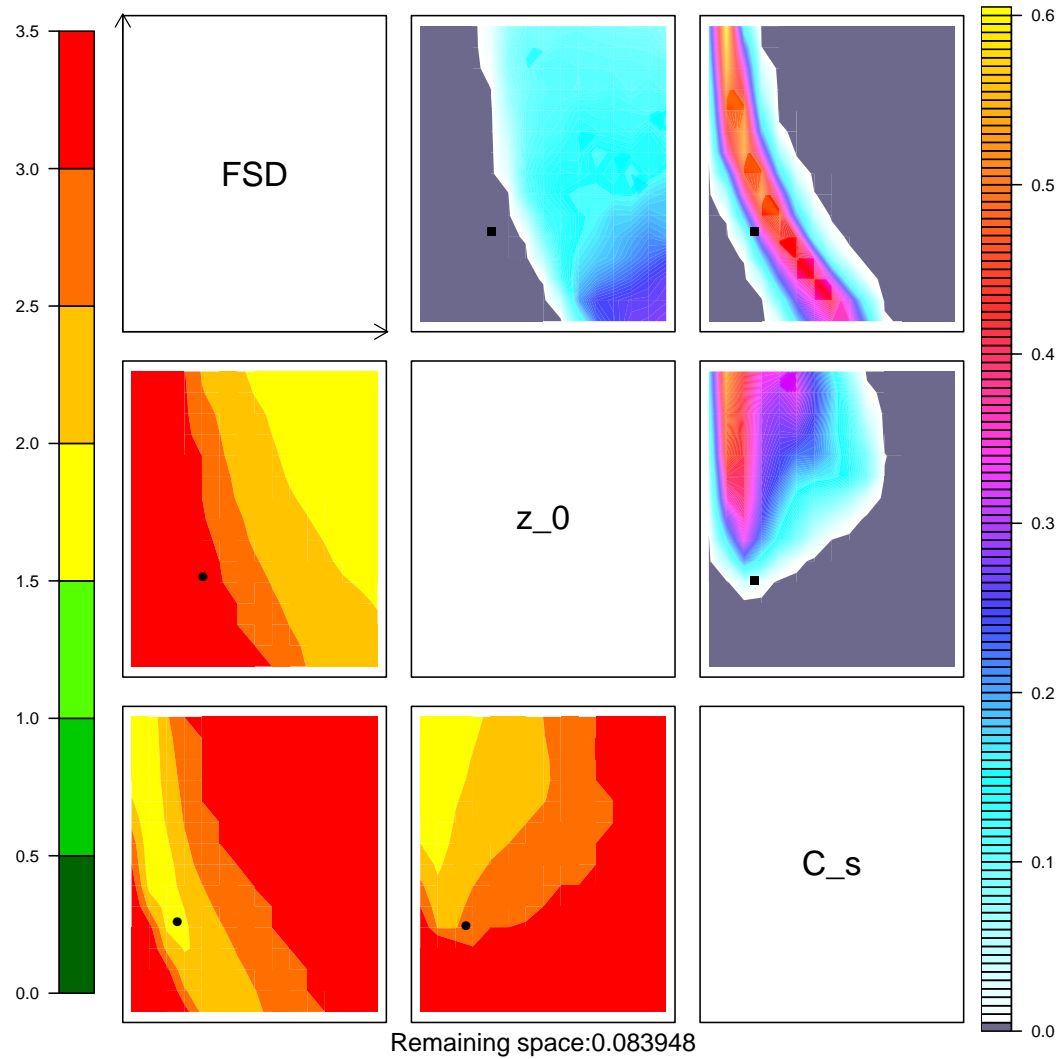
647 The three subplots that form the upper triangle of the matrix show the fraction  
 648 of points that are still in the NROY space, i.e. that verify  $I(\check{\lambda}) \leq 3$ . Each projection  
 649 shows the density of points in the NROY space in the dimension that remains once two  
 650 parameters 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 clearly 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       |

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

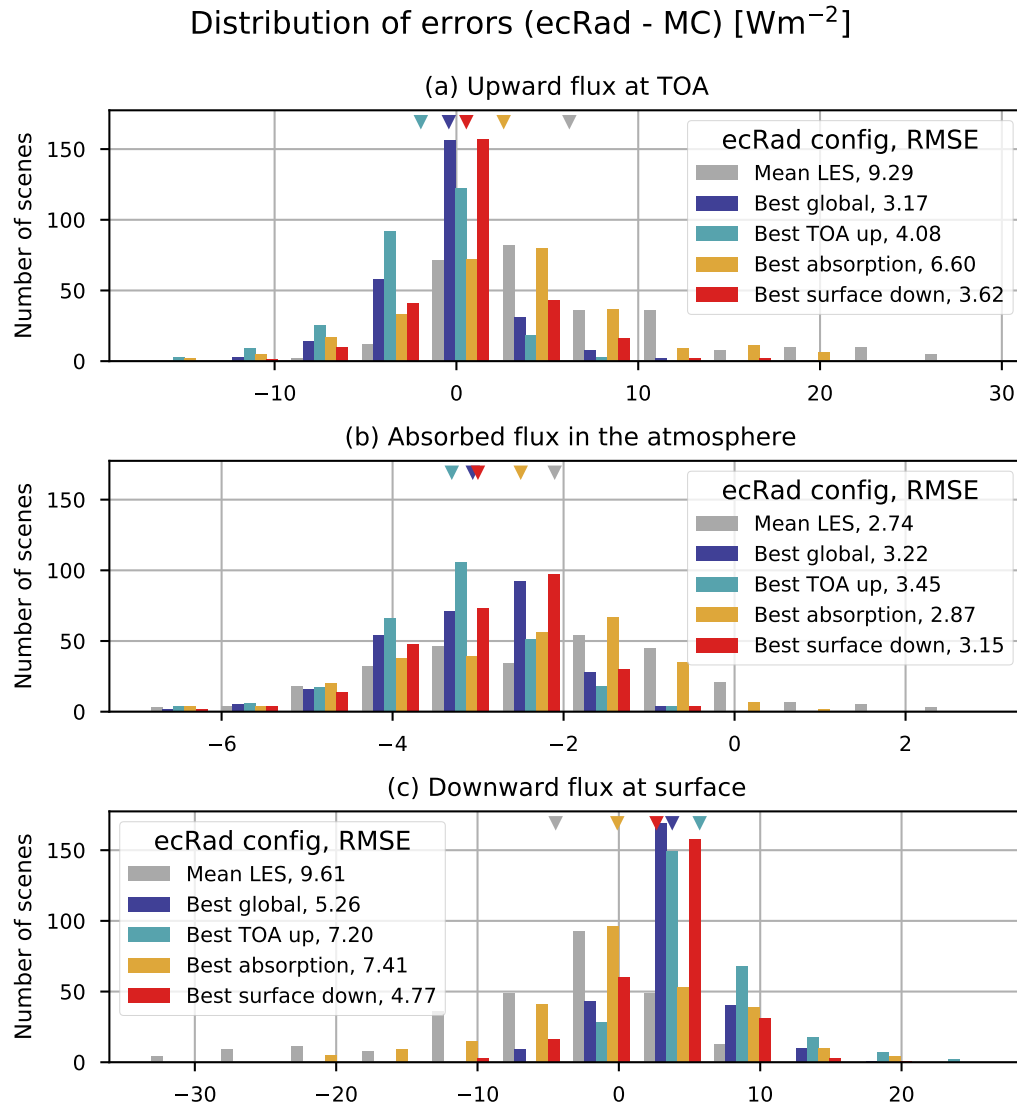
673 The quantitative information displayed in the lower triangle is strongly dependent  
 674 on the chosen tolerance to error (since it is in the expression of implausibility, see Equa-  
 675 tion 4). The variations of implausibility in the parameter space reveal more of the pa-  
 676 rameterization behaviour than the implausibility absolute values. However, the subplots  
 677 of the lower triangle show that the implausibility in the parameter space is never less than  
 678 1.5. It means that for any configuration, there is always at least one metric that is far-  
 679 ther away from its target than 1.5 times the root square sum of its uncertainties, which  
 680 is dominated by the tolerance to error at wave thirteen. It also shows that the “best”  
 681 configuration has large decorrelation lengths associated with small heterogeneous clouds,  
 682 rather than large homogeneous ones.

### 683 *4.2.3 Global improvement of fluxes using tuned configuration*

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

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

704 The fluxes at TOA and surface are systematically improved compared to the con-  
 705 figuration using the parameter values diagnosed in the LES, but all tuned configurations  
 706 are slightly worse for the absorption. The configuration corresponding to the best sur-  
 707 face score leads to the smallest global RMSE; it is the configuration with smallest and  
 708 most heterogeneous clouds. With this configuration, as well as for the two configurations  
 709 corresponding to best TOA score and best global score, the absorption bias is always neg-  
 710 ative: 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.



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

### 713 4.3 Calibration fails to reduce errors on the absorption

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

721 The reference configuration with intense 3D effects and important heterogeneity  
722 accurately reproduces the absorption dependency to solar angle but with a negative shift  
723 of 2 to 4  $\text{Wm}^{-2}$ , probably due to two-stream errors as hinted by the results of Section 3.1.  
724 Homogeneity and 1D radiation both induce errors that are larger than the “best Surf”  
725 configuration, but they are of opposite signs. Because increasing 3D effects decreases ab-  
726 sorption at low suns, the configuration with small clouds lead to even larger errors for  
727 the 77° metric. On the other hand, using error compensations from the two other pa-  
728 rameters leads to improvement for the three solar angles, even if the shape of the func-  
729 tion is wrong.

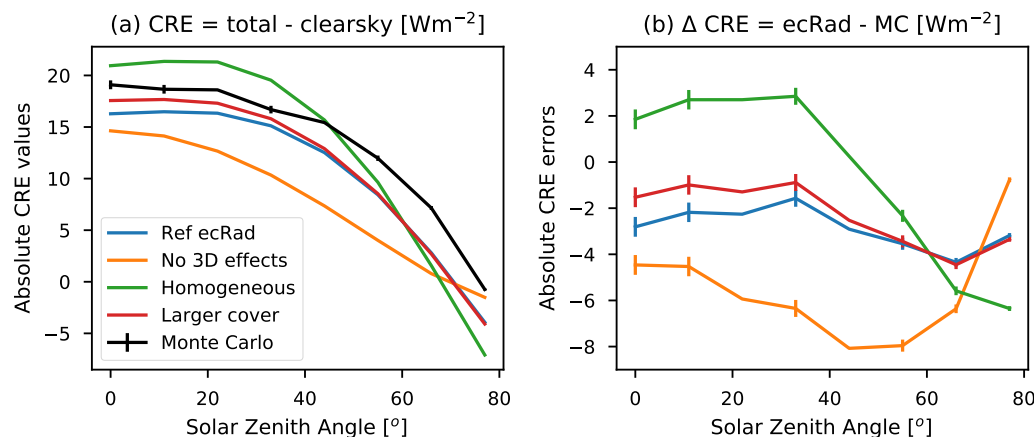
730 This highlights the importance of choosing relevant metrics for the calibration pro-  
731 cess. In particular, here, additional metrics might need to be introduced to constrain the  
732 shape of the fluxes dependency to solar angle, instead of using punctual values only. This  
733 is actually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

## 739 5 Discussions and outlooks

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

752 A key result of this study is that the parameters diagnosed in the LES fields do not  
753 belong to the final NROY space. This result strikes us as important because it questions  
754 the conceptual constraints that surround parameterization development and tuning. The  
755 main goal of parameterization development should be to derive functional forms that are  
756 able to provide accurate source terms for the explicitly resolved variables of the model  
757 over a wide range of atmospheric regimes. To achieve this, it is essential to base our de-  
758 velopments on our understanding of physical processes. However, we argue that some  
759 flexibility should be allowed in the choice of parameter values. Results reported by Bastidas

### CRE on absorbed flux in SCMS 05



**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

760 et al. (2006) and Hogue et al. (2006) also support this idea. They show that free param-  
 761 eters should be set to different values through different land surface models even though  
 762 they are supposed to have the same physical meaning. Their conclusions were limited  
 763 to so-called “functional” parameters that cannot be associated with physical measure-  
 764 ments. We argue that observational constraints on “physical” parameters should also  
 765 be alleviated. Indeed, it is most often an “effective” value of the parameters rather than  
 766 a mean observed value that is needed in the models. These can be quite different when  
 767 parameters impact the metrics in non linear ways. In the context of cloud–radiation in-  
 768 teractions, it means that the effective cloud characteristics that are appropriate to de-  
 769 rive the average radiative effect of a complex cloud field have no reason to be the detailed  
 770 characteristics averaged over the cloud population. More generally, since parameteriza-  
 771 tions are a simplification of reality, there is no fundamental reason to prefer observed pa-  
 772 rameter values. They should serve as first guesses, but the final retained values should  
 773 be the result of some model calibration. Nonetheless, a tuned value laying too far from  
 774 observations could indicate that the physical images that supported the parameteriza-  
 775 tion development are wrong or that important processes are missing.

776 Another result is that improvement of ecRad was obtained by calibrating a mean  
 777 parameter, thereby neglecting parameter variations with height and between cloud scenes.  
 778 This was probably only possible because all cloud fields used here represent cumulus clouds,  
 779 with relative resemblance between the cases, although both marine and continental clouds  
 780 were represented. An interesting follow-up would be to repeat this exercise with other  
 781 cloud types, starting with other boundary-layer clouds such as stratocumulus and transi-  
 782 tion scenes involving both cloud types. A possible diagnosis of htexpl0 might then be  
 783 that a single parameter is not able to represent different clouds. This would mean that  
 784 a sub-parameterization should be developed to make this parameter depend on atmo-  
 785 spheric conditions. Such parameterizations exist for example to predict cloud perime-  
 786 ter length in Fielding et al. (2020), or the degree of overlap in e.g. Sulak et al. (2020).  
 787 Other parameters appear in these formulations, which can in turn be calibrated using  
 788 the same procedure as described in this work.

789 A third result is that our different metrics disagree on the configuration to use. If  
 790 this disagreement had resulted in an empty NROY space for the predetermined toler-  
 791 ance, it would mean that the parameterization is deficient in parts of its expected op-  
 792 erating regime. Here, the final NROY space is not empty. However, the fact that cal-  
 793 ibration of cloud-geometry parameters fails to significantly reduce the absorption errors  
 794 suggests that either other parameters control the absorption and might deserve better  
 795 calibration, or that the formulation of radiative transfer intrinsic to the parameterization  
 796 should be revisited.

## 797 Acknowledgments

798 This work received funding from grant HIGH-TUNE ANR-16-CE01-0010. It was sup-  
 799 ported by the DEPHY2 project, funded by the French national program LEFE/INSU  
 800 and the GDR-DEPHY. Daniel Williamson was funded by NERC grant: NE/N018486/1.  
 801 Daniel Williamson and Victoria Volodina were funded by the Alan Turing Institute project  
 802 Uncertainty Quantification of multi-scale and multiphysics computer models: applica-  
 803 tions to hazard and climate models as part of the grant EP/N510129/1 made to the Alan  
 804 Turing Institute by EPSRC. All the programs, scripts, reference Monte Carlo runs and  
 805 reference LES used in the tuning process are publicly available via a Subversion through  
 806 “svn checkout <https://svn.lmd.jussieu.fr/HighTune>”; a fixed version of this code is pro-  
 807 vided under doi:XXX (place holder: IPSL data catalog). Note, however, that this tool  
 808 is a new research tool, and, as such, is still evolving. Version 1.4.0 of the ecRad pack-  
 809 age is freely available under the terms of the Apache Licence Version 2.0 at [https://](https://github.com/ecmwf/ecrad)  
 810 [github.com/ecmwf/ecrad](https://github.com/ecmwf/ecrad). This paper uses a prior version (1.3.0) available upon request  
 811 under the terms of the OpenIFS license. No major modification was made to the ecRad  
 812 radiation code between versions 1.3.0 and 1.4.0. The source code of the Monte Carlo model  
 813 that is used as a reference in this paper is available under the terms of the GNU Gen-  
 814 eral Public License at [https://gitlab.com/najdavlf/scart\\_project](https://gitlab.com/najdavlf/scart_project). The source code  
 815 of the Monte Carlo model that is used to render cloud fields is available under the terms  
 816 of GNU General Public License at <https://gitlab.com/meso-star/htrdr>. The scripts  
 817 and data that were used to produce the figures and Table 3 are available at doi:XXX (place  
 818 holder: IPSL data catalog). The authors would like to thank Igor Roffiac for extremely  
 819 fruitful discussions.

## 820 References

- 821 Barker, H. W., Cole, J. N. S., Li, J., Yi, B., & Yang, P. (2015). Estimation of er-  
 822 rors in two-stream approximations of the solar radiative transfer equation for  
 823 cloudy-sky conditions. *Journal of the Atmospheric Sciences*, *72*(11), 4053-  
 824 4074. doi: 10.1175/JAS-D-15-0033.1
- 825 Barker, H. W., Stephens, G. L., Partain, P. T., Bergman, J. W., Bonnel, B., Cam-  
 826 pana, K., ... others (2003). Assessing 1D atmospheric solar radiative transfer  
 827 models: Interpretation and handling of unresolved clouds. *Journal of Climate*,  
 828 *16*(16), 2676–2699.
- 829 Bastidas, L. A., Hogue, T. S., Sorooshian, S., Gupta, H. V., & Shuttleworth, W. J.  
 830 (2006). Parameter sensitivity analysis for different complexity land surface  
 831 models using multicriteria methods. *Journal of Geophysical Research: Atmo-*  
 832 *spheres*, *111*(D20). doi: <https://doi.org/10.1029/2005JD006377>
- 833 Bellprat, O., Kotlarski, S., Lthi, D., & Schr, C. (2012). Objective calibration of  
 834 regional climate models. *Journal of Geophysical Research: Atmospheres*,  
 835 *117*(D23). doi: <https://doi.org/10.1029/2012JD018262>
- 836 Berg, L. K., Kassianov, E. I., Long, C. N., & Mills, D. L. (2011). Surface sum-  
 837 mertime radiative forcing by shallow cumuli at the Atmospheric Radiation  
 838 Measurement Southern Great Plains site. *Journal of Geophysical Research-*  
 839 *Atmospheres*, *116*, D01202. doi: 10.1029/2010JD014593

- 840 Bony, S., Stevens, B., Frierson, D. M. W., Jakob, C., Kageyama, M., Pincus, R.,  
 841 ... Webb, M. J. (2015). Clouds, circulation and climate sensitivity. *Nature*  
 842 *Geoscience*, 8(4), 261–268. doi: 10.1038/ngeo2398
- 843 Brient, F., Couvreur, F., Villefranque, N., Rio, C., & Honnert, R. (2019). Object-  
 844 Oriented Identification of Coherent Structures in Large Eddy Simulations:  
 845 Importance of Downdrafts in Stratocumulus. *Geophysical Research Letters*,  
 846 46(5), 2854–2864. doi: 10.1029/2018GL081499
- 847 Brown, A., Cederwall, R., Chlond, A., Duynkerke, P., Golaz, J.-C., Khairoutdinov,  
 848 M., ... Stevens, B. (2002). Large-eddy simulation of the diurnal cycle of shal-  
 849 low cumulus convection over land. *Q. J. R. Meteorol. Soc.*, 128, 1075–1093.
- 850 Cahalan, R. F., Oreopoulos, L., Marshak, A., Evans, K. F., Davis, A. B., Pin-  
 851 cus, R., ... Zhuravleva, T. B. (2005). The I3RC: Bringing Together  
 852 the Most Advanced Radiative Transfer Tools for Cloudy Atmospheres.  
 853 *Bulletin of the American Meteorological Society*, 86(9), 1275–1293. doi:  
 854 10.1175/BAMS-86-9-1275
- 855 Couvreur, F., Hourdin, F., Williamson, D., Roehrig, R., Volodina, V., Villefranque,  
 856 N., ... Xu, W. (2020). Process-based climate model development harness-  
 857 ing machine learning: I. A calibration tool for parameterization improvement.  
 858 *Journal of Advances in Modeling Earth Systems*.
- 859 Dolinar, E. K., Dong, X., Xi, B., Jiang, J. H., & Su, H. (2015). Evaluation of  
 860 CMIP5 simulated clouds and TOA radiation budgets using NASA satel-  
 861 lite observations. *Climate Dynamics*, 44(7), 2229–2247. doi: 10.1007/  
 862 s00382-014-2158-9
- 863 Duan, Q., Di, Z., Quan, J., Wang, C., Gong, W., Gan, Y., ... Fan, S. (2017).  
 864 Automatic Model Calibration: A New Way to Improve Numerical Weather  
 865 Forecasting. *Bulletin of the American Meteorological Society*, 98(5), 959–970.  
 866 doi: 10.1175/BAMS-D-15-00104.1
- 867 Dufresne, J.-L., & Bony, S. (2008). An assessment of the primary sources of spread  
 868 of global warming estimates from coupled atmosphereocean models. *Journal of*  
 869 *Climate*, 21(19), 5135–5144. doi: 10.1175/2008JCLI2239.1
- 870 Fielding, M. D., Schfer, S. A. K., Hogan, R. J., & Forbes, R. M. (2020). Parametriz-  
 871 ing cloud geometry and its application in a subgrid cloud-edge erosion scheme.  
 872 *Quarterly Journal of the Royal Meteorological Society*, 146(729), 1651–1667.  
 873 doi: 10.1002/qj.3758
- 874 Golaz, J.-C., Horowitz, L. W., & Levy II, H. (2013). Cloud tuning in a coupled cli-  
 875 mate model: Impact on 20th century warming. *Geophysical Research Letters*,  
 876 40(10), 2246–2251. doi: <https://doi.org/10.1002/grl.50232>
- 877 Hinkelman, L. M., Stevens, B., & Evans, K. F. (2005). A large-eddy simulation  
 878 study of anisotropy in fair-weather cumulus cloud fields. *Journal of the Atmo-*  
 879 *spheric Sciences*, 62(7), 2155–2171. doi: 10.1175/JAS3463.1
- 880 Hogan, R. J., & Bozzo, A. (2018). A flexible and efficient radiation scheme for the  
 881 ECMWF model. *Journal of Advances in Modeling Earth Systems*. doi: 10  
 882 .1029/2018MS001364
- 883 Hogan, R. J., Fielding, M. D., Barker, H. W., Villefranque, N., & Schfer, S. A. K.  
 884 (2019). Entrapment: An Important Mechanism to Explain the Shortwave  
 885 3D Radiative Effect of Clouds. *Journal of the Atmospheric Sciences*, 76(7),  
 886 2123–2141. doi: 10.1175/JAS-D-18-0366.1
- 887 Hogan, R. J., & Illingworth, A. J. (2000). Deriving cloud overlap statistics from  
 888 radar. *Quarterly Journal of the Royal Meteorological Society*, 126(569), 2903–  
 889 2909. doi: 10.1002/qj.49712656914
- 890 Hogan, R. J., Schäfer, S. A. K., Klinger, C., Chiu, J. C., & Mayer, B. (2016, July).  
 891 Representing 3-D cloud radiation effects in two-stream schemes: 2. Matrix  
 892 formulation and broadband evaluation. *Journal of Geophysical Research:*  
 893 *Atmospheres*, 121(14), 2016JD024875. doi: 10.1002/2016JD024875
- 894 Hogan, R. J., & Shonk, J. K. P. (2013). Incorporating the Effects of 3d Radia-

- 895           tive Transfer in the Presence of Clouds into Two-Stream Multilayer Radia-  
 896           tion Schemes. *Journal of the Atmospheric Sciences*, 70(2), 708–724. doi:  
 897           10.1175/JAS-D-12-041.1
- 898 Hogue, T. S., Bastidas, L. A., Gupta, H. V., & Sorooshian, S. (2006). Eval-  
 899           uating model performance and parameter behavior for varying levels of  
 900           land surface model complexity. *Water Resources Research*, 42(8). doi:  
 901           <https://doi.org/10.1029/2005WR004440>
- 902 Hourdin, F., Mauritsen, T., Gettelman, A., Golaz, J.-C., Balaji, V., Duan, Q.,  
 903           ... Williamson, D. (2017). The art and science of climate model tun-  
 904           ing. *Bulletin of the American Meteorological Society*, 98(3), 589–602. doi:  
 905           10.1175/BAMS-D-15-00135.1
- 906 Hourdin, F., Williamson, D., Rio, C., Couvreux, F., Roebrig, R., Villefranque, N.,  
 907           ... Volodina, V. (2020). Process-based climate model development harnessing  
 908           Machine Learning. Part II: Model Calibration from Single Column to Global.  
 909           *Journal of Advances in Modeling Earth Systems*.
- 910 Iacono, M. J., Delamere, J. S., Mlawer, E. J., Shephard, M. W., Clough, S. A., &  
 911           Collins, W. D. (2008). Radiative forcing by long-lived greenhouse gases: Cal-  
 912           culations with the AER radiative transfer models. *Journal of Geophysical*  
 913           *Research: Atmospheres*, 113(D13). doi: 10.1029/2008JD009944
- 914 Jensen, J. L. W. V. (1906). Sur les fonctions convexes et les inegalits entre les valeurs  
 915           moyennes. *Acta Mathematica*, 30, 175–193. doi: 10.1007/BF02418571
- 916 Joseph, J. H., Wiscombe, W. J., & Weinman, J. A. (1976). The Delta-Eddington  
 917           Approximation for Radiative Flux Transfer. *Journal of the Atmospheric Sci-*  
 918           *ences*, 33(12), 2452–2459. doi: 10.1175/1520-0469(1976)033<2452:TDEAFR>2.0  
 919           .CO;2
- 920 Karlsson, J., Svensson, G., & Rodhe, H. (2008). Cloud radiative forcing of subtrop-  
 921           ical low level clouds in global models. *Climate Dynamics*, 30(7-8), 779–788.  
 922           doi: 10.1007/s00382-007-0322-1
- 923 Lac, C., Chaboureaud, J.-P., Masson, V., Pinty, J.-P., Tulet, P., Escobar, J., ...  
 924           Wautelet, P. (2018). Overview of the meso-nh model version 5.4 and its  
 925           applications. *Geoscientific Model Development*, 11(5), 1929–1969. doi:  
 926           10.5194/gmd-11-1929-2018
- 927 Lafore, J. P., Stein, J., Asencio, N., Bougeault, P., Ducrocq, V., Duron, J., ... de  
 928           Arellano, J. V.-G. (1997). The Meso-NH Atmospheric Simulation System. Part  
 929           I: adiabatic formulation and control simulations. *Annales Geophysicae*, 16(1),  
 930           90–109. doi: 10.1007/s00585-997-0090-6
- 931 Manners, J., Edwards, J. M., Hill, P., & Thelen, J.-C. (2017). *SOCRATES Tech-*  
 932           *nical Guide Suite Of Community RAdiative Transfer codes based on Edwards*  
 933           *and Slingo* (Tech. Rep.). FitzRoy Rd, Exeter EX1 3PB: Met Office. Re-  
 934           trieved from [http://homepages.see.leeds.ac.uk/~lecsjed/winscpuse/](http://homepages.see.leeds.ac.uk/~lecsjed/winscpuse/socrates_techguide.pdf)  
 935           [socrates\\_techguide.pdf](http://homepages.see.leeds.ac.uk/~lecsjed/winscpuse/socrates_techguide.pdf)
- 936 Marchuk, G. I., Mikhailov, G. A., Nazareliev, M. A., Darbinjan, R. A., Kargin,  
 937           B. A., & Elepov, B. S. (1980). *The Monte Carlo Methods in Atmospheric*  
 938           *Optics*. Berlin Heidelberg: Springer-Verlag.
- 939 Marshak, A., & Davis, A. (Eds.). (2005). *3D Radiative Transfer in Cloudy Atmo-*  
 940           *spheres*. Berlin Heidelberg: Springer-Verlag. (DOI: 10.1007/3-540-28519-9)
- 941 Mauritsen, T., Stevens, B., Roeckner, E., Crueger, T., Esch, M., Giorgetta, M.,  
 942           ... Tomassini, L. (2012). Tuning the climate of a global model. *Journal of*  
 943           *Advances in Modeling Earth Systems*, 4(3). doi: [https://doi.org/10.1029/](https://doi.org/10.1029/2012MS000154)  
 944           2012MS000154
- 945 Mayer, B. (2009). Radiative transfer in the cloudy atmosphere. *The European Phys-*  
 946           *ical Journal Conferences*, 1, 75–99. doi: 10.1140/epjconf/e2009-00912-1
- 947 McKee, T. B., & Cox, S. K. (1974). Scattering of Visible Radiation by Finite  
 948           Clouds. *Journal of the Atmospheric Sciences*, 31(7), 1885–1892. doi:  
 949           10.1175/1520-0469(1974)031<1885:SOVRBF>2.0.CO;2

- 950 McKee, T. B., & Klehr, J. T. (1978). Effects of Cloud Shape on Scattered Solar  
 951 Radiation. *Monthly Weather Review*, *106*(3), 399–404. doi: 10.1175/1520  
 952 -0493(1978)106<0399:EOCSOS>2.0.CO;2
- 953 Mishchenko, M., D. Travis, L., & Lacis, A. (2002). *Scattering, absorption, and emis-*  
 954 *sion of light by small particles* (Vol. 4).
- 955 Nam, C., Bony, S., Dufresne, J.-L., & Chepfer, H. (2012). The 'too few, too bright'  
 956 tropical low-cloud problem in CMIP5 models. *Geophysical Research Letters*,  
 957 *39*, L21801. doi: 10.1029/2012GL053421
- 958 Neggers, R., Duynkerke, P., & Rodts, S. (2003). Shallow cumulus convection : A  
 959 validation of large-eddy simulation against aircraft and Landsat observations.  
 960 *Quarterly Journal of the Royal Meteorological Society*, *129*, 2671-2696.
- 961 Neggers, R. A. J., Heus, T., & Siebesma, A. P. (2011). Overlap statistics of cu-  
 962 muliform boundary-layer cloud fields in large-eddy simulations. *Journal of*  
 963 *Geophysical Research: Atmospheres*, *116*(D21). doi: 10.1029/2011JD015650
- 964 Newman, W. I., Lew, J. K., Siscoe, G. L., & Fovell, R. G. (1995, 02). Systematic  
 965 Effects of Randomness in Radiative Transfer. *Journal of the Atmospheric Sci-*  
 966 *ences*, *52*(4), 427-435. doi: 10.1175/1520-0469(1995)052<0427:SEORIR>2.0.CO;  
 967 2
- 968 Pincus, R., Barker, H. W., & Morcrette, J.-J. (2003). A fast, flexible, approx-  
 969 imate technique for computing radiative transfer in inhomogeneous cloud  
 970 fields. *Journal of Geophysical Research: Atmospheres*, *108*(D13). doi:  
 971 10.1029/2002JD003322
- 972 Ramanathan, V., Cess, R. D., Harrison, E. F., Minnis, P., Barkstrom, B. R., Ah-  
 973 mad, E., & Hartmann, D. (1989). Cloud-radiative forcing and climate: Results  
 974 from the earth radiation budget experiment. *Science*, *243*(4887), 57–63. doi:  
 975 10.1126/science.243.4887.57
- 976 Salter, J. M., Williamson, D. B., Scinocca, J., & Kharin, V. (2019). Uncertainty  
 977 Quantification for Computer Models With Spatial Output Using Calibration-  
 978 Optimal Bases. *Journal of the American Statistical Association*, *114*(528),  
 979 1800-1814. doi: 10.1080/01621459.2018.1514306
- 980 Schäfer, S. A. K., Hogan, R. J., Klinger, C., Chiu, J. C., & Mayer, B. (2016, July).  
 981 Representing 3-D cloud radiation effects in two-stream schemes: 1. Longwave  
 982 considerations and effective cloud edge length. *Journal of Geophysical Re-*  
 983 *search: Atmospheres*, *121*(14), 2016JD024876. doi: 10.1002/2016JD024876
- 984 Schmidt, G. A., Bader, D., Donner, L. J., Elsaesser, G. S., Golaz, J.-C., Hannay, C.,  
 985 ... Saha, S. (2017). Practice and philosophy of climate model tuning across  
 986 six us modeling centers. *Geoscientific Model Development*, *10*(9), 3207–3223.  
 987 doi: 10.5194/gmd-10-3207-2017
- 988 Shonk, J. K. P., & Hogan, R. J. (2008). Tripleclouds: An Efficient Method for Rep-  
 989 resenting Horizontal Cloud Inhomogeneity in 1d Radiation Schemes by Using  
 990 Three Regions at Each Height. *Journal of Climate*, *21*(11), 2352–2370. doi:  
 991 10.1175/2007JCLI1940.1
- 992 Shonk, J. K. P., Hogan, R. J., Edwards, J. M., & Mace, G. G. (2010). Effect of im-  
 993 proving representation of horizontal and vertical cloud structure on the Earth's  
 994 global radiation budget. Part I: review and parametrization. *Quarterly Journal*  
 995 *of the Royal Meteorological Society*. doi: 10.1002/qj.647
- 996 Siebesma, A. P., Bretherton, C. S., Brown, A., Chlond, A., Cuxart, J., Duynkerke,  
 997 P. G., ... Stevens, D. E. (2003). A Large Eddy Simulation Intercomparison  
 998 Study of Shallow Cumulus Convection. *Journal of the Atmospheric Sciences*,  
 999 *60*(10), 1201-1219.
- 1000 Sulak, A. M., Calabrese, W. J., Ryan, S. D., & Heus, T. (2020). The contribu-  
 1001 tions of shear and turbulence to cloud overlap for cumulus clouds. *Journal of*  
 1002 *Geophysical Research: Atmospheres*, *125*(10), e2019JD032017. doi: [https://doi](https://doi.org/10.1029/2019JD032017)  
 1003 [.org/10.1029/2019JD032017](https://doi.org/10.1029/2019JD032017)
- 1004 vanZanten, M. C., Stevens, B., Nuijens, L., Siebesma, A. P., Ackerman, A. S., Bur-

- 1005 net, F., . . . Wyszogrodzki, A. (2011). Controls on precipitation and cloudiness  
1006 in simulations of trade-wind cumulus as observed during rico. *Journal of*  
1007 *Advances in Modeling Earth Systems*, 3(2).
- 1008 Várnai, T., & Davies, R. (1999). Effects of cloud heterogeneities on shortwave radi-  
1009 ation: Comparison of cloud-top variability and internal heterogeneity. *Journal*  
1010 *of the Atmospheric Sciences*, 56(24), 4206–4224.
- 1011 Villefranque, N., Fournier, R., Couvreux, F., Blanco, S., Cornet, C., Eymet, V., . . .  
1012 Tregan, J.-M. (2019). A Path-Tracing Monte Carlo Library for 3-D Radiative  
1013 Transfer in Highly Resolved Cloudy Atmospheres. *Journal of Advances in*  
1014 *Modeling Earth Systems*. doi: 10.1029/2018MS001602
- 1015 Williamson, D. (2015). Exploratory ensemble designs for environmental models  
1016 using k-extended latin hypercubes. *Environmetrics*, 26(4), 268-283. doi: 10  
1017 .1002/env.2335