Q. J. R. Meteorol. Soc. (2004), **130**, pp. 2047–2067 doi: 10.1256/qj.03.99

Stochastic generation of subgrid-scale cloudy columns for large-scale models

By PETRI RÄISÄNEN[1]*, HOWARD W. BARKER[2], MARAT F. KHAIROUTDINOV[3], JIANGNAN LI[4] and DAVID A. RANDALL[3]

[1]*Department of Physics and Atmospheric Science, Dalhousie University, Canada*
[2]*Cloud Physics Research Division, Meteorological Service of Canada, Canada*
[3]*Department of Atmospheric Science, Colorado State University, USA*
[4]*Climate Modelling and Analysis Division, Meteorological Service of Canada, Canada*

(Received 13 June 2003; revised 19 January 2004)

SUMMARY

To use the Monte Carlo Independent Column Approximation method for computing domain-average radiative fluxes in large-scale atmospheric models (LSAMs), a method is needed for generating cloudy subcolumns within LSAM columns. Here, a stochastic cloud generator is introduced to produce the subcolumns. The generator creates a cloud field on a column-by-column basis using information about layer cloud fraction, vertical overlap of cloud fraction and cloud condensate for adjacent layers, and density functions describing horizontal variations in cloud water content. The performance of the generator is assessed using a single day's worth of data from an LSAM simulation that employed a low-resolution two-dimensional cloud-resolving model (CRM) within each LSAM column (a total of ∼59 000 cloudy domains). Statistical characteristics of generated cloud fields are compared against original CRM data, and radiative-transfer biases associated with the generator are evaluated. When the generator is initialized to the greatest extent possible with information obtained from the CRM fields, overall biases are small. For example, global-mean total cloud fraction exhibits a bias of −0.004, as compared with −0.024 for maximum-random overlap (MRO) and 0.047 for random overlap. Biases in radiative fluxes and heating rates are in general $\frac{1}{4}$ to $\frac{1}{2}$ those for MRO with horizontally homogeneous clouds.

KEYWORDS: Cloud horizontal variations Cloud overlap Radiative transfer

1. INTRODUCTION

Within the current paradigm of modelling broadband radiative transfer in large-scale atmospheric models (LSAMs), unresolved variability of cloudy atmospheres is either neglected altogether or incorporated directly into one-dimensional transport solvers (e.g. Geleyn and Hollingsworth 1979; Cahalan *et al.* 1994; Oreopoulos and Barker 1999; Cairns *et al.* 2000; Li and Barker 2002; Kato 2003). As a radical alternative, Barker *et al.* (2002) and Pincus *et al.* (2003) proposed the Monte Carlo Independent Column Approximation (McICA) for estimating domain averages. McICA rests on sampling subgrid-scale columns that are unresolved by the host LSAM. The most straightforward way to apply McICA is for it to draw samples from arrays of subcolumns provided by cloud-resolving models (CRMs) that are embedded in LSAMs (Grabowski 2003; Randall *et al.* 2003). Conventional LSAMs, however, lack such information so application of McICA in them requires subgrid-scale columns to be generated using both information provided by the LSAM and additional assumptions. The objective of this paper is to present and assess a stochastic cloud generation algorithm that enables application of McICA in conventional LSAMs. This algorithm could also be useful for parametrization of other physical processes such as precipitation.

The idea of subgrid-scale cloud generation is not new. Barker *et al.* (1999) and Jakob and Klein (1999, 2000) used simple algorithms to simulate cloud fields that follow the maximum-random overlap (MRO) rule (Geleyn and Hollingsworth 1979). M. Webb and S. A. Klein have developed an International Satellite Cloud Climatology

* Corresponding author: Department of Physics and Atmospheric Science, Dalhousie University, Halifax, Nova Scotia B3H 3J5, Canada. e-mail: raisanen@mathstat.dal.ca

Project simulator* that includes options for maximum and random overlap in addition to MRO. All of these generators produce fields with horizontally homogeneous clouds. The generator presented here includes horizontal variations in cloud water and allows for a more general treatment of cloud overlap (cf. Hogan and Illingworth 2000; Bergman and Rasch 2002).

The generator is assessed using CRM data. Given the capabilities of most current computers, this can be achieved following two distinct channels: a small number of case-studies using highly resolved CRM domains; or numerous low-resolution CRM domains. The latter option is followed here. As such, eight snapshots from an LSAM simulation of a single day were used in which the LSAM ran a low-resolution two-dimensional (2D) CRM in each of its 8192 columns (Khairoutdinov and Randall 2001). Cloud property and radiative-flux errors associated with the most complete form of the generator are contrasted against those for the plane-parallel, homogeneous MRO scheme because of its current popularity among operational LSAMs (cf. Barker *et al.* 2003).

The generator is presented in section 2. Section 3 introduces the global dataset used to assess the generator. The radiative-transfer model used in the tests is described in section 4. Section 5 presents results pertaining to the performance of the generator and radiative biases that stem from it. Needs for further research and conclusions are presented in section 6. Potential extensions of the generator are discussed in the appendices.

2. Stochastic generation of subgrid-scale cloudy columns

The basic hypothesis upon which this generator is built is as follows: for computation of domain-average radiative fluxes in LSAMs, horizontal correlations in unresolved cloud structure are unimportant. This implies that the Independent Column Approximation (ICA) is adequate, and thus allows for independent generation of subcolumns. Moreover, it is assumed that each layer of each subcolumn is either filled with or free of cloud. Therefore, cloud fraction $c_{j,k}$ in the kth layer of the jth subcolumn is either 0 or 1. It is further assumed that the atmosphere is discretized into K layers, the uppermost being $k = 1$, and that $\mathcal{N}$ subcolumns are to be generated.

At the most basic level, the cloud generator is an algorithm that determines $x_{j,k} \in [0, 1]$ and $y_{j,k} \in [0, 1]$ for each subcolumn j and layer k. Properties of the generated fields depend on how $x_{j,k}$ and $y_{j,k}$ are determined as well as on conditions provided by the LSAM. The variable $x_{j,k}$ is used to decide whether the cells of subcolumns are cloudy or clear according to

$$c_{j,k} = \begin{cases} 0, & \text{for } x_{j,k} \leq 1 - C_k \text{ (clear)}, \\ 1, & \text{for } x_{j,k} > 1 - C_k \text{ (cloudy)}, \end{cases} \qquad k = k_{\text{top}}, \ldots, k_{\text{base}};\ j = 1, \ldots, \mathcal{N}; \quad (1)$$

where C_k is cloud fraction for the kth layer as provided by the LSAM, and k_{top} and k_{base} are the uppermost and lowermost layers with $C_k > 0$. The variable $y_{j,k}$ is used to determine condensate amount $w_{j,k}$ for cloudy cells following

$$y_{j,k} = \int_0^{w_{j,k}} p_k(w)\, \mathrm{d}w, \quad \forall (j, k) : c_{j,k} = 1, \qquad (2)$$

where $p_k(w)$ is a normalized probability density function for total (= liquid + ice) cloud water content w across the cloudy portion of layer k. Hence, $y_{j,k}$ is the cumulative

* http://gcss-dime.giss.nasa.gov/simulator.html

frequency distribution of w. The exact form of $p_k(w)$ is of secondary importance; it can be either non-analytic (i.e. tabulated), or idealized such as a gamma, beta, or log-normal distribution (in either complete or incomplete form).

The following subsections describe two distinct algorithms for setting $x_{j,k}$ and $y_{j,k}$. The first is referred to as the 'full generator', or FGen for short. It allows for a continuous range of cloud overlap rates as well as horizontal variations and vertical correlations in w. FGen is assessed in section 5 and recommended for use in LSAMs. The second generates horizontally homogeneous cloud fields that follow the MRO rule. These assumptions are currently very popular with LSAM groups (cf. Barker *et al.* 2003). Errors inherent to FGen are put into context by contrasting them with errors associated with MRO.

(*a*) *Full generator*

(i) *Cloud fraction and cloud overlap.* Hogan and Illingworth (2000) and Bergman and Rasch (2002) have used a generalized form of cloud overlap in which vertically projected cloud fraction $C_{k,l}$ formed by any two layers k and l is described by a linear combination of maximum and random overlap such that

$$C_{k,l} = \alpha_{k,l} C_{k,l}^{\max} + (1 - \alpha_{k,l}) C_{k,l}^{\mathrm{ran}}, \tag{3}$$

where

$$C_{k,l}^{\max} = \max(C_k, C_l), \tag{4}$$

$$C_{k,l}^{\mathrm{ran}} = C_k + C_l - C_k C_l, \tag{5}$$

and $\alpha_{k,l}$ is the overlap parameter. It is straightforward to implement this overlap model in the generator subject to two additional assumptions, the first of which is that overlap is always between random and maximum. This implies that $\alpha_{k,l} \in [0, 1]$. Second, it is assumed that overlap can be described adequately by applying (3) successively to adjacent layers and thus working directly with $\alpha_{k-1,k}$ only so that for non-adjacent layers the overlap parameter is defined implicitly as

$$\alpha_{k,l} = \prod_{m=k+1}^{l} \alpha_{m-1,m}. \tag{6}$$

Analyses of observations by Hogan and Illingworth (2000) and Mace and Benson-Troth (2002) suggest that a useful description of $\alpha_{k,l}$ (Bergman and Rasch 2002) is

$$\alpha_{k,l} = \exp\left(-\int_{z_k}^{z_l} \frac{\mathrm{d}z}{L_{\mathrm{cf}}(z)}\right), \tag{7}$$

where L_{cf} is decorrelation depth for overlapping fractional cloud, and z is altitude. Both assumptions made in the previous paragraph are perfectly consistent with (7). In actuality, negative correlations between adjacent layers (i.e. $\alpha_{k-1,k} < 0$) could be included, but the added complexity was judged unwarranted since observations indicate that $\alpha_{k,l} < 0$ occurs much less frequently for adjacent layers than for distant ones. Similarly, accounting explicitly for correlations between distant layers would represent a large jump in complexity relative to the simple approach used here.

Beginning at k_{top}, this algorithm determines $x_{j,k}$, by first assigning to each subcolumn

$$x_{j,k_{\mathrm{top}}} = \mathrm{RN1}_{j,k_{\mathrm{top}}}, \quad j = 1, \ldots, \mathcal{N}, \tag{8}$$

where $\mathrm{RN1}_{j,k_{\mathrm{top}}}$ is a unique random number distributed evenly between 0 and 1. For subsequent layers,

$$x_{j,k} = \begin{cases} x_{j,k-1}, & \text{for } \mathrm{RN2}_{j,k} \leqslant \alpha_{k-1,k}, \\ \mathrm{RN3}_{j,k}, & \text{for } \mathrm{RN2}_{j,k} > \alpha_{k-1,k}, \end{cases} \quad k = k_{\mathrm{top}}+1, \ldots, k_{\mathrm{base}};\ j = 1, \ldots, \mathcal{N}; \tag{9}$$

where $\mathrm{RN2}_{j,k}$ and $\mathrm{RN3}_{j,k}$ are new random numbers with properties like $\mathrm{RN1}_{j,k_{\mathrm{top}}}$. Note that algorithms for simulating random overlap ($\alpha = 0$) and maximum overlap ($\alpha = 1$) are special cases of (9). For random overlap $x_{j,k}$ are generated for all cells using (8), while for maximum overlap $x_{j,k_{\mathrm{top}}}$ is used for all layers in the jth subcolumn.

(ii) *Cloud condensate amount.* The algorithm used to generate vertical profiles of $y_{j,k}$ resembles closely that for $x_{j,k}$. For the uppermost cloudy layer

$$y_{j,k_{\mathrm{top}}} = \mathrm{RN4}_{j,k_{\mathrm{top}}}, \quad j = 1, \ldots, \mathcal{N}, \tag{10}$$

while for subsequent layers

$$y_{j,k} = \begin{cases} y_{j,k-1}, & \text{for } \mathrm{RN5}_{j,k} \leqslant r_{k-1,k}, \\ \mathrm{RN6}_{j,k}, & \text{for } \mathrm{RN5}_{j,k} > r_{k-1,k}, \end{cases} \quad k = k_{\mathrm{top}}+1, \ldots, k_{\mathrm{base}};\ j = 1, \ldots, \mathcal{N}; \tag{11}$$

where $r_{k-1,k}$ is the linear correlation between cloud condensate cumulative frequencies y for the *overlapping portions* of two adjacent cloudy layers. Thus, the algorithm simulates clouds with specified correlation $r_{k-1,k}$ in condensate amount by combining perfect correlation for a fraction $r_{k-1,k}$ of cloudy cells with no correlation for the remainder.

Three assumptions are involved in this algorithm. As with cloud fraction, it is assumed that $r_{k-1,k} \in [0, 1]$ and that, for non-adjacent layers k and l, condensate correlation is defined implicitly as

$$r_{k,l} = \prod_{m=k+1}^{l} r_{m-1,m}. \tag{12}$$

Observations (Hogan and Illingworth 2003) suggest that r can be approximated well by (7) but, in general, using a different decorrelation depth than that for cloud fraction.

The third assumption is that non-overlapping and overlapping portions of cloud have the same $p_k(w)$. Hence, any potential conditionalities between cloud geometry and distributions of w are neglected. In actuality, one would expect non-overlapping portions to favour smaller values of w because of mixing with clear air, and overlapped portions to favour larger values.

Presently, the cloud generator assumes (as do LSAM radiation codes) that, when liquid and ice hydrometeors co-exist in a layer, the ratio of liquid to ice condensate amounts, $w_{\mathrm{l}}/w_{\mathrm{i}}$, is horizontally invariant. This is an excellent approximation for the dataset used here to assess the generator, because w_{l} and w_{i} were diagnosed from prognostic total cloud water. Alternative methods that include horizontal variations in cloud phase were experimented with but are not reported here given our inability to justify them.

Generated cloud condensate amounts could be used in parametrization of effective cloud droplet and ice crystal sizes (see appendix A). Constant values are assumed, however, in the main part of this paper (see section 4).

(*b*) *Maximum-random overlap generator*

The magnitude of errors inherent to the full generator is put into context by contrasting them with errors associated with MRO using plane-parallel, homogeneous clouds. Following Geleyn and Hollingsworth's (1979) version of MRO, total cloud fraction for layers k through l, inclusive, is

$$C_{k\to l} = 1 - (1 - C_k) \prod_{m=k+1}^{l} \frac{1 - \max(C_{m-1}, C_m)}{1 - C_{m-1}}. \tag{13}$$

Beginning again with (8) for $k = k_{\text{top}}$, profiles of $x_{j,k}$ that correspond to (13) are generated stochastically as

$$x_{j,k} = \begin{cases} x_{j,k-1}, & \text{for } x_{j,k-1} > 1 - C_{k-1} \text{ (cloudy cell above)}, \\ \text{RN}_{j,k}(1 - C_{k-1}), & \text{for } x_{j,k-1} \leqslant 1 - C_{k-1} \text{ (cloudless cell above)}, \end{cases} \tag{14}$$

where $k = k_{\text{top}} + 1, \ldots, k_{\text{base}}$ and $\text{RN}_{j,k}$ are random numbers like those used thus far.

Because horizontally homogeneous clouds are assumed, mean cloud water content for the cloudy part of an LSAM layer is assigned to each cloudy cell in that layer. Thus, there is no need to generate $y_{j,k}$.

(*c*) *An example using a single CRM domain*

Once values of $x_{j,k}$ and $y_{j,k}$ are determined, they are run through (1) and (2) which yields the subgrid-scale fields of cloud fraction and water content. Since random numbers are used, both the FGen and MRO version of the generator produce a distribution of fields for any set of input conditions. This point is demonstrated here for a single 2D CRM field with 64 subcolumns belonging to the global climate model (GCM) dataset described in section 3. A more comprehensive assessment of the generator is given in section 5.

Figure 1(a) shows cloud occurrence for the 2D CRM field. Since generated fields lack horizontal structure, columns are reordered to facilitate visual comparison of inherent and generated fields. Figure 1(b) shows one way of reordering the 64 columns of this field according to cloud-top height. Note that application of the ICA to these fields produces identical domain-average radiative-flux profiles. Using information from this field to initialize FGen and MRO, Figs. 1(c) and (d) show corresponding single realizations of 64 subcolumns (reordered according to cloud-top height).

To demonstrate the variability of results, 5000 realizations were generated for both FGen and MRO, using either 64 or 640 subcolumns per realization. Figures 1(e)–(g) show resulting frequency distributions of total cloud fraction C_{tot}, cloudy-area mean cloud water path $\overline{W}$, and its relative standard deviation σ_W (i.e. standard deviation divided by mean). As expected, breadths of distributions decrease as the number of generated subcolumns $\mathcal{N}$ increases. While mean values of these distributions are essentially independent of $\mathcal{N}$, they do not (in general) equal corresponding CRM means. In this particular case, the FGen underestimates C_{tot} and overestimates both $\overline{W}$ and σ_W. With cloudy layers spanning a significant vertical extent, this tropical convective domain represents a more demanding test than most others in the GCM dataset.

By its very nature, the MRO model usually underestimates C_{tot} and overestimates $\overline{W}$ substantially when there are several contiguous, partly cloudy layers such as, for the example, in Fig. 1. In the majority of cases, MRO's assumption of horizontal homogeneity can be expected to produce underestimates of σ_W. These underestimates

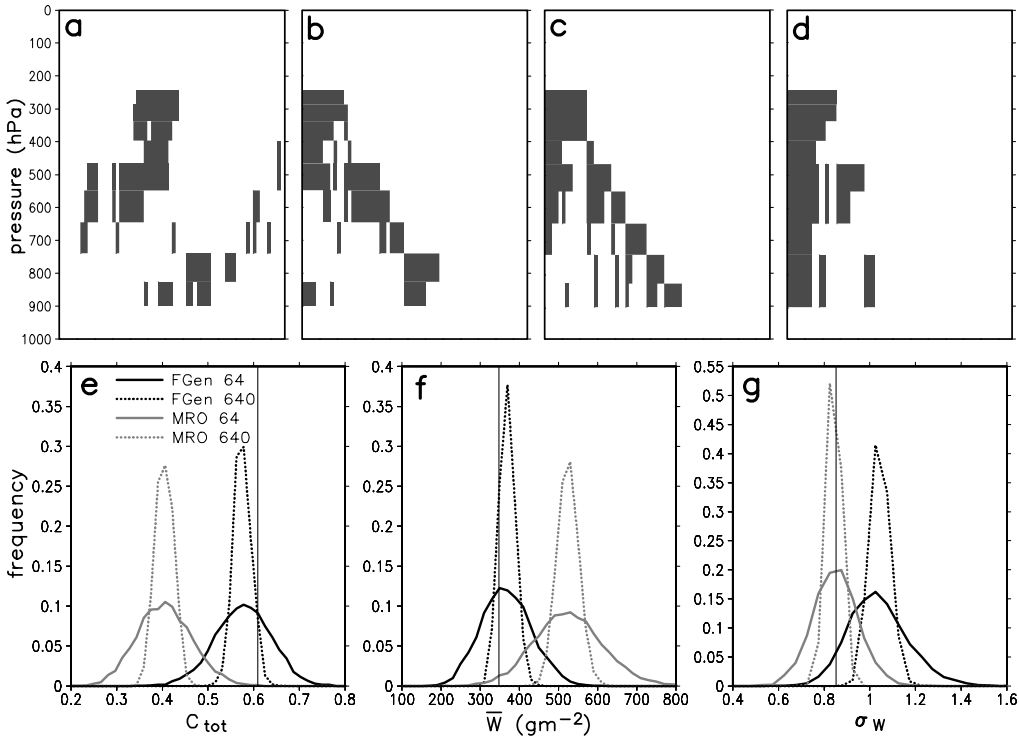

Figure 1. (a) Binary map of a two-dimensional cloud field from a cloud-resolving model (CRM) for a single global climate model column representing tropical convection at 7°S, 59°W. (b) The same field reordered according to cloud-top height. A single corresponding realization generated (c) by the full generator (FGen) and (d) by maximum-random overlap (MRO), reordered according to cloud-top height. (e)–(g) Frequency distributions of total cloud fraction $C_{\rm tot}$, mean cloudy-area cloud water path $\overline{W}$, and its relative standard deviation σ_W corresponding to 64 and 640 subcolumns per realization for FGen and MRO. Vertical lines indicate $C_{\rm tot}$, $\overline{W}$, and σ_W for the CRM cloud field.

maximize for columns that contain a single inhomogeneous cloud layer. For the multi-layer cloud field shown in Fig. 1, however, MRO produces fortuitously an excellent estimate of σ_W.

3. Global GCM/CRM dataset and initialization of the generator

(*a*) *Global GCM/CRM dataset*

The stochastic generator is assessed using data obtained from a GCM simulation that employed a low-resolution 2D CRM in each GCM column (Khairoutdinov and Randall 2001). The CRM domains consist of 24 layers on an uneven grid, and 64 columns aligned east-to-west with a grid spacing of ∼4 km (GCM grid spacing was 2.8°). Eight global snapshots from one day (1 January) are used; each separated by 3 hours, thereby constituting one diurnal cycle. This totals 65 536 GCM columns, of which ∼59 000 contain some cloudiness.

The domains range from tropical deep convection (with and without an anvil) to stratiform polar clouds, so the generator is exposed to a wide range of conditions, but not all conditions (e.g. marine boundary-layer clouds are represented poorly). While the 2D CRM framework is known to simulate deep convection well (e.g. Rotunno *et al.* 1988; Grabowski *et al.* 1998), it is not obvious that the same is true for other cloud systems.

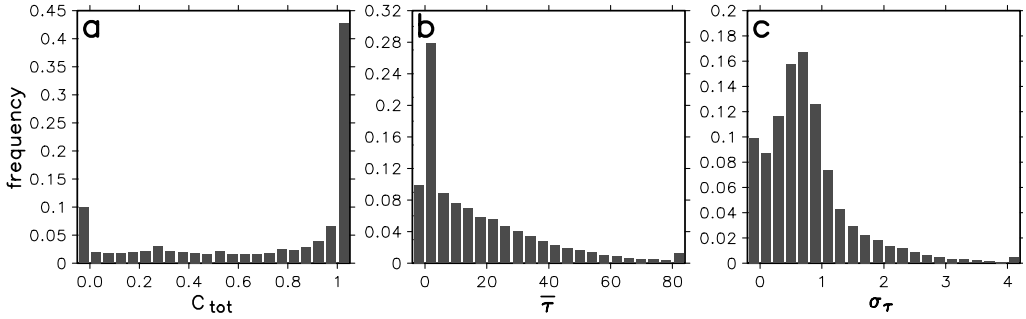

Figure 2. Frequency distributions of (a) total cloud fraction $C_{\rm tot}$, (b) mean cloud visible optical depth $\overline{\tau}$, and (c) relative standard deviation of optical depth, σ_τ, in cloud-resolving model data. The first column in each frame represents cloud-free cases ($C_{\rm tot} = 0$), while the last column represents $C_{\rm tot} = 1$ in (a), $\overline{\tau} > 80$ in (b), and $\sigma_\tau > 4$ in (c).

Another limitation may be vertical resolution which varies from ~130 m in the lowest layer to ~1.2 km in the middle and upper troposphere. While this grid spacing typifies many GCM simulations, it exceeds the geometric thickness of many real-world clouds. In particular, vertical structure of boundary-layer clouds cannot be resolved properly.

For this study it is desirable that the GCM/CRM fields bear at least some semblance to reality, but it is more important that they vary among themselves and that at least a subset of them possess considerable internal variability and thus provide a wide range of demanding tests. Figure 2 shows frequency distributions of $C_{\rm tot}$, mean cloud visible optical depth $\overline{\tau}$, and relative standard deviation of optical depth, σ_τ, for the GCM cells as computed from instantaneous CRM subcolumns. About half the GCM columns are partially cloudy, and the majority of the remainder are overcast. The distribution of $\overline{\tau}$ peaks for $\overline{\tau} < 4$, but 33% of the columns have $\overline{\tau} > 20$, and many of them are quite variable with $\sigma_\tau > 1$. For further assessments of the clouds produced by this GCM, see Khairoutdinov and Randall (2001).

Figure 3 displays effective zonal-mean values of decorrelation depth for cloud fraction and cloud condensate ($L_{\rm cf}$ and $L_{\rm cw}$, respectively) for adjacent layers. It follows from (7) that

$$\langle L_{\rm cf}\rangle = -\langle \Delta z\rangle / \ln\langle \alpha\rangle, \tag{15}$$

$$\langle L_{\rm cw}\rangle = -\langle \Delta z\rangle / \ln\langle r\rangle, \tag{16}$$

where $\langle \cdots \rangle$ denotes zonal mean, Δz is layer midpoint distance, and $\langle \alpha\rangle$ and $\langle r\rangle$ are mean values of α and r for the latitude band and layer pair considered. For a pair of layers k and l, $\langle \alpha_{k,l}\rangle$ is defined in such a way that when $\langle \alpha_{k,l}\rangle$ is applied to all GCM columns in a latitude band, correct zonal-mean cloud fraction $\langle C_{k,l}\rangle$ is obtained. When computing $\langle r_{k,l}\rangle$, cloud fractions for the overlapping part of layers k and l are used as weighting factors.

In most of the troposphere, $\langle L_{\rm cf}\rangle$ = 1–5 km, but values as large as 10 km occur in the uppermost troposphere, and values <0.5 km occur near the surface at low latitudes. Decorrelation depth for cloud condensate is about half that for cloud fraction. This agrees roughly with observational evidence from Hogan and Illingworth (2003) and other model results (R. Pincus 2003, personal communication).

(*b*) *Initializing the generator*

Both FGen and MRO are initialized with CRM data to the greatest extents possible. As such, attention is focused on the generators' capabilities rather than on quality of

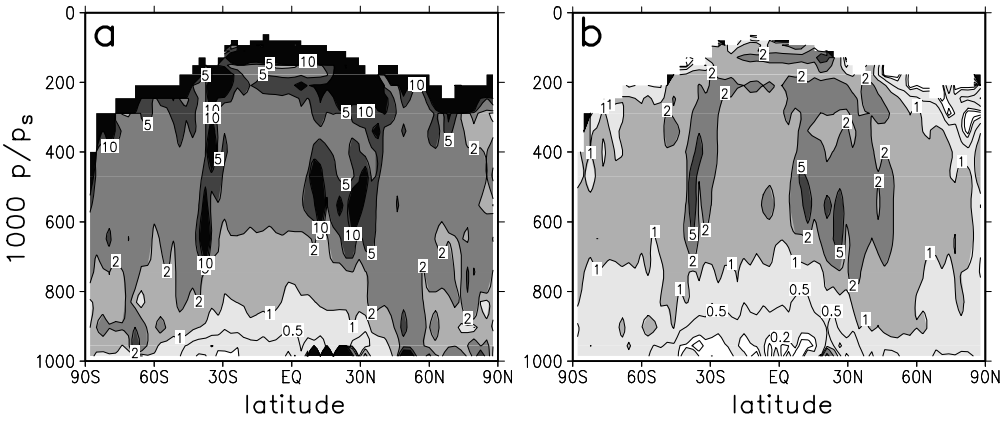

Figure 3. Zonal-mean decorrelation depths (km) for (a) cloud fraction $\langle L_{\rm cf} \rangle$ and (b) cloud condensate $\langle L_{\rm cw} \rangle$ for cloud-resolving model data used in this study.

input data. First, profiles of layer cloud fraction and cloud fraction overlap parameter $\alpha_{k-1,k}$ for adjacent layers are derived directly from CRM data. Likewise, vertical correlations $r_{k-1,k}$ of the cumulative frequency of cloud water content w are defined using CRM data by first computing the cumulative frequency of w for each cloudy cell in layers $k-1$ and k ($y_{j,k-1}$ and $y_{j,k}$). Then, isolating those subcolumns that have cloud in both layers, and using these cells only, linear correlation coefficients $r_{k-1,k}$ between $y_{j,k-1}$ and $y_{j,k}$ are determined. Following this, cumulative frequency distributions of w are tabulated and employed to convert stochastically generated values of $y_{j,k}$ into $w_{j,k}$.

Naturally, use of perfect input data overestimates the accuracy achievable using FGen in an ordinary LSAM environment where overlap parameters α and r and probability distributions of cloud condensate $p(w)$ are not readily available. However, parametrization of these quantities falls beyond the scope of this paper. Furthermore, results for both FGen and MRO neglect uncertainty in estimates of layer cloud fraction and condensate amount which is bound to infest LSAM simulations.

4. Radiative-transfer calculations

Short-wave (SW) fluxes are computed by a multi-layer two-stream approximation (e.g. Coakley *et al.* 1983), while for long-wave (LW) fluxes an emissivity-type approach with corrections for scattering is used (Li 2002). Gaseous transmittances (for H_2O, CO_2, O_3, CH_4, N_2O, $CFCl_3$ and CF_2Cl_2) are computed using the correlated-k distribution method with $K = 31$ quadrature points in cumulative probability space for the SW and $K = 46$ for the LW. Optical properties for liquid droplets (Dobbie *et al.* 1999; Lindner and Li 2000) and ice crystals (Fu 1996; Fu *et al.* 1998) are resolved into four SW bands and nine LW bands. As the CRM fields contain no information about particle size, constant values are used: an effective radius of $r_{\rm e} = 10$ μm for liquid clouds and a generalized effective diameter of $D_{\rm ge} = 50$ μm (Fu 1996) for ice clouds. Aerosol effects are also included in SW calculations. The SW component of this code participated in a recent intercomparison of radiative-transfer models (Barker *et al.* 2003) and is in accord with most other models in active service. For more details of the LW code, see Li (2002).

Temperature and specific humidity varied horizontally in the CRM fields but, for the radiation calculations performed here, they were averaged horizontally separately

for cloudless and cloudy cells. Ozone profiles came from McClatchey *et al.* (1971) and broadband SW surface albedos came from the Météo France dataset used by Räisänen (1999). In the LW, all surfaces were treated as black bodies. Horizontal and vertical distributions of four aerosol types (continental, maritime, urban, and stratospheric background) were prescribed based on Tanre *et al.* (1984).

The ICA is used to compute domain-average radiative fluxes. In the reference calculations, the subcolumns needed for ICA are taken directly from the CRM fields; otherwise they are generated stochastically.

5. Verification of the generator

This section has two subsections. The first assesses characteristics of cloud fields generated by FGen, while the second assesses spurious radiative forcings that arise from FGen fields. The focus is on the validity of FGen's assumptions about cloud field structure, and so all quantities are estimated using 10 000 generated subcolumns per cloud field (i.e. GCM column). This reduces sampling noise almost to zero.

(*a*) *Comparison of original and generated cloud fields*

The following five assumptions in FGen are capable of causing differences between generated cloud fields and the original CRM fields:

(i) Overlap parameter for cloud fraction α can be computed from (6) for non-adjacent layers;

(ii) $\alpha \geq 0$ for adjacent layers (and consequently all layer pairs);

(iii) overlap parameter for cloud condensate r can be computed from (12) for non-adjacent layers;

(iv) $r \geq 0$ for adjacent layers (and consequently all layer pairs);

(v) probability distributions of cloud condensate are independent of cloud geometry (e.g. they are identical for overlapping and non-overlapping cloud cells).

To address the impact of assumptions (i) and (ii), Fig. 4 shows global-mean α (denoted as $[\alpha]$) as a function of vertical distance expressed in model layers. On average, a distance of M layers corresponds to $\sim M$ km, even though model vertical resolution varies. As expected, $[\alpha]$ decreases as a function of increasing vertical distance so that distant layers are essentially randomly overlapped. The rate of decrease is, however, slower for generated cloud fields than for CRM data. The generator also overestimates $[\alpha]$ for adjacent layers by $\sim$0.01 because it uses $\alpha = 0$ in the event of $\alpha < 0$. The main reason for differences between FGen and CRM data, however, is that, for the present dataset, (6) tends to overestimate the degree of overlap between non-adjacent cloud layers.

Figure 4(b) shows mean cloud condensate overlap parameter $[r]$ as a function of vertical distance given in model layers. As with $[\alpha]$, the generator slightly overestimates $[r]$ compared to CRM data. Neglect of negative correlations (assumption (iv)) leads to an overestimate of $\sim$0.03 for adjacent layers. As with cloud fraction, however, the main reason for the larger difference seen for non-adjacent layers is assumption (iii): (12) tends to overestimate the horizontal correlation between cloud condensate distributions in non-adjacent layers.

It is not clear why (6) and (12) overestimate the horizontal correlations of cloud fraction and cloud condensate for non-adjacent layers. One contributing factor could

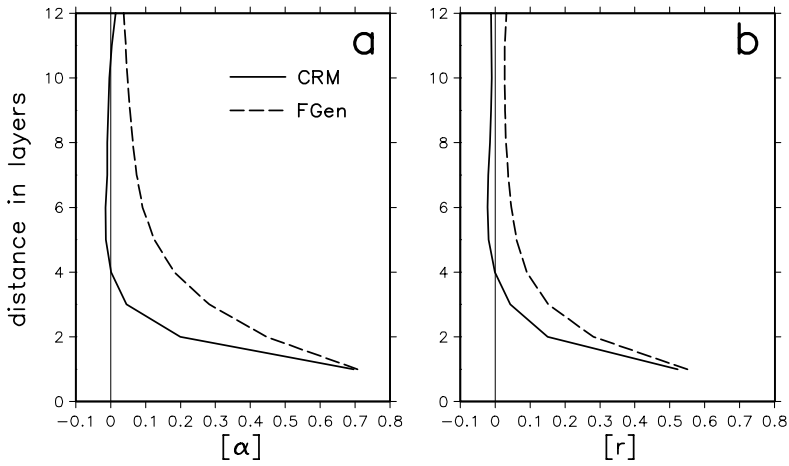

Figure 4. (a) Global-mean values of overlap parameter α for cloud fraction as a function of vertical distance in model layers (average layer thickness in the troposphere is ≈1 km) for the cloud-resolving model data (CRM) and the full generator (FGen). (b) Same as (a) but for cloud condensate overlap parameter r.

be wind shear, which was shown by Hogan and Illingworth (2003) to enhance rates at which horizontal correlation of cloud ice distributions decrease with vertical distance. It is also possible that, for convective cases, wind shear could lead often to $\alpha < 0$ for distant layers. Moreover, the limited vertical and horizontal resolutions of the present CRM dataset might also play a role. Though not shown here, generated α for distant layers usually agreed well with CRM data for the few highly resolved three-dimensional (3D) cloud fields employed by Barker *et al.* (2003). It would be pertinent, therefore, to assess the generator with an extensive set of higher-resolution CRM fields.

Figure 5 addresses assumption (v). It displays global-mean cloud condensate content for two groups of cloudy CRM cells: those adjacent to cloudy cells both above and below (group 1); all other cloudy cells, i.e. those with clear cells above and/or below (group 2). Direct comparison of mean values for the two groups is misleading because the fractions of cells in groups 1 and 2 vary from one cloud layer to another (e.g. most lower-tropospheric cloud cells belong to group 1 at high northern latitudes but to group 2 in the Tropics). Of more interest is a comparison with results computed using three simplified hypotheses. First, NOCORR assumes no correlation between cloud geometry and condensate amount. This corresponds to assumption (v) in FGen. Second, CORR assumes perfect correlation between cloud geometry and condensate amount, such that for each cloudy layer, all cells in the totally overlapped group 1 have larger w than cells in group 2. Third, NEGCORR assumes that cells in group 1 have w that are always less than those in group 2.

Not surprisingly, the NEGCORR hypothesis deviates most from CRM results, whereas NOCORR and CORR perform similarly. In the lower and middle troposphere, the assumption of no correlation between cloud geometry and condensate amount is slightly better than perfect correlation, whereas in the upper troposphere perfect correlation is closer to CRM results. While not unambiguously superior to CORR, the NOCORR assumption is preferred for the cloud generator because of its simplicity.

Errors in zonal-mean cloud field properties are addressed next. Figure 6 shows values of $\langle C_{\rm tot} \rangle$, $\langle \overline{W} \rangle$ and $\langle \sigma_W \rangle$ for CRM data along with corresponding mean bias errors (MBE) and root-mean-square errors (RMSE) for generated cloud fields relative

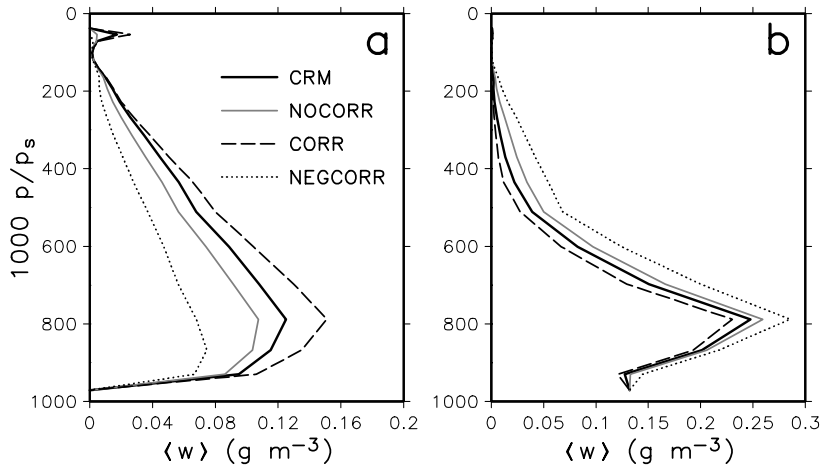

Figure 5. (a) Global-mean values of cloud condensate amount w for (a) cloudy cloud-resolving model (CRM) cells neighboured by a cloudy cell both above and below, and (b) other cloudy CRM cells. The thick solid line gives results computed directly from CRM data and the three other lines give results for the NOCORR, CORR and NEGCORR hypotheses described in the text.

to CRM data. Results for FGen are contrasted with results for MRO and random overlap (RAN). While FGen includes horizontal variations in w with overlap dictated by CRM data, MRO and RAN fields use horizontally homogenized clouds. Note that $\langle\overline{W}\rangle$ and $\langle\sigma_{\overline{W}}\rangle$ are mean values for cloudy parts of GCM columns only. In particular, biases in $\langle\overline{W}\rangle$ stem from biases in $\langle C_{\rm tot}\rangle$ as, on average, total mass of water in a GCM column predicted by all the generators is conserved.

Zonal-mean values of $C_{\rm tot}$ for FGen follow CRM values closely, although they exhibit a small negative bias of -0.004 in the global mean. This bias results from assumptions (i) and (ii) listed at the beginning of this section. Specifically, for the present dataset, (6) overestimates the degree of overlap between non-adjacent cloud layers (see earlier). Consequently, there is a small positive bias in $\langle\overline{W}\rangle$. Errors in $\langle C_{\rm tot}\rangle$, and thus in $\langle\overline{W}\rangle$, are much larger for MRO and RAN than for FGen. While MRO underestimates $\langle C_{\rm tot}\rangle$, RAN overestimates it. Root-mean-square errors in $C_{\rm tot}$, and $\overline{W}$, are likewise substantially larger for the classic overlap assumptions than for FGen.

The full generator underestimates $\langle\sigma_W\rangle$ slightly in the Tropics and overestimates it slightly at high latitudes with zonal MBEs smaller than 0.05. MRO and RAN underestimate $\langle\sigma_W\rangle$ substantially owing to their use of homogeneous clouds. In fact, the only way they produce $\langle\sigma_W\rangle > 0$ is from overlap. Values of $\langle\sigma_W\rangle$ are significantly larger for MRO than for RAN in the Tropics because of towering convective cloud fields: MRO is prone to produce overly large and small values of W in the same GCM column.

Errors in FGen's estimates of $\langle\sigma_W\rangle$ are complicated as they can arise through all assumptions (i)–(v). Additional tests (not elaborated on here), indicate that its underestimation of $\langle\sigma_W\rangle$ in the Tropics is related to the assumption that probability distributions of condensate are independent of cloud geometry. For deep tropical convection with anvils, FGen underestimates condensate amount in the cores and overestimates it in the anvils. On the other hand, the slight overestimate of $\langle\sigma_W\rangle$ at high latitudes is mainly caused by use of (12), which overestimates horizontal correlations of condensate distributions for non-adjacent cloud layers.

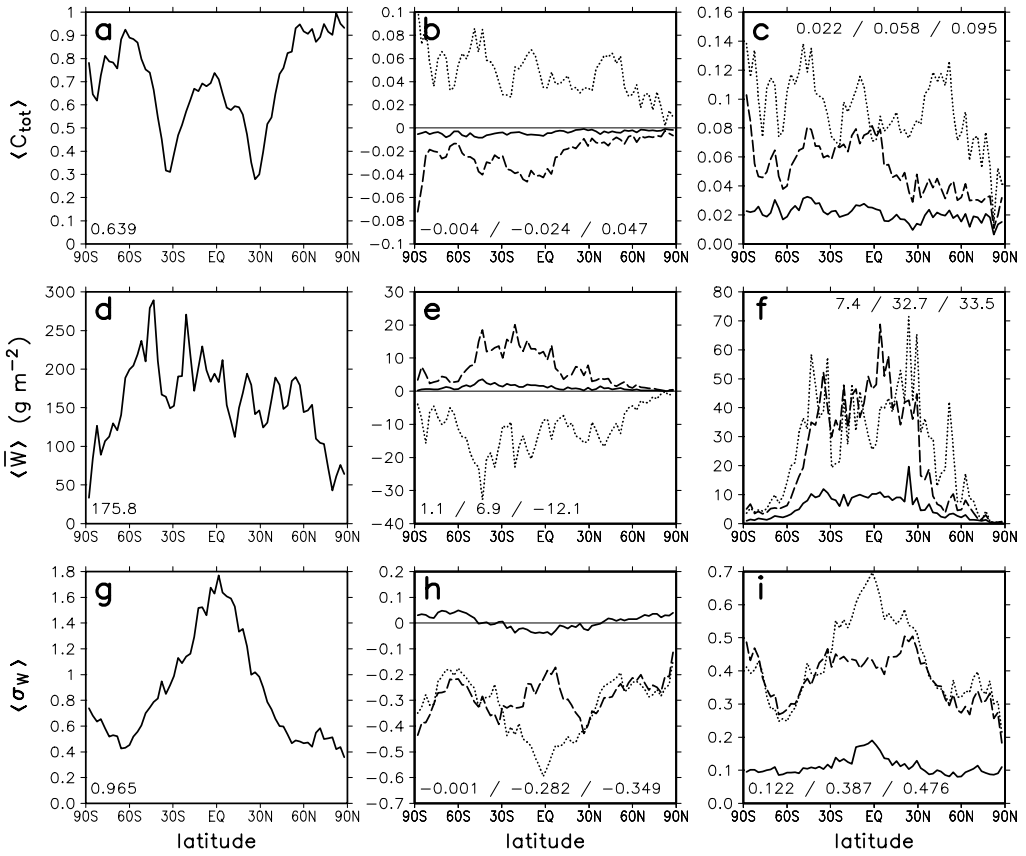

Figure 6. (a) Zonal-mean total cloud fraction C_{tot} for cloud-resolving model (CRM) data. (b) Zonal-mean bias and (c) r.m.s. errors in C_{tot} for the full generator (FGen; solid lines), maximum-random overlap (MRO; dashed lines), and random overlap (RAN; dotted lines). (d)–(f) As (a)–(c) but for mean cloud water path $\overline{W}$ (g m^{-2}) for the cloudy parts of global climate model columns. (g)–(i) As (a)–(c) but for zonal-mean relative standard deviation of cloud water path σ_W. Numbers listed on the plots are global averages for CRM data ((a), (d) and (g)) and global-mean bias errors and r.m.s. errors for FGen/ MRO/RAN. Note that MRO and RAN use horizontally homogenized clouds.

(*b*) *Radiative forcing by the generator*

Since the cloud generator is imperfect, use of its products as input for algorithms that simulate physical processes is expected to produce biases. This subsection assesses the nature and magnitude of biases in radiation fields for both FGen and MRO versions of the generator. These biases can be thought of as spurious radiative forcings induced by the generator.

The spurious radiative forcings are characterized here by global- and zonal-mean bias errors and RMSEs in net radiative fluxes (down−up) and layer heating rates. Reference results are obtained by applying the ICA to the original CRM cloud fields. Results for FGen and MRO are from ICA calculations with $\mathcal{N} = 10\,000$ generated subcolumns per CRM cloud field (i.e. GCM column). With $\mathcal{N}$ this large, sampling noise is reduced effectively to zero.

Figure 7 displays zonal-mean SW, LW and total (= LW + SW) cloud radiative forcings (CRF) for the reference calculations along with MBEs and RMSEs for FGen

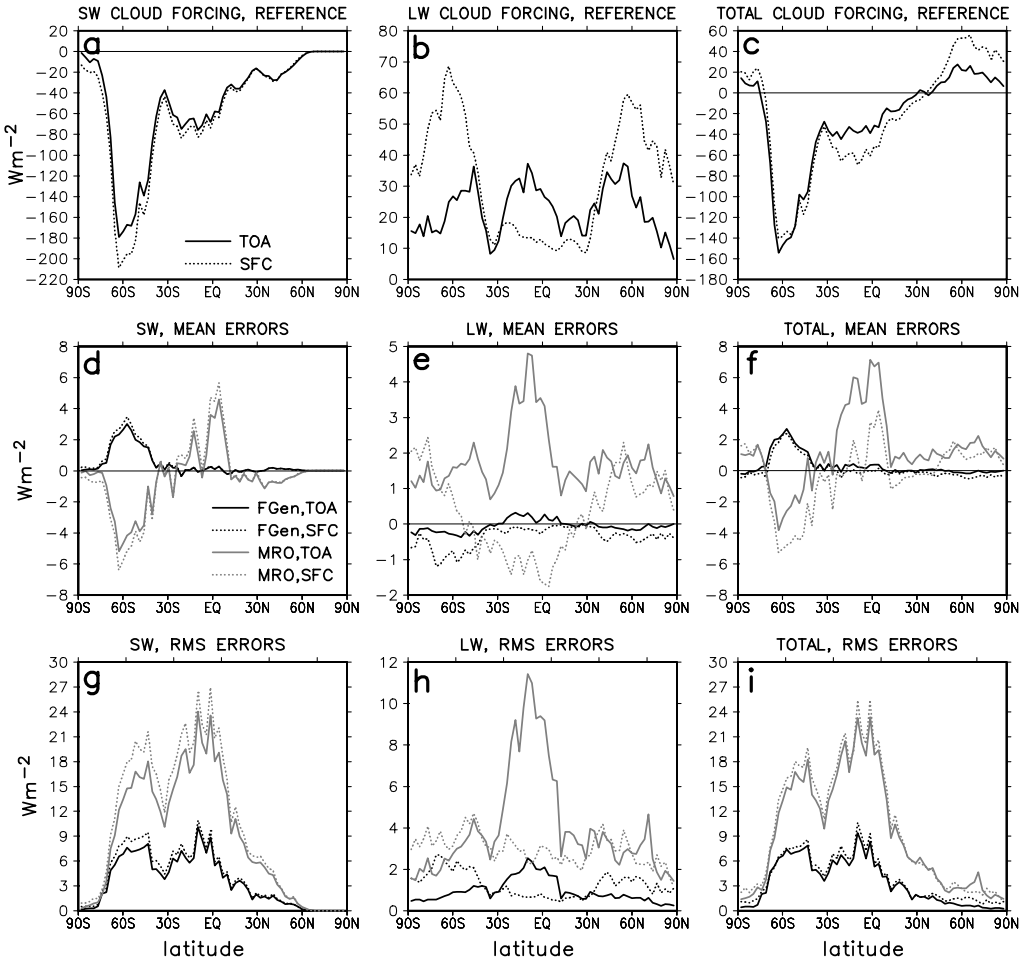

Figure 7. (a)–(c) Zonal-mean short-wave (SW), long-wave (LW) and total (= LW + SW) cloud radiative forcings for net (= down − up) fluxes at the top of the atmosphere (TOA) and at the surface (SFC) (see legend in (a)). (d)–(f) Mean bias errors in net fluxes for generalized overlap with horizontally homogeneous clouds (FGen) at the TOA and at the surface (SFC), and for maximum-random overlap with horizontally homogeneous clouds (MRO) at the TOA and at the surface (see legend in (d)). (g)–(i) As (d)–(f) but for r.m.s. errors.

and MRO. Figures 7(d)–(f) show that zonal-mean CRF (i.e. net radiative flux) errors for FGen are generally below 1 W m^{-2}. They maximize in the SW at $\sim$3 W m^{-2} at 57°S, which translates into a relative error for SW CRF of only -1.8%. This error stems from a concatenation of slightly overestimated σ_W (Fig. 6(h)) and slightly underestimated C_{tot} (Fig. 6(b)). Table 1 lists global-mean values and errors, and shows that MBEs arising from FGen are positive for net SW fluxes and negative for net LW fluxes. This is consistent with the slightly negative biases in C_{tot} in Fig. 6(b).

Looking at the MBEs in Table 1 one would conclude that the MRO model does extremely well, especially in the SW. This, however, is due to widespread cancellation of errors related to the maximum-random overlap assumption and homogeneous clouds (cf. Barker *et al.* 2003). Figures 7(d)–(f) show that these compensating errors begin to

TABLE 1. GLOBAL STATISTICS FOR RADIATIVE FLUXES

Quantity ($W\ m^{-2}$)	CRF	MBE		RMSE	
		FGen	MRO	FGen	MRO
$F_{SW,TOA}$	−54.6	0.36	−0.23	5.08	13.17
$F_{SW,SFC}$	−61.6	0.41	−0.11	5.73	15.25
$F_{LW,TOA}$	24.0	−0.01	1.98	1.28	5.57
$F_{LW,SFC}$	26.2	−0.29	−0.14	1.24	3.08
$F_{TOT,TOA}$	−30.6	0.35	1.76	4.86	13.29
$F_{TOT,SFC}$	−35.4	0.12	−0.25	5.38	14.21

CRF is global-mean cloud radiative forcing in the reference calculations. The other columns give mean bias errors (MBE) and r.m.s. errors (RMSE) for generalized overlap with horizontally variable clouds (FGen) and maximum-random overlap with horizontally homogeneous clouds (MRO). $F_{SW,TOA}$, $F_{LW,TOA}$ and $F_{TOT,TOA}$ ($F_{SW,SFC}$, $F_{LW,SFC}$ and $F_{TOT,SFC}$) are net SW, LW and total (= LW + SW) fluxes at the top of the atmosphere (surface).

unravel when viewed as zonal averages. Negative SW biases near 60°S stem from frequently overcast clouds portrayed as homogeneous, whereas positive SW biases in the Tropics come from underestimation of C_{tot} for convective cloud fields. Large positive LW biases at the TOA near the equator arise from homogenization of extensive cirrus (cf. Li and Barker 2002).

Table 1 shows that global-mean RMSEs in radiative fluxes for MRO are at least 2.5 times larger than those for FGen. Figures 7(g)–(i) indicate that the superior performance of FGen holds for all latitudes. Although SW CRF maximizes along the southern storm belt, the largest RMSEs for SW fluxes occur near the equator where cloud fields are most inhomogeneous (see Fig. 6(g)) and therefore most challenging to simulate accurately. Root-mean-square errors for LW fluxes are considerably smaller than those in SW fluxes, being only $\sim$1 W m^{-2} for FGen at most latitudes.

Figures 8(a) and (b) show zonal-mean impacts on radiative-heating rates due to clouds for the reference calculations. Figure 8(c) shows that corresponding SW heating-rate errors for FGen rarely exceed 0.02 K d^{-1}, and Fig. 8(d) shows that for the LW they are also small, save for two localized areas of positive error that reach 0.18 K d^{-1} near the surface south of 60°S, and almost 0.10 K d^{-1} in the equatorial upper troposphere. Additional tests showed that the former feature is mainly related to cloud-fraction overlap, and the latter feature can be linked to the generator's assumption that probability distributions of condensate amount are independent of cloud geometry.

Zonal-mean heating-rate errors for MRO are considerably larger than those for FGen. In the upper troposphere, homogenized clouds exposed to space produce bands of positive SW heating errors and negative LW heating errors (cf. Barker *et al.* 1999). At lower altitudes, these errors reverse sign due to excessive shielding of cloud to solar radiation, and excessive trapping of terrestrial radiation.

Figure 9 shows that FGen and MRO exhibit similar RMSE patterns for radiative-heating rates, although errors are typically 2 to 4 times larger for MRO. Generally, RMSEs for the LW exceed those for the SW, with largest errors occurring in the upper tropical troposphere and near the surface at high southern latitudes. Excessive MRO errors in the upper tropical troposphere come from horizontal homogenization of highly variable anvil clouds. The lower southern troposphere is prone to large errors due to a combination of abundant cloudiness and relatively high vertical resolution. These lead to large LW cooling-rate errors when clouds are exposed to space incorrectly.

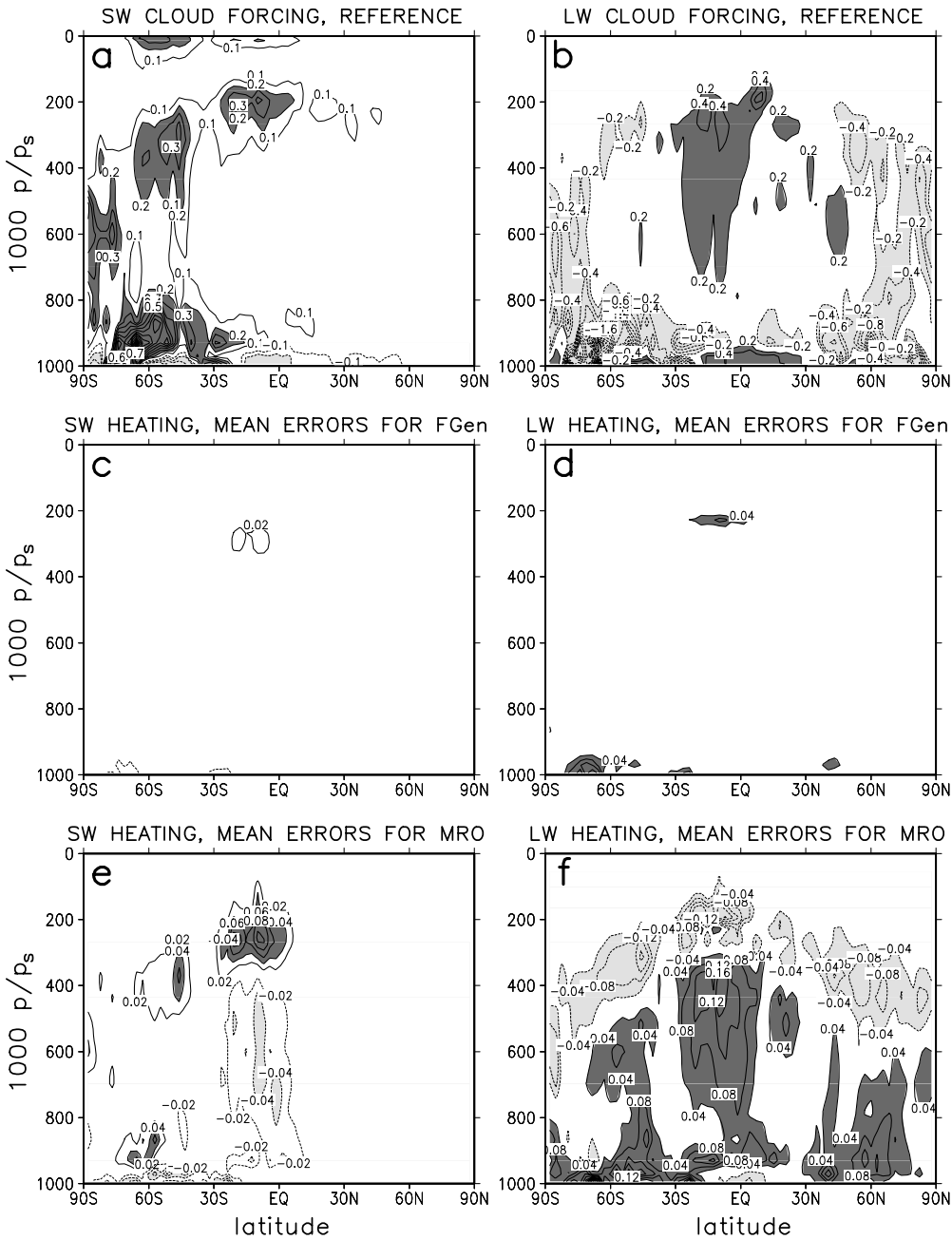

Figure 8. (a) and (b) Zonal-mean cloud radiative forcing for short-wave (SW) and long-wave (LW) heating in the reference calculations (K d^{-1}). (c)–(f) Zonal-mean errors in SW and LW heating rate for generalized overlap with horizontally variable clouds (FGen) and for maximum-random overlap with horizontally homogeneous clouds (MRO). Contour intervals are 0.1 K d^{-1} in (a), 0.2 K d^{-1} in (b), 0.02 K d^{-1} in (c) and (e), and 0.04 K d^{-1} in (d) and (f). Zero contours are omitted.

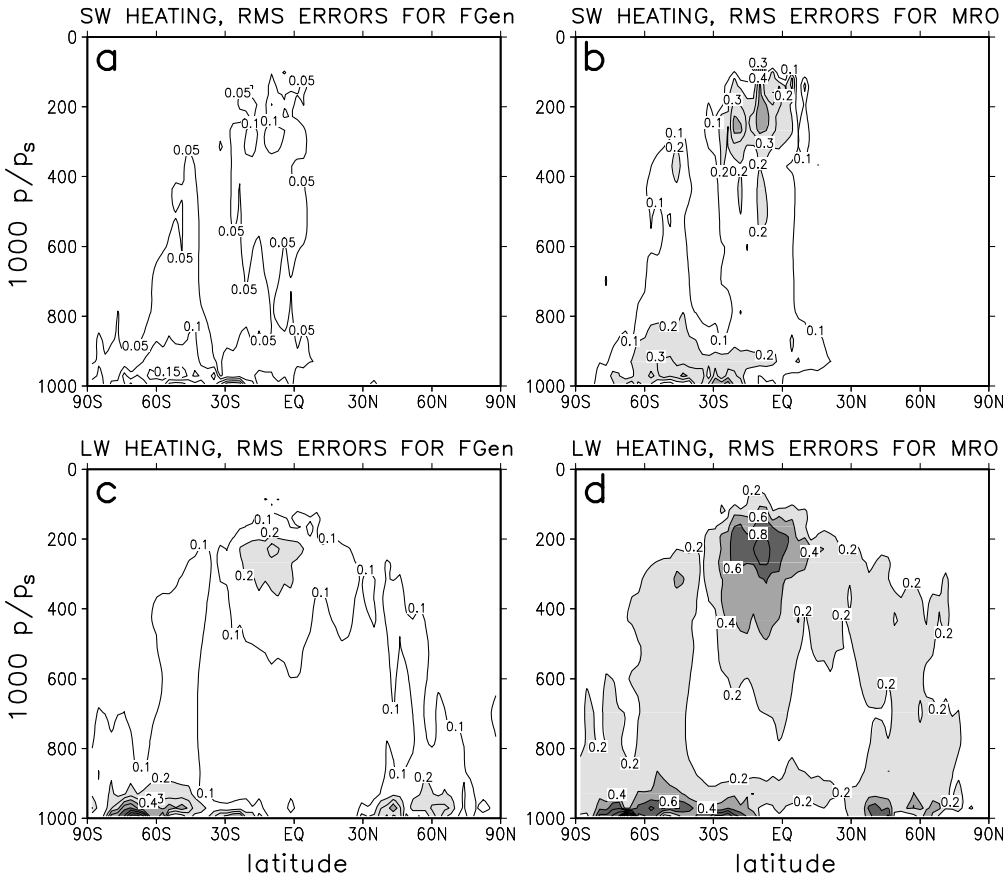

Figure 9. Zonal r.m.s. errors in short-wave heating rate for (a) generalized overlap with horizontally variable clouds (FGen), and (b) for maximum-random overlap with horizontally homogeneous clouds (MRO). (c) and (d) As (a) and (b) but for long-wave heating. Contour intervals are 0.05 K d^{-1} in (a), 0.1 K d^{-1} in (b) and (c), and 0.2 K d^{-1} in (d).

6. Conclusions and discussion

The main point of this study was to present a stochastic cloud generator whose purpose is to produce sets of subcolumns within LSAM columns. The primary motivation behind building the generator was to facilitate use of the Monte Carlo Independent Column Approximation (McICA) for radiative transfer. This generator could, however, also be of value for parametrization of precipitation, and possibly other, processes.

The generator creates cloud fields on a column-by-column basis. Since the approach is stochastic, statistical characteristics of generated fields become increasingly accurate as the number of subcolumns $\mathcal{N}$ increases. Vertical overlap of cloud fraction and cloud condensate are described with two parameters, both of which can be expressed using decorrelation depths. The following simplifying assumptions are involved:

(i) it is sufficient to consider only adjacent-layer correlations;

(ii) correlations are always non-negative; and

(iii) probability distributions of cloud water content are independent of cloud geometry (i.e. the same for overlapping and non-overlapping cloudy cells).

Performance of the cloud generator was assessed using a single day's worth of data obtained from an LSAM simulation that employed a low-resolution 2D CRM in each of its column. Biases associated with the generator are small, although there is a slight tendency towards underestimation of total cloud fraction. When the generator was initialized with CRM data and random noise was reduced to near zero, RMSEs for radiative fluxes and heating rates were generally at least 60% smaller than those for MRO of homogeneous clouds. Naturally, use of perfect input data overestimates the accuracy achievable using the generator in an ordinary LSAM environment where cloud overlap parameters and subgrid-scale distribution of cloud water content are not readily available. On the other hand, cancellation of errors related to the MRO assumption and to the neglect of cloud horizontal variations reduces, fortuitously, especially SW flux errors for the MRO model. This will be discussed in more detail in a future paper.

Uniting the generator with McICA in an LSAM has several distinct advantages over the current paradigm for modelling radiative transfer. First, it is more flexible than the conventional approach because it is far easier to define unresolved cloud structure than to interweave it into 1D radiation codes. In addition to cloud fraction and condensate amount, variations in particle size, phase and habit could be incorporated. Second, the generator could be extended to include unresolved variations in absorbing gases, aerosols and surface properties. Two simple extensions of the present generator are outlined in appendix A, while appendix B demonstrates that horizontal variations in water vapour could be accounted for by formulating the generator in terms of total atmospheric water (vapour + cloud) instead of cloud condensate amount. Third, simple radiation schemes that are free of cloud overlap, horizontal variability, and even cloud fraction, are used with McICA.

At least four areas for further research can be identified. First, overlap parameters and probability distributions of cloud condensate must be parametrized and set diagnostically in ordinary LSAMs. Since construction of these necessarily scale-dependent, conditional parametrizations will likely be an arduous task, it is wise to first assess the sensitivity of radiative transfer and LSAMs to these parameters. This is straightforward with McICA and the generator.

Second, analyses undertaken here used $\mathcal{N} = 10\,000$ subcolumns, which virtually eliminates random errors in radiative fluxes and heating rates associated with the generator. Due to computational constraints, random errors cannot be avoided in operational LSAM use. The magnitude of random errors, and their conditional dependence on cloud field and radiation code properties, need to be quantified as does their impact on LSAM simulations (cf. Pincus *et al.* 2003).

Third, it would be pertinent to test the generator's performance for parametrization of precipitation. Given the intermittent nature of precipitation processes, and the high sensitivity of parametrized large-scale precipitation to cloud overlap (Jakob and Klein 1999), this could provide a more demanding test than radiative-transfer calculations.

Fourth, the generator presented here should not be regarded as the 'final' product. Rather, it is only the first incarnation of a scheme that is, in principle, entirely open-ended (cf. the appendices).

ACKNOWLEDGEMENTS

This study was supported by grants from the US Department of Energy (Atmospheric Radiation Measurement grant DE-FG02-03ER63521) and the Modelling of Clouds and Climate Proposal funded through the Canadian Foundation for Climate and Atmospheric Sciences, the Meteorological Service of Canada, and the Natural Sciences

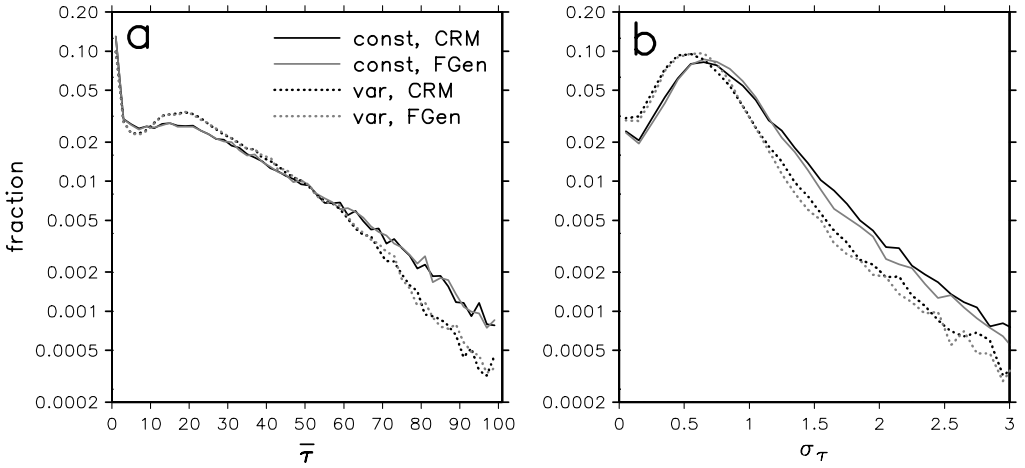

Figure A.1. (a) Frequency distributions of mean liquid-phase cloud optical depth in global climate model grid columns using cloud-resolving model data and corresponding fields produced by the full generator (bin width $\Delta\overline{\tau} = 2$). Results are shown for $r_e = 10$ μm (const) and r_e defined as in (A.1) (var). (b) Same as (a) except these are distributions of relative standard deviation of cloud optical depth (bin width $\Delta\sigma_\tau = 0.1$). See text for explanation.

and Engineering Research Council. P. Räisänen also received financial support from the Academy of Finland (project 54481). We gratefully acknowledge Robert Pincus and Philip Gabriel for helpful discussions and anonymous reviewers for their useful comments.

APPENDIX A

Two examples of extending the generator

(a) *Horizontal variability of cloud droplet size and concentration.* Aircraft observations show that there is often a clear relation between cloud liquid-water content w_l and cloud droplet effective radius r_e (e.g. Korolev *et al.* 2001). Any functional relationship between these variables can be included easily in the generator. The same is true for assumptions about effective ice-crystal size and ice-water content w_i.

To illustrate this point, Fig. A.1 shows frequency distributions of mean value and relative standard deviation of cloud liquid-phase optical depth for the GCM dataset, for the assumption of constant $r_e = 10$ μm (as used earlier), and for

$$r_e = 16.2 \times w_l^{1/3}. \tag{A.1}$$

In both cases the global-mean optical depth is the same. When r_e is assumed to increase with w_l, there are fewer thick clouds and more clouds near the mean. Variations in r_e also reduce the subgrid-scale variations of optical depth, in accord with Räisänen *et al.* (2003). The impact of changing the definition of r_e is very similar for the original CRM fields (black lines) and for fields produced by the full generator (grey lines).

(b) *Accounting for correlations between surface type and cloud.* Another issue that appears to be completely beyond the reach of the current paradigm of modelling radiative transfer in LSAMs involves variations in surface albedo that are correlated with clouds. This is a frequent occurrence for LSAM cells that straddle continental

coastal lines. Often, most notably in stratocumulus regions, cloud will be primarily over water and not over land. Current models simply construct an areal-average surface albedo and use that value as the lower boundary condition. This could be accounted for by the generator as follows.

Imagine that a fraction f_W of the underlying surface of an LSAM cell is water, and that α_W and α_L are surface albedos over water and land. Once C_{tot} has been determined, one would have to decide how to partition cloudiness between land and water so as to define fractional amounts of cloud over water C_W and land C_L. This is constrained by

$$C_{tot} = f_W C_W + (1 - f_W) C_L. \tag{A.2}$$

Then, generating a uniform random number $RN_j \in [0, 1]$ for each subcolumn j, surface albedo α_j for cloudy subcolumns would be defined as

$$\alpha_j = \begin{cases} \alpha_W, & \text{for } RN_j \leq \dfrac{f_W C_W}{C_{tot}}, \\ \alpha_L, & \text{for } RN_j > \dfrac{f_W C_W}{C_{tot}}, \end{cases} \tag{A.3}$$

whereas, for cloud-free subcolumns,

$$\alpha_j = \begin{cases} \alpha_W, & \text{for } RN_j \leq \dfrac{f_W (1 - C_W)}{1 - C_{tot}}, \\ \alpha_L, & \text{for } RN_j > \dfrac{f_W (1 - C_W)}{1 - C_{tot}}. \end{cases} \tag{A.4}$$

Generalizations to other surfaces (e.g. ice or snow and bare land or open ocean) or to systems involving more than two surfaces are straightforward.

It is acknowledged that in practice, use of this approach in an LSAM would not be straightforward since, to our knowledge, no attempts have been made to partition cloudiness within an LSAM grid cell according to surface type. Nevertheless, this example does serve to demonstrate the flexibility of the McICA-generator union.

APPENDIX B

Applying the generator to a total-water scheme

Tompkins's (2002) cloud scheme operates on total-water mixing ratio q (condensed + vapour) and uses the beta probability density function $p(q)$ to describe fluctuations in q. A slightly altered version of the generator can be applied to this type of scheme. Begin by dispensing with setting cloud fraction for subcolumn cells. As with condensed water in the main generator, total-water mixing ratios $q_{j,k}$ for all cells are generated by solving

$$y_{j,k} = \int_0^{q_{j,k}} p_k(q)\, \mathrm{d}q, \tag{B.1}$$

where $y_{j,k}$ is the cumulative frequency distribution of q, and $p_k(q)$ is the corresponding distribution of q for the kth layer. As usual, values for $y_{j,k=1}$ are drawn at random as in (10), while for all other layers

$$y_{j,k} = \begin{cases} y_{j,k-1}, & \text{for } RN1_{j,k} \leqslant r^q_{k-1,k}, \\ RN2_{j,k}, & \text{for } RN1_{j,k} > r^q_{k-1,k}, \end{cases} \quad k = 2, \ldots K; \;\; j = 1, \ldots \mathcal{N}; \tag{B.2}$$

where $\mathrm{RN1}_{j,k}$ and $\mathrm{RN2}_{j,k}$ are unique random numbers distributed evenly between 0 and 1, and $r^q_{k-1,k}$ is the linear correlation between the cumulative frequencies of q in adjacent layers. Therefore, like the main algorithm, values of $y_{j,k}$ simply jump between perfect correlation for a fraction $r^q_{k-1,k}$ of cells and no correlation for the remainder. Designating the saturation vapour mixing ratio for each layer as q^{sat}_k, cloud mixing ratio is defined as

$$q^{\mathrm{cld}}_{j,k} = \begin{cases} 0, & \text{for } q_{j,k} - q^{sat}_k \leq 0, \\ q_{j,k} - q^{\mathrm{sat}}_k, & \text{for } q_{j,k} - q^{\mathrm{sat}}_k > 0. \end{cases} \tag{B.3}$$

The appeal of this simple procedure, that appears to be so well suited to schemes based on variability of total water, is tarnished somewhat by correlations in cloud fraction overlap and condensate both being equal to $r^q_{k-1,k}$. For this reason, we opted for the generator presented in the main text.

References

Barker, H. W., Stephens, G. L. and Fu, Q. 1999 The sensitivity of domain-averaged solar fluxes to assumptions about cloud geometry. *Q. J. R. Meteorol. Soc.*, **125,** 2127–2152

Barker, H. W., Pincus, R. and Morcrette, J.-J. 2002 'The Monte Carlo Independent Column Approximation: Application within large-scale models'. Proceedings of the GCSS Workshop, 20–24 May 2002, Kananaskis, Alberta, Canada (available at http://www.met.utah.edu/skrueger/gcss-2002/Extended-Abstracts.pdf)

Barker, H. W., Stephens, G. L., Partain, P. T., Bergman, J. W., Bonnel, B., Campana, K., Clothiaux, E. E., Clough, S., Cusack, S., Delamere, J., Evans, K. F., Fouquart, Y., Freidenreich, S., Galin, V., Hou, Y., Kato, S., Li, J., Mlawer, E., Morcrette, J.-J., O'Hirok, W., Räisänen, P., Ramaswamy, V., Ritter, B., Rozanov, E., Schlesinger, M., Shibata, K., Sporyshev, P., Sun, Z., Wendisch, M., Wood, N. and Yang, F. 2003 Assessing 1D atmospheric solar radiative transfer models: Interpretation and handling of unresolved clouds. *J. Climate*, **16,** 2676–2699

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

Cahalan, R. F., Ridgway, W., Wiscombe, W. J., Gollmer, S. and Harshvardhan 1994 Independent pixel and Monte Carlo estimates of stratocumulus albedo. *J. Atmos. Sci.*, **51,** 3776–3790

Cairns, B., Lacis, A. A. and Carlson, B. E. 2000 Absorption within inhomogeneous clouds and its parameterization in general circulation models. *J. Atmos. Sci.*, **57,** 700–714

Coakley, J. A., Cess, R. D. and Yurevich, F. B. 1983 The effect of tropospheric aerosols on the earth's radiation budget: A parameterization for climate models. *J. Atmos. Sci.*, **40,** 116–138

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

Fu, Q. 1996 An accurate parameterization of the solar radiative properties of cirrus clouds for climate models. *J. Climate*, **9,** 2058–2082

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

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

Grabowski, W. W. 2003 MJO-like coherent structures: Sensitivity simulations using the Cloud-Resolving Convection Parameterization (CRCP). *J. Atmos. Sci.*, **60,** 847–864

Grabowski, W. W., Wu, X., Moncrieff, M. W. and Hall, W. D. 1998 Cloud resolving modeling of tropical cloud systems during PHASE III of GATE. Part II: Effects of resolution and the third spatial dimension. *J. Atmos. Sci.*, **55,** 3264–3282

Hogan, R. J. and Illingworth, A. J. 2000 Deriving cloud overlap statistics from radar. *Q. J. R. Meteorol. Soc.*, **128,** 2903–2909

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

Jakob, C. and Klein, S. A. 1999 The role of vertically varying cloud fraction in the parameterization of microphysical processes in the ECMWF model. *Q. J. R. Meteorol. Soc.*, **125**, 941–965

2000 A parametrization of the effects of cloud and precipitation overlap for use in general-circulation models. *Q. J. R. Meteorol. Soc.*, **126**, 2525–2544

Kato, S. 2003 Computation of domain-averaged shortwave irradiance by a one-dimensional algorithm incorporating correlations between optical thickness and direct incident radiation. *J. Atmos. Sci.*, **60,** 182–193

Khairoutdinov, M. F. and Randall, D. A. 2001 A cloud-resolving model as a cloud parameterization in the NCAR Community Climate System Model: Preliminary results. *Geophys. Res. Lett.*, **28,** 3617–3620

Korolev, A. V., Isaac, G. A., Mazin, I. and Barker, H. W. 2001 Microphysical properties of continental stratiform clouds. *Q. J. R. Meteorol. Soc.*, **127,** 2117–2151

Li, J. 2002 Accounting for unresolved clouds in a 1D infrared radiative transfer code. Part I: Solution for radiative transfer, including cloud scattering and overlap. *J. Atmos. Sci.*, **59,** 3302–3320

Li, J. and Barker, H. W. 2002 Accounting for unresolved clouds in a 1D infrared radiative transfer code. Part II: Horizontal variability of cloud water path. *J. Atmos. Sci.*, **59,** 3321–3339

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

Mace, G. G. and Benson-Troth, S. 2002 Cloud-layer overlap characteristics derived from long-term cloud radar data. *J. Climate*, **15,** 2505–2515

McClatchey, R. A., Fenn, R. W., Selby, J. E. A., Volz, F. E. and Garing, J. S. 1971 Optical properties of the atmosphere. Report AFCRL-71-0279 (available from Air Force Geophysics Laboratory, Hanscom Air Force Base, MA 01731, USA)

Oreopoulos, L. and Barker, H. W. 1999 Accounting for subgrid-scale cloud variability in a multi-layer 1D solar radiative transfer algorithm. *Q. J. R. Meteorol. Soc.*, **125,** 301–330

Pincus, R., Barker, H. W. and Morcette, J.-J. 2003 A fast, flexible, approximate technique for computing radiative transfer in inhomogeneous cloud fields. *J. Geophys. Res.*, **108,** D13, doi:10.1029/2002JD003322

Räisänen, P. 1999 Effect of vertical resolution on cloudy-sky radiation calculations: Tests with two schemes. *J. Geophys. Res.*, **104,** 27407–27419

Räisänen, P., Isaac, G. A., Barker, H. W. and Gultepe, I. 2003 Solar radiative transfer for stratiform clouds with horizontal variations in liquid-water path and droplet effective radius. *Q. J. R. Meteorol. Soc.*, **129,** 2135–2149

Randall, D. A., Khairoutdinov, M. F., Arakawa, A. and Grabowski, W. W. 2003 Breaking the cloud-parameterization deadlock. *Bull. Am. Meteorol. Soc.*, **84,** 1547–1564

Rotunno, R., Klemp, J. B. and Weisman, M. L. 1988 A theory for strong, long-lived squall lines. *J. Atmos. Sci.*, **45,** 463–485

Tanre, D., Geleyn, J.-F. and Slingo, J. M. 1984 First results of the introduction of an advanced aerosol-radiation interaction in the ECMWF low resolution global model. Pp. 133–177 in *Aerosols and their climatic effects*. Eds. H. E. Gerber and A. Deepak, A. Deepak Publishing, Hampton, Virginia, USA

Tompkins, A. M. 2002 A prognostic parameterization for the subgrid-scale variability of water vapor and clouds in large-scale models and its use to diagnose cloud cover. *J. Atmos. Sci.*, **59,** 1917–1942