Overlap assumptions for assumed probability distribution function cloud schemes in large-scale models

Robert Pincus,[1] Cécile Hannay,[1,2] Stephen A. Klein,[3,4] Kuan-Man Xu,[5] and Richard Hemler[3]

[1] Cloud vertical structure influences the fluxes of precipitation and radiation throughout the atmosphere. This structure is not predicted in large-scale models but is instead applied in the form of "overlap assumptions." In their current guise, overlap assumptions apply to the presence or absence of clouds, and new data sets have led to the development of empirical formulations described by exponential decay from maximum to random overlap over a characteristic length scale. At the same time, cloud parameterizations in many large-scale models have been moving toward "assumed PDF" schemes that predict the distribution of total water within each grid cell, which will require overlap assumptions that may be applied to cells with specified internal variability. This paper uses a month-long cloud-resolving model simulation of continental convection to develop overlap assumptions for use with assumed PDF cloud schemes in large-scale models. An observing system simulation experiment shows that overlap assumptions derived from millimeter-wavelength cloud radar observations can be strongly affected by the presence of precipitation and convective clouds and, to a lesser degree, by limited sampling and reliance on the frozen turbulence assumption. Current representations of overlap can be extended with good accuracy to treat the rank correlation of total water in each grid cell, which provides a natural way to treat vertical structure in assumed PDF cloud schemes. The scale length that describes an exponential fit to the rank correlation of total water depends on the state of the atmosphere: convection is associated with greater vertical coherence (longer scale lengths), while wind shear decreases vertical coherence (shorter scale lengths). The new overlap assumptions are evaluated using cloud physical properties, microphysical process rates, and top-of-atmosphere radiative fluxes. These quantities can be reproduced very well when the exact cloud structure is replaced with its statistical equivalent and somewhat less well when the time mean vertical structure is imposed. Overlap formulations that treat total water can also be used to determine the variability in clear-air relative humidity, which might be used by convection and aerosol parameterizations.

1. Vertical Structure and Cloud Subgrid-Scale Variability in Large-Scale Models

[2] Global weather forecast and climate models predict the evolution of the atmosphere by computing changes in the energy, momentum, and air and water mass budgets at many points around the planet. The governing equations in these models contain source and sink terms representing processes both explicitly resolved by the model (such as advection) and those that are not resolved, but are instead parameterized (such as convection). The rates at which these processes proceed depends, of course, on the state of the atmosphere: that is, the values of energy, momentum, and mass in each cell.

[3] Many processes represented in a model are local, in that the process rate in one cell is independent of the rate in other cells. Others, however, including radiation and precipitation/evaporation, are nonlocal because they depend on the state of the atmosphere throughout each grid column. Rain falling into subsaturated clear sky evaporates, for example, while the same amount of rain falling into another cloud collects smaller droplets and produces even more rain, at the cost of depleting the lower cloud's water content.

---

[1]Climate Diagnostics Center, NOAA–Cooperative Institute for Research in Environmental Sciences (CIRES), Boulder, Colorado, USA.

[2]Now at National Center for Atmospheric Research, Boulder, Colorado, USA.

[3]Geophysical Fluid Dynamics Laboratory, NOAA, Princeton, New Jersey, USA.

[4]Now at Atmospheric Sciences Division, Lawrence Livermore National Laboratory, Livermore, California, USA.

[5]NASA Langley Research Center, Hampton, Virginia, USA.

Similarly, because the incremental change of albedo with cloud optical depth decreases as the optical depth increases, two partially cloudy layers reflect less solar radiation if one cloud obscures the other than if both layers are visible from the top of the atmosphere. To compute the large-scale effects of precipitation and radiation the degree to which clouds and precipitation are vertically aligned with each other must be known. This structure is determined by cloud-scale dynamics that are not resolved by global or other large-scale models, so the calculations are closed using rules called "overlap assumptions."

[4] Overlap assumptions are applied to instantaneous profiles of domain-averaged cloud properties to infer sub-grid-scale structure. In most current large-scale models these cloud properties include the portion of each grid cell occupied by cloud (the "cloud fraction") and the mean condensate (liquid and ice) concentration. Given these two variables, overlap assumptions apply to the presence or absence of cloud in each layer, which we call "occurrence overlap." Occurrence overlap assumptions can be used to construct explicit subcolumns from a profile of cloud properties. Given N partially cloudy layers, for example, overlap assumptions determine what fraction of the column's area is occupied by each of the 2^N possibilities of cloud/clear sky in each layer. Profiles of precipitation [*Jakob and Klein*, 2000] or radiative fluxes [*Collins*, 2001] can then be calculated for each possible configuration of clouds and the results weighted by the area fraction to provide domain mean fluxes. A variation on this technique is to construct a population of subcolumns stochastically such that each layer of each subcolumn is homogeneous (that is, completely clear or cloudy, with uniform cloud properties) while ensuring that the statistics of a large collection of subcolumns reproduces both the profiles of cloud properties and the overlap assumptions. This very flexible approach has been used in radiation calculations [*Räisänen et al.*, 2004] and for diagnostic purposes, notably the ISCCP simulator [*Webb et al.*, 2001; *Yu et al.*, 1996].

[5] To date, large-scale models have typically used one of three very simple assumptions: random, maximum, and the popular maximum-random [*Geleyn and Hollingsworth*, 1979], the latter of which is supported, to some degree, by observations [*Tian and Curry*, 1989]. Recently obtained long-term radar records, however, indicate that clouds are less vertically coherent than the maximum-random assumption implies. These new observations suggest that occurrence within vertically continuous cloud layers decays inverse-exponentially from maximum to random overlap as the vertical distance separating cloud layers increases, with a scale length of several kilometers [*Hogan and Illingworth*, 2000; *Mace and Benson-Troth*, 2002]. Clouds produced by relatively fine-scale cloud-resolving models behave in much the same way [*Oreopoulos and Khairoutdinov*, 2003; *Räisänen et al.*, 2004]. The characteristic length scale in both observed and modeled clouds varies widely depending on location and time of year, but does not appear to depend primarily on height within the atmosphere.

[6] This new description of occurrence overlap arrives just as interest is shifting toward cloud schemes that account for variability of cloud condensate within model layers. Models that predict only the mean value of cloud properties in each cell are subject to uncomfortably large biases in many process rates, including radiation and precipitation [*Cahalan et al.*, 1994; *Pincus and Klein*, 2000]. The biases introduced by unresolved subgrid-scale inhomogeneity appear to be one of the main reasons that large-scale models need to be tuned, i.e., why physical parameters in the model must be changed from known, reasonable values to unrealistic values in order to reproduce a realistic climate [*Rotstayn*, 2000]. One promising way around this problem are schemes which predict the horizontal probability distribution function (PDF) of total water within each model grid cell, from which the PDF of condensate is inferred [*Tompkins*, 2002]. These schemes are sometimes called "statistical cloud schemes," although they might more accurately be called "assumed PDF cloud schemes."

[7] Assumed PDF cloud schemes require more general overlap assumptions than currently exist. If the amount of condensate varies within each grid cell, overlap assumptions must express not only the likelihood of occurrence, but also the degree to which condensate concentration is correlated in the vertical. Radar observations show that condensate behaves much like occurrence overlap, changing from perfectly correlated to completely uncorrelated roughly exponentially as the separation between layers increases [*Hogan and Illingworth*, 2003], although the length scales are typically shorter than those for occurrence overlap. Cloud-resolving models produce similar structures. It is possible to incorporate this correlation, along with specified subgrid-scale distributions of condensate, into stochastically generated subcolumns [*Räisänen et al.*, 2004]. It is not clear, however, how one would reconcile an approach that uses separately defined correlations for cloud occurrence and condensate overlap with the total water distributions predicted by an assumed PDF cloud scheme.

[8] In this paper we develop overlap specifications for cloud schemes that predict the distribution of total water within each model grid cell. The descriptions are based on a month-long simulation made with a cloud-resolving model. We treat the domain of the cloud-resolving model as a single large-scale model grid cell within which we have perfect information about the horizontal and vertical distribution of clouds and water vapor at every snapshot time. On the basis of these descriptions, we examine the degree to which overlap assumptions derived from cloud radar data might be expected to agree with those derived from model snapshots, then extend current representations of overlap to specify the vertical organization of total water when the horizontal distribution is specified. We explore some of the factors that determine the degree of vertical coherence of total water, and examine how accurately column integrated cloud properties and precipitation and radiation fluxes can be reconstructed based on our representations of overlap.

2. Sources of Information for Overlap Parameterizations

[9] We seek a description of the vertical structure of clouds within a large-scale model's grid column; that is, a description that can be applied instantaneously to describe the horizontal and vertical structure within a domain several tens or hundreds of kilometers across. Direct observations of this structure as it occurs in nature are hard to come by, so to date most descriptions of overlap have relied on two
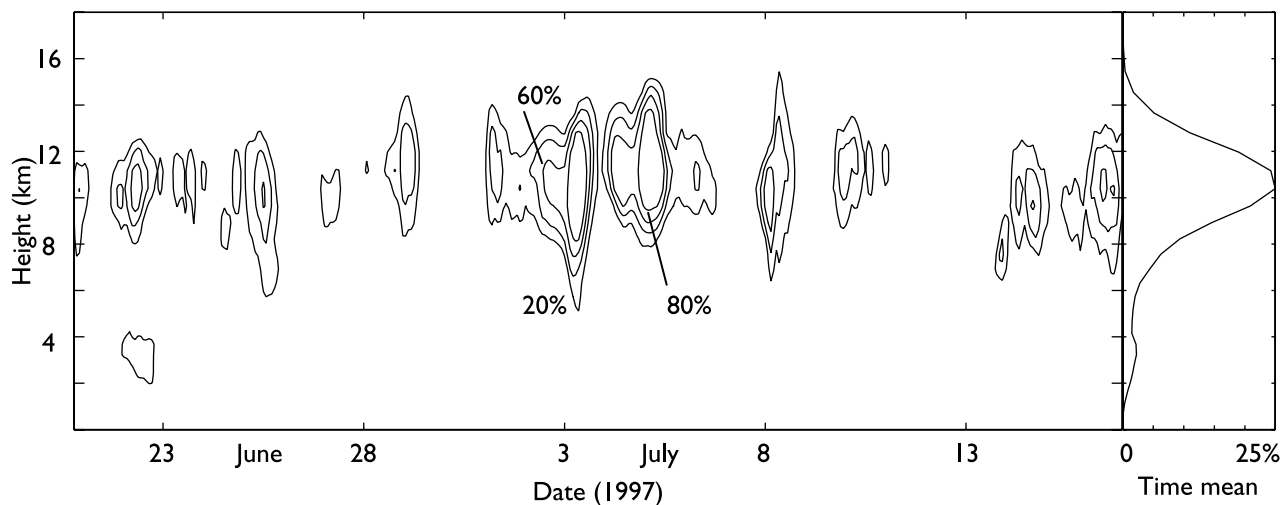
Figure 1. Time-height cross section of domain mean cloud fraction during the last 26 days of a 29 day simulation of deep convection over a continental site and a profile of the time mean cloud fraction. Few clouds occurred in the first 3 days. The majority of clouds are anvils and other upper level clouds produced by detrainment from deep convection. Contour levels are every 20%.

sources: time series obtained by vertically pointing cloud radars, and the fields produced by fine-scale models.

[10] Cloud-sensing (millimeter wavelength) radars have been deployed over the last decade in many locations around the globe. Observations from these radars (and from lidar, in thin clouds) are the only long-term direct measurement of cloud vertical structure in the real atmosphere, and can be used to estimate overlap occurrence as well as the correlation of condensate. They are, however, somewhat difficult to interpret. Radars measure the distribution of all forms of condensed water in the atmosphere, while most large-scale models make distinctions between precipitation, clouds directly associated with convection ("convective clouds"), and large-scale ("stratiform") clouds, with overlap being applied separately to each category. Ambiguities also arise from the way the instruments are operated. The vast majority of radar observations are made from the ground with the instrument pointing directly upward, which yields a time series of profiles of cloud properties from which statistics describing spatial structure (like overlap) are inferred by accumulating point observations over time. This approach relies heavily on the frozen turbulence assumption, since the observations include both evolution and advection. Additionally, inferences of spatial structure from time series almost always equate a constant time interval with a fixed horizontal domain size, even though wind speeds vary with time and height, and overlap parameters derived from observations made by vertically pointing radars are known to be sensitive to the time window chosen [see *Hogan and Illingworth*, 2000, Table 1; *Mace and Benson-Troth*, 2002, Table 3]. Vertical and horizontal structure may be obtained more readily from radars operated in scanning mode, but these observations are few and far between.

[11] Cloud fields simulated by fine-scale models (i.e., cloud resolving models, or CRMs) have complementary strengths and weaknesses. They rely on inexact subgrid-scale parameterizations embedded in the cloud model and

have finite spatial resolution, and so may not resolve all the important circulations. This means that there is no guarantee that the cloud structures produced by a fine-scale model are faithful to nature. On the other hand, fine-scale models can provide instantaneous snapshots of two- or three-dimensional cloud structure in a domain about the size of a large-scale model grid column, and this structure is exactly what a large-scale model tries to simulate each time it applies an overlap assumption. Furthermore, it is straightforward to distinguish precipitation from stratiform and convective clouds (though this has not always been done), and it is possible to examine the vertical structure of quantities (i.e., total water concentration) that are difficult to measure directly.

[12] Here we use a 29 day simulation of summertime deep convection over the central United States to quantify overlap. The runs are made with the UCLA/CSU cloud system model [*Krueger*, 1988; *Xu and Krueger*, 1991] forced by a variational analysis of observations made from 19 June to 17 July 1997 [*Zhang et al.*, 2001] at the ARM Southern Great Plains site near Lamont, Oklahoma. The model is configured as a two-dimensional 512 km domain with horizontal grid spacing of 2 km and 35 vertical levels on a stretched grid. Calculations use a bulk microphysics scheme with prognostic equations for two species of cloud condensate (liquid and ice) and three species of precipitation (rain, snow, and graupel). Snapshots of the atmospheric state are recorded every five minutes. Most of the clouds are produced by convection and detrained into the free atmosphere as anvils, although some (poorly resolved) shallow convection is also simulated. The domain-averaged cloud fraction as a function of time and height is shown in Figure 1. Additional details about the simulation are given by *Xu et al.* [2002]. Extensive evaluation against observations are reported by *Luo et al.* [2003], who show the simulation accurately reproduces the occurrence of cumulonimbus clouds and their associated anvils, but is less successful at predicting cirrus clouds associated with

large-scale advection. Cloud microphysical and macrophysical properties in the simulations are also comparable with those obtained from cloud radar observations.

3. Estimates of Overlap From Radar Observations and From Fine-Scale Model Fields

[13] Overlap assumptions in large-scale models determine the instantaneous three-dimensional structure of cloudiness given a profile of the atmospheric state. The overlap derived from radar observations, which is typically determined using methods introduced by *Hogan and Illingworth* [2000] and described more fully below, is not quite the same quantity. In this section we use the cloud structure produced by the cloud resolving model to assess the accuracy with which the overlap needed by a large-scale model can be determined from radar observations.

3.1. Occurrence Overlap of Clouds and Precipitation as Would Be Inferred From Radar Observations

[14] We begin by constructing a "radar's eye view" of occurrence overlap in the model simulations. We extract the central column from the CRM domain at each model snapshot time (i.e., we sample 12 columns per hour) and compute the reflectivity profile that would be observed by a cloud-sensing radar. These instruments are sensitive to all hydrometeors (i.e., both clouds and precipitation), so we determine the reflectivity in each grid cell due to all forms of condensed water using the power law fits of reflectivity to concentration for each species described by *Luo et al.* [2003]. As in that paper, a cell is considered "cloudy" if the reflectivity exceeds -40 dBz, and the end result is a binary (yes/no) profile of cloudiness every five minutes.

[15] We determine overlap from these profiles following the techniques introduced by *Hogan and Illingworth* [2000]. We compute the hourly-averaged hydrometeor layer fraction c_i for every layer i by counting the fraction of cloudy cells in each layer during the hour. We use the term "hydrometeor fraction" to emphasize that this measure includes both clouds and precipitation.

[16] For each pair of layers (i, j) we compute the combined fraction C_{true} in each hour by counting the number of profiles in which hydrometeors exist in either or both of layer i and j. We also compute the theoretical combined fractions $C_{max} = \max(c_i, c_j)$ and $C_{ran} = c_i + c_j = c_i c_j$ that would be observed if the hydrometeors obeyed the maximum or random overlap assumptions. The combined hourly fractions are then averaged over the length of the simulation, excluding times when the hydrometeor fraction in either layer is exactly zero or one (in which case the overlap is irrelevant) and including only times when the hydrometeor fraction in all intervening levels is greater than zero (i.e., the hydrometeors are "contiguous"). The three time-averaged fractions are used to define an overlap parameter by solving

$$\overline{C_{true}} = \alpha \overline{C_{max}} + (1 - \alpha)\overline{C_{ran}} \qquad (1)$$

for the weighting α [*Hogan and Illingworth*, 2000]. The variable α is sometimes called the "overlap parameter," and has the value $\alpha = 1$ when layers are maximally overlapped and $\alpha = 0$ when layers are randomly overlapped. It may take

on negative values if the true combined fraction for a pair of layers is greater than would be computed from the random overlap assumption. We examine the value of α as a function of the distance Δz separating layers i and j, and estimate a characteristic length scale z_0 for α by fitting the data to the relationship

$$\alpha(\Delta z) = \exp(-\Delta z/z_0) \qquad (2)$$

using Gauss-Newton nonlinear least squares. The goodness of this fit can be evaluated by considering the normalized error variance [see, e.g., *DelSole and Chang*, 2003], the mean of the squared residuals divided by the variance of α; good fits have normalized error variance values much less than 1.

[17] Figure 2a shows the overlap parameter α determined from (1) for each pair of layers for all hydrometeors, including all clouds and precipitation. (We exclude pairs of layers with five or fewer observations in the course of the simulation so that estimates of α from each layer pair are made from a large enough sample.) Values of α decrease roughly exponentially with distance for the first five or six kilometers before becoming very broadly distributed about a value somewhat greater than 0. This behavior is consistent with radar observations [*Mace and Benson-Troth*, 2002] and other cloud resolving simulations [*Oreopoulos and Khairoutdinov*, 2003] at this location in this season. Hydrometeors are observed in continuous layers as deep as 13,660 m. The value of z_0 is 3676 m (see Table 1). Because layers separated by distances greater than a few kilometers are essentially uncorrelated, we also examine the value of z_0 estimated using only those layer pairs separated by less than 5 km. This does not change the estimate of z_0 very much but does improve the quality of the fit.

[18] In order to use the radar-derived overlap in a large-scale model, one has to make three assumptions: (1) that the overlap determined from all the radar-sensed clouds and precipitation may be equated with the overlap of large-scale clouds in a large-scale model, (2) that cloud fraction may be determined in every layer of the atmosphere simultaneously by averaging over a fixed time interval, regardless of varying wind speed or direction, and (3) that overlap determined from cloud fractions averaged over long times may be equated with the temporally averaged value of overlap determined instantaneously. Assumption 1 can be relaxed by using additional information from the cloud resolving simulations. This is relevant because most large-scale models assume that both convective clouds and precipitation are each maximally overlapped, and apply a separate assumption to stratiform (or "large-scale") clouds, while the simulated radar measurements lump all three kinds of hydrometeors together. We refine our calculations in two stages to make them more applicable to large-scale models. First, we identify clouds (as distinct from hydrometeors) as being those CRM grid cells containing nonzero amounts of cloud liquid and ice. Then we use the algorithm of *Xu* [1995] to find and remove the convective columns at each time step (these comprise about 4.7% of the total), leaving only stratiform clouds. The overlap behavior for all clouds and for stratiform clouds, respectively, is shown in Figures 2b and 2c. Clouds are substantially less vertically coherent than the mixture of clouds and precipitation ($z_0 = $
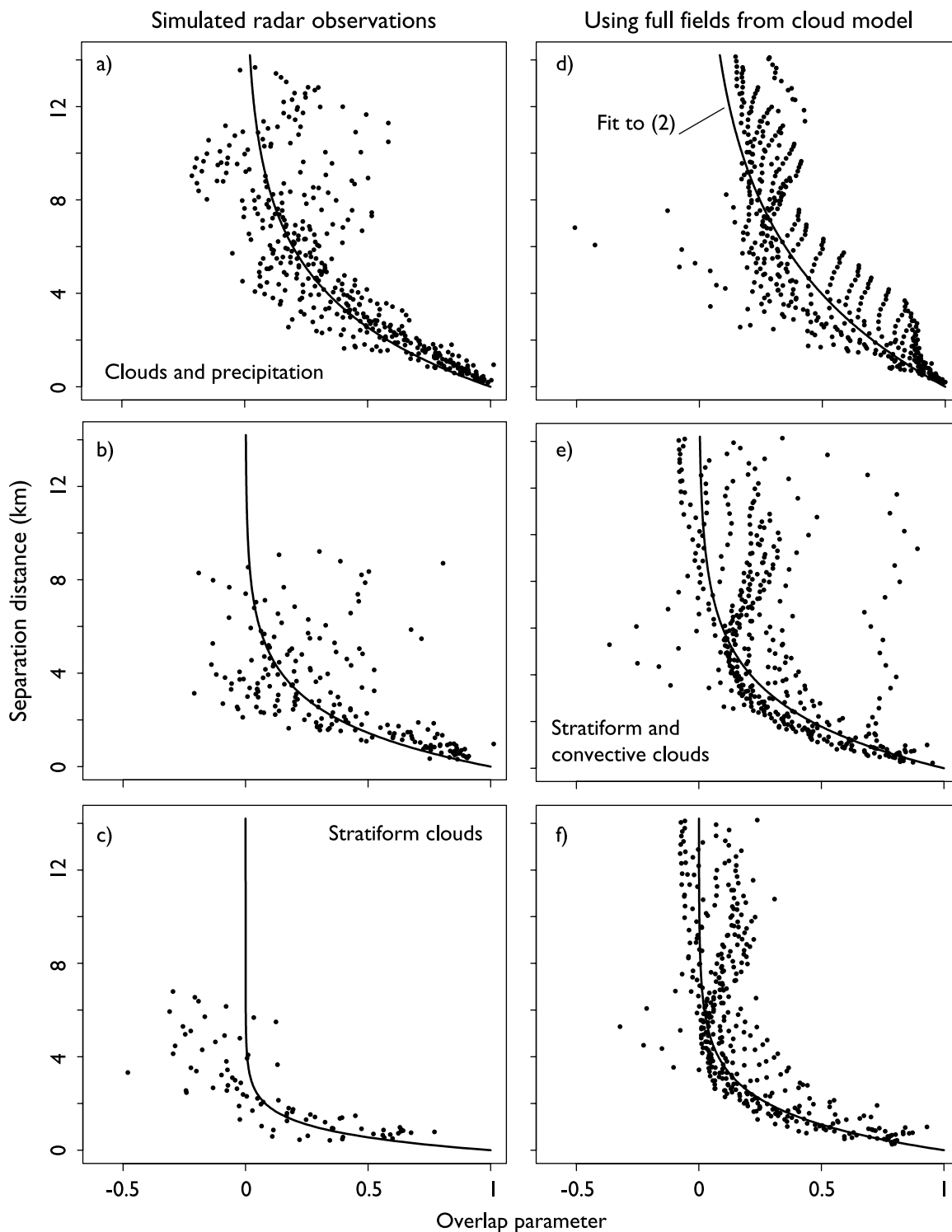
Figure 2. Occurrence overlap as a function of separation distance in a cloud-resolving model simulation of summertime continental deep convection. Overlap is characterized by the weight α describing the similarity of the observed overlap to maximum ($\alpha = 1$) and random ($\alpha = 0$) overlap. The rows show (a and d) the overlap for all clouds and precipitation as would be observed by a cloud radar, (b and e) the overlap for all clouds, and (c and f) the overlap for stratiform clouds alone. Figures 2a–2c use only information from the central column of the cloud resolving model domain, which has been processed to mimic radar observations, while Figures 2d–2f use information from the full model domain. Fits to the parameter z_0 in (2) using nonlinear least squares are indicated with lines.

Table 1. Scale Length z_0 for Occurrence Overlap Estimated by Fitting (2) to Cloud Resolving Model Simulations of Deep Convection[a]

	α Derived From Central Column		$\alpha(\overline{C_{true}}, \overline{C_{max}}, \overline{C_{ran}})$[b]		α Averaged Over Time[c]	
	All Δz	$\Delta z < 5$ km	All Δz	$\Delta z < 5$ km	All Δz	$\Delta z < 5$ km
Hydrometeors	3676 (0.21)	3569 (0.18)	6339 (0.28)	5944 (0.39)	5704 (0.30)	5241 (0.43)
All clouds	2085 (0.41)	1990 (0.29)	2569 (0.58)	2245 (0.33)	2812 (0.68)	2313 (0.33)
Contiguous stratiform clouds	793 (0.43)	795 (0.42)	1580 (0.24)	1561 (0.22)	1697 (0.33)	1634 (0.24)
All stratiform clouds	782 (0.66)	781 (0.50)	1481 (0.25)	1472 (0.26)	1560 (0.32)	1524 (0.28)

[a]The normalized error variance for each fit is shown in parentheses. Occurrence overlap is estimated for all clouds and precipitation, as would be observed by a radar, as well as for convective and stratiform clouds and for stratiform clouds alone. Estimates made from a time series extracted from the model's central column (imitating radar observations) differ substantially from those that account for the whole domain because the sample from the central column is so small. When cloud fraction is computed from the spatial variability at each time step, estimates of z_0 do not depend strongly on the order of averaging (i.e., whether z_0 is computed using (1), shown in the $\alpha(\overline{C_{true}}, \overline{C_{max}}, \overline{C_{ran}})$ columns, or with (3), shown in the α averaged over time columns). Layers separated by distance larger than a few kilometers are essentially uncorrelated, so restricting the fit to layers separated by less than 5 km does not affect z_0 but does improve the quality of the fit.
[b]Equation (1) applied to CRM fields.
[c]Equation (3) applied to CRM fields.

2085 m), which reflects the fact that heavy convective precipitation falls more or less straight down in this low-shear environment (i.e., precipitation is maximally overlapped, as large-scale models assume). Stratiform clouds are in turn less vertically coherent than the mixture of stratiform and convective clouds ($z_0 = 793$ m). This is because the algorithm of *Xu* [1995] identifies convective columns as those in which updrafts are strong, and clouds in these columns tend to be more vertically extensive than clouds subject to less vigorous vertical motions. The sensitivity of the scale length to the exact definition of "cloud" suggests that results for overlap obtained by radars in convective regions may be very sensitive to the subset of data chosen.

3.2. Occurrence Overlap of Stratiform Clouds Inferred Using Two-Dimensional Cloud Structure

[19] The overlap estimates in section 3.1. followed *Hogan and Illingworth* [2000] and so rely on assumptions about inferring layer and combined cloud fractions from a time series of profiles obtained at a single point (assumption 2), and about the ability to infer overlap by time-averaging the cloud fraction observations (assumption 3). Both assumptions are related to time-averaging, and neither is necessary when computing overlap from time-varying cloud-resolving model fields. Because the spatial variability is resolved, cloud fraction can be computed for every model layer at each time step, and combined cloud fractions and α computed for every model layer pair at each time step. That is, the cloud resolving model fields let us replace (1) with

$$\alpha = \frac{1}{N} \sum_{k}^{N} \frac{C_{true,k} - C_{ran,k}}{C_{max,k} - C_{ran,k}} \qquad (3)$$

where the layer cloud fractions and true, random, and maximum combined fractions for each pair of layers are computed at each time step k, and the overlap parameter computed from these combined fractions and averaged over time. As a helpful reviewer pointed out, (1) provides the best fit to time-averaged combined cloud fraction, while (3) the best fit time to time-averaged values of α. Radar observations could also be processed using (3), though this still requires accumulating observations over some length of time to compute a cloud fraction.

[20] We repeat the calculation of overlap for each of the three categories described in the last section (hydrometeors,

clouds, and stratiform clouds, are shown in Figures 2d, 2e, and 2f, respectively) using (3) applied to the full CRM fields. In all cases the length scales inferred from full CRM cloud fields are substantially longer than those inferred from the simulated radar observations, which we attribute to a combination of limited sampling and the application of the frozen turbulence assumption to the time series of profiles extracted from the central column. The cloud structure between any pair of layers varies with time as clouds evolve, and Figure 3 illustrates the variability of overlap in stratiform clouds by showing the number of times that a value of α occurs for a given separation distance during the course of the simulation.
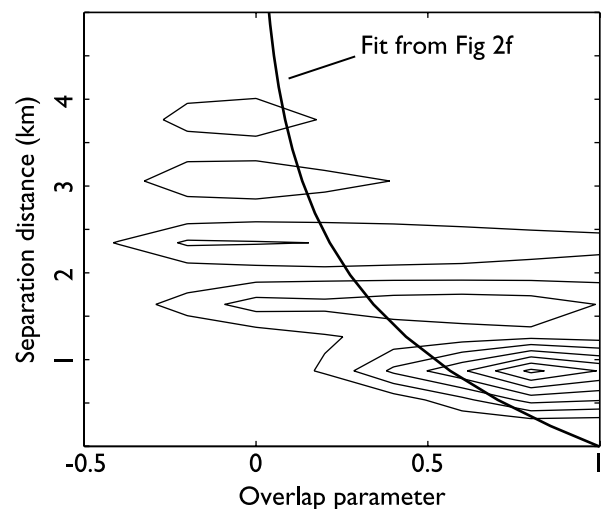

Figure 3. Number of 5-min intervals during the 29 day simulation during which pairs of layers separated by a given distance (the y axis) had a given value of the overlap parameter α. The overlap parameter is computed only for stratiform clouds in continuous layers. Also shown is the fit to (2) computed by averaging the value of α over time for each pair of layers with separation distances less than 5 km; this calculation underestimates the value of α (i.e., the degree of vertical coherence) for scales less than about 1.5 km. Contours are drawn for stratiform clouds in continuous layers every 1000 occurrences.

[21] We can use the high density of observations available from the full CRM fields to test the validity of assumptions 2 and 3, i.e., that overlap parameters determined from time-averaged cloud fractions using (1) reproduce time-averaged overlap parameters computed with (3). We make one set of calculations in which layer cloud fraction c_i and combined cloud fractions C_{true}, C_{max}, and C_{ran} are computed from the cloud model fields at each time step, then averaged over the duration of the simulation before determining the overlap structure (i.e., assumption 2 is relaxed but assumption 3 is retained, shown as the $\alpha(\overline{C_{\text{true}}}, \overline{C_{\text{max}}}, \overline{C_{\text{ran}}})$ columns of Table 1). We make another set of calculations in which α is computed for each layer pair at each time step and then averaged (i.e., both assumptions 2 and 3 are relaxed, shown in Figures 2d, 2e, and 2f and the α averaged over time columns of Table 1). As Table 1 shows, estimates of the overlap parameter made using (1) are quite similar to those made using (3), suggesting that assumption 3 works well.

4. Vertical Structure of Total Water

[22] The spatial distribution of clouds, including occurrence overlap, reflects the structure of the underlying total water field: clouds persist only where the total water concentration exceeds the local saturation vapor pressure. Assumed PDF cloud schemes (sometimes called statistical cloud schemes [e.g., *Tompkins*, 2002]) are built on this idea. These parameterizations assume a form for the distribution of total water q_t within each grid cell, then predict the evolution of this distribution based on sources and sinks linking the PDF to other (parameterized) physical processes. Cloud properties in each grid cell, including cloud fraction and the distribution of condensate, are diagnosed from the PDF of total water and the mean thermodynamic state. Sources and sinks for various parameters of the distribution may behave differently in clear and cloudy skies, but clouds are otherwise treated as a by-product of the total water distribution.

[23] Overlap assumptions for assumed PDF schemes are therefore applied to columns in which the PDF of total water in each layer is already specified. We suggest that overlap assumptions for these schemes might be formulated in terms of the rank correlation R_{q_t} of total water between layers. This quantity expresses the degree to which each part of the total water distribution in one layer is vertically aligned with the same part of the distribution in other layers. The rank correlation between two layers is high, for example, when the relatively dry and moist parts of the domain in one layer, respectively, are spatially correlated with the relatively dry and moist parts of the other layer, regardless of the actual values of total water in either layer. Equation (2) can be generalized to treat R_{q_t}:

$$R_{q_t}(\Delta z) = \exp(-\Delta z / z_0) \qquad (4)$$

[24] We estimate z_0 in (4) from the rank correlation of total water in the fields produced by the cloud resolving model. We calculate the rank of each cell's value of total water in the cumulative probability distribution function (CDF) of total water for each layer at each time step. (The rank is defined so that the cell with the smallest value of total water in that layer at that time step has rank 1, the cell

with the next smallest value has rank 2, and so on.) We then compute R_{q_t} from the model fields directly (rather than using, say, Spearman's formula) by calculating the spatial correlation of the rank for each pair of layers at each time step. (If, for example, the cells were arranged in space so that the lowest value in one layer was in the same horizontal position as the lowest value of the other layer, and similarly for every other value of rank, the rank correlation would be 1.) We include all nonconvective columns, including those with clouds in discontinuous layers and those that don't contain clouds at all, consistent with an assumed PDF cloud scheme.

[25] Figure 4 (left) shows the time-averaged rank correlation of total water as a function of the separation distance between layer pairs separated by less than 5 km. (The rank correlation at greater separation distances is near 0.) Figure 4 (right) shows the number of occurrences of each value of rank correlation as a function of separation distance. The solid line shows a fit to (4), which yields a value of $z_0 = 1263$ m with a normalized error variance of 0.091. This underestimates the rank correlation at small separation distances, however. When we fit (4) to the values of rank correlation obtained from neighboring pairs of layers (denoted as circles in Figure 4) the scale length increases to $z_0 = 2341$ m with a normalized error variance of 0.201 (dashed line). We explore the impact of emphasizing short- and/or long-range correlations in section 6. The quality of the fit to total water is as good or better than the fits to clouds alone.

5. On What Does Vertical Structure Depend?

[26] The overlap assumptions currently used in large-scale models are applied uniformly in every column at every time step. Before a parameterization based on (4) can be implemented (as in section 6, for example) the characteristic length scale z_0 for the rank correlation of total water must be specified. One could use a single fixed scale length for the rank correlation of total water between every pair of levels in every column at all times. However, radar observations in a range of locations [*Mace and Benson-Troth*, 2002] and cloud-resolving models run as a "super-parameterization" within a global model [*Räisänen et al.*, 2004] show that the length scales for cloud occurrence and the correlation of condensate vary with location, season, and, at any particular moment, height in the atmosphere. Some amount of the variability seen in vertical correlation is almost certainly related to the state of the atmosphere, including the profiles of temperature, humidity, wind, and so on. In the context of a large-scale model, the scale length might be parameterized to depend on both the large-scale state (i.e., the values of prognostic quantities within the grid cell) and the state of other parameterized subgrid-scale processes (i.e., the amount of turbulence within each layer.) In our simulations almost all non-convective clouds are produced by detrainment from deep convection, so we focus on the relationship between vertical structure and two quantities: wind shear and the strength of convection.

[27] We identify a set of observations (layer pairs at specific times) that are strongly affected by shear or convection. The amount of shear is defined as the absolute
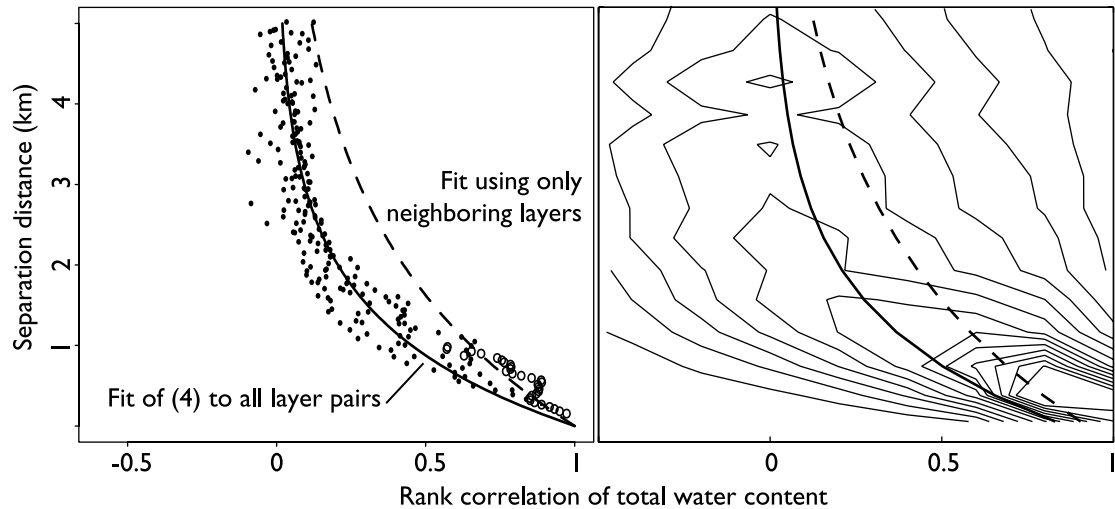
Figure 4. Rank correlation of total water in nonconvective columns, including both cloudy and clear layer pairs. (left) Relationship between each pair of layers averaged over the length of the cloud-resolving model simulation. Neighboring layers are shown as circles. (right) Contours showing the number of 5-min intervals during the simulation during which pairs of layers separated by a given distance (the y axis) had a given value of rank correlation of total water (x axis). Contours are in steps of 5000. Solid lines show a fit to (4) for all layer pairs shown in Figure 4 (left); dashed lines show a fit to those pairs of layers adjacent to one another.

value of the difference between the domain-averaged horizontal wind speeds in the two layers, which is dominated by the applied large-scale forcing. This definition of shear could be determined from the large-scale wind fields in a global model. (Because the CRM is two-dimensional, we cannot examine the ways directional wind shear might affect overlap properties.) The strength of convection is computed as the average convective mass flux between each pair of layers at each time step. (The convective mass flux is defined as the product of the reference density, the mean updraft velocity in the convective columns, and the fraction of the domain deemed convective.) Convective mass flux is a diagnostic quantity in many large-scale models. We build the cumulative distribution from the values of shear and convection strength for each layer pair at each time during the simulation. Observations at the extremes of the distribution of convective mass flux are considered to be strongly or weakly affected by convection, and similarly for shear.

[28] Figure 5 (left) shows the time mean average rank correlation of total water as a function of separation distance for those observations in the top and bottom 10% of the cumulative distribution of shear; Figure 5 (right) shows the 10% of observations when convection is strongest. Wind shear reduces vertical organization [*Hogan and Illingworth*, 2003]: the correlation length when shear is small ($z_0 = 1541$ m, shown in black) is only slightly larger than that computed from the full set of observations ($z_0 = 1263$ m), but decreases dramatically ($z_0 = 731$ m, in blue) in the most strongly sheared layers. Strong convection, on the other hand, produces abundant upper level anvils, but does not have a large effect on low-level clouds. This is evident when scale lengths are calculated for high- and low-level layers separately. In those observations most affected by convection, the scale length for the rank correlation of total water

for layer pairs above 6 km (in blue) is $z_0 = 3159$ m, while layers below 2 km are unperturbed (in red; $z_0 = 1269$ m), producing an intermediate value ($z_0 = 1826$ m; in black) for the whole domain. Figure 5 suggests that (4) might be useful in a variety of circumstances if z_0 can be adjusted appropriately, though the value of z_0 might be expected to depend in many ways on the state of the atmosphere.

6. Evaluating Overlap Assumptions

[29] How can the success of an overlap parameterization be judged? Overlap assumptions help determine column-integrated cloud statistics such as cloud fraction, and they are used by large-scale models to provide the information needed to compute radiation fluxes and precipitation/evaporation profiles. Here we evaluate the skill of various overlap assumptions by testing their ability to reproduce column-integrated cloud physical quantities, microphysical process rates, and top-of-the-atmosphere radiative fluxes computed from the original CRM fields.

[30] For each overlap assumption we construct a set of columns corresponding to each cloud model snapshot in which the horizontal variability (i.e., the PDF of total water in each layer) is sampled directly from the corresponding cloud model field, but the vertical structure is determined by the overlap assumption. We construct as many synthesized columns at each time step (256) as are contained in the original model fields.

[31] We test two overlap assumptions based on the rank correlation of total water between layers: (1) a "best case" scenario, in which the rank correlation between each neighboring pair of layers in each snapshot is taken directly from the cloud-resolving model fields, and (2) using (4) with values of z_0 (for all layer pairs and for neighboring layers)
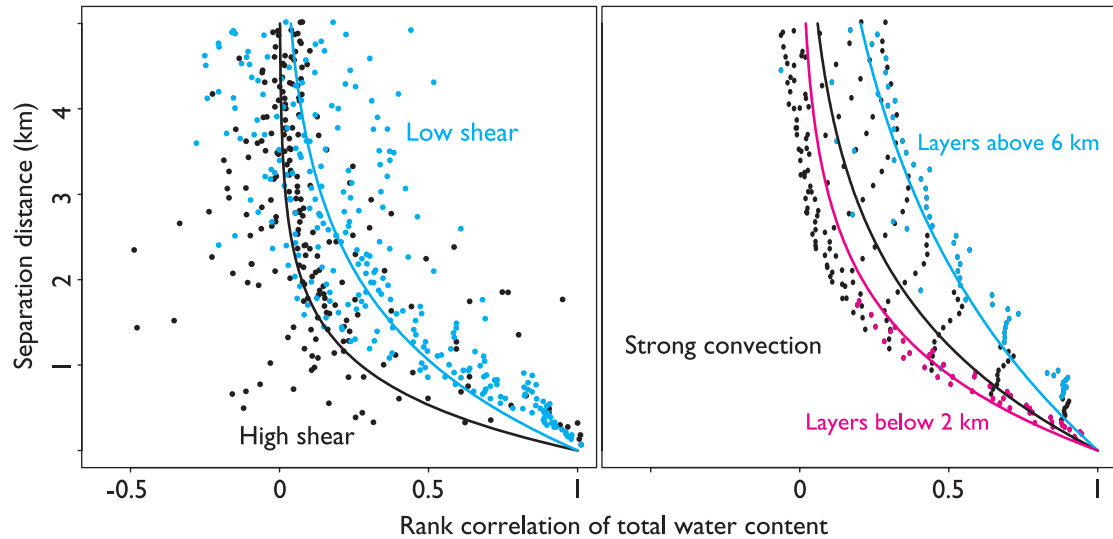
Figure 5. The dependence of rank correlation of total water on (left) wind shear and (right) the strength of convection. Wind shear is measured as the absolute difference between the domain-averaged horizontal wind speed between any pair of layers; convection strength is quantified using the average convective mass flux between each pair of layers. These diagnostic quantities are computed for every layer pair at every time step, and the rank correlation of total water computed for those layer pairs and times in the top and bottom 10% of the distribution. The times of greatest shear (in blue) are contrasted with times of no shear (in black); layers with large differences in wind speed are substantially less organized in the vertical. Convection acts to organize layers in the vertical, increasing the correlation length scale, but the effect on organization is more pronounced in the upper part of the troposphere. At times of strong convection, layers above 6 km (blue) have much longer correlation length scales than when convection is weak, while organization in the lower atmosphere (layers below 2 km, in red) is less sensitive to the strength of convection. These sensitivities moderate one another, so that the organization of the atmosphere as a whole (black) is only slightly sensitive to the convection. The best fit to (4) for each set of observations is shown as a solid line with the same color as the data points to which it is fit.

determined from Figure 4. We also include three overlap assumptions used in large-scale models (maximum, random, and maximum-random). These are extended to treat total water in the same way that (4) is a generalization of occurrence overlap expressed by (2), i.e., the rank correlation of total water between pairs of layers is 1 for maximum overlap and 0 for random overlap.

[32] We create sets of sample columns following the "cloud generator" described by *Räisänen et al.* [2004]. Within each column, we generate a random number $R1(z_i)$ for each layer z_i, uniformly distributed between 0 and 1. This number determines the rank (position within the CDF) of total water for the grid cell. We control the rank correlation between each pair of layers, where applicable, by mixing maximum and random overlap: we generate a second random number $R2(z_i)$ and compare this to rank correlation between layer i and layer $i + 1$; if $R2(z_i)$ is less than the rank correlation we set $R1(z_i) = R1(z_{i+1})$. (Choosing $R2(z_i)$ is not necessary when the maximum or random overlap assumptions are used, since rank correlation is, by definition, 1 for maximum overlap and 0 for random overlap.) This process proceeds from top to bottom.

[33] At each time step, the rank in each sample column is replaced by the value of total water taken from the PDF within that layer, so that a cell in which $R1(z_i) = 0.5$ will be assigned the median value of total water in layer i at the

time of the model snapshot. The proportion of ice and liquid, if applicable, from the original CRM cell are retained. An example of this process is shown in Figure 6 for a single cloud-resolving model snapshot.

[34] For each set of columns at each time step we compute domain mean vertically integrated cloud physical quantities, including the total vertically projected cloud fraction and the domain mean condensate (liquid plus ice) water path. We calculate domain-averaged process rates by applying the parameterizations from the Geophysical Fluid Dynamics Laboratory's AM2 global atmospheric model [*GFDL Global Atmospheric Model Development Team*, 2004] to each column and averaging the results. Microphysical process rates (column integrated ice settling and column integrated snow sublimation) are computed using the cloud scheme, which includes the cloud microphysics of *Rotstayn* [1997] and the fall speed/ice water content relationships from *Heymsfield and Donner* [1990]. (This is certainly not the best way to determine precipitation in a CRM, but it does provide a useful test of vertical structure.) Top-of-atmosphere reflected solar and outgoing longwave radiation are computed independently in each subcolumn assuming a solar zenith angle of 60°.

[35] We gauge the success of a given overlap assumption by comparing column-integrated cloud properties and domain mean process rates determined from the original fields
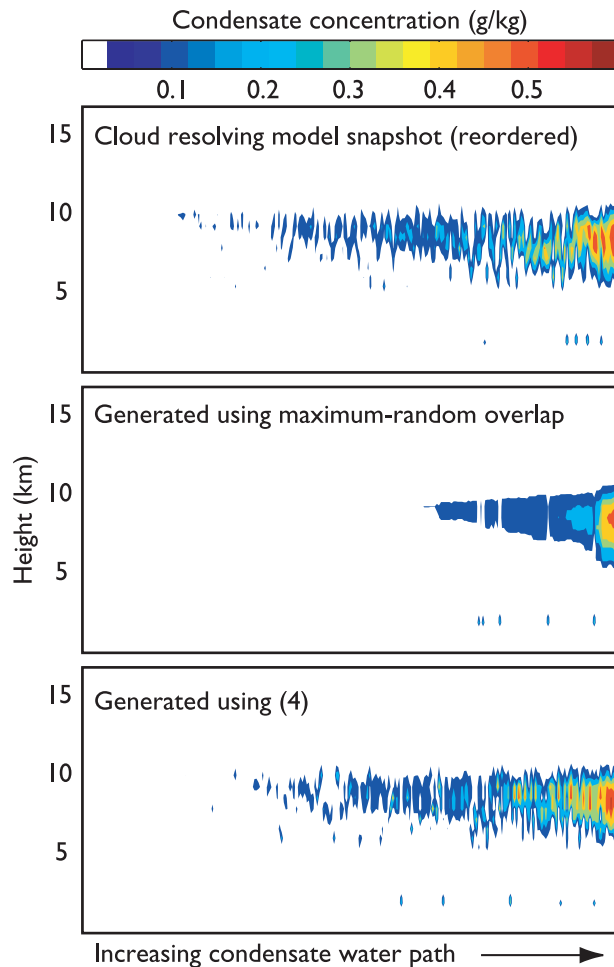
Figure 6. Cloud subgrid-scale structure (top) as produced by a cloud-resolving model and (middle and bottom) as created by two overlap assumptions. The columns in Figure 6 (top) are taken from a single time step of the cloud-resolving model and have been reordered so that the total condensed (ice plus liquid) water path increases from left to right. Figures 6 (middle) and 6 (bottom) show columns created randomly so as to reproduce the distribution of total water (including cloud fraction and mean condensate amount) from the cloud resolving model in each layer but with vertical structure imposed with overlap assumptions. Figure 6 (middle) uses the maximum-random overlap assumption, while Figure 6 (bottom) uses (4) at a scale length of 2341 m. Column integrated quantities show better agreement with the original fields when (4) is used.

ment in cloud properties and process rates provided by the best case scenario would be possible if all dependencies of vertical structure on atmospheric state were identified and correctly parameterized in a large-scale model.

[36] Assumptions using less detailed information (i.e., (4) using a constant scale length) lead to greater errors. Errors are smallest for many quantities when overlap is prescribed by (4) with $z_0 = 2341$ (the value inferred using neighboring pairs of layers in Figure 4), though the maximum or maximum-random overlap assumptions also perform well for some calculations. This reflects the fact that the various quantities used as performance metrics are sensitive in varying degrees to the vertical distribution of cloudiness. Outgoing longwave radiation, for example, depends mostly on the overlap in the first cloudy pair of layers, since OLR is unaffected by low clouds if the high clouds are opaque. Total cloud fraction, on the other hand, can remain sensitive to overlap assumptions over many partially cloudy model layers. Equation (4) describes the vertical structure at all separation distances, but underestimates short-range correlations even when the value of z_0 derived from neighboring layers is used (see Figure 4). This causes errors in those quantities (like OLR) that are most sensitive to short-range correlations. One could try to develop an alternate overlap assumption to address this failing. On the other hand, the best case scenario produces uniformly small errors, which suggests that effort might be better spent learning how to relate the scale length in (4) to the state of the atmosphere.

7. Implementation and Applications of Overlap Assumptions for Total Water in Global Models

[37] Parameterizations like (4) that depend explicitly on physical distances have two important advantages over, say, maximum-random overlap schemes: column-integrated statistics (e.g., total cloud cover) are not sensitive to the model's vertical resolution, and the parameterization's behavior varies smoothly with some adjustable parameter. This adaptability offers a way to tie vertical structure to the state of the atmosphere. We expect that overlap assumptions might strike a reasonable balance between complexity and accuracy by implementing (4) but allowing the scale length to change from layer to layer and place to place depending on the values of resolved and parameterized processes.

[38] Our results for the characteristic scale length and its dependence on large-scale wind shear and the amount of convection provide a starting point for parameterizations in global models, but are limited because we have relied on a specific two-dimensional simulation in a particular climate regime. Relationships between atmospheric state and vertical structure might be determined more generally by further analysis of measurements by ground-based active profiling instruments and by examining cloud-resolving model simulations of a wider variety of environments. In addition, two rich new sources of information may make the problem easier. Cloud-resolving models embedded as superparameterizations in global models [*Khairoutdinov and Randall*, 2001] see a wide range of large-scale conditions. In addition, spaceborne profilers on the upcoming CloudSat platform [*Stephens et al.*, 2002] will offer a view of clouds in the real atmosphere, providing snapshots more closely analogous to the state of a large-scale model, and avoid the pitfalls

(neglecting convective columns) at each time step with those determined from the synthesized fields. Bias and root-mean-square errors are shown in Table 2. The best case scenario indicates the lower bound of errors for overlap parameterizations expressed as layer-by-layer rank correlations. Errors in cloud properties and process rates arise in these calculations because we use a finite, if large, number of columns, which gives rise to sampling errors, and because the rank correlation of total water does not always exactly reproduce the overlap structure in the cloudy parts of the cloud-resolving model domain. The level of agree-

Table 2. Bias and Root-Mean-Square Errors in Cloud Physical Properties, Precipitation/Evaporation Rates, and Top-of-Atmosphere Broadband Radiative Fluxes for Five Overlap Assumptions[a]

Quantity	CRM	Best Case	(4) With $z_0 = 2341$ m	(4) With $z_0 = 1263$ m	Random	Maximum	Maximum-Random
Cloud fraction	42.2%						
Bias		−0.9	**2.2**	6.6	16.2	−9.8	−7.1
RMS		5.1	**7.7**	11.4	22.1	14.0	10.6
Mean condensate water path	10.3 g/m^2						
Bias		−0.8	−1.9	−2.8	−4.1	3.4	**−1.2**
RMS		3.7	**4.8**	5.7	7.3	8.2	5.4
Ice settling	524×10^{-6} mm/day						
Bias		7	**173**	342	790	−198	−255
RMS		181	413	678	1410	**347**	451
Snow sublimation	503×10^{-6} mm/day						
Bias		14.3	**182**	353	807	−190	242
RMS		188	427	687	1439	**356**	434
Reflection	148 W/m^2						
Bias		2.7	**10.6**	20.1	45.5	−15.5	−17.3
RMS		18.1	**27.2**	38.6	73.1	30.1	31.5
OLR	250 W/m^2						
Bias		−1.7	−4.8	−8.4	−17.4	**3.8**	7.1
RMS		6.8	10.5	14.7	26.2	**8.1**	12.0

[a]The errors are for quantities computed from synthetic cloud fields whose vertical structure is specified by the overlap assumption, as compared to the rates computed from the original cloud resolving model fields in which vertical structure arises from resolved cloud-scale dynamics. The distribution of total water in the synthetic fields matches that in the original fields; in the "best case" scenario the rank correlation for each pair of neighboring layers at each time is also taken from the original fields. The smallest mean and RMS errors in each row (excepting the "best case scenario") are bold. No overlap assumption produces the smallest errors for every test, though the parameterization described in (4) with $z_0 = 2341$ m has the lowest bias for many quantities.

associated with time averaging. Long time records and dense observations from both these sources may make it simpler to determine the first-order connections between the vertical structure and the state of the atmosphere.

[39] We have explored how overlap assumptions developed for cloud occurrence might be extended for use with cloud schemes that predict the distribution of total water in each model layer. To date, overlap assumptions have been applied only to cloud-related calculations, but this restriction is not necessary if overlap assumptions are formulated in terms of total water. The sets of columns discussed in section 6 vary in total water, which means that the relative humidity in clear air varies from place to place. Large-scale models using assumed PDF cloud schemes can also produce columns with variable relative humidity, and these might be used in physical parameterizations beyond those related to clouds and radiation. Aerosol deliquescence could be calculated separately in each column, for example, and the overlap assumption will play a role in producing a distribution of aerosol optical depths, just as it does in helping determine the distribution of liquid water path in clouds. Columns with variable relative humidly might also be used as inputs to convection schemes. If the scale length used by the overlap assumption increases during deep convection, some sample columns would have deep humidity anomalies and thus favor further deep convection via enhanced buoyancy. This local positive feedback might usefully represent the effects of the mesoscale organization of convection.

References

Cahalan, R. F., W. Ridgway, W. J. Wiscombe, T. L. Bell, and J. B. Snider (1994), The albedo of fractal stratocumulus clouds, *J. Atmos. Sci.*, *51*, 2434–2455.

Collins, W. D. (2001), Parameterization of generalized cloud overlap for radiative calculations in general circulation models, *J. Atmos. Sci.*, *58*, 3224–3242.

DelSole, T., and P. Chang (2003), Predictable component analysis, canonical correlation analysis, and autoregressive models, *J. Atmos. Sci.*, *60*, 409–416.

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

GFDL Global Atmospheric Model Development Team (2004), The new GFDL global atmospheric and land model AM2/LM2: Evaluation with prescribed SST simulations, *J. Clim.*, *17*, 4641–4673, doi:10.1175/JCLI-3223.1.

Heymsfield, A. J., and L. J. Donner (1990), A scheme for parameterizing ice-cloud water-content in general-circulation models, *J. Atmos. Sci.*, *47*, 1865–1877.

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

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

Jakob, C., and S. A. Klein (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.

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

Krueger, S. K. (1988), Numerical-simulation of tropical cumulus clouds and their interaction with the subcloud layer, *J. Atmos. Sci.*, *45*, 2221–2250.

Luo, Y. L., S. K. Krueger, G. G. Mace, and K. M. Xu (2003), Cirrus cloud properties from a cloud-resolving model simulation compared to cloud radar observations, *J. Atmos. Sci.*, *60*, 510–525.

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

Oreopoulos, L., and M. Khairoutdinov (2003), Overlap properties of clouds generated by a cloud-resolving model, *J. Geophys. Res.*, *108*(D15), 4479, doi:10.1029/2002JD003329.

Pincus, R., and S. A. Klein (2000), Unresolved spatial variability and microphysical process rates in large-scale models, *J. Geophys. Res.*, *105*, 27,059–27,065.

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

Rotstayn, L. D. (1997), A physically based scheme for the treatment of stratiform clouds and precipitation in large-scale models. 1. Description and evaluation of the microphysical processes, *Q. J. R. Meteorol. Soc.*, *123*, 1227–1282.

Rotstayn, L. D. (2000), On the "tuning" of autoconversion parameterizations in climate models, *J. Geophys. Res.*, *105*, 15,495–15,507.

Stephens, G. L., et al. (2002), The CloudSat mission and the A-train—A new dimension of space-based observations of clouds and precipitation, *Bull. Am. Meteorol. Soc.*, *83*, 1771–1790.

Tian, L., and J. A. Curry (1989), Cloud overlap statistics, *J. Geophys. Res.*, *94*, 9925–9935.

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.

Webb, M., C. Senior, S. Bony, and J. J. Morcrette (2001), Combining ERBE and ISCCP data to assess clouds in the Hadley Centre, ECMWF and LMD atmospheric climate models, *Clim. Dyn.*, *17*, 905–922.

Xu, K. M. (1995), Partitioning mass, heat, and moisture budgets of explicitly simulated cumulus ensembles into convective and stratiform components, *J. Atmos. Sci.*, *52*, 551–573.

Xu, K. M., and S. K. Krueger (1991), Evaluation of cloudiness parameterizations using a cumulus ensemble model, *Mon. Weather Rev.*, *119*, 342–367.

Xu, K. M., et al. (2002), An intercomparison of cloud-resolving models with the atmospheric radiation measurement summer 1997 intensive observation period data, *Q. J. R. Meteorol. Soc.*, *128*, 593–624.

Yu, W., M. Doutriaux, G. Seze, H. LeTreut, and M. Desbois (1996), A methodology study of the validation of clouds in GCMs using ISCCP satellite observations, *Clim. Dyn.*, *12*, 389–401.

Zhang, M. H., J. L. Lin, R. T. Cederwall, J. J. Yio, and S. C. Xie (2001), Objective analysis of ARM IOP data: Method and sensitivity, *Mon. Weather Rev.*, *129*, 295–311.

────────────────

C. Hannay, National Center for Atmospheric Research, Boulder, CO 80307, USA.

R. Hemler, Geophysical Fluid Dynamics Laboratory, NOAA, 201 Forrestal Road, Princeton, NJ 08542, USA.

S. A. Klein, Atmospheric Sciences Division, Lawrence Livermore National Laboratory, 7000 East Avenue, Livermore, CA 94550, USA.

R. Pincus, Climate Diagnostics Center, NOAA-CIRES, 325 Broadway, R/CDC1, Boulder, CO 80305, USA. (robert.pincus@colorado.edu)

K.-M. Xu, NASA Langley Research Center, Hampton, VA 23681, USA.