Evaluating statistical cloud schemes: What can we gain from ground-based remote sensing?

V. Grützun,[1,2] J. Quaas,[3] C. J. Morcrette,[4] and F. Ament[2]

Received 21 January 2013; revised 2 September 2013; accepted 4 September 2013; published 19 September 2013.

[1] Statistical cloud schemes with prognostic probability distribution functions have become more important in atmospheric modeling, especially since they are in principle scale adaptive and capture cloud physics in more detail. While in theory the schemes have a great potential, their accuracy is still questionable. High-resolution three-dimensional observational data of water vapor and cloud water, which could be used for testing them, are missing. We explore the potential of ground-based remote sensing such as lidar, microwave, and radar to evaluate prognostic distribution moments using the "perfect model approach." This means that we employ a high-resolution weather model as virtual reality and retrieve full three-dimensional atmospheric quantities and virtual ground-based observations. We then use statistics from the virtual observation to validate the modeled 3-D statistics. Since the data are entirely consistent, any discrepancy occurring is due to the method. Focusing on total water mixing ratio, we find that the mean ratio can be evaluated decently but that it strongly depends on the meteorological conditions as to whether the variance and skewness are reliable. Using some simple schematic description of different synoptic conditions, we show how statistics obtained from point or line measurements can be poor at representing the full three-dimensional distribution of water in the atmosphere. We argue that a careful analysis of measurement data and detailed knowledge of the meteorological situation is necessary to judge whether we can use the data for an evaluation of higher moments of the humidity distribution used by a statistical cloud scheme.

Citation: Grützun, V., J. Quaas, C. J. Morcrette, and F. Ament (2013), Evaluating statistical cloud schemes: What can we gain from ground-based remote sensing?, *J. Geophys. Res. Atmos.*, *118*, 10,507–10,517, doi:10.1002/jgrd.50813.

1. Introduction

[2] Feedbacks between clouds and the climate system have been identified as one of the greatest challenges in refining climate change projections [*LeTreut et al.*, 2007]. Though many processes in clouds are well known, details about the cloud physics remain a challenge and are not completely understood. Even if they were, their inclusion in global climate models (GCMs) is not straight forward since there is a huge gap between the scales of the physics and the model grid: Cloud physical processes start in the range of nanometers to millimeters for cloud nuclei and droplets and range up to many kilometers for cloud systems. Global climate models have horizontal grid lengths of hundreds of kilometers, and regional climate models use grid lengths of tens of kilometers. Thus, clouds are normally subgrid-scale features in these models and need to be parameterized.

[3] Of major interest in the simulation of climate is the interaction of clouds with radiation. Thus, a correct prediction of cloud cover within the grid boxes is essential. A common approach to calculating the grid box mean cloud cover is to assume that there is a subgrid-scale distribution of humidity within the grid box and to calculate the cloud cover from the portion of the distribution that is above saturation. The distribution may be a simple top hat function with a fixed distribution width as in *LeTreut and Li* [1991], or a more sophisticated distribution described by a beta function or multiple Gaussians. Such sophisticated schemes predict higher moments like variance and skewness and are called statistical cloud schemes [e.g., *Tompkins*, 2002]. They represent the cloud cover in a more physical way and have the potential to describe the subcloud variability as well. Some of these schemes even use joint distributions of total water, temperature, and the vertical wind speed [e.g., *Golaz et al.*, 2002; *Zhu and Zuidema*, 2009] and thus attempt to capture the physics of clouds even better. However, there is no detailed knowledge about the exact dependence of variance and skewness of the total water distribution on cloud processes such as turbulence and convection and many closure assumptions have to be made.

[1]Max Planck Institute for Meteorology, Hamburg, Germany.
[2]Meteorological Institute, University of Hamburg, Hamburg, Germany.
[3]Leipzig Institute for Meteorology, University of Leipzig, Leipzig, Germany.
[4]Met Office, Exeter, UK.

Corresponding author: V. Grützun, Max Planck Institute for Meteorology, Bundesstr. 53, 20146 Hamburg, Germany. (verena.gruetzun@zmaw.de)

©2013. American Geophysical Union. All Rights Reserved.
2169-897X/13/10.1002/jgrd.50813

[4] To be confident that such sophisticated cloud schemes work properly, a thorough evaluation of the prognostic higher-order moments of variance and skewness is desirable. A simple way to estimate these higher moments would be to apply high-resolution modeling and perform a statistical analysis of the model variables. The downside of this method is that a very high resolution in the order of 10–100 m is needed to simulate all relevant small-scale features of a cloud field. For one, extensive realistic weather simulations on such scales are not available yet. Another, even at such high resolutions, still some processes such as cloud microphysics and turbulence remain partly subgrid scale and need to be parameterized. We would prefer to access the statistics through high-resolution measurements of the atmosphere.

[5] Suitable observation data and methods to do so are scarce. One approach is to investigate the statistics of observed column-integrated quantities like water vapor and cloud water or ice from e.g., MODIS [*King et al.*, 2003], but these products cannot give detailed information about the deficiencies on the process level, because they lack the required vertical resolution. Vertically resolved information at a global scale, in turn, has so far only be used in a simplified framework [*Quaas*, 2012]. Data on cloud properties at certain heights with high temporal resolution can be obtained from aircraft measurements [*Duynkerke et al.*, 1999; *Tompkins*, 2003; *Boutle et al.*, 2013]. If airborne remote sensors are present, a high vertical resolution can be obtained. However, flights cannot deliver long time series and long statistics of the situation but only "snapshots" on certain flight legs.

[6] Longer time series originate from long-term ground-based remote sensors, such as lidar, radar, and microwave radiometers at certain locations. Combinations of various instruments as described, e.g., in *Löhnert et al.* [2004] yield additional quantities, which cannot directly be measured. The high temporal and vertical resolution of such measurements seems to have a unique potential to gain statistical information about the atmospheric state to evaluate statistical cloud schemes. The great challenge in using these data, however, is to compare time statistics of point measurements, i.e., comparably tiny vertical columns measured from one location at the ground, with the spatial snapshot statistics from the huge model grid boxes.

[7] Such an approach has been used by many authors in order to evaluate first-order moments such as the grid box mean of variables like cloud cover, liquid water path, and liquid water content [e.g., *Mace et al.*, 1998; *Hinkelman et al.*, 1999; *Hogan et al.*, 2001; *van Meijgaard and Crewell*, 2005; *Bouniol et al.*, 2010; *Morcrette et al.*, 2012]. It is not clear, however, whether the same approach can be used to evaluate higher-order moments such as variance and skewness of the moisture distribution since only few studies have been done on the applicability of the methods, and they indeed focus on the comparability of space-borne or quasi-instantaneous 2-D line measurements with 3-D quantities [e.g., *Astin et al.*, 2001; *Uttal and Kropfli*, 2001; *Hennemuth et al.*, 2005].

[8] In this paper, we investigate in detail the comparability of high temporal resolution single-column data with high spatial resolution but instantaneous three-dimensional data. We explore the potential use of ground-based remote-sensing data to evaluate statistical cloud schemes by using a "perfect model approach." We employ the mesoscale, partly cloud-resolving COSMO-DE model [*Baldauf et al.*, 2011] of the Deutscher Wetterdienst (DWD, German Weather Service) as a "virtual reality" and simulate two cases. The first case is a simulation of the weather over Germany on 16 June 2009, a day with synoptic forcing, which resulted in a cloudy day with some moderate rain. Our second case is 19 September 2012, a postfrontal case with locally initiated convective clouds and no rain in our region of interest. We concentrate on one COSMO-DE grid point serving as "virtual supersite" and construct a larger box around it, corresponding to a grid column of the global circulation model ECHAM6 (European Centre/Hamburg (GCM model)) [*Stevens et al.*, 2013]. Through COSMO-DE's high spatial and temporal resolution, the model offers both good spatial statistics within the region representing the GCM grid box and good temporal statistics, i.e., a statistics over a certain length of a time series, with a high vertical resolution at the COSMO-DE grid point. Thus, the virtual reality setup allows us to mimic the situation we would have when evaluating a climate model with ground-based remote sensors.

[9] Regardless of whether the COSMO-DE model captures the correct weather situation and all the subgrid-scale processes well or not, we get a data set from the simulation which is in itself consistent. In our virtual reality, we have a perfectly simulated model grid box (i.e., the spatial statistics in the global grid box) which is entirely consistent with the observations (i.e., our virtual measurement). Thus, any deviation between the statistics derived from the virtual observations to those derived from the virtual GCM grid box can exclusively be attributed to a lack of representativity when using the time-height data.

[10] The paper is organized as follows: we describe the methods we used in section 2, then we show and discuss our results and present idealized artificial cases for clarification of the results in section 3, and provide final thoughts and conclusions in section 4.

2. Methods

2.1. What Do We Want to Gain From the Evaluation?

[11] To evaluate methods of testing statistical cloud schemes with observational data, we first have to ask ourselves what we actually need as data. Naturally, we are interested in the (total) cloud cover, which is the "end result" of our scheme. We can get the cloud cover from satellites. We need to use the respective satellite forward operator in the model and keep in mind the specific challenges of the satellite retrieval and the model operator.

[12] However, once we realize that the cloud cover is not well simulated, we would like to figure out why this is the case. We can get an estimate by analyzing the error region-wise [*Weber et al.*, 2011], knowing that, for example, in the tropics, convective processes are the dominant processes [e.g., *Morcrette*, 2012]. But still, ideally, we would like to assess the error on the process level to reformulate and correct the terms in the equations, and for this we need detailed vertically resolved data.

[13] For an evaluation on the process level we need to focus on the specific distribution moments used in the

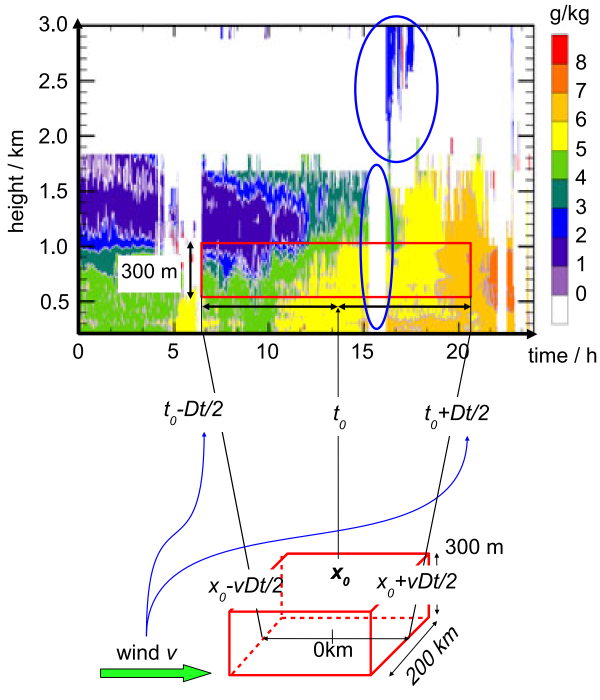

Figure 1. Retrieval of temporal statistics from time series corresponding to spatial statistics within the model grid box. (top) Example measurement of water vapor mixing ratio from a differential absorption lidar located in Hamburg. The blue circle marks areas where the data have to be excluded or corrected. The red "time-height box" inside the plot corresponds to the (bottom) red global model grid box assuming a constant wind speed *v*.

scheme. In the case of the Tompkins scheme [*Tompkins*, 2002], for example, we need the mean, the variance, and the skewness of the horizontal distribution of total water on each model level. Ideally, this would be a spatial statistics at a specific time, but since we have no access to the full 3-D information of the atmosphere from observations, time statistics over high-resolution ground-based measurements are often thought to be the way to get an idea about how the model is behaving.

[14] The direct approach to compare the time series statistics with the spatial statistics is to "just wait long enough." For the time of the spatial output from the model, we go to the same time in the time series and put a time window around it; the duration of which depends, for example, on the wind speed. This is illustrated in Figure 1 at the example of a differential absorption lidar (DIAL) data set measured in Hamburg (H. Linné, personal communication, 2010).

[15] Because the time window is chosen to be symmetric around the base time, the time t_0 corresponds to the center point of the grid box. Assuming a grid box width of, e.g., 200 km and a wind speed of 10 m/s, the width of the time window would be $\approx$ 5.6 h. Figure 2 shows an example of such a retrieval with a fixed window size of $\approx$ 5.6 h for a DIAL data set from 16 May 2006, measured in Hamburg. We notice that the general shape of the beta distribution employed by *Tompkins* [2002] fits the data well, apart from where there are double peaks in the data set. In addition we find negative skewness, for example, for 0.6–0.9 km at time 5.6–11.2 h, which is not allowed by the current implementation of that cloud scheme, and which seems to point to errors in the scheme. The data look tempting to be used for a thorough evaluation of our cloud scheme, because the distributions look so nice and seemingly realistic already.

[16] On a closer look, immediately two challenges arise: first, the wind speed might not be the same on all levels. This could easily be corrected by choosing different time window lengths at the different levels [*Illingworth et al.*, 2007]. However, this assumption would only work if the mean wind vector is a fair representation of the true wind over the time window. We ignore this problem in this study, since the following point is much more severe and more difficult to tackle: While the spatial data set is a snapshot in time, the time series obviously undergoes a time evolution due to, e.g., the diurnal cycle or a changing weather situation, which is not negligible in all cases. Also, our DIAL represents only a tiny column of the GCM grid box. How representative is one location of an entire GCM grid box whose sides are in the range of tens to hundreds of kilometers? Although this has not been clarified, line and point measurements have

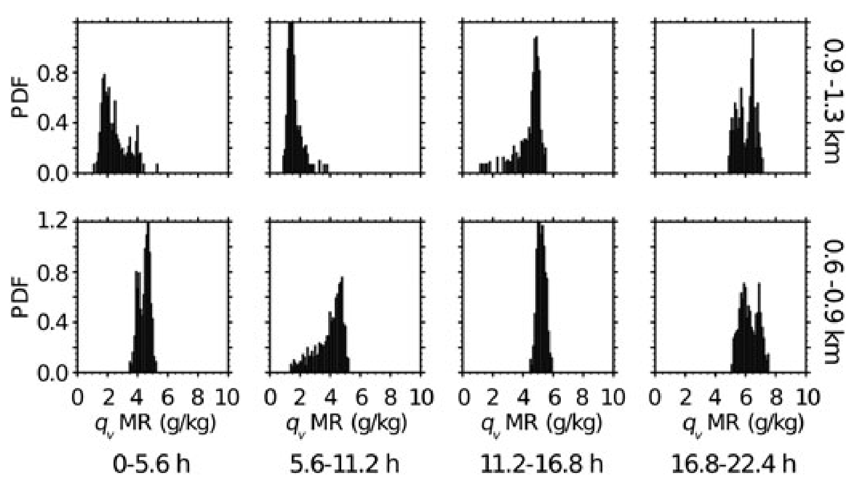

Figure 2. Example distributions of water vapor mixing ratio (*qv*) retrieved from the DIAL measurement shown in Figure 1 for (top and bottom) different heights and (from left to right) different times t_0.

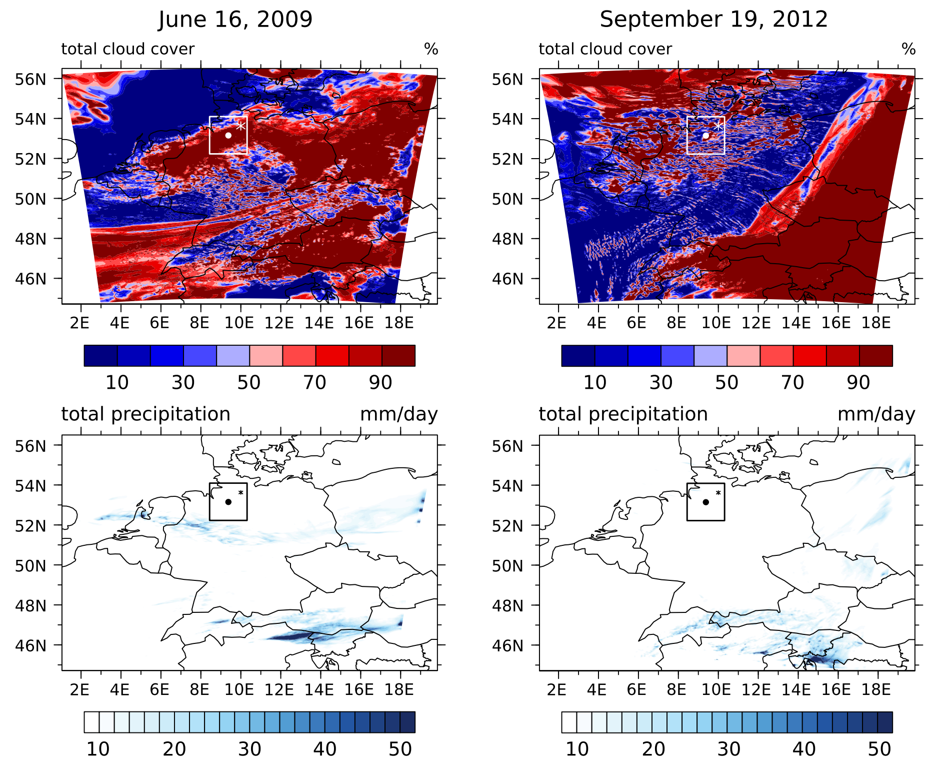

Figure 3. (top left) COSMO-DE cloud cover for 16 June 2006 after 12 h of simulation and (bottom left) accumulated precipitation after 24 h. The boxes over Northern Germany denote the grid box of ECHAM6 (T63 resolution) containing Hamburg, which we use for our evaluation, the dot marks the center of the box, and the asterisk Hamburg, respectively. The cloud cover and precipitation for 19 September 2012 are shown in the right column.

frequently been compared with model data representing spatial distributions in entire grid boxes. In the following, we investigate this issue and analyze the usability of point or line measurements for evaluating GCM subgrid-scale total water variability.

2.2. Testing the Method

[17] COSMO-DE [*Baldauf et al.*, 2011] is the operational weather forecast model of DWD. The model covers Germany and the surrounding regions within Europe and has a horizontal resolution of approximately $2.8 \times 2.8\,\mathrm{km}^2$ ($0.025° \times 0.025°$, see Figure 3). Hybrid terrain-following Gal-Chen coordinates [*Gal-Chen and Sommerville*, 1975] are used with a vertical resolution of about 20 m at the bottom, up to 1 km at the top of the atmosphere and 50 height levels reaching up to 22 km.

[18] The atmospheric profiles obtained at one COSMO-DE grid box with a high temporal resolution are referred to as "virtual supersite" for simplicity from now on, though we can not only retrieve quantities a real supersite would see but also the cloud water, vertical and horizontal winds, temperature, and any desired quantity a "virtual experimentalist" might wish for. The virtual supersite observes the atmospheric column in 25 s intervals, which corresponds to the COSMO-DE model time step.

[19] For our investigation, we chose the COSMO-DE model grid box containing the city of Hamburg, Germany, at 53.6° N, 10.0° E as location for the virtual supersite (indicated in Figure 3). The corresponding ECHAM6 grid box on the resolution of T63 is located between 52.2 and 54.1° N ($\approx$ 211 km), and between 8.4 and 10.3° E ($\approx$ 127 km). It contains 3336 COSMO-DE grid points and is indicated as white and black boxes, respectively (Figure 3). We also evaluate the data at the center point of this global box, being located at 53.1° N, 9.4° E. The region around Hamburg is convenient for several reasons: First, a real DIAL was located at Hamburg for some years providing us with measurements which initiated this study on the method itself. Second, the area around Hamburg is relatively flat (the maximum difference in COSMO-DE elevation within the ECHAM6 global box is about 270 m, with the highest elevation in the Southern edge of the box), such that there are no distracting effects from local orographic phenomena nor from numerical issues associated with terrain-following coordinates, which can cause large variances if not treated with care.

[20] The quantity we are going to look at is the total water mixing ratio, i.e., the ratio of the mass of total water to the mass of the dry air, because this is the most important quantity for determining the cloud-related process for the calculation of cloud cover in a statistical cloud scheme. For these schemes, of particular interest is its subgrid-scale (to the global model) distribution.

[21] In this study we focus on the mean, variance, and skewness (first, second, and third moment) of the distribution. For discrete values these are defined as

$$\text{mean} \;\; \bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i, \tag{1}$$

$$\text{variance} \ \ \nu = \frac{1}{N-1}\sum_{i=1}^{N}(x_i-\bar{x})^2\,, \tag{2}$$

$$\text{skewness} \ \ \zeta = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{x_i-\bar{x}}{\sqrt{\nu}}\right)^3, \tag{3}$$

with x the quantity which is evaluated, x_i the values at the different times or locations, and N the number of points in the data set.

[22] The first case chosen for this simulation was 16 June 2009. This day was a cloudy and also partly rainy summer day over Europe, subject to a large-scale forcing, with a cloud system slowly moving over Germany from the northwest to the southeast. No precipitation fell in the ECHAM6 grid box domain on this day and only a negligible amount of ice clouds was present for some time at a small number of COSMO-DE grid points, near the upper edge of our considered height range of 4 km. We choose such a situation to exclude the influence of complex microphysics on the development of the distribution. Figure 3 shows the total cloud cover after 12 h of simulation and the accumulated precipitation after 24 h.

[23] The second case is 19 September 2012. This is a postfrontal case and our focus is on locally initiated convective clouds. As in the previous case, no precipitation fell in our target domain and only a small amount of ice occurs at a small number of grid points below 4 km height. This case was chosen as an example for a situation with a variability on smaller scales, which appears more favorable for a successful application of our evaluation method. The respective cloud cover after 12 h of simulation and the accumulated precipitation after 24 h are also shown in Figure 3.

3. Results

3.1. Perfect Model Approach to Evaluation

[24] In the following, we first focus on 16 June 2009, which is the synoptically forced day. Figure 4 shows the time-height evolution of the mean, variance, and skewness of the spatial total water distribution within the virtual ECHAM6 box derived from all the COSMO-DE grid boxes inside it. The distribution moments have been calculated for each of the model output time steps (every 15 min). The black contours show the condensed water mixing ratio, i.e., the location of the clouds. Remembering that we look at the model as a virtual reality, these plots are what a statistical cloud scheme should produce if it worked perfectly. These statistics are also what observations would need to give us in order to be of use for evaluating the cloud scheme.

[25] As mentioned before, the direct way to retrieve a statistics is to wait some time. The estimation of a correct waiting time is challenging. In the following we chose different waiting times, which are constant with height. Assuming a grid box size of 200 km, they correspond to wind speeds of 40 m/s (1.4 h), 10 m/s (5.6 h), and 5 m/s (11.2 h). We shift the center point of the box through the whole time series with a 15 min time step, which corresponds to the spatial model output time step.

[26] Figure 5 shows the statistics which were derived from the virtual time series at the Hamburg grid point at 53.6° N, 10.0° E of the COSMO-DE model (cf. Figure 3). The white spaces at the beginning and end of the day correspond to half a time window and are due to the choice of the center point to evaluate the statistics inside this window. For a first evaluation it would be desirable to get at least the time-height structures of the mean and variance and the sign of the skewness right compared to the spatial statistics shown in Figure 4. The larger the time window, the smoother the evolution of the moments, which is reasonable because (a) the statistics are based on a much larger data set if we include more times and (b) the time windows overlap more the larger they are. Thus, data points can contribute to and influence several subsequent time windows.

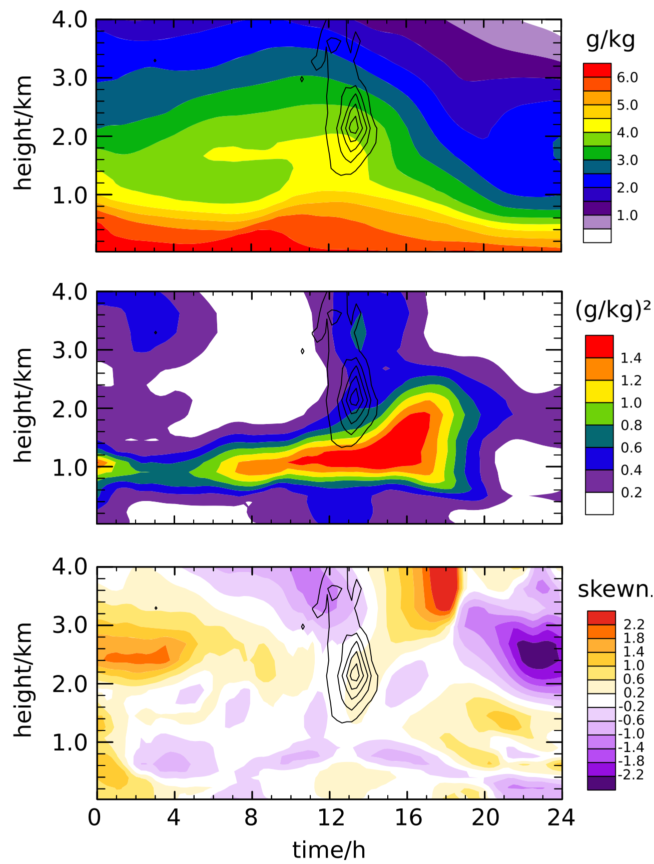

Figure 4. Time-height slices of spatial statistics of total water mixing ratio in the virtual global grid box for 16 June 2009: (top) mean, (middle) variance, and (bottom) skewness, retrieved from the 15-minutely spatial output of COSMO-DE. The black contours show the grid box mean cloud water mixing ratio with a contour spacing of 0.005 g/kg.

[27] From the short time window, even though the mean value is captured remarkably well, we cannot gain much information about the higher moments of variance and skewness. There are only a few times and heights with noticeable variance, and the skewness is quite noisy. However, structures are noticeable in the skewness, like the development of the boundary layer over the day. The higher accuracy of the mean values becomes clear from the definitions of the three moments as given in equations (1)–(3). We see that the distance of a single point to the mean value comes into play to the power of two in the variance and to the power of three in the skewness. Thus, the higher moments are much more sensitive to outliers.

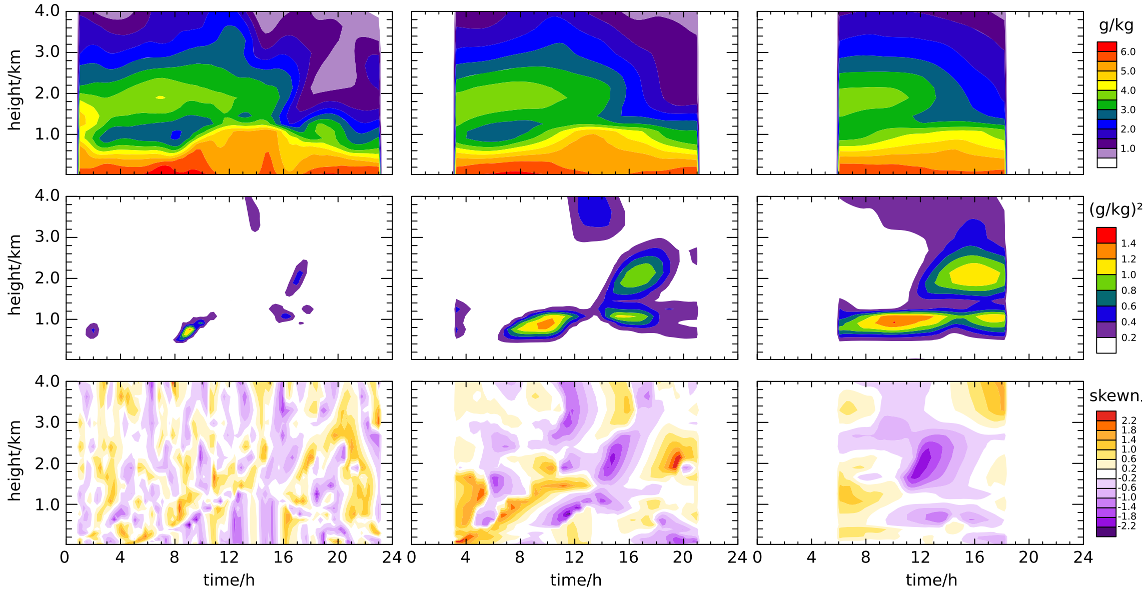

Figure 5. Time-height slices of time statistics for 16 June 2009: (top row) mean, (middle row) variance, and (bottom row) skewness for the location of Hamburg. A 15 min time resolution was chosen corresponding to the spatial output time increment of COSMO-DE. Results are shown for a time window of (left column) 1.4 h, (middle column) 5.6 h, and (right column) 11.2 h.

[28] For the 5.6 h time window the structures are clearer. Also, at this window size larger areas of noticeable variance occur, and its overall structure exhibits some similarities with the reference spatial statistics shown in Figure 4. The quantitative values are at least in the desired range to match the reference spatial statistics. Comparing the skewness, we notice that time, space, and even the sign of the skewness do not match well. Hoping the statistics become more reliable as we use a longer time window, we also show, as an example, statistics from a 11.2 h time window, which represents a length twice as large as our grid box for the assumed wind of 10 m/s. However, save for the fact that the statistics become less noisy, which is especially evident in the skewness, we do not get a much better estimate of the higher moments at all.

[29] From these first results another challenge arises: The values of the moments are highly sensitive to the time window. So an accurate window size is important. The only quantity which is available to translate time into space is the wind speed. A detailed knowledge of the wind profile however, will not necessarily help us to sample the global grid box decently. In meteorological situations with highly variable wind speed, either in terms of magnitude or direction, we may not sample a fraction of the air mass which is representative for the whole grid box.

[30] Looking at Figure 3 again, one might argue that the Hamburg grid box of the COSMO-DE model is a bad choice for the evaluation, since it is located near the upper right corner of our global ECHAM6 grid box. Thus we also evaluated the time series at the center point of the box, and the respective results are shown in Figure 6. There is only a negligible improvement of the results, and they remain unsatisfactory for our purposes of evaluating a statistical cloud scheme on the process level.

[31] In the model world we have the opportunity to circumvent the time versus space problem by introducing a wide network of virtual supersites. Figure 7 shows results from a network of 100 supersites evenly spaced in the Hamburg global grid box, sampling the spatial distribution at each time step of the model. The statistics is done over these 100 supersites for each model output time step. The structures in the mean, variance, and skewness from the spatial statistics are captured remarkably well. Even the sign of the skewness is captured well in the lower two kilometers but misses some features in the upper part of the height range shown. However, due to the few points in space, i.e., a poor statistics where each point has a big influence on the higher moments, the variance is heavily overestimated. In principle, a dense network of around 100 observational sites located within a region comparable to a GCM grid box would be a step in the right direction, but in reality such a high number of facilities measuring high vertical resolution atmospheric profiles would be far too expensive to implement.

[32] So far, we have only focused on one of our cases, which contained a large-scale cloud system moving over Germany and found that the evaluation method did not work very well. We now look at our case from 19 September 2012, where the clouds are locally initiated. The cloud field in our target ECHAM6 grid box in this case varies on smaller scales and has more random features than the previous case (Figure 3). The hypothesis is that such a situation is more favorable for an application of the method. From the previous results we saw that the short time window resulted in a noisy statistics and that the long time window did not give better results than the medium-sized one. Thus, for this case we only show the medium time window of 5.6 h.

[33] Figure 8 shows the spatial statistics from the ECHAM6 grid box domain, the temporal statistics from the Hamburg grid point, and the temporal statistics from the center point of our considered box. Again, the mean is captured well, although some structure is missed by the virtual supersite at the Hamburg grid point. The spatial variance in this

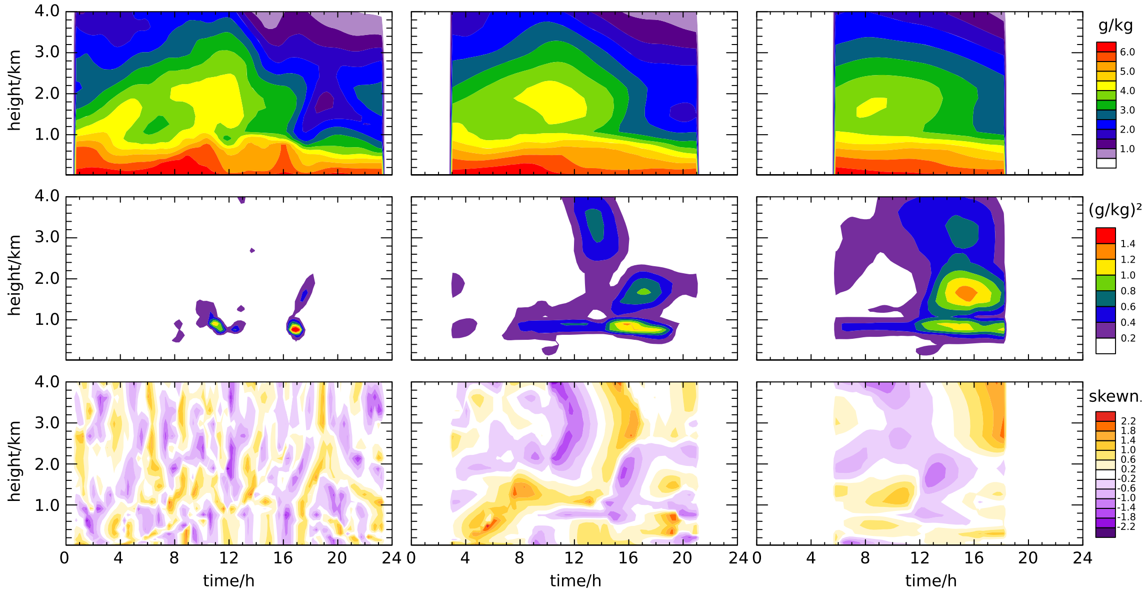

Figure 6. Same as in Figure 5 but for the center point of the box.

case is smaller than in the previous case. Some basic features of its structure are found in the temporal statistics of both virtual supersites, but the similarities are not striking. At both sites, the maximum variance is underestimated, and for the center point site the variance looks slightly worse than the one from the Hamburg site.

[34] The Hamburg supersite captures some features of the skewness structure, which is a mainly positive skewness above 1 km. The positive skewness below 1 km, however, is missed. The clear distinction between positive skewness above 1 km and the negative one below is not evident from the center point supersite. While in the first case both sites gave at least similar structures, here we are faced with a new problem: the observed structure differs depending on where we put our supersite. Even though this case is characterized by a much more randomly distributed convective cloud field compared to our first case, again we do not find reliable statistics from our supersites, which could be used to evaluate the global model's statistical cloud scheme.

[35] The operational COSMO-DE model, which we chose to obtain near-reality situations for our study has one great disadvantage: its spatial resolution of 2.8×2.8 km^2 limits our potential choices for a suitable case, since not all small-scale features of clouds and the moisture field can be resolved at this scale. Especially, COSMO-DE is not able to simulate shallow cumuli explicitly but parameterizes them. A field of shallow cumuli or stratocumuli, however, would have a great potential to be stationary long enough to allow for an evaluation of the cloud scheme with ground-based remote sensing. By applying a realistic high-resolution model with horizontal grid spacings in the order of 10–100 m we could gain an even better horizontal statistics and also the possibility to find and classify situations in which our method worked. This, however, is not the main purpose of this paper. We intend to create an awareness of the problems arising when ground-based remote-sensing data, especially higher-order moments of their statistics, are used for comparisons with models.

3.2. Artificial Test Cases

[36] To make it clearer why the assessment of higher-order moments from ground-based remote sensing is so challenging, we created four artificial examples of highly idealized cloud fields (Figure 9). The colors represent different random numbers between 0 and 1 of arbitrary unit, which have a triangular distribution as for example shown in Figure 9a) next to the plane. Let us assume the numbers represent a humidity field and the plane represents our global grid box, in which we want to evaluate the spatial distribution.

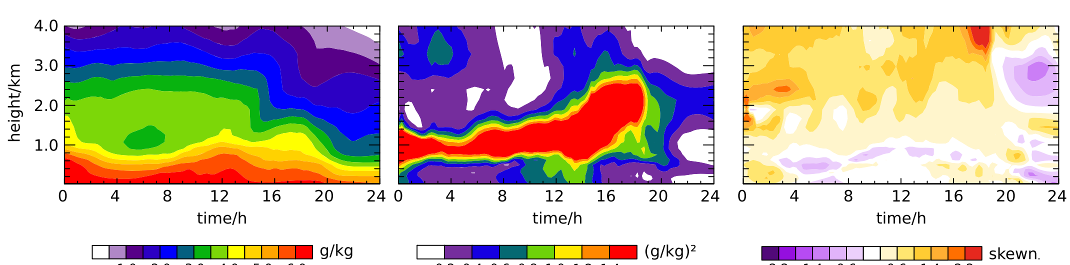

Figure 7. Time-height slice of statistics from 100 supersites with a time resolution of 15 min for 16 June 2009. (left) Mean, (middle) variance, and (right) skewness.

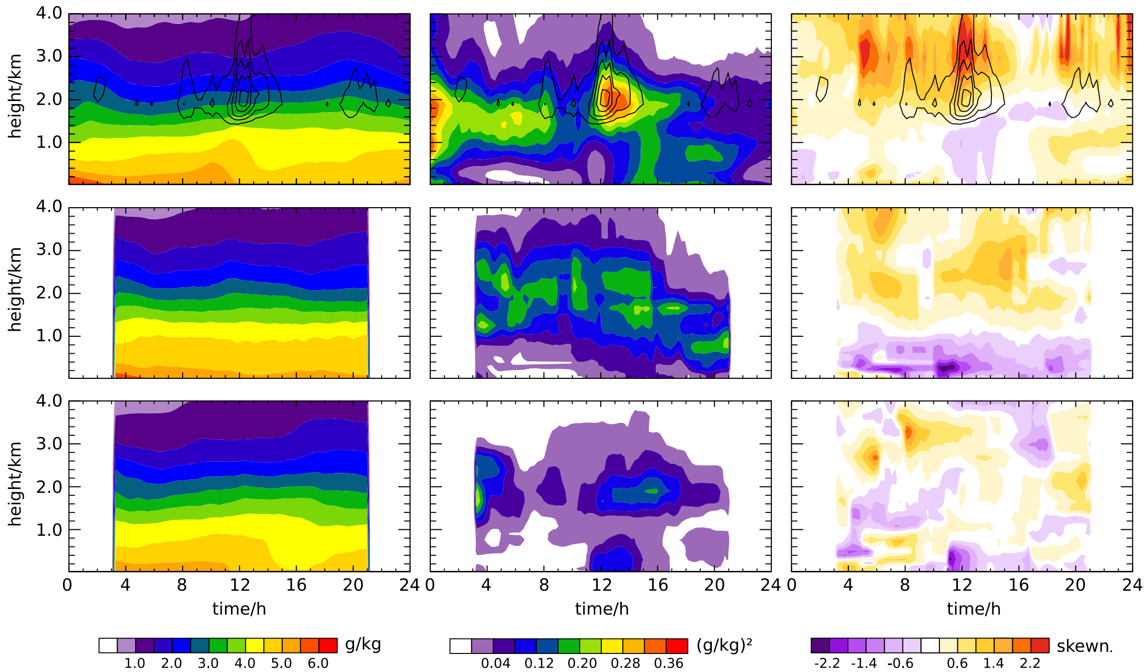

Figure 8. Statistics obtained from the simulation of 19 September 2012. (top row) Spatial statistics from the ECHAM6 grid box; black contours show the grid box mean cloud water mixing ratio with a contour spacing of 0.005 g/kg. (middle row) Temporal statistics for the location of Hamburg and (bottom row) temporal statistics for the center point of the ECHAM6 grid box. (left) mean, (middle) variance, and (right) skewness.

[37] Four realizations of the same triangular probability density distribution are shown. In Figure 9a the numbers have been randomly distributed over the area; in Figure 9b they have been sorted by size, which may be representative for a spatial gradient. These two cases stand for the best and the worst case. Figure 9c may represent a schematic view of a frontal system, where the upper half of the random numbers is randomly distributed in the northern part of the plane, and the lower half of the numbers is randomly distributed in the southern half of the plane. In Figure 9d, these two halves are distributed checkerboard-like, representative for a schematic view of a cumulus cloud field. Of course, these different realizations are a very crude approximation of reality, but they can serve to explain some reasons for the shortcomings of the method.

[38] Let us assume the weather situation does not change with time except for translations in one direction, i.e., the box is moving with the mean flow and the mean flow does not change with time. We can then simulate a virtual supersite measurement of this humidity field by taking the values along a straight line across the area. Immediately, it becomes clear how much the results of such a measurement depend on the underlying situation. In case of Figure 9a the orientation of the line does not matter, and as long as we have a large enough sample, we will get a very good estimate of the spatial distribution. The diagonal line (3) has the same length as the other two lines, so the statistics is equally good (or bad). The resulting (not normalized) distributions from the different virtual measurements are sketched in the diagrams next to the plane.

[39] In Figure 9b we have a very different situation. If we follow the north-south line, the distribution might be well met (line and diagram (1)), but if we go along the east-west line, one single, narrow peak will be the result (2). Its position will depend on where we put the line. Even the diagonal line will miss out the tails of the distribution, since we were aiming for the same sample length as in the north-south and east-west line (3). Thus, we would get quite a wrong view of the distribution if we looked in the wrong direction or with the wrong winds.

[40] Of course, the latter cases are very unrealistic. But also in the slightly more realistic cases in Figures 9c and 9d the retrieved distribution depends on where we put the line. Again, the north-south lines for both cases estimate the distribution well (diagrams (1)). In Figure 9d, which stands for a schematic cumulus case, also the east-west line gives good results (2), while in Figure 9c, a schematic frontal case, it depends on its position in the plane whether we sample the upper (2) or the lower half (3) of the distribution. Here, a diagonal line will give a good estimate (4). The diagonal line in Figure 9d demonstrates that even in such comparably well mixed cases, we might be very unlucky and miss out part of the distribution (3), e.g., in cloud streets.

[41] From these simple examples it becomes clear that (1) without the knowledge of the meteorological situation and (2) a good planning of the tracks through the respective situation, we do not have a chance to retrieve the right 2-D spatial distribution from a line or point measurement. We should also keep in mind that these examples do not even include a time evolution of the cases as before but only a transport of a static field over the virtual supersite. A time evolution of the system complicates things even further. The difficulty when using ground-based remote sensing to retrieve the distributions is that we have no chance to plan

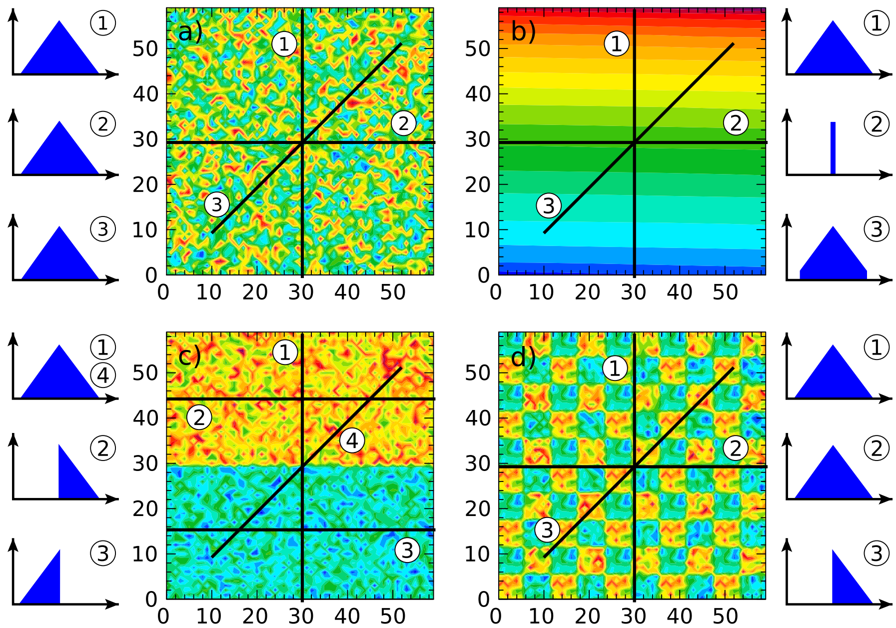

Figure 9. Different representations of a horizontal triangular probability density distribution of random numbers between 0 and 1 within a plane: (a) randomly distributed, (b) sorted, (c) 0 to 0.5 randomly distributed in the southern half of the plane, 0.5 to 1 randomly distributed in the northern half of the plane, and (d) similar, but the different ranges are distributed checkerboard-like. The lines correspond to tracks through the plane, and the distributions derived from these tracks are shown in the diagrams next to the plane. Axis labels are in arbitrary units.

our track—we have to take what comes with the mean flow and the local wind—and that most systems do not remain in a stable state long enough.

3.3. Do We Have to Throw Away Our Measurements?

[42] No. The point we are making is that we have to deal with our data very carefully and keep the specific meteorological situation in mind instead of blindly using time series to retrieve distribution moments higher than the mean, which was captured decently. It is likely that the higher moments can be reasonably well captured when the atmospheric conditions are such that the cloud and humidity fields are self-similar with respect to being rotated relative to the direction of the mean flow. This will for example be the case for unorganized fair weather cumulus, a large number of which will be present within a region comparable to the size of a GCM grid box. There are severe challenges to estimating the higher-order moments of the humidity distribution, however, if the cloud and humidity fields exhibit any organization on scales comparable to the size of the grid box or if they are in some way organized relative to the direction of the mean flow.

[43] So far, we have looked into comparably short times in the range of hours, trying to investigate on a moment-by-moment way if the model works correctly, and we found that a direct translation from time into space is not adequate to evaluate the higher distribution moments of variance and skewness. Let us now think of the spatial pattern of the total water as only one representation of the respective distribution. As we saw in Figure 9, there are many ways this distribution may be realized in a certain box. On long time scales, being in the order of years, we may then find an ensemble of different spatial representations of the same distributions, where each ensemble member is found at a different time. If we have a remote-sensing instrument at one location, we then have a fair chance to sample the whole distribution by sampling the same point each time the distribution is found. We call this "ensemble method" or a probabilistic approach to the evaluation of statistical cloud schemes.

[44] First tests using the perfect model approach, i.e., using a long-term model run and a high number of data sets, which are generated consistently with that simulation, show a great potential of this method, though challenges arise with regard to building the ensembles and the recognition of model errors. The method is currently further investigated.

4. Conclusions

[45] Statistical cloud schemes are increasingly used in the global modeling of the atmosphere. They are based on a prognostic horizontal distribution of the total water mixing ratio in a grid box, where moments such as variance and/or skewness are prognostic quantities. These schemes have the great potential to represent the cloudy part of the model grid box in a much more physical way than simple relative humidity schemes and are scale adaptive at the same time, which makes them especially useful for the new model generation employing local grid refinement.

[46] The downside, however, is that the schemes have, so far, been hardly evaluated since the full three-dimensional total water in the atmosphere cannot be measured. Either we have a sufficient horizontal resolution, e.g., from satellites, or a sufficient vertical resolution, but not both from the same instrument. Thus, we cannot directly assess the

specific processes which are responsible for failures of the cloud schemes, which may be evident from, e.g., a wrong cloud cover compared to satellite measurements. Long-term ground-based remote-sensing data sometimes appear to be the solution to this problem due to their high time and height resolution.

[47] In this paper we explored the potential of ground-based remote sensing of a vertical column of atmospheric total water to evaluate statistical cloud schemes in global models by employing the so-called perfect model approach. We have used the COSMO-DE model [*Baldauf et al.*, 2011] of DWD and simulated one case with a large-scale forcing with moderate rain and a cloud deck moving over Germany and a second, postfrontal case with locally initiated convective clouds. From these simulations we got spatial statistics in a box which represents a global model grid box and within that box a high resolution and height resolved "virtual measurement" time series which is entirely consistent with the distribution within the box. The data sets were consistent, and the working hypothesis was that if a method for evaluation of a statistical cloud scheme with ground-based remote-sensing measurements was viable, then using it with these entirely consistent data sets would result in perfect model agreement. However, we find large errors in the statistics we gain from the virtual global box and from the virtual supersites which stem from the method.

[48] First of all, for both cases we find that the statistics we gain from the virtual supersite time series strongly depends on the waiting time we employ. Short waiting times naturally lead to very noisy statistics. However, the mean values of the distribution are already captured remarkably well. But even if we can get rid of the noise by choosing a longer waiting time, the higher moments of the distribution do not compare well with the ones from the spatial distribution obtained from the global grid box, because long waiting times tend to include too much time evolution and thus are not representative of a spatial snapshot anymore. The variance exhibits at least quantitative values in the right range, but the time-height structure does not resemble the one from the spatial statistics very well.

[49] The skewness in the synoptically forced case of 16 June 2009 shows some basic features such as the deepening of the boundary layer, which is comforting—but neither the size nor the sign of the skewness is captured correctly. The system is not ergodic, and the time and spatial statistics cannot be assumed to be the same. For the locally forced case of 19 September 2012, we do not find such pronounced features, but at least some similarities of the skewness structures for one of the supersites. The other one, however, differs strongly and it became evident that the location where we measure is also an important factor in the evaluation of the statistics. As a result, an evaluation of statistical cloud schemes is not possible with this approach.

[50] Also, even if a time evolution of the system is neglected, comparing 1-D (lines) versus 2-D data (planes) can already cause very large errors. By using artificial fields generated from the same triangular distribution, we have shown that the line statistics differ strongly, depending on the underlying spatial pattern of the quantity and the direction of the line.

[51] Ground-based remote-sensing data, however, can be used on much longer time scales in a probabilistic approach, which we call ensemble method. The method is based on a categorization of the distributions found in the model with regard to the variance or skewness, where the times that category is matched is building the ensemble. We are currently investigating this basic method, and first results show its great potential to evaluate statistical cloud schemes on long time scales.

[52] **Acknowledgments.** This work was funded by the "Climate System Analysis and Prediction" (CLISAP) excellence cluster of the University of Hamburg; the COSMO-DE model was provided by the German Weather Service (Deutscher Wetterdienst, DWD), and computing time was provided by the German Climate Computing Centre (Deutsches Klimarechenzentrum, DKRZ). We thank Axel Seifert, DWD, for his help with choosing our second case and our reviewers for their valuable comments.

References

Astin, I., L. Di Girolamo, and H. M. van de Poll (2001), Bayesian confidence intervals for true fractional coverage from finite transect measurements: Implications for cloud studies from space, *J. Geophys. Res.*, *106*, 17,303–17,310.

Baldauf, M., A. Seifert, J. Forstner, D. Majewski, M. Raschendorfer, and T. Reinhardt (2011), Operational convective-scale numerical weather prediction with the COSMO model: Description and sensitivities, *Mon. Wea. Rev.*, *139*, 3887–3905.

Bouniol, D., et al. (2010), Using continuous ground-based radar and lidar measurements for evaluating the representation of clouds in four operational models, *J. Appl. Meteor.*, *49*, 1971–1991.

Boutle, I. A., S. J. Abel, P. G. Hill, and C. J. Morcrette (2013), Spatial variability of liquid cloud and rain: Observations and microphysical effects, *Q. J. Roy. Meteorol. Soc.*, doi:10.1002/qj.2140.

Duynkerke, P. G., P. J. Jonker, A. Chlond, M. C. Van Zanten, J. Cuxart, P. Clarke, E. Sanchez, G. Martin, G. Lenderink, and J. Teixeira (1999), Intercomparison of three- and one-dimensional model simulations and aircraft observations of stratocumulus, *Bound.-Layer Meteor.*, *92*, 453–487.

Gal-Chen, T., and R. C. J. Sommerville (1975), On the use of a coordinate transformation for the solution of the Navier-Stokes equations, *J. Comput. Phys.*, *17*, 209–228.

Golaz, J.-C., V. E. Larson, and W. R. Cotton (2002), A PDF-based model for boundary layer clouds. Part I: Method and model description, *J. Atmos. Sci.*, *59*, 3552–3571.

Hennemuth, B., A. Weiss, J. Bösenberg, D. Jacob, H. Linné, G. Peters, and S. Pfeifer (2005), Quality assessment of water cycle parameters in REMO by radar-lidar synergy, *Atmos. Chem. Phys.*, *8*, 287–308.

Hinkelman, L. M., T. P. Ackerman, and R. T. Marchand (1999), An evaluation of NCEP Eta model predictions of surface energy budget and cloud properties by comparison with measured ARM data, *J. Geophys. Res.*, *104*, 19,535–19,594.

Hogan, R. J., C. Jakob, and A. J. Illingworth (2001), Comparison of ECMWF winter season cloud fraction with radar-derived values, *J. Appl. Meteorol.*, *40*, 513–525.

Illingworth, A. J., et al. (2007), CloudNet - continuous evaluation of cloud roles in seven operational models using ground-based observations, *B. Am. Meteorol. Soc.*, *88*, 883–898.

King, M. D., W. P. Menzel, Y. J. Kaufman, D. Tanre, B. C. Gao, S. Platnic, S. A. Ackerman, L. A. Remer, R. Pincus, and P. A. Hubanks (2003), Cloud and aerosol properties, precipitable water, and profiles of temperature and water vapor from MODIS, *IEEE T. Geosci. Remote Sens.*, *41*, 442–458.

LeTreut, H., and Z.-X. Li (1991), Sensitivity of an atmospheric general circulation model to prescribed SST changes: Feedback effects associated with the simulation of cloud optical properties, *Clim. Dyn.*, *5*, 175–187.

LeTreut, H., R. Somerville, U. Cubasch, Y. Ding, C. Mauritzen, A. Mokssit, T. Peterson, and M. Prather (2007), Historical overview of climate change, in *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, edited by S. Solomon et al., Cambridge Univ. Press, Cambridge, United Kingdom and New York, NY, USA.

Löhnert, U., S. Crewell, and C. Simmer (2004), An integrated approach toward retrieving phyically consistent profiles of temperature, humidity, and cloud liquid water, *J. Appl. Meteorol.*, *43*, 1295–1307.

Mace, G. G., C. Jakob, and K. P. Moran (1998), Validation of hydrometeor occurrence predicted by the ECMWF model using millimeter wave radar data, *Geophys. Res. Let.*, *25*, 1645–1648.

Morcrette, C. J. (2012), Prognostic-cloud-scheme increment diagnostics: A novel addition to the case-study tool kit, *Atmosph. Sci. Lett.*, *13*, 200–207, doi:10.1002/asl.380.

Morcrette, C. J., E. J. O'Connor, and J. C. Petch (2012), Evaluation of two cloud parameterization schemes using ARM and Cloud-Net observations, *Q. J. R. Meteorol. Soc.*, *138*, 964–979.

Quaas, J. (2012), Evaluating the critical relative humidity as a measure of subgrid-scale variability of humidity in general circulation model cloud cover parameterizations using satellite data, *J. Geophys. Res.*, *117*, D09208, doi:10.1029/2012JD017495.

Stevens, B., et al. (2013), The atmospheric component of the MPI-M Earth System Model: ECHAM6, *J. Adv. Model. Earth Syst.*, *5*, 146–172, doi:10.1002/jame.20015.

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

Tompkins, A. M. (2003), Impact of temperature and humidity variability on cloud cover assessed using aircraft data, *Q. J. R. Meteorol. Soc.*, *129*, 2151–2170.

Uttal, T., and R. A. Kropfli (2001), The effect of radar pulse length on cloud reflectivity statistics, *J. Atmos. Oceanic Technol.*, *18*, 947–961.

van Meijgaard, E., and S. Crewell (2005), Comparison of model predicted liquid water path with ground-based measurements during CLIWA-NET, *Atmos. Res.*, *75*, 201–226.

Weber, T., J. Quaas, and P. Räisänen (2011), Evaluation of the statistical cloud scheme in the ECHAM5 model using satellite data, *Quart. J. Roy. Meteor. Soc.*, *137*, 2079–2091.

Zhu, P., and P. Zuidema (2009), On the use of PDF schemes to parameterize sub-grid clouds, *Geophys. Res. Lett.*, *36*, L05807, doi:10.1029/2008GL036817.