Using ARM Data to Model Cloud Overlap with Copulas

P. Norris[1], L. Oreopoulos[2], A. Hou[2], W.-K. Tao[2], X. Zeng[1]

1. UMBC-GEST, 2. NASA-GSFC

Motivation

The realism of GCM cloudy radiation and precipitation estimates is currently limited by a poor representation of clouds. Specifically, both the horizontal subgrid variability of cloud properties and its vertical correlation with other cloudy layers (cloud overlap) needs improvement.

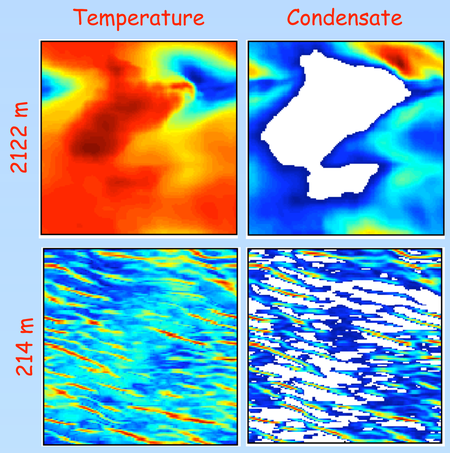
A new statistical approach

Recent work (e.g., Pincus et al., Räisänen et al.) has modeled the rank correlation in condensate between adjacent pairs of cloudy layers. We have extended this work by considering the rank correlation between all pairs of cloudy layers, via statistical functions called copulas.

Copulas in the News

Copulas have gained some recent notoriety (e.g. WIRED, "The Secret Formula that Destroyed Wall Street", 3/09) as a cause of the "economic meltdown." In reality, copulas serve a very useful role in the analysis of multivariate correlations if used properly, not treated as "black boxes".

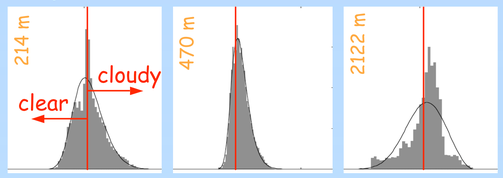
Analysis of Cloud Resolving Model Simulations



High and low layers from a CRM simulation (Goddard Cumulus Ensemble model, 128x128 km^2). A frontally-lifted large-scale cloud overlies a cold dry layer penetrated by warm, moist convective bands from a wetted surface.

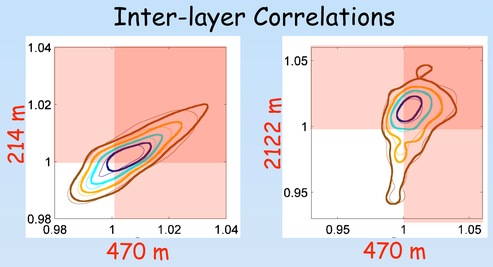
Reduce Complexity

Layer Distributions of Saturation Ratio



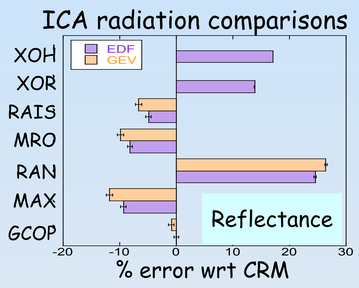
Step I: For each layer, the total saturation ratio ($S=q_T/q_S$) data are ranked to the range (0, 1) by two methods: (a) by fitting the frequency distribution of S with an analytic PDF [the Generalized Extreme value (GEV) distribution above]. Notice the pronounced skewness; (b) by empirically ranking the S values via their sorted index [the so-called Empirical Distribution Function (EDF)].

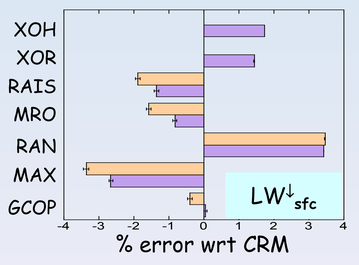
Inter-layer Correlations



Step II: the inter-layer S correlations are modeled with a Gaussian copula (a type of joint distribution in rank-space). The blue, cyan, orange and brown lines contain 20, 40, 60 & 80% of the joint S probability. The thick lines are for the CRM data and the thin lines are for synthetic data regenerated from the Gaussian copula. The agreement is good. The clear, light pink, and dark pink areas are clear in both layers, cloudy in one layer, and cloudy in both layers.

ICA radiation comparisons





Generate synthetic clouds from copula and compare with original CRM clouds

Step III: Calculate radiation (Independent Column Approximation) for original CRM clouds and copula regenerated synthetic clouds. Copula versions were built for maximum overlap (MAX), random overlap (RAN), maximum random overlap (MRO, Geleyn & Hollingsworth), and Räisänen-like generalized overlap (total water version). Two exact overlaps, XOH (in-cloud Homogenized) and XOR (in-cloud Randomized), are shown for comparison. The Gaussian copula (GCOP) performs very well.

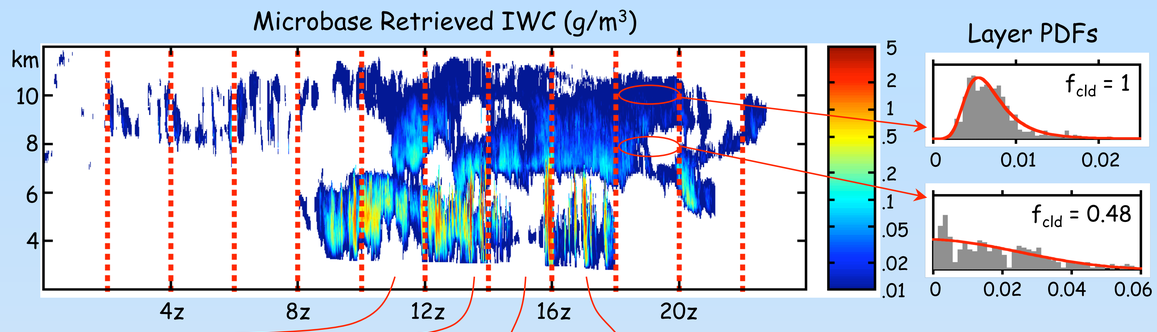
Conclusions

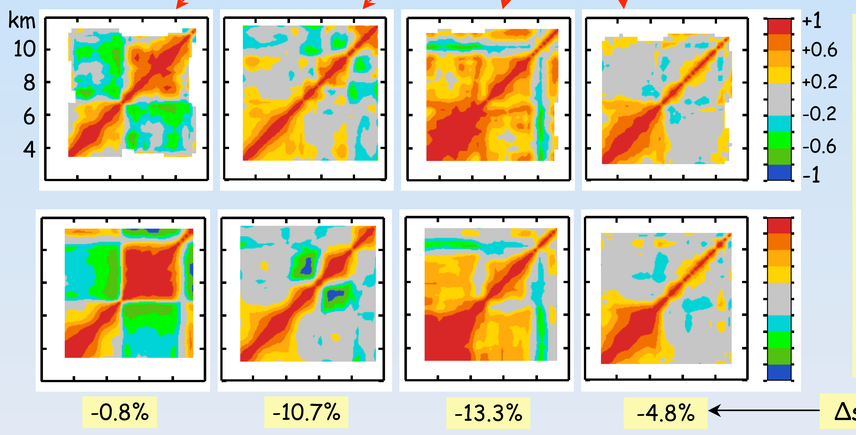
We have built a Gaussian copula-based cloud generator that models correlations (and anticorrelations) between arbitrary separated layers. The model performs well & handles skewed layer distributions naturally. For more details see Norris *et al.*, QJRMS, **134**: 1843-1864 (2008).

Analysis of ARM Microbase Data

In the case of Microbase IWC retrievals, only the condensed tail of the total water distribution is observable. To what extent can this data be used to constrain the total water PDFs and their rank correlations between layers? In the case study below, Microbase IWC retrievals for 1/1/04 are segmented into 2hr (72km @ 10m/s) windows, of similar scale to a GCM gridcolumn. Layers are fitted with skewed GEV distributions, assuming the PDF smoothly extends into the clear fraction and that its probability mass is replicated. Preliminary inspection of the fits look reasonable, but quantitative analysis indicates that better modeling of upper tail behavior may be needed. A fully bounded distribution such as the Beta distribution may be required. For now, we will use the EDF-based ranks not the GEV-based ranks in the copula analysis that follows.

Microbase Retrieved IWC (g/m^3)





Interlayer Correlations. Top: rank correlations of IWC for horizontal positions where both layers are cloudy. Bot: Gaussian copula correlation matrices for a "total-water-like" variable (see below). Note the anticorrelation between upper and lower clouds in Segment 6.

In the lower panel GCOP correlation matrices, for each layer pair, a GCOP fit maximizes the likelihood of observing the "both-cloudy" IWC points (in the dark pink region in Step II to the left) and the probability masses in the "both-clear" and "one-layer clear" (light pink) regions. This enables easy regeneration of a complete synthetic cloud field. The Δs_{Wc}(%) are the mean biases in the sdev of IWP for the regenerated clouds compared to the ARM clouds. Further work on the copula analysis is required.

Many thanks to Maureen Dunn and Mike Jensen for providing the Microbase data.

Conclusions & Future Work

Preliminary studies indicate that it may be possible to use a "total-water-like" copula model to fit condensate-only observations when there is sufficient cloud fraction. More work is needed to study the most appropriate marginal PDFs and copula over longer sequences of data with differing synoptic conditions. Analysis and parameterization of the correlation matrices is also needed.