GEOPHYSICAL RESEARCH LETTERS, VOL. 30, NO. 21, 2082, doi:10.1029/2003GL018367, 2003

Objective identification of cloud regimes in the Tropical Western Pacific

Christian Jakob[1]
Bureau of Meteorology Research Centre, Melbourne, Australia

George Tselioudis
NASA GISS/Columbia University Dept. of Applied Physics, New York, USA

Received 8 August 2003; revised 5 September 2003; accepted 16 September 2003; published 1 November 2003.

[1] Identifying cloud regimes and their role in the climate system can serve a multitude of purposes, ranging from a better understanding of clouds to guiding field experiments to improving the representation of clouds in models. This study describes early results in identifying cloud regimes from ISCCP data using cluster analysis. A simple algorithm for cloud regime identification is introduced and applied to data in the Tropical Western Pacific region. Four major cloud regimes, namely a shallow cumulus regime, a transparent isolated cirrus regime, thick cirrus with convection and a deep and probably organized convective regime are identified and their frequency of occurrence is quantified. The use of the regime information for various applications is discussed and the use of regime classifications for representativeness studies is presented using the ARM TWP sites as an example. *INDEX TERMS:* 0320 Atmospheric Composition and Structure: Cloud physics and chemistry; 3374 Meteorology and Atmospheric Dynamics: Tropical meteorology; 3360 Meteorology and Atmospheric Dynamics: Remote sensing; 3314 Meteorology and Atmospheric Dynamics: Convective processes. **Citation:** Jakob, C., and G. Tselioudis, Objective identification of cloud regimes in the Tropical Western Pacific, *Geophys. Res. Lett.*, *30*(21), 2082, doi:10.1029/2003GL018367, 2003.

1. Introduction

[2] The important role of clouds in the climate system is undisputed and many efforts are being undertaken to improve their representation in general circulation models (GCMs). In order to improve models it is first necessary to assess where they fail. A large number of approaches to make such assessments exist; ranging from the analysis of long-term climate simulations to the use of models of specific cloud processes. *Jakob* [2003] has recently concluded that in order to link the various efforts it is necessary to analyze the overall climate simulations of clouds in terms of the prevalent cloud regimes by means of composite averaging. Many recent studies have made use of such techniques using both observations and GCMs [e.g., *Lau and Crane*, 1995; *Tselioudis et al.*, 2000; *Webb et al.*, 2001; *Norris and Weaver*, 2001; *Tselioudis and Jakob*, 2002; *Williams et al.*, 2003]. These studies used various indicators such as surface pressure anomalies or vertical velocity to first identify "dynamical" regimes and then showed by using cloud observations provided by the International Satellite Cloud Climatology Project [ISCCP; *Rossow and Schiffer*, 1991] that these dynamical regimes could be linked with distinct cloud regimes.

[3] These findings seem to indicate that the average cloudiness observed in a given region, rather then consisting of a mixture of random cloud fields, is actually composed of a limited number of distinct cloud regimes, each linked to certain characteristics of the atmosphere. The major limitations of previous studies are that they either do not consider dynamical regimes at all or require a priori knowledge about a dynamical regime classification. Many of the cited studies are also restricted to the extra-tropics.

[4] Here we report on first results of a technique that aims to identify cloud regimes based on observed cloud information alone. Joint histograms of cloud-top pressure (CTP) and cloud optical thickness (τ) available three-hourly from ISCCP are used to identify cloud regimes in the tropical Western Pacific (TWP) by means of cluster analysis. It is shown that this simple approach leads to a physically interpretable identification of cloud regimes, which can be used in a number of applications. Section 2 briefly describes the data and the cluster algorithm used. Section 3 contains the main findings of the study and section 4 provides examples for the application of these results. Conclusions and a brief outlook to future work are presented in section 5.

2. Data and Techniques

[5] The ISCCP data set provides three-hourly global information on several cloud parameters retrieved from a combination of visible and infrared channels on both polar-orbiting and geostationary satellites. Amongst other products the ISCCP D1 data set used here provides joint histograms of the frequency of occurrence (FOCC) of clouds with a certain CTP and τ in grid boxes of ca. 280 × 280 km [e.g., *Tselioudis et al.*, 2000]. An example for the structure of such a histogram is shown in Figure 1. The figure shows the mean CTP-τ histogram averaged for one year (1999) over all grid boxes in a region located in the TWP (130° to 170°E, 10°N to 10°S). The contoured field is the FOCC of combinations of CTP and τ. The strong relationship of the location of FOCC maxima in this diagram to prevailing cloud types is evident. For this tropical area we can identify a high frequency of cirrus, both transparent and opaque, located in the top left corner

---

[1] BMRC, GPO Box 1289K, Melbourne, VIC 3001, Australia.

0094-8276/03/2003GL018367$05.00

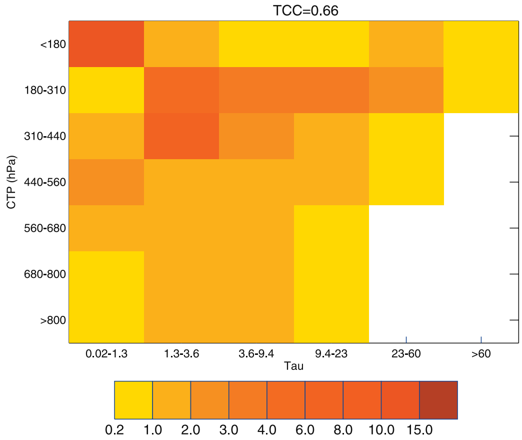

Figure 1. Mean CTP-τ histogram averaged for 1999 over all grid boxes in the TWP (130° to 170°E, 10°N to 10°S). Colors indicate the RFO in each CTP-τ class.

of the diagram, a fair number of occurrences of deep convective clouds with high optical thickness and high cloud tops (top right corner), as well as shallower, less optically thick clouds (bottom left part). The integral over the entire histogram yields the total cloud cover (TCC), which has a value of 0.66.

[6] The basic question we want to address in identifying cloud regimes is whether the annually and spatially averaged cloud field depicted in Figure 1 is composed of a collection of random (in space and time) cloud situations or of distinct and recurring cloud regimes. The statistical method we choose in addressing this question is cluster analysis. As its name suggests, cluster analysis searches for possible "clusters" in a data set usually by evaluating a measure of distance between its individual data points. Note that in this case a "data point" is an individual CTP-τ histogram forming a vector comprised of its 42 classes. About 160,000 histograms can be found in the TWP area in 1999 by treating each ISCCP gridpoint at each time as an independent data entry. They form the basis of the analysis. The particular clustering algorithm used is that of KMEANS clustering [*Anderberg*, 1973]. This algorithm iteratively searches for a predefined number (k) of clusters using the following technique: 1) k elements of the data set of size N are used as clusters of one member each; 2) each of the remaining N-k elements are assigned to the cluster with the nearest (in a Euclidian distance sense) centroid whereby after each assignment the centroid of the gaining cluster is recalculated; and 3) after all elements have been assigned the centroids found in step 2) are used as new seed points and the algorithm is iterated. In this study a varying number of clusters is used, while ten iterations have been found to be sufficient for the convergence of the algorithm.

3. Typical Cloud Regimes in the Tropical Western Pacific

[7] The KMEANS algorithm is applied to all histograms in the TWP area for the year 1999. The key results of the analysis are the mean histogram within each cluster (cluster centroid), the number of cases within each cluster, which is used to calculate the relative frequency of occurrence (RFO) of the identified cloud regime, and the TCC within each cluster. These parameters will be used below to describe the results of this study.

[8] Given that the cluster algorithm requires the number of clusters to be specified, it is logical to analyze the evolution of the results as this number is increased. Figure 2 shows the cluster means (centroids) identified when two clusters are searched for. The RFO and TCC of each of the identified cloud regimes are indicated at the top of each panel. Even this most basic separation of the data set into two parts delivers results that are physically interpretable. The first cloud regime represents 69% of the samples and is characterized by mostly transparent cirrus and low thin clouds with very little evidence of deep convection. The mean TCC of the regime is just above 0.5. The second regime has a TCC of close to 1 and is characterized by a significant number of deep convective pixels as well as a high frequency of high top clouds with medium and low optical thickness, most likely anvil and cirrus outflow from convection. This regime represents 31% of the cases sampled. The separation achieved in the two clusters is well in line with our expectation of the presence of convectively active and suppressed conditions known to exist in the tropics.

[9] Division into more clusters leads to an interesting further separation of those two basic cloud regimes into sub-regimes. Figure 3 shows the results for a four-cluster analysis. It is evident that clusters 1 and 3 result from a separation of cluster 1 of the two-cluster analysis (Figure 2), while clusters 2 and 4 originate from splitting cluster 2. This is confirmed by the three-cluster analysis (not shown), in which cluster 1 of the two-cluster analysis is split while cluster 2 remains intact. It is worthwhile noting that each of these analyses is carried out independently using the full data set, lending some credit to the robustness of the results. The cloud regimes identified in the four-cluster analysis are 1) a regime dominated by shallow clouds of medium optical thickness and low TCC (0.4); 2) a high-TCC (0.9) regime dominated by cirrus of measurable optical thickness in the

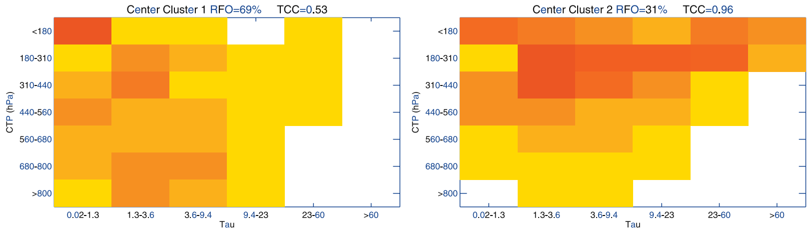

Figure 2. CTP-τ histograms of the centroids of a two-cluster analysis using all histograms for 1999 in the TWP (130° to 170°E, 10°N to 10°S).

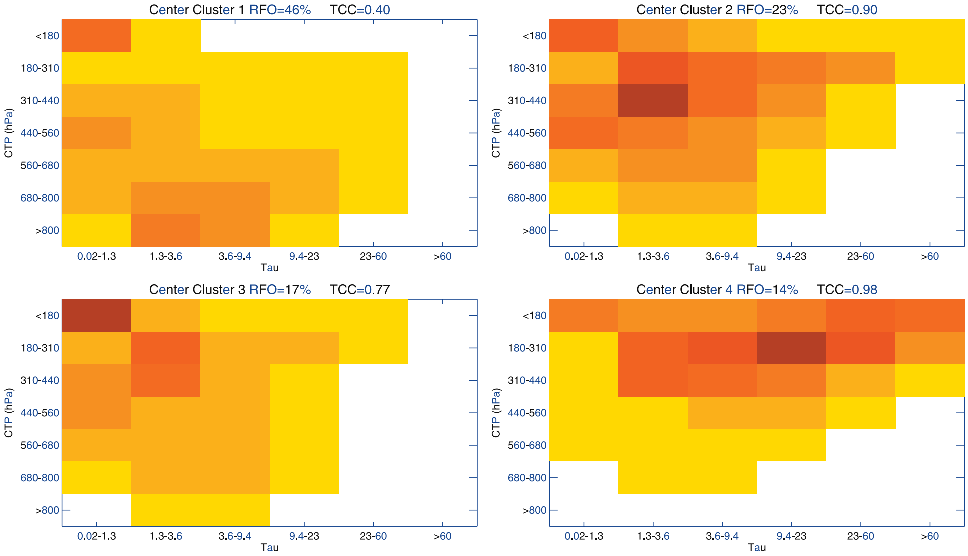

Figure 3. CTP-τ histograms of the centroids of a four-cluster analysis using all histograms for 1999 in the TWP (130° to 170°E, 10°N to 10°S).

presence of some deep convection; 3) a regime dominated by transparent cirrus with little coincident deep convection and a TCC of about 0.77 and; 4) a strongly convective regime with a a TCC of close to 1, dominated by optically thick high-top clouds, most likely to be organized convective systems with significant stratiform anvil coverage. The four regimes occur in 46%, 23%, 17%, and 14% of the cases respectively. Increasing the number of identified clusters to five (not shown) leads to a further subdivision of regime 2 identified above mainly by CTP. Increasing the number of clusters even further (not shown) does not lead to the identification of significantly different regimes, but rather focuses on some of the details within regimes and delivers clusters with very low RFO. We therefore (somewhat subjectively) choose to use four cloud regimes for the remainder of this study. To our knowledge there is no completely objective way for deciding the optimal number and statistical significance of the clusters chosen. We repeated the analysis for other time periods and for monthly and regional subsets of the data used here (not shown) with negligible changes in the results, lending some credit to our choice.

[10] In summary we can conclude that for the TWP the cluster analysis method is able to provide physically interpretable cloud regimes. The most frequent cloud regime is a shallow convective regime, while deep convection is present in a dominant way in less than 40% of all cloud scenes. This may be indicative of two facts. First, the background state of the tropical atmosphere is one of subsiding motion disturbed by intermittent convection. Second, deep and optically thick convective systems are relatively rare and small compared to the cirrus outflow they produce. Both of these findings do not come as a surprise but apart from being able to reveal them, the technique proposed here also allows for a quantitative assessment of the relative role of each of the regimes.

4. Discussion

[11] We have shown that it is possible to derive information on cloud regimes and their occurrence in the TWP based on ISCCP cloud observations using a simple cluster analysis technique. A valid question to ask at this point is of what use this knowledge is.

[12] Firstly there is useful information in knowing that clouds are organized into regimes and in quantifying their RFO. Combined with other data such as radiative fluxes and precipitation, the importance of each regime for the hydrological and energy cycles can be assessed providing focus for the planning of field experiments and modeling efforts.

[13] Another major aim of this and similar work is to aid the development of cloud representations in atmospheric models. Imagine a model exhibits a certain error structure in radiation at the top of the atmosphere in an area of the TWP. Combining the analysis of the radiation error with information on the ISCCP cloud regimes, it would be possible to decompose the model error into three components; an erroneous representation of the cloud regime; an incorrect representation of the radiative effects of an individual cloud regime; and an incorrect distribution of the RFO of the cloud regimes. Depending on which of the three components dominates the model error, very different action in improving the model is required. Hence, supplying this type of information to model developers is extremely useful.

[14] A further, equally important application of cloud regime information is in assessing the representativeness of local measurements such as taken during field experiments. To exemplify this application we have used the four cloud regimes above and evaluated their RFO at two TWP measurement sites deployed by the U.S. Department of Energy's (DOE) Atmospheric Radiation Measurement Program [ARM; *Stokes and Schwartz*, 1994] on Manus (2.1°S, 147.4°E) and Nauru (0.5°S, 166.9°E) islands. This enables us to assess the representativeness of those sites for the cloud regimes occurring in the wider TWP area. For this purpose the following technique is applied. First, the ISCCP histograms for 1999 are retrieved for the ISCCP grid box containing each site. Around 1,500 cloud situations are available at each site from this data set. Second, the Euclidian distance of each histogram to each of the four

cluster centroids identified for the TWP is calculated. Each histogram is then placed into the regime for which this distance is smallest. Having assigned all histograms to one of the four regimes, the RFO of the regime at the site and its average TCC are calculated and displayed in Table 1.

Table 1. RFO and TCC (in parentheses) for the four TWP cloud regimes for the entire TWP and for individual ISCCP boxes located around the ARM sites on Manus and Nauru Island

Cluster No	TWP	Manus	Nauru
1	46% (0.40)	34% (0.33)	73% (0.26)
2	23% (0.90)	21% (0.91)	10% (0.85)
3	17% (0.77)	33% (0.76)	16% (0.68)
4	14% (0.98)	12% (0.99)	1% (0.98)

[15] The shallow cloud regime is the dominant regime at both ARM sites, just as in the entire TWP region. However, it occurs on fewer than average occasions at Manus (34% instead of 46%). At Nauru the shallow cloud regime is by far the most frequent (73%) and the low TCC (0.26) suggests strongly suppressed conditions at Nauru. This is explained by the location of Nauru in the relatively cold sea surface temperature (SST) region at the eastern edge of the TWP warm pool and confirms the usefulness of the clustering method in delivering physically sensible results. The next most frequent regime at both ARM sites is that dominated by transparent high-level cirrus (Regime 3). It occurs 33% of the time at Manus, almost twice as frequently as in the TWP, and 16% of the time at Nauru. This is indicative of the closer proximity of the Manus site to convective activity. The two deep convective regimes, one dominated by cirrus the other by optically thicker clouds, occur with about average frequency at Manus and are extremely rare at Nauru. This again reflects the location of the two sites with respect to the SST distribution in the TWP. It is worthwhile stressing that the results here are only representative for an individual year (1999) and one would expect different RFO at the sites in other years, in particular during El Nino conditions. Nevertheless, since 1999 is probably close to a “normal” year with respect to ENSO we can still draw some conclusions from our results.

[16] Probably the most important one is that between them the tropical ARM sites cover all cloud regimes typically found in the TWP. Nauru appears to be a good site for the study of shallow clouds, although known island effects may make it difficult to use the ARM data alone for that purpose. A second regime suitable for study at this site is that of mostly thin or transparent cirrus, although Manus exhibits a higher RFO of this regime. Cloud regimes involving significant amounts of deep convection are best studied at Manus, at least for years with normal SST conditions.

[17] The conclusions drawn above demonstrate how the cloud regime identification enables us to place individual measurement sites into a large-scale context. This is of importance since it allows for a more quantitative generalization of the results of detailed process studies at these sites with respect to their role in climate.

5. Conclusions and Future Work

[18] A cluster algorithm has been applied to ISCCP histograms of cloud-top pressure and cloud optical thickness over the TWP with the aim of identifying typical cloud regimes in that region. It has been shown that this algorithm is capable of robustly identifying physically interpretable cloud regimes. Four major regimes have been identified for the TWP region. Several applications for the classification, in particular in model evaluation and representativness studies, have been suggested. In an example, the representativeness of the ARM TWP sites at Manus and Nauru for cloud regimes occurring in the broader TWP region has been investigated. We can conclude that all cloud regimes identified in the TWP do occur over the ARM sites, albeit with very different frequencies of occurrence mostly related to the location of the sites with respect to the SST distribution in the region.

[19] The success of the simple cluster analysis technique in identifying structures in the ISCCP histograms is encouraging. An obvious extension of the work presented here is the inclusion of other geographical areas and longer time periods in the analysis with the aim of studying other cloud regimes and to establish a global picture of their distribution and occurrence. At the ARM sites themselves the work will be extended into studying the radiative and cloud signatures of the different regimes as revealed by the extensive array of ARM instruments. Finally, the application of regime classifications has proven useful for model evaluation in the past and our findings provide a natural path to an extension of previous studies to tropical latitudes.

[20] **Acknowledgments.** We would like to thank William Rossow (GISS), Wasyl Drosdowsky (BMRC) and Debbie Hudson (BMRC) for their helpful discussions and comments on the work carried out. We thank Doerte Jakob for providing us with the cluster analysis code. The comments of Mark Webb (UKMO) and an anonymous reviewer have enabled us to improve the original manuscript. We are grateful for the support of this work by the U. S. Department of Energy under contract LANL-23662-001-013T and the NASA Radiation Sciences Program managed by Dr. Donald Anderson.

References

Anderberg, M. R., Cluster analysis for applications, Academic Press, New York, 359 pp., 1973.

Jakob, C., An Improved Strategy For The Evaluation Of Cloud Parametrizations In GCMs, *Bull. Amer. Meteorol. Soc.*, in press, 2003.

Lau, N.-C., and M. W. Crane, A satellite view of the synoptic-scale organization of cloud properties in Midlatitude and Tropical circulation systems, *Mon. Wea. Rev.*, *107*, 1984–2006, 1995.

Norris, J., and C. Weaver, Improved Techniques for Evaluating GCM Cloudiness Applied to the NCAR CCM3, *J. Clim.*, *14*, 2540–2550, 2001.

Rossow, W. B., and R. A. Schiffer, ISCCP cloud data products, *Bull. Amer. Meteorol. Soc.*, *72*, 2–20, 1991.

Stokes, G. M., and S. E. Schwartz, The Atmospheric Radiation Measurement (ARM) Program: Programmatic background and design of the Cloud and Radiation Test Bed, *Bull. Amer. Meteorol. Soc.*, *75*, 1201–1221, 1994.

Tselioudis, G., Y. Zhang, and W. B. Rossow, Cloud and radiation variations associated with northern midlatitude low and high sea level pressure regimes, *J. Clim.*, *13*, 312–327, 2000.

Tselioudis, G., and C. Jakob, Evaluation of midlatitude cloud properties in a weather and a climate model: Dependence on dynamic regime and spatial resolution, *J. Geophys. Res.*, *107*, 4781, doi:10.1029/2002JD002259, 2002.

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

Williams, K. D., M. A. Ringer, and C. A. Senior, Evaluating the cloud response to climate change and current climate variability, *Clim. Dyn.*, *20*, 705–721, 2003.

C. Jakob, Bureau of Meteorology Research Centre, Melbourne, Australia. (c.jakob@bom.gov.au)

G. Tselioudis, NASA GISS/Columbia University Dept. of Applied Physics, New York, USA.