7A.6 ARCHETYPAL TRMM RADAR PROFILES IDENTIFIED THROUGH CLUSTER ANALYSIS

Dennis J. Boccippio *
NASA / Marshall Space Flight Center, MSFC, Alabama

1. INTRODUCTION

It is widely held that identifiable 'convective regimes' exist in nature, although precise definitions of these are elusive. Examples include land / ocean distinctions, break / monsoon beahvior, seasonal differences in the Amazon (SON vs DJF), etc. These regimes are often described by differences in the realized local convective spectra, and measured by various metrics of convective intensity, depth, areal coverage and rainfall amount.

Objective regime identification may be valuable in several ways: regimes may serve as natural 'branch points' in satellite retrieval algorithms or data assimilation efforts; one example might be objective identification of regions that "should" share a similar Z-R relationship. Similarly, objectively defined regimes may provide guidance on optimal siting of ground validation efforts. Objectively defined regimes could also serve as natural (rather than arbitrary geographic) domain "controls" in studies of convective response to environmental forcing.

Quantification of convective vertical structure has traditionally involved parametric study of prescribed quantities thought to be important to convective dynamics: maximum radar reflectivity, cloud top height, 30-35 dBZ echo top height, rain rate, etc. Individually, these parameters are somewhat deficient as their interpretation is often nonunique (the same metric value may signify different physics in different storm realizations). Individual metrics also fail to capture the coherence and interrelationships between vertical levels available in full 3-D radar datasets.

An alternative approach is discovery of natural partitions of vertical structure in a globally representative dataset, or "archetypal" reflectivity profiles. In this study, this is accomplished through cluster analysis of a very large sample (O[10^7]) of TRMM-PR reflectivity columns. Once achieved, the rain-conditional and unconditional "mix" of archetypal profile types in a given location and/or season provides a description of the local convective spectrum which retains vertical structure information. A further cluster analysis of these "mixes" can identify recurrent convective spectra. These are a first step towards objective identification of convective regimes, and towards answering the question: "what are the most convectively similar locations in the world"?

2. CLUSTER ANALYSIS OF VERTICAL PROFILES

A cluster analysis simply answers the question: Given a set of n-parameter descriptions of individual cases or instances, find x natural clusters, or partitions, of these cases in the n-space. The number of clusters requested, x, is arbitrary and must be prescribed, althogh iterative examination of analyses while varying x can often reveal when too many or too few clusters are sought. In the case of clustering reflectivity profiles, a reasonable goal is separation into clusters which appear to indicate clearly different convective or microphysical states. In this study, we use the Interactive Data Language (IDL) CLUSTER_WTS() and CLUSTER() routines, which perform a 'flat' (non-hierarchical) clustering using a (transparent) neural network-based optimization engine. The routines accept a standardized array of multiparameter case descriptors (see below) and yield an assignment of each case into one of x discovered clusters.

4 million TRMM-PR profiles (a subsampling of 2 years of data) containing echo in or adjacent to a precipitating column (a 4 km x 4 km PR pixel) are used to identify natural cluster centroids. For simplicity during this proof-of-concept study, only profiles occurring in locations with surface temperature > 16C are considered (loosely, 'warm season' profiles). The PR reflectivity-vs-altitude information is remapped to reflectivity-vs-temperature profiles using the closest prior 6-hrly NCEP reanalysis temperature profile. Reflectivities at 31 levels from 16C to -65C are used as inputs to the cluster analysis. Additionally, the TRMM 2A23 convective/stratiform/other classifiers, and their confidence levels, are used as inputs, as well as the 2A23 bright band detection classifier. Each reflectivity column thus contains n=37 descriptors: 31 reflectivity values, "is convective", "is stratiform", "is other" binary flags, convective and stratiform confidence (ordinal from 0-10) and a "has bright band" binary flag. Since the 2A23 convective/stratiform classifier considers both vertical and horizontal information, the cluster analysis does incorporate some horizontal structure information in addition to the vertical profile data. The input to the cluster analysis is thus a 37 column x 4,000,000 row array, standardized along columns. Theoutput is an assignment of each column into one of x=25 clusters (25 being subjectively selected after examination of x=2,3,...29,30 analyses). Once the cluster centroids are identified, the entire ~70,000,000 columns in the 2-yr dataset are assigned to clusters (Fig 1).

In Fig. 1, the clusters are subjectively grouped (i.e., not a result of the cluster analysis itself) into 'familiar' profile types: stratiform, 'weak aloft', convective and anvil. For the geographic convective spectrum analysis below, the clusters are consolidated into 12 subsets (outline boxes in Fig. 1), merging clusters which appear phsyically similar and only differ in, e.g., their depth. Fig. 1 also documents the percentage of columns in each cluster in which the various 2A23 classifiers are active (stratiform, convective, 'other', and bright band).

The spatial distributions of occurrence of these profile types is informative in and of itself. Fig. 2 shows maps of

* Corresponding author address: Dennis J. Boccippio, NASA/MSFC SD-60, Marshall Space Flight Center, AL 35812; e-mail: Dennis.Boccippio@nasa.gov

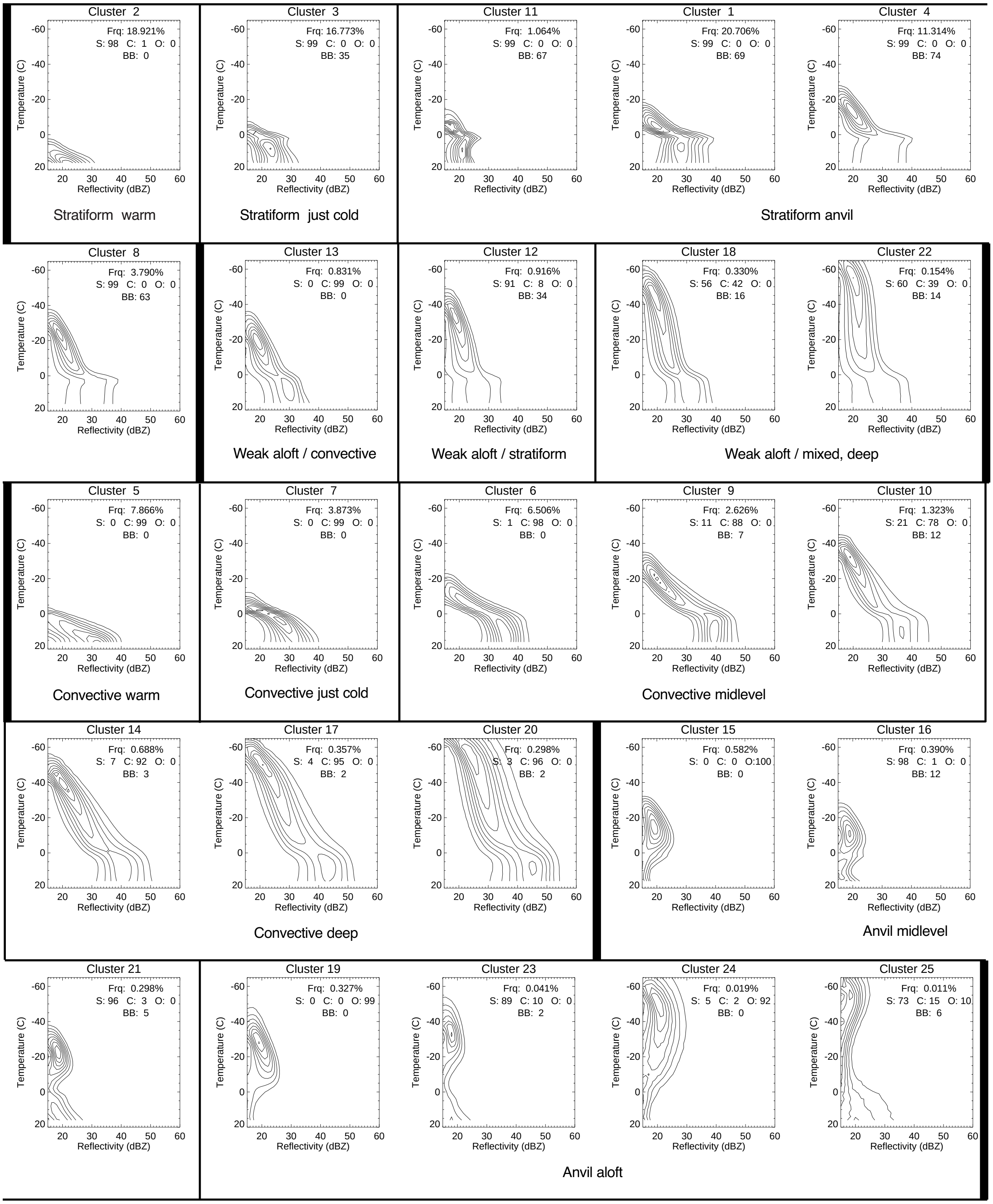

Fig. 1: TRMM PR vertical column clusters identified from a 37-parameter, 25-cluster cluster analysis, trained on 4 million columns then applied to a 2-yr 70 million column dataset. The conditional reflectivity spectra at each temperature level in each cluster are contoured; contours are 8 uniformly-spaced levels from 0 to the maximum conditional frequency in each plot. Upper right insets show (top): frequency of this cluster in the entire warm-season dataset, (middle): percentage of columns in this cluster classified by 2A23 as stratiform, convective or 'other', (bottom): percentage of columns in this cluster classified by 2A23 as having bright band. Boxes show subjective reduction into 12 ":consolidated" clusters.

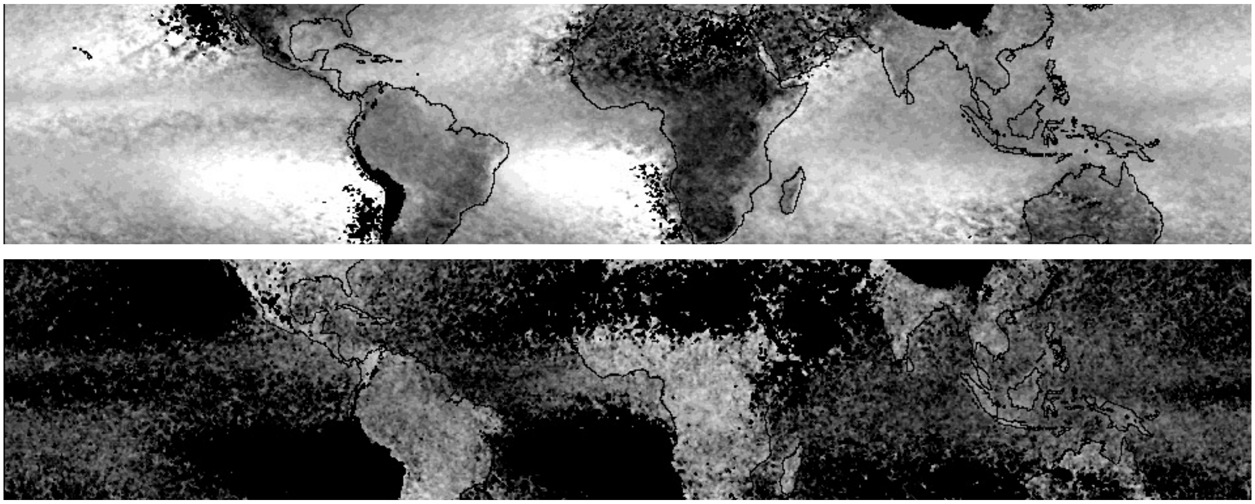

Fig. 2: Conditional frequencies of occurrence of the “warm stratiform” profile type (cluster 2 in Fig. 1) and the “deep convective” profile types (clusters 14+17+20 in Fig. 1). Black denotes lowest frequency of occurrence; white, highest.

the rain-conditional frequency of occurrence of the “warm stratiform” and “deep convective” reduced clusters. The “warm stratiform” cluster, as expected, is dominant over cold ocean gyres (likely stratocumulus) and the “deep convective” over tropical continents. In the “warm stratiform” frequency map, several features are of note: (1) an apparent marine layer is found penetrating the west coast of Africa, (2) most of Africa is ‘depleted’ of shallow warm stratiform rain, relative to South America, (3) very high resolution features, such as the Red Sea, are discernible. The fact that spatially coherent variability exists in these frequencies of occurrence suggests that using them to describe the local “convective spectrum” is a legitimate endeavor.

3. CLUSTER ANALYSIS OF CONVECTIVE SPECTRA

As noted above, the 25 “discovered” clusters have been reduced to 12 “physically distinct” clusters (by simply combining similar clusters). The set of 12 rain-conditional frequecies and 12 unconditional frequencies (24 total) at each map grid point (e.g., in Fig 2) provide the inputs to a further cluster analysis to identify recurring convective spectra. Note that in this study, the inputs are annualized frequencies, so the unconditional inputs (half of the total) are somewhat influenced by seasonality; future analyses will be performed with seasonal inputs.

Convective spectrum cluster analyses were performed requesting x=2,3,...25 clusters. To mitigate noise in the input data (e.g., as in Fig. 2), 2.5 deg spatial averaging was applied to the “raw” profile frequencies of occurrence, which were recorded on a 0.5 deg resolution grid (as in Fig. 2). Maps of these analyses are diffficult to render in grayscale or contour form and not shown here (although their application is shown in section 4, below).

4. SIMILARITY MAPS

The vertical profile and convective spectrum cluster analysis results are immediately capable of answering the question: For a reference location, where else in the world is most convectively similar to this location? (with emphasis on vertical structure and using a nonparametric approach). This question is particularly important for interpretation and application of field program or ground validation data. As such, two “similarity metrics” for a number of “interesting” locations (e.g., locations of high-profile field campaigns) have been computed.

The first metric (left column of Fig. 3) actually ignores the convective spectrum cluster analysis, but uses the same 24-parameter descriptor of each location. It is simply the distance, in 24-space, of each map point’s parameters to the reference location’s parameters (again, these parameters are profile type frequency of occurrence). The second matric (right column of Fig. 3) is defined from a set of 23 (x=2,3...25) convective spectrum cluster analyses (as in Section 3). This metric is simply the number of times each map point occurs in the same cluster as the reference point, in this set of 23 analyses. This metric thus considers not only the multidimensional distance between each location’s spectra, but also the tendency of some spectra to recur (cluster) within that n-space.

The two approaches both yield useful information (Fig. 3). For the three oceanic domains, the distance metric shows that the reference domains are broadly representative of many oceanic regions, although all are most similar to “transitional” regions of the warm pool or ITCZ. The cluster metric refines this interpretation. The GATE domain is perhaps most representative of the west Pacific warm pool (its spectrum is both fairly similar in the 24-input space and unique enough to recur in the same cluster often), while the COARE domain is more representative of the transition/edge of the warm pool, and the Kwajelein domain is even further on the ‘periphery’ of both the warm pool and other convectively active oceanic regions. The TRMM-LBA (Rondonia, Brazil) occurs in the same cluster as many of the maritime continent islands. The ‘front range’ of the Congo basin (the location of the world’s extremum in annual lightning production) exhibits a fairly unique spectrum, matched only by the Colombian highlands (another global

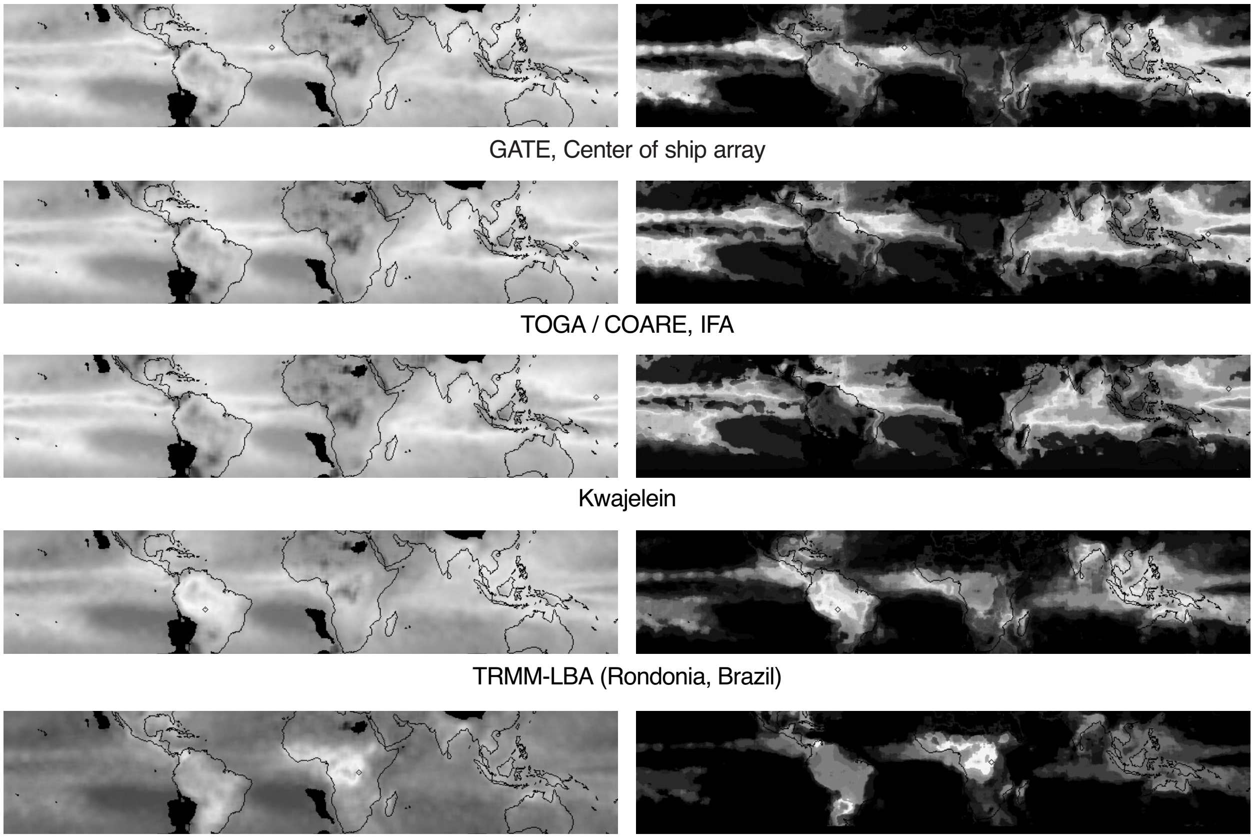

Fig. 3: “Similarity maps” showing metrics of the degree of convective spectrum similarity between the entire tropics and various reference locations (on an annualized basis). Left column denotes the 24-dimensional distance metric, with white areas “most similar”. Right column shows the number of times in the 23 spectrum cluster analyses each location occurred in the same location as the reference location; again, white is “most similar”.

lightning extremum) and some of southern South America. Similarity maps (not shown) were also generated for the southern Great Plains in the U.S., and showed most similarity to the northern African ITCZ / Sahel region, where MCS’ are common. Also examined (and not shown) were the EPIC (East Pacific) domain, Darwin Australia, Socorro New Mexico, the SPCZ, the southeast Pacific cold gyre, Brazil, Dallas Texas, Orlando Florida, and the Colombian highlands. These results are available at: http://homepage.mac.com/wxguyinal/Cluster/Cluster.htm.

The applications of these plots are immediately obvious. For example, from a climatological standpoint, Z-R relationships trained using data from the reference location should, in principle, be most accurate in the most similar regions of the maps. Similarly, regions of similarity may serve as useful control domains within which to examine, e.g., PR and TMI retrieval algorithm discrepancies. Finally, in concert with environmental data as inputs, objectively defined convective spectrum clusters can serve as inputs to cluster analyses which isolate true convective regimes, where a regime could be reasonably described by its combination of forcing, convective response (spectrum) and adjusted state.

5. CONCLUSIONS

This ‘proof of concept’ study demonstrates that automated typing of radar vertical columns, using a somewhat nonparametric approach which retains vertical structure information, is feasible. Once profiles have been classified, the classes can be used to help classify storms (an application not discussed here). The spatio-temporal frequencies of occurrence of these ‘archetypal’ profiles provide a multiparameter description of the realized convective spectrum. Since this spectrum varies greatly across the globe, a similarly nonparametric approach (cluster analysis) can [should] be used to isolate recurring convective spectra, a first step in objective identification of convective regimes. Objective regime identification may be useful in algorithm development, ground validation, satellite retrievals, data assimilation, and empirical or theoretical studies of convective response to environmental forcing.

6. ACKNOWLEDGEMENTS

Data in this study are from the University of Utah ‘reduced Level 1’ dataset, developed by S. Nesbitt and E. Zipser. This research was funded under NRA-99-OES-03.