



## Convective cloud identification and classification in daytime satellite imagery using standard deviation limited adaptive clustering

Todd A. Berendes,<sup>1,2</sup> John R. Mecikalski,<sup>3</sup> Wayne M. MacKenzie Jr.,<sup>2</sup>  
Kristopher M. Bedka,<sup>4</sup> and U. S. Nair<sup>2</sup>

Received 17 April 2008; revised 25 June 2008; accepted 9 July 2008; published 25 October 2008.

[1] This paper describes a statistical clustering approach toward the classification of cloud types within meteorological satellite imagery, specifically, visible and infrared data. The method is based on the Standard Deviation Limited Adaptive Clustering (SDLAC) procedure, which has been used to classify a variety of features within both polar orbiting and geostationary imagery, including land cover, volcanic ash, dust, and clouds of various types. In this study, the focus is on classifying cumulus clouds of various types (e.g., “fair weather,” “towering,” and newly glaciated cumulus, in addition to cumulonimbus). The SDLAC algorithm is demonstrated by showing examples using Geostationary Operational Environmental Satellite (GOES) 12, Meteosat Second Generation’s (MSG) Spinning Enhanced Visible and Infrared Imager (SEVIRI), and the Moderate Resolution Infrared Spectrometer (MODIS). Results indicate that the method performs well, classifying cumulus similarly between MODIS, SEVIRI, and GOES, despite the obvious channel and resolution differences between these three sensors. The SDLAC methodology has been used in several research activities related to convective weather forecasting, which offers some proof of concept for its value.

**Citation:** Berendes, T. A., J. R. Mecikalski, W. M. MacKenzie Jr., K. M. Bedka, and U. S. Nair (2008), Convective cloud identification and classification in daytime satellite imagery using standard deviation limited adaptive clustering, *J. Geophys. Res.*, *113*, D20207, doi:10.1029/2008JD010287.

### 1. Introduction

[2] The main motivation for this paper is the development, evaluation and application of a “cumulus cloud mask” (CCM) as a means to isolate only cumulus clouds in a given satellite scene. The approach used to develop the CCM is a very flexible algorithm, and this is only one potential application (with others to be discussed near the end of this paper). For this study, Geostationary Environmental Operational Satellite (GOES; GOES 12 in particular), the Moderate Resolution Infrared Spectrometer (MODIS), and Meteosat Second Generation’s (MSG) Spinning Enhanced Visible and Infrared Imager (SEVIRI) instruments are used. This paper investigates new techniques for detection of features using data from these three meteorological satellites via an unsupervised clustering approach. The identification procedure is performed through the use of statistical clustering which invokes an expert labeling of defined “clusters” of pixels with similar

properties in the visible (VIS) and infrared (IR) spectrum. Clusters that are identified as cumulus clouds in various stages of growth are labeled as such across multiple satellite scenes. Once trained, the algorithm is automated to highlight only cumulus across any satellite image the method is applied to. In the present study, GOES 12, MSG SEVIRI, and MODIS imagery for similar scenes are presented to show the relative portability of this method to various satellite platforms.

[3] As proof of functionality, the CCM has been tested for several years as part of the Satellite Convection Analysis and Tracking (SATCAST) System. SATCAST currently provides 0- to 1-h forecasts of convective initiation (CI) using Imager data from GOES satellites [Mecikalski and Bedka, 2006]. SATCAST relies on determining the location of cumulus in varying stages of growth, from “fair weather” to “towering” (e.g., cumulus humulus). Once these clouds have been identified (among the many other cloud types within a visible satellite image), focused monitoring of the Lagrangian evolution of 1-km-resolution cumulus can be done as they grow, glaciate, precipitate, and produce lightning. Cumulus cloud tracking is accomplished using the Bedka and Mecikalski [2005] methodology for producing very dense, “mesoscale” atmospheric motion vectors (AMVs). The Bedka and Mecikalski [2005] AMV procedure follows from the more traditional “cloud motion” winds work of Velden *et al.* [1997, 1998] and Dunion and Velden [2002].

<sup>1</sup>Information Technology and Systems Center, University of Alabama in Huntsville, Huntsville, Alabama, USA.

<sup>2</sup>Earth System Science Center, University of Alabama in Huntsville, Huntsville, Alabama, USA.

<sup>3</sup>Atmospheric Sciences Department, University of Alabama in Huntsville, Huntsville, Alabama, USA.

<sup>4</sup>Cooperative Institute for Meteorological Satellite Studies, University of Wisconsin-Madison, Madison, Wisconsin, USA.

[4] Several aspects of the GOES data stream are used within SATCAST to form satellite infrared–CI relationships: cloud growth rates via IR cloud top temperature ( $T_B$ ) changes  $\partial(10.7 \mu\text{m } T_B)/\partial t$ , and glaciation via  $10.7 \mu\text{m } T_B$ ,  $13.3\text{--}10.7 \mu\text{m}$ , and  $\partial(13.3\text{--}10.7 \mu\text{m})/\partial t$  [Mecikalski and Bedka, 2006]. Currently, the SATCAST procedure is being tested within several systems that monitor for CI on 0- to 2-h timescales as part of the Advanced Satellite Aviation Weather (ASAP) initiative [Mecikalski et al., 2007]. These include the “AutoNowcaster” [Mueller et al., 2003] and the deterministic 0- to 2-h Tactical Convective Weather Forecast (TCWF) as part of the FAA’s Corridor Integrated Weather System (CIWS) currently operated by MIT Lincoln Laboratory in a concept exploration mode [Evans et al., 2004]. The CIWS system is used by air traffic personnel to improve traffic flow on jet routes, and to provide proactive reroutes when convective weather impacts their capacity [Wolfson et al., 2004]. SATCAST is also being developed for use with the Meteosat Second Generation (MSG) satellite. Again, proof of concept for the CCM, as presented herein, partly comes from the use of CI nowcasts by the SATCAST algorithm within existing, sophisticated systems, as well as by various end users including regional National Weather Service Forecast Offices (NWSFO). We refer the reader to the works Mecikalski and Bedka [2006] and Mecikalski et al. [2007] for examples of SATCAST. Validation of SATCAST has recently been performed by Mecikalski et al. [2008].

[5] The clustering algorithm developed here is tested on the VIS and IR channels on GOES (4 IR channels), MODIS (18 IR channels of the possible 36 available) and MSG (9 IR channels) data as a means of comparing how this method may be applied to real-time imagery.

[6] This paper proceeds as follows: Section 2 provides background into classification methods designed for use with satellite-based data. Section 3 describes the data used in this study, while section 4 outlines the methodology. Examples are presented in section 5 and the study is discussed and concluded in section 6.

## 2. Background

[7] Using aircraft and satellite observations and numerical cloud models, a number of studies have examined various aspects of cloud field structure [Plank, 1969; Welch et al., 1988; Cahalan and Joseph, 1989; Ramirez and Bras, 1990; Sengupta et al., 1990; Joseph and Cahalan, 1990; Weger et al., 1992; Zhu et al., 1992; Kuo et al., 1993; Nair et al., 1998; Han and Ellingson, 1999; Nair et al., 2001]. Identification of a particular cloud type (or types within a given class of clouds—stratus, cumulus, cirrus) has many potential applications. One, as noted above, is in aviation systems that attempt to forecast for hazards related to thunderstorms. These include the AutoNowcaster and CIWS to name a couple. The hazards include turbulence, strong wind shear, reduced visibility and heavy precipitation. Other likely areas in which delineation of atmospheric is important include the identification of aerosols and volcanic ash clouds, given that aerosols and ash come in a variety of forms and compositions. Aerosols have proven effects on human and plant health [Pope et al., 1995], while volcanic ash is an aviation safety hazard [Casadevall, 1993].

[8] There are other areas that would benefit from improved understanding of boundary layer cumulus cloud field structure, such as parameterization of shallow convection in general circulation models (GCMs). Since shallow convection is strongly influenced by surface characteristics, areas of significant land use change are expected to show indications of local climate modification [Nair et al., 2000]. Statistical descriptions of the temporal variability of cloud field properties, useful for parameterization in larger-scale models, are not currently available [Joseph and Cahalan, 1990; Nair et al., 1998; Han and Ellingson, 1999]. Studies suggest microphysical properties of cloud fields may be linked to atmospheric dynamic and thermodynamic structure [Sykes and Henn, 1989; Weckwerth et al., 1997; Rosenfeld et al., 2008]. Knowledge of such relationships would be useful for parameterization of cloud field characteristics in large-scale models, as well as for short-term (0–6 h) weather forecasting. Satellite data are ideal to further our understanding of the issues related to boundary layer cumulus cloud field characteristics. In particular, Landsat, ASTER and Ikonos imagery can be used to resolve high-resolution features (30–100 m), data from the MODIS have capabilities to resolve features from (250 m to 1 km), and geostationary data can resolve cloud properties at 1–4 km with the added value of measuring their temporal variations. Note that the analysis of large volumes of data is necessary to develop good statistical descriptions of cloud field structure. This requires the use of automated methods for identifying cumulus cloud fields in satellite imagery.

[9] Supervised classification techniques utilizing neural networks, maximum likelihood, and other algorithms have proven successful in cloud detection and classification [Baum et al., 1997; Berendes et al., 1999]. Traditional supervised techniques require the construction of a large database of training samples that must be selected from imagery and correctly labeled by an expert analyst. Accurate and consistent labeling of training samples can be a tedious, time-consuming, and error-prone process. As a basis for improved efficiency, studies that capitalize on the textural and spectral signatures of clouds have been used to classify cloud high-resolution Landsat multispectral imagery [see, e.g., Chen et al., 1989; Wielicki and Welch, 1986; Welch and Wielicki, 1989]. However, the classification schemes presented in these studies suffer from the limitation that they are trained for a specific environment, such as for maritime [Bankert, 1994], tropical [Shenk et al., 1976; Inoue, 1987], or polar [e.g., Ebert, 1987, 1989; Key and Barry, 1989; Key, 1990; Welch et al., 1989, 1992; Rabindra et al., 1992; Tovinkere et al., 1993] regions. Peak and Tag [1992] developed a method of cloud classification based on a method using a cloud classification database, image segmentation and neural network identification. The Peak and Tag [1992] approach uses VIS and IR data from GOES and the advanced very high resolution radiometer (AVHRR) satellites. Tag et al. [2000] described an automated cloud classification system using AVHRR data for use in the interpretation of synoptic-scale events as an aid in forecasting. Their approach is based on a 1-nearest neighbor classifier, in which comparisons are made to a cloud type database selected by human experts, henceforth comprising a supervised classification system. A method similar to the

*Tag et al.* [2000] approach has been implemented on real-time GOES imagery (see <http://www.nrlmry.navy.mil>).

[10] Unsupervised classification techniques utilize various distance or similarity metrics in order to segment imagery into spatially or spectrally similar groups or clusters. Cluster membership statistics are computed from sample imagery and clusters are determined on the basis of a cluster distance or similarity metric. Resultant clusters are initially unlabeled and must be identified and labeled by an expert.

[11] The Standard Deviation Limited Adaptive Clustering (SDLAC) technique is an iterative, statistically based method that is similar to the ISODATA clustering algorithm [see *Gonzalez and Wintz*, 1977]. However, unlike ISODATA, SDLAC does not require initial guesses for cluster centers. The SDLAC algorithm implements an iterative standard deviation threshold adjustment along with several iterative adjustments to the cluster centers. In essence, new clusters are created during each iteration of the algorithm as the cluster centers and thresholds are refined. The SDLAC method has been applied successfully for detection and classification purposes. A convective cloud mask has been created using GOES data. SDLAC has been used on MODIS to detect volcanic ash plumes and to identify tropopause penetration from deep convective updrafts (also known as “overshooting tops”).

[12] The physical information one can glean from the clustering algorithm is the composition of the cluster centers themselves. For example, the “cumulus cloud” class is defined by a set of clusters, with each cluster consisting of a mean and standard deviation for each channel (or feature, i.e., channel difference, ratio, texture, etc.). Upon decomposition of a cluster, the range of each individual feature is determined. One particular “cumulus” cluster may have a high GOES channel 1 (visible data) variability, with a medium standard deviation and a simultaneously large channel 1 contrast value and a large standard deviation. This type of information could be interpreted physically, yet in practice it may be difficult to know which channels are most important in defining a particular class.

[13] One of the properties of SDLAC is that several clusters may be formed for the same class. Sometimes the defining difference between clusters is produced by variations in a particular feature not of interest for the given application. For instance, in the CCM, there is no need to differentiate between land, water, and other background classes based upon reflectance since we are only concerned with clouds; however, since reflectance is a feature, extraneous clusters are defined because of variations of reflectance of which the land surface plays a part. Fortunately, extraneous clusters can either be merged together into meaningful classes or removed during the cluster labeling process.

[14] The SDLAC clustering algorithm is an unsupervised classification technique that groups statistically similar pixels in the data. Our implementation of the algorithm is not designed to determine which features are most important (physically) for a given class, but simply finds the best fit for each cluster on the basis of the clustering algorithm. It is important to note that other algorithms are much better suited for direct physical interpretations, such as ontologies

and decision trees, where the classification is defined physically from the start.

[15] Feature selection is one way to algorithmically determine important features for defining specific classes. This method provides insight (thresholds, ranges, etc.) into the effect of specific features on the classification accuracy. In order to do feature selection, however, one requires labeled training data, something that is also a requirement for a supervised technique. Clustering algorithms can be used to effectively “bootstrap” training data, yet that is beyond the scope of what we were trying to accomplish in this study. The strength of the clustering algorithm is its automatic nature that does not require a preconceived definition of each class. Clustering allows pixels to be grouped “naturally” according to the input features used. Clusters can then be grouped together into logical classes, which is much easier than selecting training samples for individual classes as required for supervised learning techniques.

### 3. Data

[16] The data used in this study from GOES 12 are one VIS channel (central wavelength  $0.65 \mu\text{m}$ ) and four IR channels (3.9, 6.5, 10.7, and  $13.3 \mu\text{m}$ ). All GOES 12 infrared data are linearly interpolated to 1-km resolution (away from the native 4-km IR resolution) so all spectral channels are the same spatial resolution for processing. On SEVIRI, all 12 channels are used, with one broadband VIS channel (central wavelength  $0.7 \mu\text{m}$ ) and 8 IR and 3 visible channels (0.6, 0.8, 1.6, 3.9, 6.2, 7.3, 8.7, 9.7, 10.8, 12.0, and  $13.4 \mu\text{m}$ ). The SEVIRI broadband visible channel is at the native 1-km resolution, and all other channels are replicated to 1 km resolution. From MODIS, the 1-km-resolution channels used consist of the following wavelengths: 0.64, 1.37, 3.75, 3.96, 6.72, 7.3, 8.5, 9.7, 11.0, 12.0, and  $13.3 \mu\text{m}$ . The MODIS channels were chosen to be similar to the channels available in SEVIRI. The use of all 1-km data allows us to compare clustering, cloud identification results without resolution-related issues.

[17] In addition to the spectral channels, a new feature was created by computing an  $11 \times 11$  neighborhood grey level difference vector (GLDV) based spatial texture for each pixel in the high-resolution visible channel of each instrument. Grey level differences are computed over the  $11 \times 11$  pixel grid using a distance of 1 pixel with angles of  $0^\circ$ ,  $45^\circ$ , and  $90^\circ$ , producing a grey level difference vector containing the counts of each grey level difference normalized by the number of differences computed. The “contrast” measure is defined as

$$CON = \sum_{i=1}^n i^2 * GLDV(i), \quad (1)$$

where  $n$  is the number of grey level differences. The GLDV contrast feature is useful for finding cumulus and stratocumulus clouds [Welch et al., 1988].

[18] The training phase of the classifiers utilizes imagery that contains clouds of various types with emphasis on convective clouds. The training data sets for the classifiers consist of daytime imagery from over 450 GOES 12 scenes,

42 MODIS scenes, and 75 MSG scenes. The MODIS classifier was compared to coincident GOES 12 and SEVIRI images as a means of demonstrating that similar cloud identification can be obtained from two different sensors with different spectral channels.

#### 4. Methodology

[19] Image clustering using SDLAC begins with the selection and computation of image features. Features are imager data specific, but may include imager channels, channel differences and ratios, vegetation indices, textures, and other derived features. Features are normalized to a fixed range (0–255) on the basis of either a computed or specified physical range.

[20] SDLAC clustering of a satellite scene is performed on a pixel-by-pixel basis. First, a subset of the pixels in the scene is selected for processing. The first pixel processed forms the seed for the first cluster. As additional pixels are processed, they are tentatively added to each existing cluster. Tentative feature means  $\mu_k^*(q)$  and standard deviations  $\sigma_k^*(q)$  are computed for each cluster  $k$  according to

$$\begin{aligned}\mu_k^*(q) &= \frac{1}{n_k + 1} [X_q(i^*, j^*) + n_k \mu_k(q)] \\ \sigma_k^*(q) &= \sqrt{\frac{1}{n_k - 1} \left[ n_k \sigma_k^2 + \frac{n_k}{n_k + 1} (X_q(i^*, j^*) - \mu_k(q))^2 \right]}\end{aligned}\quad (2)$$

where  $k$  is the cluster,  $n_k$  is the number of pixels in cluster  $k$ ,  $q$  is the feature (i.e., channel), and  $X_q(i^*, j^*)$  is the normalized value of feature  $q$  at pixel  $(i^*, j^*)$ . The tentative feature means and standard deviations are used as a basis for constructing clusters and determining cluster membership.

[21] The SDLAC method does not require any initial guesses at the number and position of the image clusters. Instead of initial cluster guesses, initial standard deviation thresholds  $\theta(q)$  are specified for each feature. The standard deviation thresholds are arbitrary numbers between 0 and 255. Large  $\theta(q)$  values produce fewer clusters with larger standard deviations while smaller  $\theta(q)$  values produce more numerous, tighter clusters. Typically, smaller  $\theta(q)$  values are chosen initially, and during the iterative clustering algorithm the  $\theta(q)$  values are gradually increased until a stable cluster set is formed.

[22] The standard deviation limiting function,  $\Theta_k(q)$ , one of the two metrics [Levine and Shaheen, 1981] used to determine cluster membership, is defined by

$$\Theta_k(p) = S_k(q)\theta(q), \quad (3)$$

where

$$S_k(q) = \left[ 1 - \frac{\sigma_k^*(q)}{C(q)} \right].$$

Also,  $C(q)$  is a scale factor chosen to insure the ratio in  $S_k(q)$  is  $< 1$ , and  $\theta(q)$  is an adapted standard deviation threshold.

[23] We compute  $\Theta_k(q)$  by scaling the standard deviation thresholds,  $\theta(q)$ , by a standard deviation limiting scale factor  $S_k(q)$ . As the standard deviation gets larger,  $S_k(q)$  decreases the  $\theta(q)$  values, preventing larger clusters from

absorbing too many pixels. The decrease of  $\theta(q)$  has the effect of starting new nearby clusters which effectively splits larger clusters. Conversely, when the standard deviation is low, the smaller cluster will retain most the  $\theta(q)$  value and new clusters will not form as easily.

[24] The change in cluster means,  $\Delta\mu_k(q)$ , is the other clustering metric used in conjunction with  $\Theta_k(q)$ . It is defined by

$$\Delta\mu_k(q) = |X_q(i^*, j^*) - \mu_k(q)| \quad (4)$$

A single clustering iteration is accomplished by processing all of the selected pixels in a scene using the following algorithm.

[25] 1. For each selected pixel  $X_q(i^*, j^*)$ , compute tentative feature means  $\mu_k^*(q)$  and standard deviations  $\mu_k^*(q)$ .

[26] 2. Compute standard deviation limiting function,  $\Theta_k(q)$ .

[27] 3. Compute change in cluster means,  $\Delta\mu_k(q)$ .

[28] 4. Determine cluster membership as follows: (1) If  $\Delta\mu_k(q) \leq \Theta_k(q)$  for all features  $q$ , only in cluster  $k$ , merge pixel with cluster  $k$  and update cluster statistics:  $\Delta\mu_k(q) = \mu_k^*(q)$  and  $\sigma_k(q) = \mu_k^*(q)$ . (2) If more than one cluster satisfies case 1, merge to cluster, producing minimum  $\Sigma[\Delta\mu_k(q)]^2$  and update cluster statistics as in case 1. (3) If no cluster satisfies case 1 or 2, create a new cluster (seed).

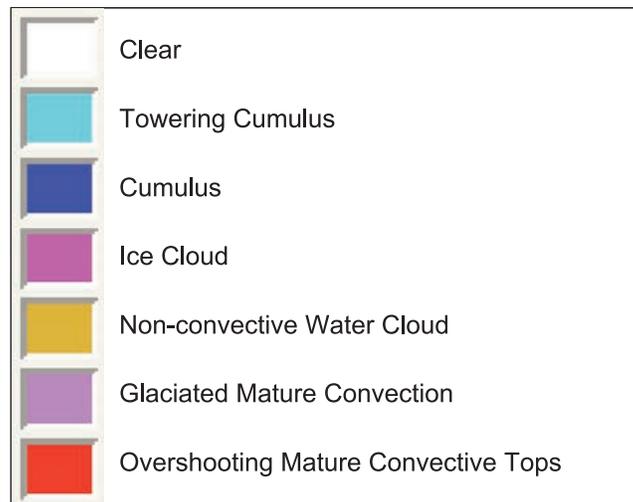
[29] Steps 1–4 are repeated until all pixels are processed.

[30] After a single pass of the clustering algorithm, we are left with a set of clusters based upon our arbitrary initial  $\theta(q)$  values. Since the SDLAC algorithm does not initially have knowledge of the actual distribution of the clusters in the image, we must devise a method of determining optimal clustering performance.

[31] Like most clustering techniques, SDLAC is an iterative algorithm. The stability of the clusters is evaluated by counting the number of pixels that have changed cluster membership since the previous iteration.

[32] At the end of each iteration, the following procedure is then applied: (1) For each cluster update, feature means  $\Delta\mu_k(q)$  and standard deviations  $\sigma_k(q)$  using only pixels are added to the clusters during the current iteration; the updated clusters are used as seeds for the next iteration. (2) Small clusters containing less than a specified number of pixels are removed. (3) Clusters with standard deviations of zero in several features usually represent image noise or data errors and are removed. (4) The standard deviation thresholds,  $\theta(q)$ , are incremented to aid convergence. (5) If the percentage of pixels that change cluster membership since the previous iteration is less than a specified amount, the image has been clustered successfully and iterations are stopped. (6) Otherwise, continue iterations until the image is successfully clustered or fails after the number of iterations exceeds a specified value.

[33] In practice, it is usually possible to achieve a successful clustering of an image by adjusting initial  $\theta(q)$  and increment values. The initial  $\theta(q)$  and increment values are data specific, and must be adjusted by experimentation to achieve a good clustering. If the initial values are too low, extraneous clusters are formed which may slow down the convergence process. If the initial values are too high, small but important clusters may be merged with larger clusters



**Figure 1.** Classes defined by SDLAC convective cloud mask.

and class resolution may be lost. Since the values  $\theta(q)$  can be incremented during iteration, it is very effective to select slightly lower initial values and allow the algorithm to find the optimal value during the iteration process.

[34] The final step in the creation of a cluster-based classifier is the “blending” of the clusters from individual images into a single combined cluster set. By considering clusters as two separate normal distributions, the mean and standard deviation update equations for the merging of cluster  $l$  into cluster  $k$  can be written as

$$\begin{aligned} \mu_k^*(q) &= \frac{(\mu_k n_k + \mu_l n_l)}{n_k + n_l} \\ \sigma_k^*(q) &= \sqrt{\frac{(n_k + n_l)[n_k(\sigma_k^2 + \mu_k^2) + n_l(\sigma_l^2 + \mu_l^2)] - (\mu_k n_k + \mu_l n_l)^2}{(n_k + n_l)^2}} \end{aligned} \quad (5)$$

[35] The new mean and standard deviation update equations are applied and blending of the clusters proceeds using the same iterative algorithm that is applied to cluster the pixels in individual scenes. The resulting cluster set represents all of the scenes of the training image set and provides the basis for the creation of a nonscene specific classifier.

[36] After the clustering is completed, clusters must be identified and labeled. Labeling of the clusters is done by a human expert, using the Interactive Visualizer and Image Classifier for Satellites (IVICS) visualization software [Berendes *et al.*, 2001]. Physical classes of interest are identified in the imagery and each cluster is labeled as one of those class types. Many clusters may be combined into a single physical class, and unneeded clusters may remain unlabeled or may be removed entirely.

[37] Finally, the labeled cluster set is used to classify imagery. Image classification is accomplished by computing the mean and standard deviation update equation (1) and applying the cluster membership functions (2) and (3) using the labeled cluster set as initial clusters. Each individual pixel in the image is assigned the class label of the cluster to which it would merge. The standard deviation limiting function,  $\Theta_k(q)$ , can either be applied or ignored. If  $\Theta_k(q)$

is applied, the pixel can be labeled as “unknown” if it would form a new cluster based upon the logic in step 4 of the clustering algorithm. Conversely, if  $\Theta_k(q)$  is ignored, pixels are simply labeled by the cluster which minimizes  $\Sigma[\Delta\mu_k(q)]^2$ , ensuring that every pixel is labeled as a class.

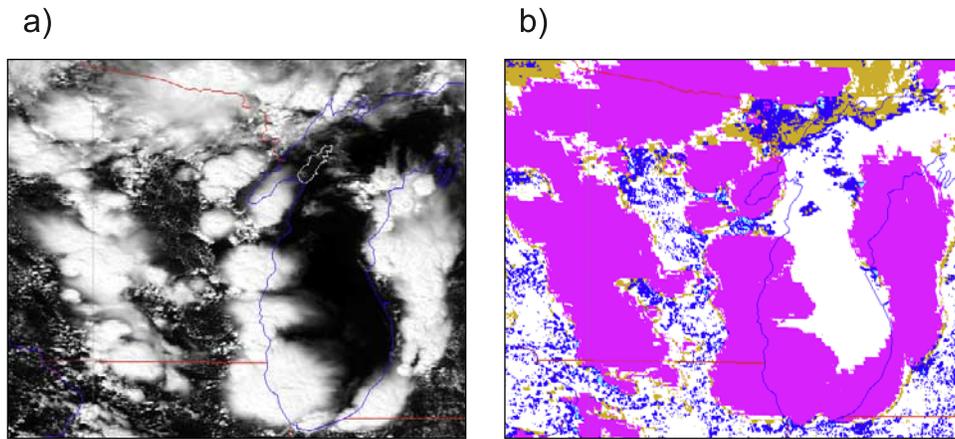
[38] It has been shown that spatial textures produce valuable features for cloud detection and classification [Sengupta *et al.*, 1990]. The gray level difference vector (GLDV) method is a computationally efficient and effective method that has been used successfully in cloud classification algorithms [Chen *et al.*, 1989]. The GLDV “contrast” measure applied to the GOES visible channel enhances small-scale spatial variations of reflectance that are indicative of cumulus clouds. Therefore, the GLDV contrast of the visible GOES 12, MODIS and MSG channel is used as an additional clustering feature. This textural derived “channel” has proven to be a critical component for identifying daytime cumulus cloud types.

[39] The iterative SDLAC algorithm can be tuned to produce very stable clusters using a wide variety of satellite data types and can be utilized as a tool for cloud identification (e.g., *cumulus* identification for convection initiation purposes). Applications of the SDLAC algorithm are discussed in subsequent sections.

[40] There are limits of applicability that are dependent upon the representative imagery used in the training process. For example, if you used only polar imagery to create and label the clusters, the classifier would not be applicable to desert imagery simply because there would be no clusters of the desert classes represented in the polar clusters. The SDLAC algorithm can be applied with a very specific focus (i.e., trained for a specific geographic region, season, or classes) or it can be applied in a very general way (i.e., trained using a broad set of imagery representing multiple location, seasons, etc.)

## 5. Examples

[41] The SDLAC clustering algorithm for GOES 12, MODIS and SEVIRI sensors is applied to produce a convective cloud mask for each data type. The clustering



**Figure 2.** (a) GOES 12 visible image centered over Lake Michigan taken at 2100 UTC 3 August 2003 and (b) color-coded SDLAC cloud mask (refer to Figure 1 for color code labels).

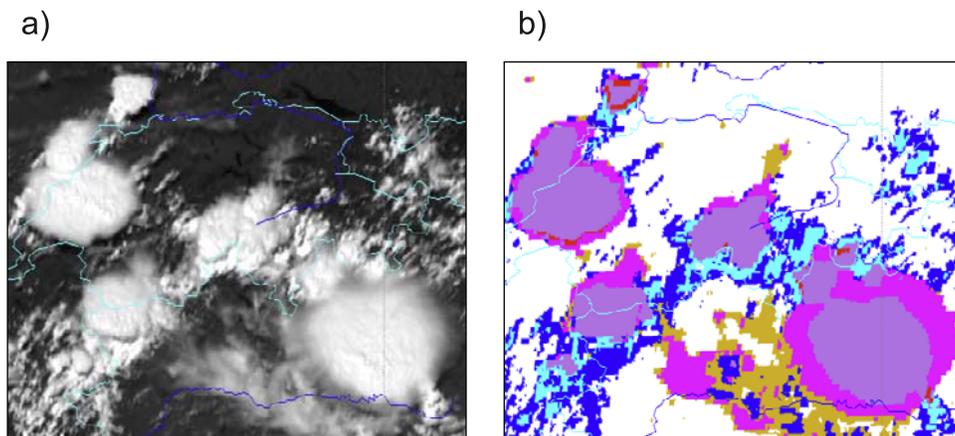
algorithm was applied to a set of images from each sensor and the resulting clusters were labeled as belonging to seven different classes. Figure 1 shows a list of the classes along with associated colors that will be used in subsequent figures. The classes were defined on the basis of visually identifiable features in the image. Note that the main goal was to identify convective clouds (“towering cumulus” and “cumulus”) in various stages for use in convective now-casting systems. Toward that end, minimal effort was made to distinguish cloud types beyond convective types and they have been grouped into general categories such as “ice cloud” and “nonconvective water cloud.” Two additional classes, “glaciated mature convection” and “overshooting convective tops,” are also being examined for possible application in turbulence studies.

[42] Figure 2a shows an example GOES channel 1 visible image over Lake Michigan and the resulting color-coded SDLAC image can be seen in Figure 2b. The primary function of this mask (i.e., the detection of convective clouds) is represented by the dark blue and cyan areas in Figure 2b. Visual comparison of Figures 2a and 2b shows that the convective classes (“cumulus” and “towering cumulus”) are well detected. Large pink areas of ice cloud

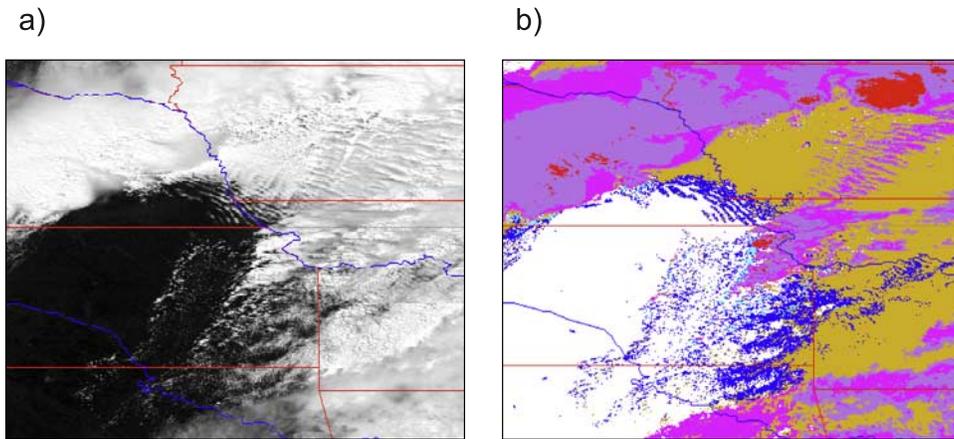
can also be seen in Figure 2. The ice cloud class represents primarily diffuse (visually fuzzy) ice clouds that are warmer than glaciated convective tops and produce a low GLDV contrast value. No glaciated mature convection or overshooting tops are visible in the scene.

[43] An MSG SEVIRI example over northern Italy is shown in Figure 3 with channel 1 shown in Figure 3a and the SDLAC results in Figure 3b. Visual examination shows that similarly to the GOES case, the MSG SDLAC mask detects convective clouds well. Additionally, the MSG example contains area labeled “glaciated mature convection” which represent visually diffuse ice clouds at very low temperatures above a large convective cloud. Some of the “bumpy” texture of the cumulus cloud can be seen through the ice cloud because of gravity waves from the convective updraft penetrating the tropopause. The additional channels in the MSG SEVIRI sensor may be providing more information about cloud phase to help make a better distinction than the GOES version.

[44] MODIS allows for an examination into the benefits into increased spectral channels, and 1-km infrared spatial resolution. Figure 4a shows an image over Kansas along with the corresponding SDLAC mask results in Figure 4b.



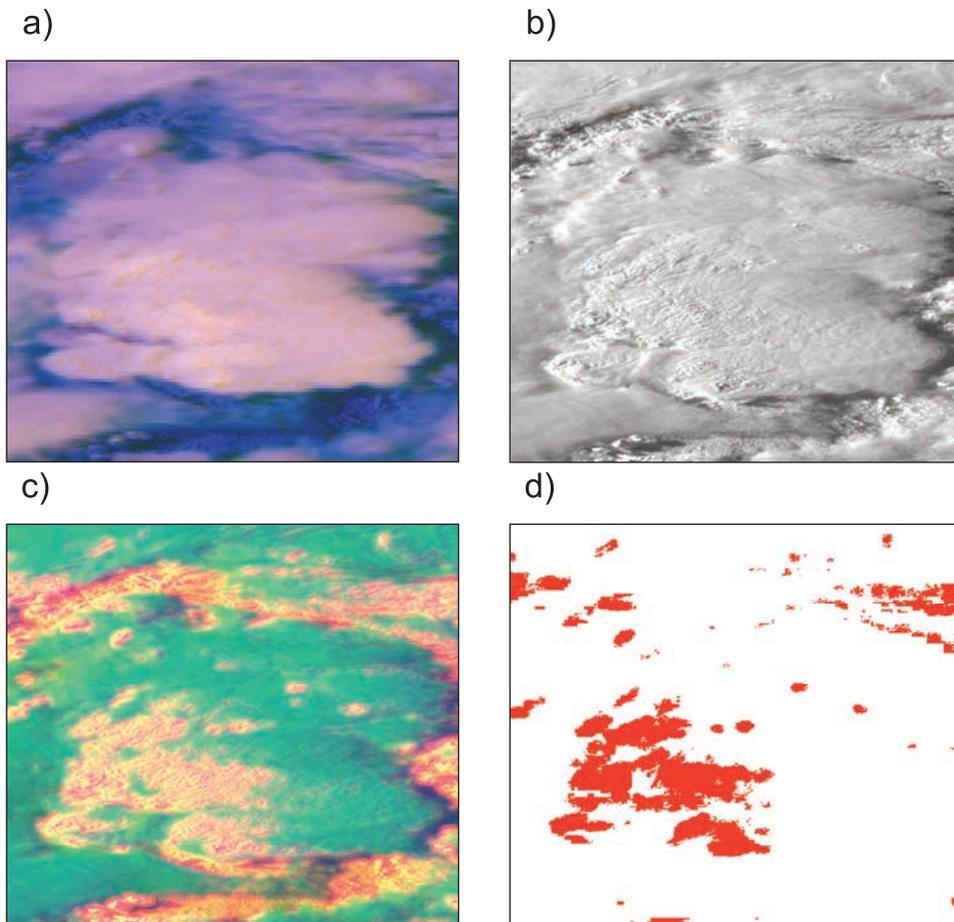
**Figure 3.** (a) MSG SEVIRI image centered over northern Italy taken at 1310 UTC 25 June 2006 and (b) SDLAC convective cloud mask (refer to Figure 1 for color code labels).



**Figure 4.** (a) MODIS image centered over northern Kansas taken at 1815 UTC 4 May 2003 and (b) SDLAC convective cloud mask (refer to Figure 1 for color code labels).

As in the MSG and GOES cases, the smaller convective clouds shown in blue in Figure 4b are well represented. Large areas of “nonconvective water cloud,” in this case stratocumulus, are shown in dark yellow. The stratocumulus

is warmer with lower GLDV texture. The mature convective areas are present in the anvil of the large thunderstorm. The upper right corner of Figure 4b shows a red area of “overshooting mature convective tops,” characterized by



**Figure 5.** Example of an “overshooting top” mask for active convection. Data are from GOES 12: (a) a three-band enhanced color composite with channel 4 inverted in red, channel 3 in green, and channel 1 in blue; (b) an enhanced channel 1 visible image; (c) a three-band enhanced color composite with GLDV texture of channel 1 in red, channel 1 in green, and channel 2 inverted in blue; and (d) the “overshooting top” mask with red indicating tops. Data are from 2315 UTC 10 May 2003 over Illinois.

extremely low temperature with high GLDV textures and generally high channel 1 ( $0.6 \mu\text{m}$ ) reflectance. The additional channel information present in MODIS data may aid in the detection of overshooting tops.

[45] The tropopause penetrations within deep convection, that is, “overshooting mature convective tops” are also detectable in GOES and MSG data using the SDLAC method. A version of the SDLAC clustering mask was produced specifically for the purpose of detecting overshooting tops in GOES imagery. Figure 5 shows an example of GOES imagery with active convection over Illinois. Figure 5a shows a three-band color composite of the active system while Figure 5b shows the same area in the visible channel. Areas of overshooting tops are very cold and have a rougher textured appearance due to the convection protruding above the cirrus anvil. Figure 5c shows a three-band color composite with channel 1 in green, channel 2 inverted in blue, and the GLDV texture of channel 1 in red. Using this color scheme, the areas of high GLDV texture are clearly visible as red and yellow areas. The highly textured areas correspond to the overshooting tops and smaller convective clouds. Figure 5d shows the overshooting tops detected by the SDLAC algorithm in red. Clearly, the overshooting tops correspond well with the areas of high channel 1 GLDV texture, but the smaller convective clouds are eliminated, probably on the basis of temperature. Accurate detection of overshooting tops allows for identification of strong convection, and can be used as a tool for identification for the potential for aircraft turbulence. This has a strong application for aviation safety [see *Mecikalski et al.*, 2007]. There appears to be some areas shown which may not be overshooting tops, however maybe gravity waves or other areas that may be slightly penetrating the tropopause. It is important to note that the areas highlight will be highly turbulent and thus still important to aviation safety.

## 6. Conclusions

[46] SDLAC is a cluster-based cloud classification technique that is currently being used within an operational framework. While it is difficult to validate a “cloud mask” or other cloud classifications because of the subjective nature of cloud classification and the lack of “truth” data, the efficacy of the SDLAC convective cloud mask can be seen by subjective examination of the imagery. Additionally, the operational results of the SATCAST algorithm objectively show that the SDLAC convective cloud mask is accurately detecting the convectively active clouds [Mecikalski et al., 2008], which was the primary design objective. The general nature of the SDLAC clustering algorithm produces a classification technique that is not satellite platform dependent. As satellite technology improves (as in the case of GOES R [Schmit et al., 2005]), this method allows easy addition of the new spectral channels for possible improvement to cloud classification.

[47] Using many cases within different seasons and different times, the statistical information of each cluster becomes robust. In our examples with GOES and MSG, we used sufficient cases over various seasons and viewing angles, reducing dependencies on time of the year and location. This enhances the utility of the SDLAC method

as a tool for cloud identification within automated systems such as SATCAST. The SDLAC algorithm has been successfully applied as a detection algorithm for overshooting convective cloud tops. This detection capability has a potential aviation safety application for turbulence detection.

[48] **Acknowledgments.** This research was supported by National Aeronautical Space Administration grants NAS5-31718 and NAS1-98131. Additional funding has been provided by NASA grant NNX07AFF14G. The GOES 12 data were acquired by GOES ground station at Global Hydrology and Climate Center (GHCC), Huntsville, Alabama. The MSG data were kindly provided by the National Environmental Satellite, Data, and Information Service (NESDIS) center in Camp Spring, Maryland.

## References

- Bankert, R. L. (1994), Cloud classification of AVHRR imagery in maritime regions using a probabilistic neural network, *J. Appl. Meteorol.*, *33*, 909–918.
- Baum, B. A., V. Tovinkere, J. Titlow, and R. M. Welch (1997), Automated cloud classification of global AVHRR data using a fuzzy logic approach, *J. Appl. Meteorol.*, *36*, 1519–1540.
- Bedka, K. M., and J. R. Mecikalski (2005), Application of satellite-derived atmospheric motion vectors for estimating mesoscale flows, *J. Clim. Appl. Meteorol.*, *44*, 1761–1772.
- Berendes, T. A., K. S. Kuo, A. M. Logar, E. M. Corwin, R. M. Welch, B. A. Baum, A. Pretre, and R. C. Weger (1999), A comparison of paired histogram, maximum likelihood, class elimination, and neural network approaches for daylight global cloud classification using AVHRR imagery, *J. Geophys. Res.*, *104*, 6199–6213, doi:10.1029/98JD02584.
- Berendes, T. A., R. M. Welch, U. S. Nair, and D. A. Berendes (2001), Interactive Visualizer and Image Classifier for Satellites (IVICS), *Eos Trans. AGU*, *82*(20), Spring Meet. Suppl., Abstract U21A-16.
- Cahalan, R. F., and J. H. Joseph (1989), Fractal statistics of cloud fields, *Mon. Weather Rev.*, *117*, 261–272, doi:10.1175/1520-0493(1989)117<0261:FSCOF>2.0.CO;2.
- Casadevall, T. J. (1993), Volcanic ash and airports—Discussion and recommendations from the Workshop on Impacts of Volcanic Ash on Airport Facilities, *U.S. Geol. Surv. Open File Rep.* 93–518, 52 pp.
- Chen, D. W., S. K. Sengupta, and R. M. Welch (1989), Cloud field classification based upon high spatial resolution textural features: 2. Simplified vector approaches, *J. Geophys. Res.*, *94*(D12), 14749–14765, doi:10.1029/JD094iD12p14749.
- Dunion, J. P., and C. S. Velden (2002), Application of surface-adjusted GOES low-level cloud-drift winds in the environment of Atlantic tropical cyclones, Part I: Methodology and validation, *Mon. Weather Rev.*, *130*, 1333–1346.
- Ebert, E. (1987), A pattern recognition technique for distinguishing surface and cloud types in the polar regions, *J. Appl. Meteorol.*, *26*, 1412–1427.
- Ebert, E. E. (1989), Analysis of polar clouds from satellite imagery using pattern recognition and a statistical cloud analysis scheme, *J. Appl. Meteorol.*, *28*, 382–399.
- Evans, J. E., S. Allan, and M. Robinson (2004), Quantifying delay reduction benefits for aviation convective weather decision support systems, paper presented at the 11th Conference on Aviation, Range and Aerospace, Am. Meteorol. Soc., Hyannis, Mass., 3–8 October.
- Gonzalez, R. C., and P. Wintz (1977), *Digital Image Processing*, 431 pp., Addison-Wesley, Reading, Mass.
- Han, D., and R. G. Ellingson (1999), Cumulus cloud formulations for longwave radiation calculations, *J. Atmos. Sci.*, *56*, 837–851, doi:10.1175/1520-0469(1999)056<0837:CCFFLR>2.0.CO;2.
- Inoue, T. (1987), A cloud type classification with NOAA 7 split-window measurements, *J. Geophys. Res.*, *92*(D4), 3991–4000.
- Joseph, J. H., and R. F. Cahalan (1990), Nearest neighbor spacing of fair weather cumulus clouds, *J. Appl. Meteorol.*, *29*, 793–805, doi:10.1175/1520-0450(1990)029<0793:NNSOFW>2.0.CO;2.
- Key, J. (1990), Cloud cover analysis with Arctic advanced very high resolution radiometer data: 2. Classification with spectral and textural measures, *J. Geophys. Res.*, *95*(D6), 7661–7675.
- Key, J., and R. Barry (1989), Cloud cover analysis with Arctic AVHRR data: 1. Cloud detection, *J. Geophys. Res.*, *94*(D15), 18521–18535.
- Kuo, K. S., R. M. Welch, and R. C. Weger (1993), The three-dimensional structure of cumulus clouds over the ocean: 1. Structural analysis, *J. Geophys. Res.*, *98*, 20685–20711, doi:10.1029/93JD02331.
- Levine, M. D., and S. I. Shaheen (1981), A modular computer vision system for picture segmentation and interpretation, *IEEE Trans. Pattern Anal. Mach. Intel.*, *3*, 540–556.

- Mecikalski, J. R., and K. M. Bedka (2006), Forecasting convective initiation by monitoring the evolution of moving convection in daytime GOES imagery, *Mon. Weather Rev.*, *134*, 49–78, doi:10.1175/MWR3062.1.
- Mecikalski, J. R., et al. (2007), Aviation applications for satellite-based observations of cloud properties, convective initiation, in-flight icing, turbulence and volcanic ash, *Bull. Am. Meteorol. Soc.*, *88*, 1589–1607, doi:10.1175/BAMS-88-10-1589.
- Mecikalski, J. R., K. M. Bedka, S. J. Paech, and L. A. Litten (2008), A statistical evaluation of GOES cloud-top properties for nowcasting convective initiation, *Mon. Weather Rev.*, in press.
- Mueller, C., T. Saxen, R. Roberts, J. Wilson, T. Betancourt, S. Dettling, N. Oien, and J. Yee (2003), NCAR Auto-Nowcast system, *Weather Forecasting*, *18*, 545–561.
- Nair, U. S., R. C. Weger, K. S. Kuo, and R. M. Welch (1998), Clustering, randomness, and regularity in cloud fields: 5. The nature of regular cumulus cloud fields, *J. Geophys. Res.*, *103*, 11363–11380, doi:10.1029/98JD00088.
- Nair, U. S., R. M. Welch, R. O. Lawton, and R. A. Pielke Sr. (2000), Impact of land use on Costa Rican regional climate, in *IGARSS 2000: IEEE 2000 International Geoscience and Remote Sensing Symposium*, vol. V, pp. 1993–1995, IEEE Press, Piscataway, N. J.
- Nair, U. S., R. M. Welch, R. O. Lawton, and R. A. Pielke Sr. (2001), Impact of land surface heterogeneity on the spatial organization of cumulus clouds, paper presented at 12th Symposium on Global Change and Climate Variations, Am. Meteorol. Soc., Albuquerque, N. M., 14–18 Jan.
- Peak, J. E., and P. M. Tag (1992), Toward automated interpretation of satellite imagery for Navy shipboard applications, *Bull. Am. Meteorol. Soc.*, *73*, 995–1008, doi:10.1175/1520-0477(1992)073<0995:TAIOSI>2.0.CO;2.
- Plank, V. G. (1969), The size distribution of cumulus clouds in representative Florida populations, *J. Appl. Meteorol.*, *8*, 46–67, doi:10.1175/1520-0450(1969)008<0046:TSDOCC>2.0.CO;2.
- Pope, A. C., M. J. Thun, M. M. Namboodiri, D. W. Dockery, J. S. Evans, F. E. Speizer, and C. W. Heath (1995), Particulate air pollution as a predictor of mortality in a prospective study for US adults, *Am. J. Respir. Crit. Care Med.*, *151*, 669–674.
- Rabindra, P., S. K. Sengupta, and R. M. Welch (1992), An interactive hybrid expert system for polar cloud and surface classification, *Environmetrics*, *3*, 121–148.
- Ramirez, J. A., and R. L. Bras (1990), Clustered or regular cumulus cloud fields: The statistical character of observed and simulated cloud fields, *J. Geophys. Res.*, *95*, 2035–2045, doi:10.1029/JD095iD03p02035.
- Rosenfeld, D., W. L. Woodley, A. Lerner, G. Kelman, and D. T. Lindsey (2008), Satellite detection of severe convective storms by their retrieved vertical profiles of cloud particle effective radius and thermodynamic phase, *J. Geophys. Res.*, *113*, D04208, doi:10.1029/2007JD008600.
- Schmit, T. J., M. M. Gunshor, W. P. Menzel, J. J. Gurka, J. Li, and A. S. Bachmeier (2005), Introducing the next-generation advanced baseline imager on GOES-R, *Bull. Am. Meteorol. Soc.*, *86*, 1079–1096.
- Sengupta, S. K., R. M. Welch, M. S. Navar, T. A. Berendes, and D. W. Chen (1990), Cumulus cloud field morphology and spatial patterns derived from high-resolution satellite imagery, *J. Appl. Meteorol.*, *29*, 1245–1267, doi:10.1175/1520-0450(1990)029<1245:CCFMAS>2.0.CO;2.
- Shenk, W. E., R. J. Holub, and R. A. Neff (1976), A multispectral cloud type identification method developed for tropical ocean areas with Nimbus-3 MRIR measurements, *Mon. Weather Rev.*, *104*, 284–291.
- Sykes, R. I., and D. S. Henn (1989), Large-eddy simulation of turbulent sheared convection, *J. Atmos. Sci.*, *35*, 25–39.
- Tag, P. M., R. L. Bankert, and L. R. Brody (2000), An AVHRR multiple cloud-type classification package, *J. Appl. Meteorol.*, *39*, 125–134, doi:10.1175/1520-0450(2000)039<0125:AAMCTC>2.0.CO;2.
- Tovinkere, V., M. Penaloza, A. Logar, J. Lee, R. Weger, T. Berendes, and R. Welch (1993), An intercomparison of artificial intelligence approaches for polar scene identification, *J. Geophys. Res.*, *98*(D3), 5001–5016.
- Velden, C. S., C. M. Hayden, S. J. Nieman, W. P. Menzel, S. Wanzong, and J. S. Goerss (1997), Upper-tropospheric winds derived from geostationary satellite water vapor observations, *Bull. Am. Meteorol. Soc.*, *78*, 173–195.
- Velden, C. S., T. L. Olander, and R. M. Zehr (1998), Development of an objective scheme to estimate tropical cyclone intensity from digital geostationary satellite infrared imagery, *Weather Forecasting*, *13*, 172–186.
- Weckwerth, T. M., J. W. Wilson, R. M. Wakimoto, and N. A. Crook (1997), Horizontal convective rolls: Determining the environmental conditions supporting their existence and characteristics, *Mon. Weather Rev.*, *125*, 505–526, doi:10.1175/1520-0493(1997)125<0505:HCRDTE>2.0.CO;2.
- Weger, R. C., J. Lee, T. Zhu, and R. M. Welch (1992), Clustering, randomness and regularity in cloud fields: 1. Theoretical considerations, *J. Geophys. Res.*, *97*, 20519–20536.
- Welch, R. M., and B. A. Wielicki (1989), Reflected fluxes for broken clouds over a Lambertian surface, *J. Atmos. Sci.*, *46*, 1384–1395, doi:10.1175/1520-0469(1989)046<1384:RFFBCO>2.0.CO;2.
- Welch, R. M., K. S. Kuo, B. A. Wielicki, S. K. Sengupta, and L. Parker (1988), Marine stratocumulus cloud fields off the coast of southern California observed using LANDSAT imagery: Part 1. Structural characteristics, *J. Appl. Meteorol.*, *27*, 341–362, doi:10.1175/1520-0450(1988)027<0341:MSCFOT>2.0.CO;2.
- Welch, R., M. Navar, and S. Sengupta (1989), The effect of spatial resolution upon texture-based cloud field classifications, *J. Geophys. Res.*, *94*(D12), 14767–14781.
- Welch, R., S. Sengupta, A. Goroch, P. Rabindra, N. Rangaraj, and M. Navar (1992), Polar cloud and surface classification using AVHRR imagery: An intercomparison of methods, *J. Appl. Meteorol.*, *31*, 405–420.
- Wielicki, B. A., and R. M. Welch (1986), Cumulus cloud properties derived using Landsat satellite data, *J. Clim. Appl. Meteorol.*, *25*, 261–276, doi:10.1175/1520-0450(1986)025<0261:CCPDUL>2.0.CO;2.
- Wolfson, M. M., B. E. Forman, K. T. Calden, W. J. Dupree, R. J. Johnson Jr., R. A. Boldi, C. A. Wilson, P. E. Bieringer, E. B. Mann, and J. P. Morgan (2004), Tactical 0–2 hour convective weather forecasts for FAA, paper presented at 11th Conference on Aviation, Range, and Aerospace Meteorology, Am. Meteorol. Soc., Hyannis, Mass.
- Zhu, T., J. Lee, R. C. Weger, and R. M. Welch (1992), Clustering, randomness, and regularity in cloud fields: 2. Cumulus cloud fields, *J. Geophys. Res.*, *97*(D18), 20537–20558.

K. M. Bedka, Cooperative Institute for Meteorological Satellite Studies, University of Wisconsin-Madison, Madison, WI 53706, USA.

T. A. Berendes, Information Technology and Systems Center, University of Alabama in Huntsville, Huntsville, AL 35805, USA.

W. M. MacKenzie Jr. and U. S. Nair, Earth System Science Center, National Space Science and Technology Center, University of Alabama in Huntsville, 320 Sparkman Drive, Huntsville, AL 35805, USA. (wayne.mackenzie@uah.edu)

J. R. Mecikalski, Atmospheric Sciences Department, University of Alabama in Huntsville, Huntsville, AL 35805, USA.