JOURNAL OF GEOPHYSICAL RESEARCH, VOL. 114, D06207, doi:10.1029/2008JD011005, 2009

Statistical approaches to error identification for plane-parallel retrievals of optical and microphysical properties of three-dimensional clouds: Bayesian inference

P. Gabriel,[1] H. W. Barker,[2] D. O'Brien,[1] N. Ferlay,[3] and G. L. Stephens[1]

Received 19 August 2008; revised 3 December 2008; accepted 7 January 2009; published 25 March 2009.

[1] This paper addresses the effects of three-dimensional (3-D) radiative transfer on the retrieval of optical depth for inhomogeneous stratiform liquid water clouds from passive satellite imagery. A nonparametric Bayesian classifier is developed to identify locations in a scene where plane-parallel retrievals fail to meet the requirements of a criterion that dictates a specified level of accuracy. Receiver operating characteristics are introduced that provide useful metrics that assess the quality of the error identification procedure as functions of illumination-viewing geometry. By fixing droplet effective radii, distributions of errors for retrieved optical depth are estimated at a scale of 120 m. These estimates suggest the best performance that can be expected for optical depth retrievals when 3-D radiative transfer cannot be ignored. The developments in this paper were made possible through the use of Monte Carlo radiative transfer simulations on stratiform clouds that were generated by a cloud system-resolving model. Plane-parallel retrievals employ the CloudSat optical depth retrieval algorithm.

Citation: Gabriel, P., H. W. Barker, D. O'Brien, N. Ferlay, and G. L. Stephens (2009), Statistical approaches to error identification for plane-parallel retrievals of optical and microphysical properties of three-dimensional clouds: Bayesian inference, *J. Geophys. Res.*, *114*, D06207, doi:10.1029/2008JD011005.

1. Introduction

[2] A key parameter that determines how clouds influence Earth's radiation budget, hydrological cycle, and climate in general is their optical depth. Hence, it is essential that global climate models and numerical weather prediction models simulate spatiotemporal distributions of cloud optical depth with some degree of accuracy. In order to verify how well these models do this requires estimates of cloud optical depth from global observations. To date, the most cost-effective means of obtaining such information is from imagery obtained by satellites equipped with passive sensors that measure emitted terrestrial and reflected solar radiation. Because these sensors are relatively economical, they will almost certainly continue to be employed for the foreseeable future. However, despite advances in the accuracy and resolution of multispectral imagers, explicit reliance on plane-parallel solutions of the radiative transfer equation (RTE) compromises the utility of satellite-inferred cloud optical depths. This is because most clouds possess substantial variability near photon mean-free path lengths which are often close to the resolution of satellite imagery.

[3] Virtually all operational radiative transfer models used in remote sensing are built on the assumption that clouds are locally plane-parallel. Moreover, they are applied directly to passive radiances regardless of pixel resolution and spatial variations in radiances. The common practice is to discretize cloudy regions into a grid of columns. Within columns, clouds are considered as horizontally infinite uniform sheets, but whose properties can vary vertically. Hence, through neglect of three-dimensional transport, columns are assumed not to interact radiatively. Cloud properties are often deduced by inverting these simple models initialized by measured radiances. While three-dimensional effects can be reduced by using spectral intervals where absorption is relatively strong, the price paid is loss of sampling volume and sensitivity. Thus, ignoring information contained in the spatial distribution of satellite radiances leads to retrievals that are suboptimal, exacerbating the difficulties in attaining the primary goals.

[4] While several approaches to three-dimensional forward radiative transfer modeling exist, the retrieval problem lacks a formal theory that will permit practical inversion of such a model. At any time, optical properties of clouds are dependent on three spatial coordinates. Even if a cloudy region is discretized into finite, internally uniform cells, the dimensionality would still be immense and the problem ill-posed; if a solution exists, it would not be unique. Furthermore, the results would also depend on the size of the cells (i.e., spatial resolution). Numerical inversion of the plane-parallel equation of transfer, which only depends on the vertical coordinate, is also an ill-posed problem. To retrieve column optical depth or effective radius of the droplet

[1]Department of Atmospheric Science, Colorado State University, Fort Collins, Colorado, USA.
[2]Cloud Physics and Severe Weather Section, Environment Canada, Toronto, Ontario, Canada.
[3]Laboratoire d'Optique Atmospherique, Université des Sciences et Technologies de Lille, Villeneuve d'Ascq, France.

0148-0227/09/2008JD011005$09.00

distribution within a cloudy layer, additional constraints must be imposed to obtain a stable solution. For example, in the CloudSat optical depth and droplet effective radius retrieval algorithm, also known as the "2btau algorithm," the optimal estimation method described by *Rogers* [2004] is employed, driven by MODIS radiances (OEM, http://cloudsat.cira.colostate.edu/ICD/2B-TAU/2B-TAU_PDICD_3.0.pdf). A priori constraints in the form of a covariance matrix and a mean state vector for the optical depth and effective radius must be introduced with the additional assumption that errors are distributed normally. A useful feature of this method is that measurement and model errors can be introduced so that uncertainties in the retrieved products can be calculated. However, this feature cannot be relied upon to account for errors introduced by neglecting three-dimensional radiative transfer.

[5] Studies of retrieval errors for optical depth and effective radius of liquid water clouds when 3-D effects are prevalent have been performed by several authors. For instance, *Loeb and Coakley* [1998] concluded that even for overcast marine stratus layers, the frequently cited "prototypical plane-parallel cloud," inferred optical depths are overestimated systematically as solar zenith angle increases. They attributed this to undulating cloud tops. Such cloud top structures affect the accuracy of the retrievals of optical depth because they introduce view angle dependence. A number of other studies have also demonstrated how 3-D bias errors in optical depth vary according to cloud morphology [e.g., *Iwabuchi and Hayasaka*, 2002], view and solar geometry [e.g., *Loeb and Davies*, 1997], and measurement resolution [*Heidinger and Stephens*, 2000]. These results, among others, motivated this series of papers whose objectives are to quantify the errors caused by the neglect of 3-D radiative transfer, identify regions in a scene where plane-parallel theory can be used, and develop methods that can be used even when 3-D effects are extant.

[6] Before describing our approach to this problem, a perusal of the literature reveals that other methods have been developed to identify signatures when 3-D transport is significant. One promising method, based on absorption spectroscopy was proposed by [*Stephens and Heidinger*, 2000]. Their method uses oxygen A-band radiances to compute the photon mean path lengths that serve as measures of spatial homogeneity. The advantage of this method is that it unambiguously detects regions of shadows, as opposed to dark regions of low optical depth and is less empirical than geometrical methods like those of *Varnai and Marshak* [2002]. As yet, however, there are no measurements to assess their technique; though this might change with the Orbiting Carbon Observatory mission (OCO, http://oco.jpl.nasa.gov).

[7] The primary objectives just mentioned are elaborated on further in section 2. Data used throughout these analyses are described in section 3. Section 4 describes the use of Bayes' Theorem in the error identification problem. Results of using the Bayesian classifier are discussed section 5. Section 6 introduces the notion of a receiver operating characteristic (ROC), a graphical representation of the information contained in a 2×2 outcome matrix which enumerates the results of applying a test to determine the state of a system. Only four possibilities exist: a test indicates the system is/is not in a particular state while in reality, the system is/is not in that state. The ROC will be used to define and set thresholds in identifying where in a scene plane-parallel retrievals will likely fail. Section 7 discusses the connection between the ROC and retrieval error caused by three-dimensional radiative transfer. Concluding comments are made in section 8.

2. Objectives

[8] The undeniable practicality of plane-parallel theory, the computational efficiency of freely available corresponding computer codes, and the intractability of explicitly addressing three-dimensional effects suggest the continuance of conventional plane-parallel retrieval algorithms. Hence, it is prudent to establish a mechanism for defining accurate error estimates owing to the neglect of three-dimensional transport.

[9] When three-dimensional radiative transfer is extant, alternates to straight inversion of the 1-D RTE are statistical retrieval techniques based on 3-D radiative transfer calculations, at multiple wavelengths, for particular genres of clouds as simulated by cloud-resolving models (CSRMs). Such retrievals would displace the problem from being one of radiation to one of cloud dynamics since the fidelity of the cloud model used to provide the fields, and not the three-dimensional transfer, is the factor that will limit retrieval accuracy. In this paper we formulate a solution to the error identification problem wherein portions of images that plane-parallel retrievals are likely to be unreliable, according to some specified criteria, are signaled in a binary plot. Forthcoming extensions of this paper, herein referred to as parts 2 and 3, will develop different statistical methodologies that exploit the main ideas presented here to accommodate the effects of 3-D transport in retrievals of optical depth with the goals of reducing the bias and variance as well as solving the error identification problem.

[10] In practice, achieving these objectives will require a ring of concepts that connect dynamical modeling of clouds, three-dimensional radiative transfer, and cloud observations for verification of results. This aspect of the problem can be stated as follows: operate on fields produced by CSRMs with a 3-D Monte Carlo photon transport algorithm to produce simulated radiances; for the error identification problem, act on these radiances with plane-parallel retrieval algorithms; and assess error biases and dispersions in the retrieved products. The error distribution is determined using all cells in an image where retrievals are performed and characterizes differences between the retrieved and true optical or microphysical properties.

[11] The nonparametric statistical methods described here, and in parts 2 and 3, differ in formulation, but enjoy a common input, and hereafter are referred to as "statistical inference engines," since, in passing from sample data to generalizations, a certain degree of uncertainty will necessarily be introduced. At this juncture, atmospheric science overlaps with computer science, mathematics, and statistics in its search to construct algorithms, or inferential engines, that can be taught by example but generalize effectively. By detecting significant patterns in the available data, a system can be constructed to make predictions about new data coming from the same source. It is in this sense that we

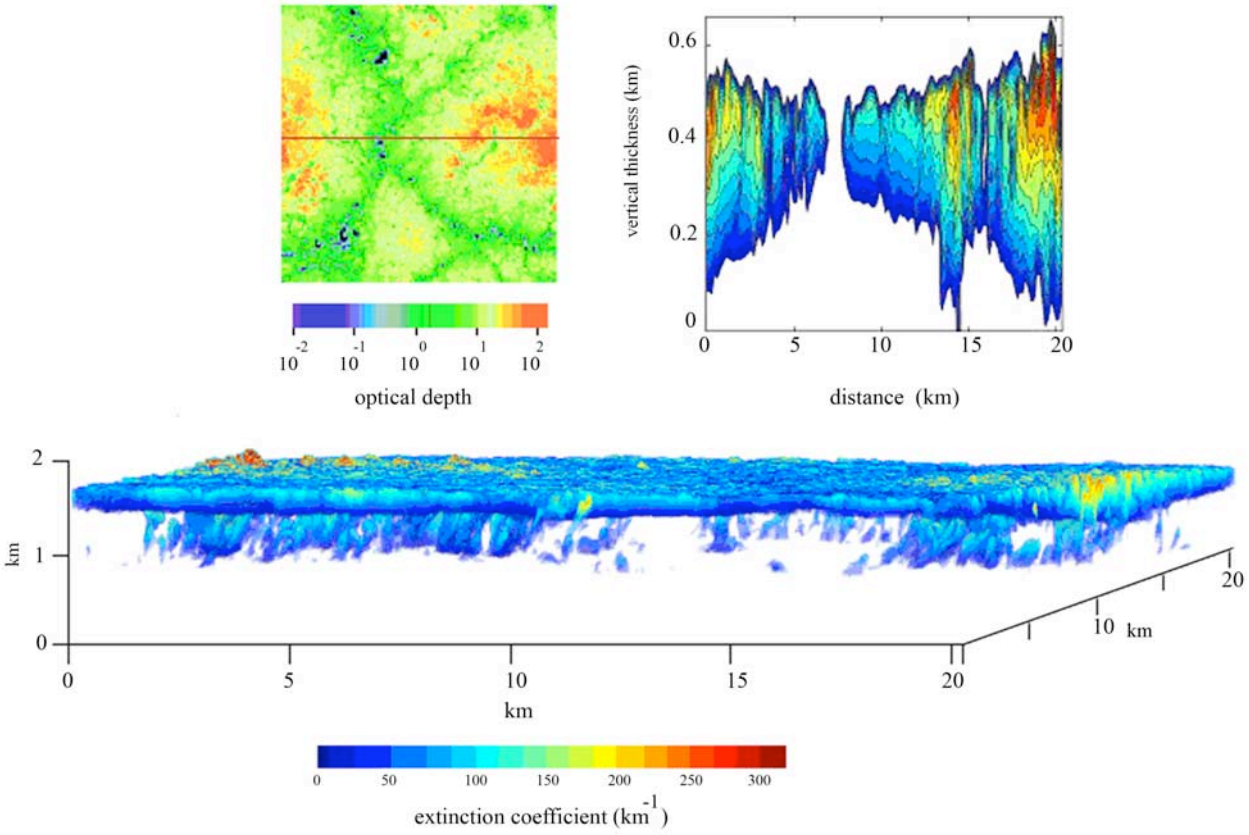

Figure 1. (top left) Spatial distribution of the CSRM's optical depth calculated for a wavelength of 0.8 μm and fixed effective radius of 10 μm. (top right) Two-dimensional cross section of extinction along the transect as indicated on the optical depth plot. (bottom) Volumetric, three-quarters view of the extinction field. The color bar is also applicable to the extinction cross section.

mean that a system has acquired generalization power by "learning" something about the source generating the data.

[12] For the sake of clarity and brevity, this paper discusses the use of Bayes' Theorem for the error identification problem only. Part 2 will use neural networks to test independently the criterion employed here for error identification by inferring optical depth without use of plane-parallel radiative transfer and then compares the bias and variance of the resulting error distribution to the one obtained from plane-parallel retrievals. It also extends the criterion in a simple way that enhances the reduction of the bias of the retrieved optical depths significantly as well as reducing the variance. Neural networks are not trained to solve the error identification problem caused by 3-D effects, but are used as "pseudoclassifiers" to indicate the improvement that might be expected in error identification if 3-D effects were accounted for by the retrieval. Neural networks were also used to explore the effects on retrievals of optical depth introduced by averaging radiances over different scales. Finally, results obtained from the Bayesian and neural network inference engines are compared to those obtained using support vector machines, the subject of part 3. This engine will be used to further analyze the accuracy of error predictions using the criterion studied here and to obtain estimates of optical depth using the extended criterion to be discussed in part 2.

[13] We begin with the Bayesian inference engine because it not only has pedagogical value, but because it is useful in its own right as a classifier, as well as introducing the notion of the ROC which will also be of fundamental importance in part 3 in connection with support vector machines. In effect, the approach taken in this paper to classify retrieval error serves as a feasibility study; a precursor of the more advanced methods described in parts 2 and 3.

3. Synthetic Data

[14] Throughout the course of this study, output from a bulk-microphysical CSRM [*Stevens et al.*, 2002] coupled with assumptions made about particle number concentration was used to calculate the extinction field at a nominal wavelength of 0.8 μm. The field is illustrated in Figure 1 (top left). Figure 1 (top right) shows the vertical cross section of extinction coefficient along the red transect as shown. Figure 1 (bottom) shows the 3-D volumetric description of extinction coefficient. A fixed gamma-type particle size distribution was ascribed with an effective radius of 10 μm and effective variance of 0.1 for the entire cloud. The Monte Carlo radiative transfer technique of *Barker et al.* [2003] was used to calculate radiances for several combinations of viewing and illumination geome-

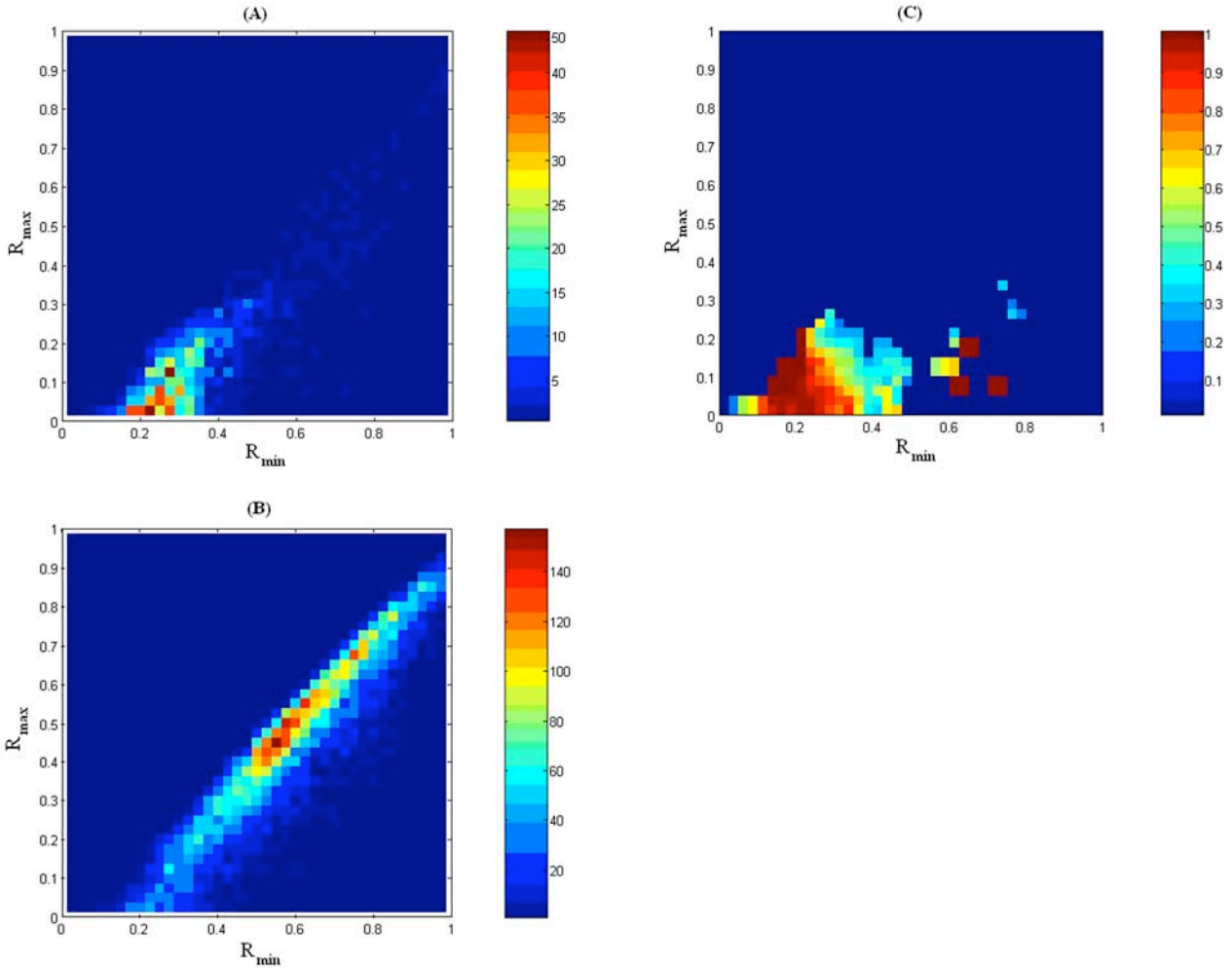

Figure 2. (a) Frequency of occurrence of the extremal radiances conditioned on failed retrievals. (b) Joint histogram of the extremal radiances. (c) Posterior probability from which a threshold must be chosen. The solar zenith angle is 0°, and nadir view and the observer-solar azimuth angle is 0°.

tries. The number of photons used to calculate radiances was 10^{10}. Prior to performing Monte Carlo calculations the original CSRM data were averaged from 40-m horizontal grid spacing to 120 m. The vertical resolution of the data was 20 m. This averaging scale was selected because it just exceeds the smoothing scale described by *Davis et al.* [1997] as calculated using mean cloud physical thickness of about 440 m, asymmetry factor of ~0.85 (used to define a Henyey-Greenstein phase function), and mean optical thickness of 14.7. The optical depth standard deviation is 12.9, indicating that this cloud is extremely inhomogeneous, as is visually apparent.

4. Application of a Bayesian Classifier to the Error Identification Problem

[15] Computations were performed using spectral directional radiances contained within a movable square template, referred to here as a window. The results presented here are for a window consisting of 3 × 3 elements. All elements were of size equal to that of the cloud data, 120 m. Although 5 × 5 windows were also explored, we noted results comparable to or less accurate than using the smaller window. Windows were common to all inference engines. In this paper only the minimum and maximum or extremal radiances falling within a window are used. In parts 2 and 3, optical depths are also retrieved in addition to performing error identification. In those works either the extremal radiances or all radiances in a window were used. The hypothesis common to this and the other two papers is that "small" variations in the radiances inside the window will statistically maximize the likelihood of successful plane-parallel retrievals. Alternatively stated, our criterion of spatial uniformity uses two random variables: the minimum and maximum radiances contained within a window. If these radiances are identical, the window is deemed uniform, and a retrieval based on plane-parallel theory applied to the central pixel is likely to be accurate. As will be demonstrated here and in parts 2 and 3, numerical experiments consistently show that these extrema are useful and robust indicators of inhomogeneity.

[16] In this section the Bayesian inference engine will be used as a binary classifier to solve the error identification problem caused by 3-D effects. Bayes' theorem can be

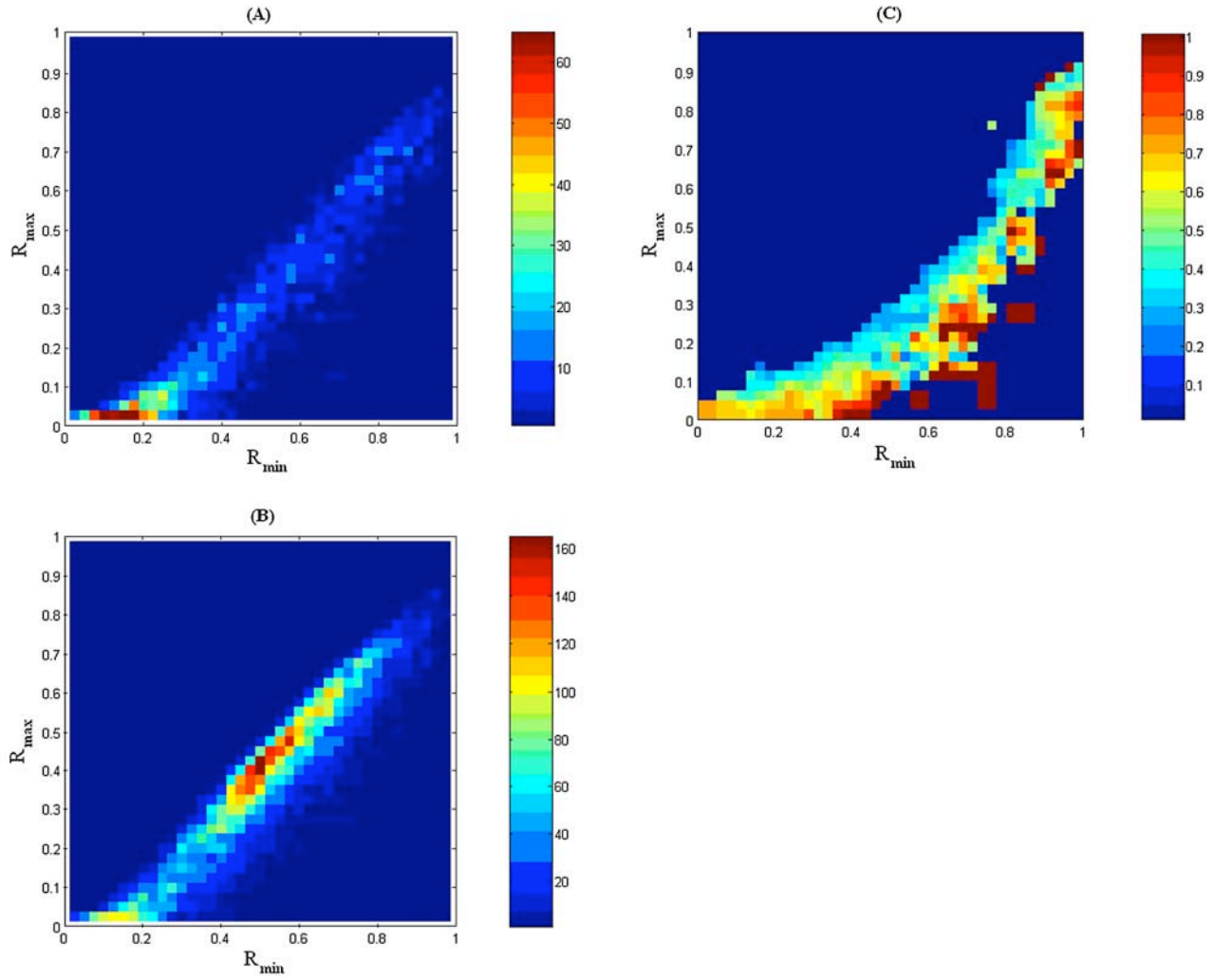

Figure 3. As in Figure 2 but with a solar zenith angle of 30°, observation angle of 30°, and observer-solar azimuth angle of 0°.

found in almost any textbook on probability and statistics [e.g., *Duda and Hart*, 1973; *Duda et al.*, 2001]. We summarize briefly the statement of the theorem as follows.

[17] Consider a sample of data which we designate as "R" consisting of R_{min} and R_{max}. Let H be the hypothesis that R is to be associated with either "successful" or "failed" retrievals designated as "S" and "F," respectively. We wish to determine the conditional probability P (H|R) that the hypothesis H holds, given the observed R. Henceforth we shall generally use "P" to denote probabilities and "p" to denote probability density functions. The latter are nonnegative and integrate to unity, although the function p(x) can exceed unity. The quantity P (H|R), called the posterior determines the probability of a successful or failed retrieval. In contrast, P(H), called the prior or a priori probability, for any given data record is the probability of a successful or failed retrieval in the data record. The posterior probability is based on more information (such as background knowledge) than is the prior probability which is independent of R. The quantity p(R|H), called the posterior probability density of R conditional on H, is the probability of observing R given that we know the status of the retrieval (i.e., whether it has succeeded or failed). Similarly, the quantity p(R), called the prior probability density of R, is the probability of observing R_{min} and R_{max} jointly within a window. Bayes' theorem is useful in that it provides a way of calculating P(H|R) from P(H), p(R), and p(R|H). Since we wish to relate the probability of a failed retrieval (H = F) conditioned on a given set of extremal values {R = (R_{min},R_{max})} to the probability of observing these values, but conditional on a failed retrieval, Bayes' Theorem reads

$$P(F|R_{min}, R_{max}) = p(R_{min}, R_{max}|F)P(F)/p(R_{min}, R_{max}). \quad (1)$$

The application of (1) involves first the creation of tables of $p(R_{min},R_{max}|F)$, P(F) and $p(R_{min},R_{max})$ from a training set comprising half of the radiances for the synthetic stratiform cloud shown in Figure 1. Examples of such tables used for training are illustrated in Figures 2, 3, and 4. Figures 2a, 3a, and 4a show the cooccurrences of (R_{min},R_{max}) associated with failed retrievals. Figures 2b, 3b, and 4b show the joint histogram of (R_{min},R_{max}). Posterior probabilities shown in Figures 2c, 3c, and 4c calculated directly from these tables

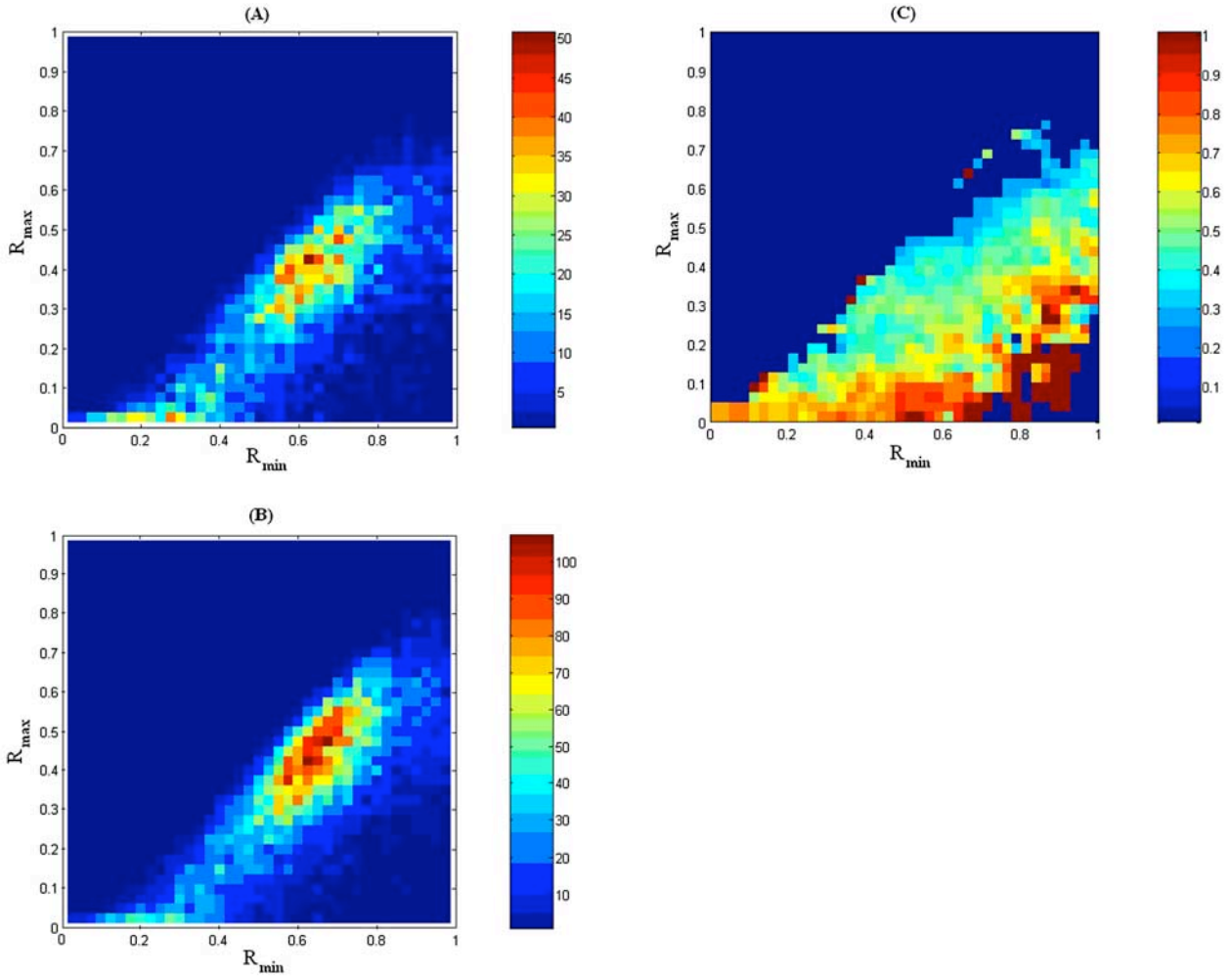

Figure 4. As in Figure 2 but with a solar zenith angle of 60°, observation angle of 30° and observer-solar azimuth angle of 0°.

exhibit the complexity of the solution space. The fundamental problem is the selection of a probability threshold such that if $P(F|R_{min},R_{max})$ exceeds this threshold, the retrieval at the pixel of interest is classified as failed. Since the values of the combinations of (R_{min},R_{max}) in the training and test data sets can be very different, calculation of the posterior probabilities was carried out using a Parzen Kernel Estimator. This procedure, applicable to nonparametric distributions uses the precomputed tables to interpolate the extremal radiances in the untrained data to the conditional and joint probability densities required by equation (1). A clear exposition of the latter is given by *Duda and Hart* [1973] and *Duda et al.* [2001]. It is observed that, although both $p(R_{min},R_{max}|F)$ and $p(R_{min},R_{max})$ can exceed unity, their ratio is less than unity because the former density function is narrower than the latter owing to the fact that it uses additional information in its construction. To create a binary image that discriminates retrieval success from failure, we must, however, define a threshold to impose on the probabilities as has been mentioned. We will shortly describe how this threshold can be determined in an objective, optimal way. Looking ahead, we develop the ROC here because it is also used to determine the quality of the solution to the error identification problem computed by the support vector machine, to be described in part 3.

[18] The reason for using the ROC is that it is insufficient to merely assign pixels in "accept" or "reject" categories; the ROC permits an objective evaluation of the classifier and provides a criterion for comparing the performance of different algorithms. In part 3, the performance of the Bayesian and support vector machine classifiers will be compared. The latter also computes decision thresholds from the ROC but does not use conditional probability densities.

5. Results of Applying Bayesian Classification to Synthetic Radiance Data

[19] We begin by considering Figure 5a. For this case, radiances were computed for an overhead sun and nadir observation. Figure 5b displays the spatial distribution of error. We have defined the retrieval error as the base 10 logarithm of the ratio of retrieved (t) to true optical depth (T) because of the large dynamic range spanned by this quantity and as an aid in plotting. This error criterion is not "tuned" for any particular application, but used here merely

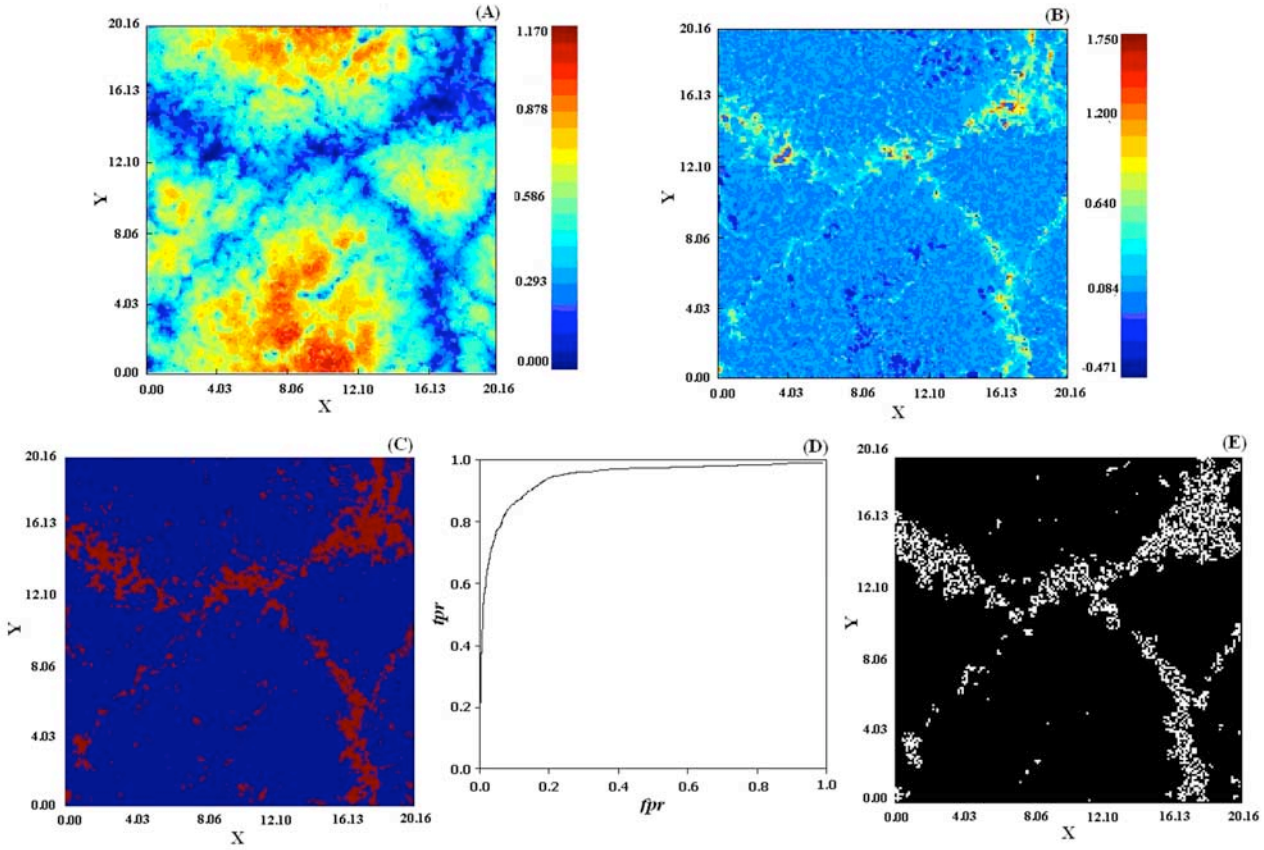

Figure 5. (a) Scene radiance (image size is 168 × 168 pixels, 120 m per pixel; coordinates X and Y are in kilometers) for the case described in Figure 2. (b) Base 10 logarithm of the ratio of retrieved (t) to true optical depth. (c) Retrieval error $\varepsilon > 0.25$ (see text). (d) Receiver operating characteristic which indicates the accuracy of the classifier and is used to determine operating thresholds. (e) Predictions of optical depth errors using the Bayesian classifier, to be compared to Figure 5c. The mottled appearance of Figure 5e is the result of excluding the training set.

to demonstrate the concept. For example, if the retrieved optical depth is 10^{-1} but the true optical depth is 10^{-2}, the factor of 10 appears large but depending on the application, may be practically negligible. In the radiance scene shown in Figure 5a, only 70 pixels out of 28,224 had both zero radiance and optical depth and 198 pixels had optical depths less than 0.1. Pixels where T = 0 were flagged as errors to prevent division by zero in the logarithmic ratio. Hence results connected with the identification of failed optical depth retrievals in an image are slightly exaggerated, reflecting the assumptions used to train the classifier.

[20] In this and subsequent analyses the 2btau algorithm was set to use a Henyey Greenstein phase function with g = 0.85, discretized into 32 streams, excluding the observation angles with single scattering albedo set to unity and zero surface albedo. By setting the aforementioned parameters equal to those used to calculate 3-D radiances, we eliminate uncertainties in the surface and optical properties and can study how spatial inhomogeneities in the cloud field affect retrievals from the 2btau algorithm or by the table lookup technique employed by MODIS.

[21] In practical applications, the approach taken here would have to be extended to include at least two wavelengths from which both optical and microphysical properties can be inferred. However, through a judicious selection of wavelengths, the radiances would be nearly independent, resulting in the possibility of factorizing the probability density functions. The relevance of this section is not minimized by the use of a single wavelength as it foretells the likelihood of success in the more general case, indeed of the success of plane-parallel identification in images using the technique described in part 3 which also lends itself well to the use of multiple wavelengths.

[22] Figure 5c maps where the error ε, defined by $\varepsilon = |\mathrm{Log}_{10}\ t/T|$ exceeds 0.25. This choice of ε identifies optical depth retrieval errors exceeding about ±78%. Figure 5e indicates the locations where the algorithm predicts retrieval errors are likely. Here, for the sake of simplicity, we will assume that 3-D effects will introduce errors of similar magnitude to both MODIS and CloudSat retrieval methods, since the forward models in both cases are plane-parallel, though even here, differing in detail.

[23] This matter will be discussed in greater depth in the conclusions. Since the CloudSat retrieval algorithm was readily available, we used it to perform the numerical experiments. We postpone the discussion of how retrieval errors were computed until the ROC is introduced in section 6.

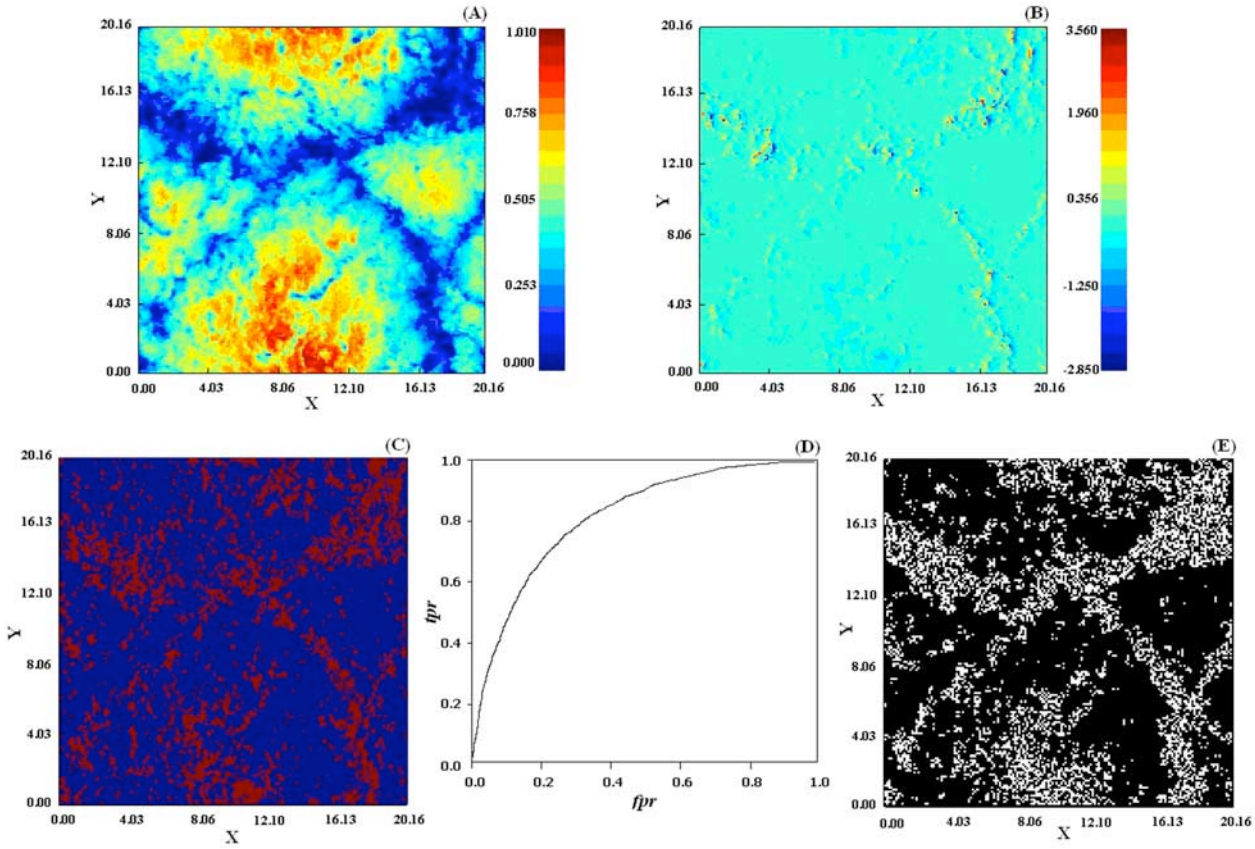

Figure 6. As in Figure 5 but with a 30° solar zenith angle, 30° observation zenith angle, and an observer-solar azimuth angle of 0°.

[24] The effects of changing solar zenith angle and viewing angle on the predicted retrieval error locations are illustrated in Figures 6 and 7. Properties of the error distributions of the retrieved optical depths as they are affected by viewing and solar zenith angles are discussed in section 7.

6. Using the ROC

[25] The ROC has its origins in the development of radars where it was applied to the processing of signals to ascertain the presence of targets in weak, noisy echoes via the setting of discrimination thresholds. Signal echoes exceeding these thresholds were classified as targets. The criterion was strongly motivated by the desire to understand how the U.S. radar "receiver operators" missed Japanese aircraft during the attack on Pearl Harbor. The criterion finds extensive usage in psychophysics [*Green and Swets*, 1966; *Swets*, 1973], machine learning [*Provost et al.*, 1998; *Hand and Till*, 2001], evidence-based medicine such as radiology [*Hanley and McNeil*, 1982; *Swets*, 1988], and in the atmospheric sciences [*Olson*, 1965; *Mason*, 1979, 1982; *Murphy and Winkler*, 1987; *Doswell et al.*, 1990; *Harvey et al.*, 1992; *Wilks*, 1995; *Mason and Graham*, 1999].

[26] In the biomedical field, interest in the ROC centers on issues connected with sensitivity (a measure of how well a test can pick out patients with a disease) and specificity (the ability of a test to pick out patients who do not have the disease). Interest is not necessarily focused in the predictive probability that someone has a disease, but rather on a test (e.g., the concentration of white blood cells in checking for leukemia) and a test threshold that allows accurate identification and treatment of diseased individuals. The uses of ROC curves, which assess the discriminatory power of diagnostic tests in correctly classifying diseased and nondiseased individuals, are designed to address these kinds of questions. The ROC concept can be understood by considering Figure 8, which shows the probabilities of pixels in an image associated with successfully classified and failed optical depth retrievals. The horizontal axis is the quantity $P(F|R_{min},R_{max})$, a threshold that for the moment is specified arbitrarily. The solid vertical line corresponds to some threshold that demarcates the decision boundary. For (R_{min},R_{max}) combinations resulting in $P(F|R_{min},R_{max}) > \gamma$, the decision would be to classify the central pixel in the window in the failed category.

[27] Errors are classified as either a miss or a false alarm. A miss occurs when the classifier fails to detect error in the retrieval, while a false alarm occurs when a classifier predicts an incorrect retrieval when in fact the retrieval was successful. Likewise, there are two types of correct classifications: a hit or a correct reject. In a hit, the classifier correctly predicts the occurrence of a retrieval failure, whereas in a correct rejection the classifier indicates the absence of retrieval error. The analogy between a physician

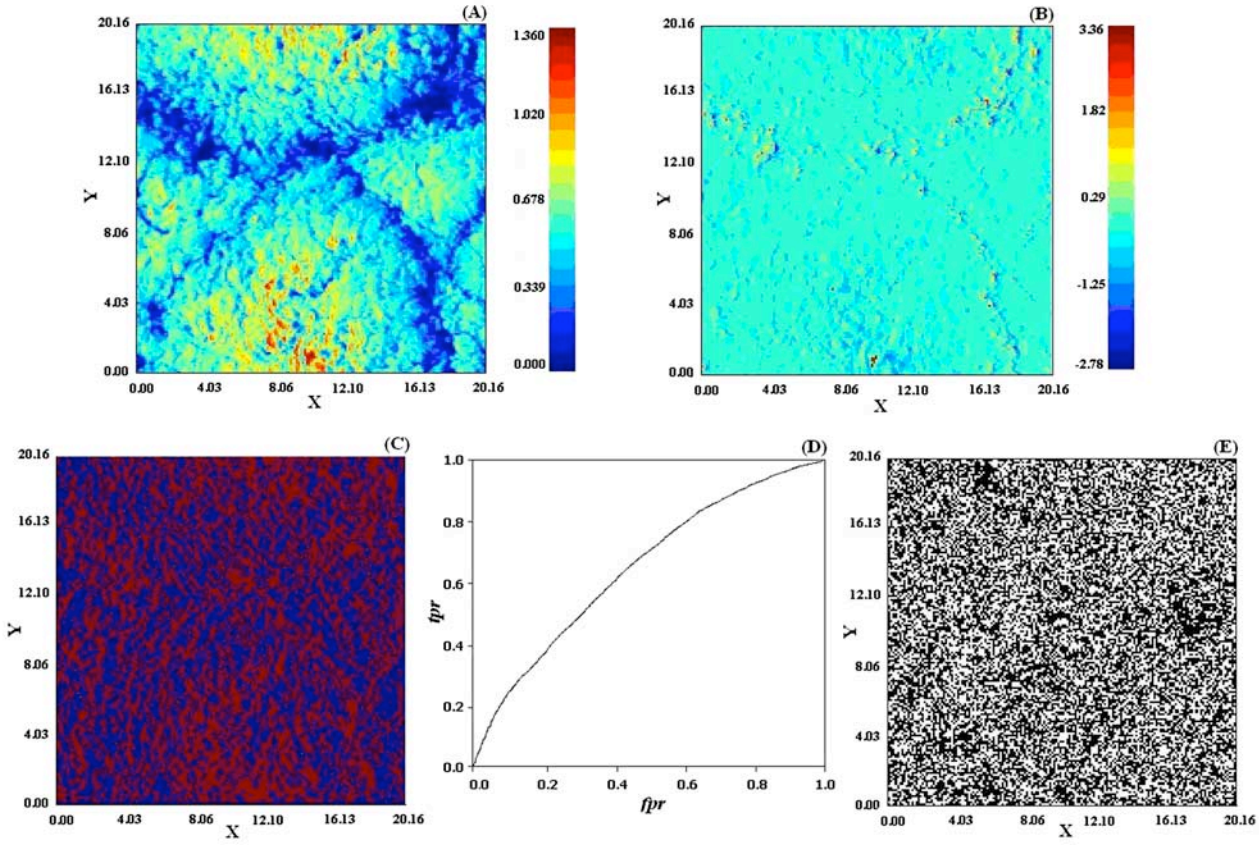

Figure 7. As in Figure 5 but with a 60° solar zenith angle, 30° observation zenith angle, and an observer-solar azimuth angle of 0°.

and patient is clear; the physician must inform the patient of the presence or absence of a disorder which the patient may or may not have. By setting the test threshold to a low value, the number of hits increases as do the number of false alarms. Setting the threshold too high will result in a large number of misses, but fewer false alarms. No matter where the test threshold is set, the incurrence of error is certain, reflecting our inability to separate classes perfectly. The frequency of error is a function of how far apart the means of the distributions are apart from one another as well as the width of the distributions. Narrow distributions separated by and large distances would result in a classifier having the low error rate. The question of threshold selection will be discussed in this section after describing how the ROC is calculated.

[28] In the context of our problem, T_{ROC} is a threshold used to discriminate between two kinds of predictions that the classifier can make. A "true positive rate" (tpr) is a situation that is predicted to have $\varepsilon > 0.25$ when in fact $\varepsilon > 0.25$. A "false positive rate" (fpr) is a situation that is predicted to have $\varepsilon > 0.25$ when in fact $\varepsilon < 0.25$. The ROC is a plot of tpr versus fpr as shown in Figure 9 (left). Any point on this curve is uniquely determined by the threshold. The best possible prediction would produce a single point in the upper left corner of the ROC space, where only true positives and no false positives are found. If the predictor were completely random, it would result in a 45 degree line from the horizontal with intercept at the origin, often called the "line of no discrimination" since increasing the threshold would admit as many true as false positives. Below this line, the predictor would consistently give incorrect classifications. In fact, prediction would improve by inverting the classifications.

[29] The determination of a threshold that is used to decide whether a pixel in a scene is to be placed in the "accept" or "reject" categories is based on a criterion that accounts for the importance of false alarms and misses. Many criteria are possible and several examples are discussed by *Peterson et al.* [1954]. Decision errors may have greater consequences in one application than in another. Bayes' rule is versatile in that it can accommodate costs of making incorrect decisions. For example, costs can be ascribed to a false alarm or a miss, designated here as $C_{F,S}$ and $C_{S,F}$ respectively. Costs can also be ascribed to correct classifications; these are designated by $C_{S,S}$ and $C_{F,F}$. The first subscript in $C_{i,j}$ $\{i,j\} \in S,F\}$ indicates the hypothesis chosen and the second, the hypothesis that was true. Generally, costs are set according to the requirements of an application and their effect is to introduce a bias. Expressed in terms of the likelihood ratio, the decision rule can be written as [*Van Trees*, 2001]

$$L(R_{\min}, R_{\max}) = \frac{P(R_{\min}, R_{\max}|F)}{P(R_{\min}, R_{\max}|S)} \begin{matrix} <^{F} \\ >^{S} \end{matrix} \frac{P(S)\left(C_{F,S} - C_{S,S}\right)}{P(F)\left(C_{S,F} - C_{F,F}\right)}, \quad (2)$$

where the inequalities dictate rejection if the left side is less than the right, acceptance otherwise.

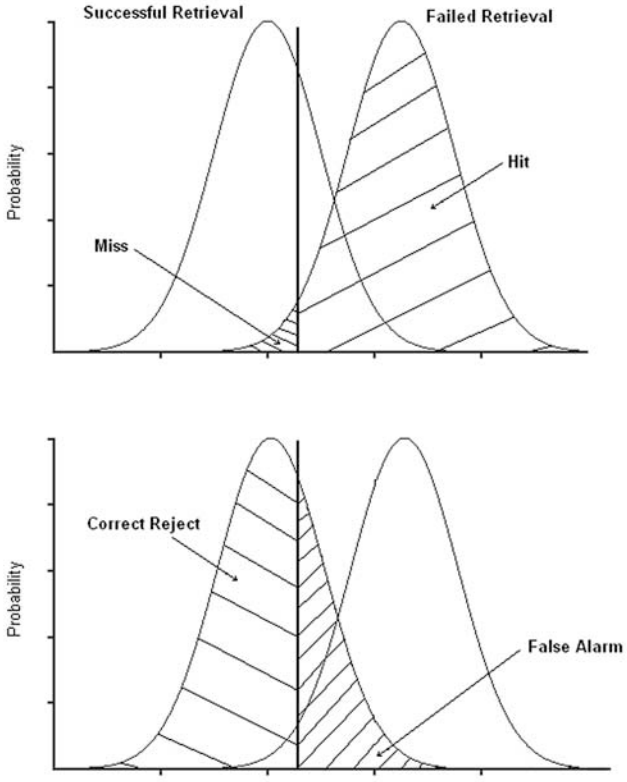

Figure 8. (top) Vertical axis refers to the probability of successful or failed retrievals. The horizontal axis designates the test threshold. At the position indicated by the heavy vertical line, we are assured of identifying a large number of failed retrievals or hits as indicated by the shaded area to the right of the line and a few misses to the left. (bottom) A desirable high correct rejection rate but perhaps an unacceptably large false alarm rate. The overlapping distributions introduce inevitable ambiguity in classification.

[30] The importance of an incorrect decision is assessed in terms of its consequential costs, relative to making a correct decision. In our application, these features are embodied within the concept of the likelihood ratio as given by equation (2) which can be connected to the ROC analytically. It can be proved [*Van Trees*, 2001; *Marques de Sá*, 2007] that the derivative of the ROC is equal to the likelihood ratio. The optimum threshold for the two class problem is given by

$$\frac{d(\text{tpr})}{d(\text{fpr})} = \frac{(\text{false positive cost} = \text{C}_{\text{FS}})}{(\text{false negative cost} = \text{C}_{\text{SF}})}\left[\frac{1-P}{P}\right], \qquad (3)$$

where P is the prior probability of failure and the costs of correct classification $C_{S,S} = C_{F,F} = 0$. Insightful discussions regarding the introduction of costs are given by *Zweig and Campbell* [1993]. The point on the ROC plot where a tangent line having a slope equal to that given by (3) is the optimal threshold for the given costs and prior.

[31] When the number of failed and successful retrievals are equal, (P = 0.5) and for equal misclassification costs, the resulting slope of the tangent line is unity. A slope of unity can also be obtained by setting $\text{C}_{\text{F,S}} = (1 - \text{P})^{-1}$ and $\text{C}_{\text{S,F}} = \text{P}^{-1}$, equivalent to setting the cost ratio $\frac{\text{C}_{\text{F,S}}}{\text{C}_{\text{S,F}}} = \frac{\text{P}}{(1-\text{P})}$, which is the condition employed in the analyses.

[32] We also explored the case when $\text{C}_{\text{F,S}} = \text{C}_{\text{S,F}}$ but only mention that the results of these experiments showed a consistent tendency of the classifier to report a larger false positive rate than for the unit-slope classifier with the aforementioned cost ratio.

[33] To motivate using the unit slope classifier, consider points A, B and C and the tangent lines at those points as shown in Figure 9. Derivatives of the ROC d(tpr)/d(fpr) less than unity (e.g., at point C) indicate that small changes in tpr will change fpr by a significant amount. Similarly, points on the ROC where derivatives exceed unity (e.g., point A)

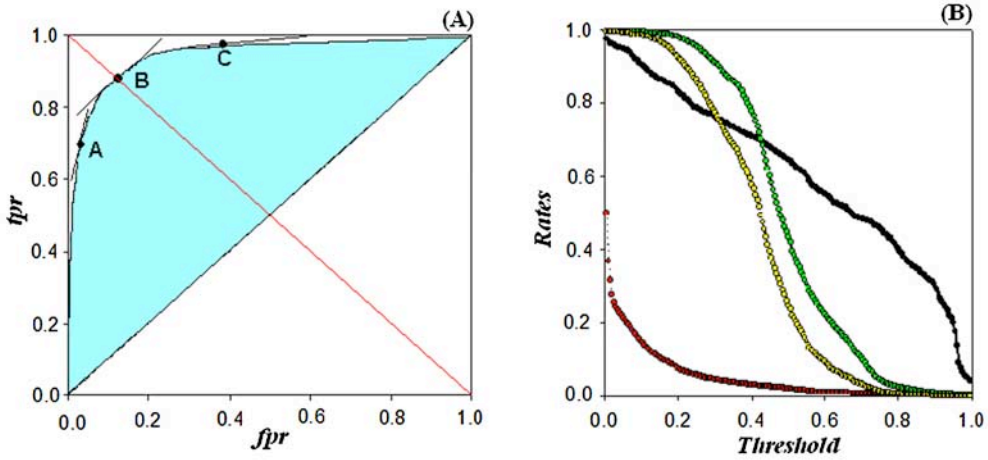

Figure 9. (a) ROC curve of Figure 5. The area in blue is used as a metric of accuracy of the classifier (see section 7 for discussion). The 45° black, major diagonal line is the line of no discrimination. Points A, B, and C are used to explain properties of the slopes of the ROC curve computed for solar zenith angle of 0° and nadir observation. (b) Variation of tpr and fpr values, designated by the rates ordinate with threshold. The black tpr and red fpr curves associated with Figure 5 depict an accurate classification; the green tpr and yellow fpr curves, associated with Figure 7, are very similar, suggesting that the ROC will lie near the line of no discrimination.

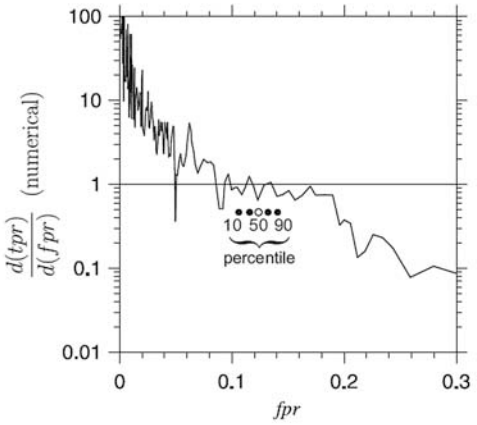

Figure 10. Derivative of the ROC curve computed for solar zenith angle of 0° and nadir observation. The noisiness of the derivative is clearly evident, even though the corresponding ROC in Figure 5 appears to be smooth. The horizontal line depicts the tangent for the unit-slope classifier. The solid circles represent the 10th, 30th, 70th, and 90th percentiles used to analyze the results of the method described fully in Appendix A. The open circle is the median solution to the optimal operating point, or threshold, fpr = 0.123, which is comparable to the solution obtained by the Euclidean distance approach described in the text.

indicate that small changes in fpr will cause significant changes in tpr. Thus the point on the ROC where its derivative is unity (e.g., point B) is optimal in the sense that small perturbations about this point will not introduce significant changes to the threshold. Point B is located at the intersection of the minor diagonal, depicted by the red line with the ROC. From this point, a line parallel to the fpr axis is drawn intersecting the tpr axis. This point of intersection gives tpr. In a similar way, fpr may be found. In general, prior to the taking of derivatives, the ROC must be plotted. For example, to compute the ROC associated with Figure 5, one proceeds iteratively. First, $P(F|R_{min},R_{max})$ is calculated from the training set. This probability is compared with a (variable) threshold inside a loop. If it exceeds the threshold, both tpr and fpr are computed, as these quantities change with the threshold. The resulting plot is shown in the lower central row of Figure 5. The ROC plots shown in Figures 6 and 7 are calculated for the indicated illumination and viewing geometries. It is seen that as retrieval performance deteriorates, the ROCs approach the major 45° diagonal. The physical reasons for this behavior are further discussed in section 7.

[34] The inherent nature of ROCs results in their derivatives being generally very noisy. For instance, Figure 10 shows the numerical derivative of ROC shown in Figure 5. Clearly, accurate estimation of fpr given a particular slope is extremely difficult. Filtering the ROC data helps attenuate noise, but since differentiation amplifies high-frequency noise, it is often unsatisfactory. Differentiation of noisy functions is known to be an ill-posed problem. While the literature abounds with techniques for data smoothing and differentiation [e.g., *Wood*, 1982; *Woltring*, 1985, 1986], we offer yet two methods that exploit the symmetry and smoothness of ROCs encountered in this application. In part 3 a different way of computing the ROC is employed that uses results of a nonlinear regression in connection with support vector machines.

[35] An inspection of the data used to produce the plots shown in Figures 5, 6, and 7 reveals the ROC curves are nearly symmetrical about the minor diagonal. For symmetrical ROCs, a chord of the ROC perpendicular to the minor diagonal will be bisected by that diagonal. The minor diagonal delineates the points of no bias for the ROC. Points to the right of the minor diagonal indicate a bias toward increasing failure rates; those to the left of the diagonal indicate a bias favoring a reduction in the false alarm rate.

[36] A priori, such near symmetry was not expected and is probably due to the widths of the distributions illustrated in Figure 8 being about the same. For the moment, we shall require the curves to be symmetrical, but shortly relax this constraint. For reasons discussed above in connection with the selection of point B in Figure 9 as the point least sensitive to perturbations, we observe that symmetry about the minor diagonal places that point closest to (0,1), associated with the absence of false alarms. Numerical experiments show that for the special case of unit slope, the (fpr, tpr) combination that minimizes the Euclidean distance to (0,1) is a good approximation of the operating point. This simple metric leads to the development of a robust, efficient algorithm that can be automated and was used to determine the thresholds for the cases described in this text. We tested the accuracy of this metric by comparing its results to those of a much more general algorithm fully described in the appendix and in general, the agreement was excellent as can be see in Table A1 in Appendix A. The more general method does not require symmetry in the ROC and can accommodate any slope; important considerations for choices of cost functions other than the ones used here. In essence, this algorithm approximates the ROC over an interval using a Padé approximant. Sampling data points over an interval repeatedly leads to the calculation of new Padé coefficients that allow the derivative to be calculated analytically. The optimal fpr from the ensemble of solutions is obtained by analyzing the median which is robust to outliers. The method was also applied to very general ROCs that were scanned from the paper of *Zweig and Campbell* [1993] in a separate study not documented in this paper.

7. Understanding Retrieval Error in Terms of Conditional Probabilities and the ROC

[37] In section 4, we proposed the hypothesis that extremal radiances contained within a grid of a given size are useful metrics of inhomogeneity. Here we wish to explore the consequences of that hypothesis by connecting the pdfs calculated by Bayes' theorem to the computed ROCs. To motivate this discussion, we first consider the simplest situation: a hypothetical plane-parallel atmosphere. In this case, because the radiances are spatially uniform for a given observer, sun geometry, the minimum and maximum radiances within a (movable) grid of any specified size

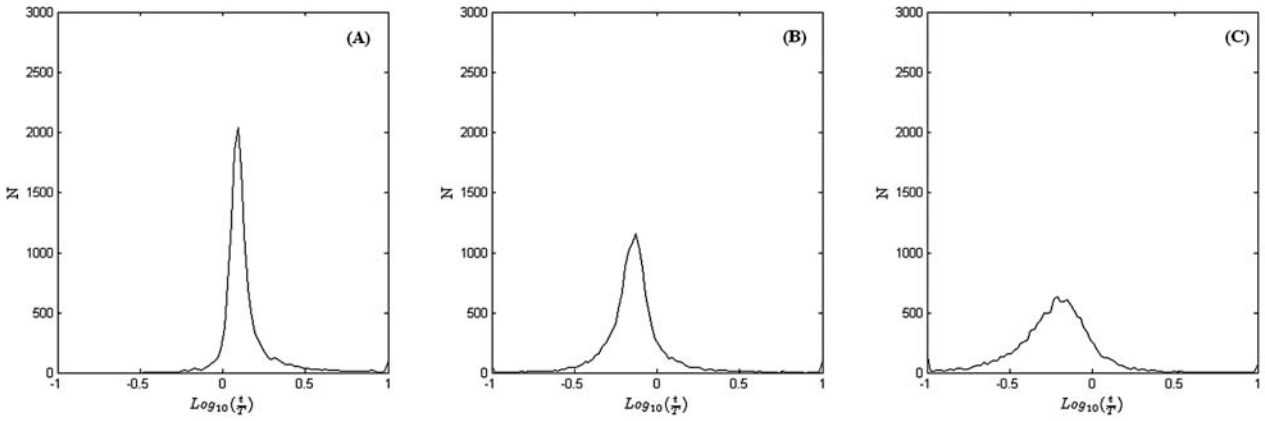

Figure 11. Distribution of optical depth error. Cases depicted in Figures 11a, 11b, and 11c were computed for the observer-solar geometries described in the captions of Figures 5, 6, and 7, respectively.

would be identical; hence a plot of R_{min} versus R_{max} will be a single point. Thus, plane-parallel retrievals will succeed since $P(F|R_{min}, R_{max}) = 0$ and the ROC plot is the single point at (0,1).

[38] Next we consider noninteracting plane-parallel cells of equal size, for example arranged in the pattern of a checkerboard. If the size of the moving grid exactly equals the size of the cells there would be no retrieval error. A plot of R_{min} versus R_{max} is a line of unit slope passing through the origin. However, the joint frequency histogram of (R_{min}, R_{max}) along that line would exhibit variation. In this case too, the ROC is a single point at (0, 1).

[39] If the radiances within the moving grid are now averaged over two or more cells, as might be necessary in a retrieval, then errors would be generated, because $R_{min} \neq R_{max}$ within the grid. Thus, there exists a single critical scale where there is no retrieval error for this situation. Averaging of the radiances within the moving grid would introduce nonuniqueness, since now a mean radiance would not necessarily correspond to a single optical depth. The retrieval error would be clearly dependent on the averaging scale. Points would start filling the region below the 45° line, resulting in the generation of a nontrivial ROC plot. As the number of errors increases, we might expect the ROC to start approaching the line of no discrimination assuming there were as many correct as incorrect retrievals.

[40] In three dimensions, averaging is still present and its action more subtle than the example just given. The emerging remotely sensed radiance over a given column in a scene is no longer simply related to the optical depth beneath that column, but a weighted average of photons entering that column from (theoretically) all locations in the cloud. The weights or contributions of photons from distant locations to the total radiance exiting at the observation angle over the column of interest will be small owing to their large probability of out-scattering caused by the presence of holes or by scattering by the intervening medium. However, this averaging of photons does not necessarily produce unique radiances and is portrayed by the ROC as a loss of discriminatory power between failed and successful retrievals through its curvature, recalling how "failure" has been previously defined.

[41] The location of the failures is predicted accurately by the Bayesian approach, while a quantitative measure of its accuracy is provided by the ROC curve. In an independent set of tests using neural networks, and by comparison to the support vector machine classifier to be discussed in parts 2 and 3 of the series, we show that the accuracy of the simple Bayesian classifier is of practical utility.

[42] In Figure 11, histograms of retrieval errors show the presence of a bias that may be either positive or negative. Figures 11a−11c, the displacements of the peaks from the origin are +0.123, −0.161 and −0.262 units respectively or at the peaks, t = 1.33T, 0.69T and 0.55T, suggesting that, with the exception of the largest in-magnitude displacement, the bias may be more detrimental to applications requiring optical depths of order unity.

[43] For example a 40% error on an optical depth of unity is radiatively more significant than on an optical depth of 10. Figures 11a−11c also show that the error histograms become wider as both the viewing angle and solar zenith angle increase. The exacerbation of retrieval error is with increasing solar zenith angle is consistent with the findings of *Loeb and Davies* [1996] and *Loeb and Coakley* [1998]. These results underscore the importance of the solar geometry and viewing angles associated with 3-D transfer and the effects they exert on retrievals as measured by both bias and variance in the retrieved optical depth error distributions. In part 2, retrievals of optical depth using neural networks employing extremal radiances show the utility of these radiances by virtually eliminating all bias even for case C, making the distribution of error more symmetrical and reducing the variance most noticeably for that case.

[44] An additional benefit derived from using ROC curves is that it allows direct comparisons of the effectiveness of different tests (with different units) by transforming the power of each test into one common unit, the area under the curve defined here, for example, as the area of the blue region in Figure 9 incremented by a half. The increment accounts for the area of the triangular region under the major diagonal. This means that, for instance, if instead of using extremal radiances, one used the mean and standard deviation in a moving window, the area under the curve could be used to objectively compare the performance of the classifiers. This area, a nonparametric statistic can be shown

Table 1. Areas Under ROC Curve for Indicated Solar and Viewing Zenith Angles[a]

Reference Geometry	Area
$\theta_0 = 0$, $\theta_v = 0$	**0.94**
$\theta_0 = 0$, $\theta_v = 30$	0.86
$\theta_0 = 0$, $\theta_v = 60$	0.71
$\theta_0 = 30$, $\theta_v = 0$	0.89
$\theta_0 = 30$, $\theta_v = 30$	**0.81**
$\theta_0 = 30$, $\theta_v = 60$	0.64
$\theta_0 = 60$, $\theta_v = 0$	0.67
$\theta_0 = 60$, $\theta_v = 30$	**0.65**

[a]Solar zenith angle, θ_0; viewing zenith angle, θ_v. Scores in bold letters, top, middle, and bottom, were computed using Figures 5, 6, and 7, respectively. Other scores are associated with the results of analyses of cases not presented are reference points for part 3.

to be equivalent to the Mann-Whitney test [*Mann and Whitney*, 1947] which tests for difference in median between the scores obtained by the two categories (i.e., ''successful'' or ''failed'' retrievals) considered in this study.

[45] The area under the ROC curve provides a useful measure of the probability that when a random selection of a correct and incorrect classification is made, the predictor (or classifier) will assign a higher score to the correct than to the incorrect classification. The areas presented in Table 1 calculated from the ROCs, are in excellent agreement with our subjective sense of quality of the predicted problem locations given by Figures 5–7. In addition to Figures 5–7, several ROCS with different viewer-illumination geometries were analyzed but not presented owing to lack of space. The areas for these cases are reported here because they will be compared to those computed from an entirely different method based on support vector machines in part 3.

8. Concluding Comments and Remarks on Bayesian Classification

[46] At this point some observations and comments are in order. To begin, we remark that the Bayesian classifier is optimal if the statistics of the process are known. The probabilities arising in the analyses cannot be represented by parametric forms such as Gaussians, so computation of the probability densities is necessary. Use of nonparametric statistics is actually a feature of the classifier. It is also, however, a potential source of difficulty if the joint probability density function is of high dimensionality, as this would introduce sampling problems. In this application, nonparametric statistics are still useful because the probability densities required by the classifier depend on only two features: R_{min} and R_{max}. In the development of the Bayesian inferential engine, no attempt was made to estimate cloud optical depth mainly because of the difficulty in comparing the results of such attempts to those where more than two radiances could be employed. While it is theoretically possible to use more than two radiances to perform optical depth retrievals, evaluation and storage of such higher-dimensional probability densities would be impractical. The task of retrieving optical depths without using plane-parallel theory is left to parts 2 and 3 whose inferential engines do not require probability distributions as inputs and thus can more thoroughly characterize the accuracy of retrievals employing extremal radiances.

[47] Equation (1) is implicitly dependent on the observer's zenith angle, solar zenith angle and the azimuth between the observer's viewing plane and the Sun. Thus, at a given scale, all the probabilities for different viewing and solar geometries must be computed and stored in a database. The availability of fast computers and mass storage technologies make this a feasible undertaking for a finite number of discrete angles. In practical applications, it would be necessary to use data computed for angles most closely matching those observed. Monte Carlo simulations, that exploit parallelism for the sake of efficiency, would be suitable for this task. Alternatively, it would be necessary to interpolate observed radiances to those in the database; a procedure described by *Cornet et al.* [2004, 2005] that relies on the use of neural networks. Since parts 1–3 are exploratory, focusing on understanding the issues connected with the use of different inferential engines in retrievals, their use in error classification and ascertaining the best accuracy that can be expected by such methods, interpolation of radiances is not attempted.

[48] The boundary layer stratiform clouds considered here are comprised entirely of liquid water droplets. Such clouds are common and climatologically significant because they reflect much solar radiation yet their temperature is often close to that of the surface; thus they exert a cooling effect. The morphology of stratiform clouds is not of singular character as can be confirmed by even casual observation. Although internally inhomogeneous, we require these clouds to occupy a physically large area and lack well defined structures such as streets, ideally, to be statistically spatially homogeneous. This is reasonable since we expect cloud streets to exhibit strongly oriented radiances, where rotating the cloud field, while maintaining observer-Sun geometry, affects radiance patterns. Furthermore, radiances would also be sensitive to other uncontrollable factors such as cloud height and street spacing. Sensitivity to rotation would introduce errors in the generalization, affecting the practicality of using inference engines. We assert that for stratiform clouds having the properties stated, inference engines will be relatively insensitive to rotations of the cloud field, provided that the sample of cloud fields is sufficiently large. Radiances will still, however, exhibit a dependence on azimuth; a moment's reflection should convince the reader that this must be the case, since such dependence applies even to plane-parallel media owing to highly anisotropic scattering phase functions.

[49] This study, as well as parts 2 and 3, focuses exclusively on liquid water clouds since their particle size distributions and shape are much more constrained than they are for cirrus clouds [*Cooper et al.*, 2006] for example. In fact, uncertainties in radiances due to ice crystal shape can compete with uncertainties in the radiance due to spatial inhomogeneities [*Tsay et al.*, 1996]. On the contrary, for clouds comprising water droplets, while the effects of nonsphericity are eliminated, uncertainty is introduced through the specification of particle size distribution which governs extinction. For such clouds, it seems that a thorough study into the effects of particle size distribution uncertainty on retrievals is lacking; but according to *Arduini et al.* [2005], uncertainty in the variance of droplet size distribution affects retrievals of effective radius more strongly than retrievals of optical depth. The operational 2btau CloudSat retrieval also calculates the effective radius for an assumed lognormal droplet size distribution because the effective

radius acts as a constraint to reduce the error in the optical depth retrieval (R. Austin, personal communication, 2008).

[50] It might be argued that use of CSRM (and other model) data as proxies for naturally occurring stratocumulus clouds is a fundamental weakness of all inferential methods. On the other hand, all retrievals, including data assimilation, assume that the forward models used are "reasonable" approximations to processes occurring in the atmosphere. In such models, certain outputs approximate the true state of nature more than others. Otherwise, what purpose would a cloud model serve if it did not capture some essential characteristics of real clouds?

[51] In using any inferential method a certain model dependency is expected. For example, the spatial statistics of the cloud field will affect the Bayesian classifier in a way that is unknown since three dimensional radiative transfer, used in training, is sensitive to cloud morphology, considered here to be random functions. We do not view these factors as fundamental limitations, but rather as avenues requiring further exploration. For instance, how does the accuracy of the classifier change when it is trained on one model and applied to the output of another? Such information may be useful in answering how different assumptions in models (e.g., subgrid turbulent closures, cloud microphysics, etc) affect the statistical properties of simulated clouds. Even if the Bayesian classifier is robust to variations in cloud spatial statistics, which is necessary for it to be operationally useful, it would ultimately have to be verified, perhaps by in situ measurements of cloud microphysical and optical properties over some limited domain, in conjunction with satellite radiances. Clearly, this task cannot be performed routinely. While it is true that real and model clouds differ, the question here is one of extent. In this study, what is of interest is knowledge of how model departures from real clouds affect the error statistics of retrieved quantities. An analysis of such departures would benefit this work as well as both the modeling and observational communities by establishing a useful symbiosis that seems, at present, to be missing.

Appendix A

[52] It has already been shown that derivatives calculated from ROCs obtained through the analysis of data can be very noisy. It is, however, possible to develop an algorithm that exploits the relative smoothness of ROCs typically found in this application. The method approximates a portion of the ROC, over the range Δfpr that contains $\frac{d(tpr)}{d(fpr)} = s$ with the Padé approximant

$$\mathrm{tpr}(\mathrm{fpr}) = \frac{\sum_{n=0}^{N-1} a_n (\mathrm{fpr})^n}{\sum_{n=0}^{N} b_n (\mathrm{fpr})^n}, \tag{A1}$$

where a_n and b_n are coefficients with $b_N = 1$ [*Brezinski*, 1973]. First, define the width of the range Δfpr that (A1) is to be generated along with the value of s. Then, using the computed values of the ROC, find the smallest value of fpr = α such that

$$\frac{\mathrm{tpr}(\beta) - \mathrm{tpr}(\alpha)}{\beta - \alpha} < s,$$

where $\beta = \alpha + \Delta$fpr. The interval $[\alpha,\beta]$ will contain the solution. As 2N pairs of (fpr,tpr) are required to solve for a_n and b_n in (A1), subdivide $[\alpha,\beta]$ into 2N equal subranges of width Δx separated by intervals $\xi\Delta x$ where $\xi < 1$. Hence,

$$\Delta x = \frac{\Delta \mathrm{fpr}}{[2N + (2N-1)\xi]}.$$

From each of the 2N subranges, using a uniform distribution within each subrange, randomly select a value of fpr, and from that, choose the closest $(\mathrm{fpr}_i,\mathrm{tpr}_i)_{i = 1,\ldots,N}$, from the computed values of the ROC. Using the 2N ordered pairs, solve for the coefficients in (A1) [*Brezinski*, 1973]. Then, differentiating (A1), solve

$$\frac{\left(\sum_{n=0}^{N} b_n (\mathrm{fpr})^n\right)\left(\sum_{n=0}^{N-1} n a_n (\mathrm{fpr})^{n-1}\right) - \left(\sum_{n=0}^{N} n b_n (\mathrm{fpr})^{n-1}\right)\left(\sum_{n=0}^{N-1} a_n (\mathrm{fpr})^n\right)}{\left(\sum_{n=0}^{N} b_n (\mathrm{fpr})^n\right)^2} - s = 0, \tag{A2}$$

for fpr. If it turns out that the denominator in the first term of (A2) has a root in $[\alpha,\beta]$, simply discard it and randomly select another four $(\mathrm{fpr}_i, \mathrm{tpr}_i)$.

[53] It was found, by experimentation that N = 2 is adequate for the purpose at hand. As such, (A2) reduced to the quartic equation

$$s(\mathrm{fpr})^4 + 2sb_2(\mathrm{fpr})^3 + \left[s\left(b_2^2 + 2b_1\right) + a_2\right](\mathrm{fpr})^2 + (2sb_1b_2 + 2a_1)(\mathrm{fpr}) + \left(sb_1^2 - b_1a_2 + b_2a_1\right) = 0, \tag{A3}$$

which can be solved readily in closed form for fpr.

[54] What this technique does is approximate the monotonically increasing, piecewise-continuous ROC over the range $[\mathrm{fpr}_1,\mathrm{fpr}_{2N}]$ by a continuous function, with continuous derivative, and solve for the point whose tangent has a slope s. This is achieved using information about the ROC from a neighborhood around the solution. One does not want to make Δfpr too small for too little information about the ROC will be sampled. Similarly, making Δfpr too large will often force the fit to be too approximate near the solution. The solution depends on Δfpr, ξ and the 2N random samples drawn from the subranges. By employing the bootstrap technique, that is, performing this process many times using randomly sampled values of Δfpr, ξ, and ordered pairs, it is possible to estimate uncertainties on the estimate of fpr at slope s with the estimated value itself being the median of the resulting distribution of solutions.

[55] The bootstrap portion of the solution uses a large number of sample solutions of (A3). One hundred samples were found to be sufficient (though the routine is very fast

and, in all likelihood, 1000 samples could be tolerated easily in an operational setting). During bootstrap sampling, Δfpr and ξ were produced by

$$\Delta\text{fpr} = 0.1 + (0.3 - 0.1)\mathbf{R} \tag{A4}$$

$$\xi = 0.3 + (0.7 - 0.3)\mathbf{R}, \tag{A5}$$

where **R** is a uniform random deviate on [0,1]. Of the two, the solution depends much more on Δfpr. Figure A1 shows how several percentiles of the solution depend on Δfpr assuming (A5) and s = 1 for the ROC shown in Figure 5. Figure 10 shows the bootstrap-generated distribution of solutions assuming (A4), (A5), and s = 1 (of the samples used to produce these estimates, ~2.5% of them were rejected owing to a root in the denominator of (A2) inside $[\alpha,\beta]$). The median is 0.123 while its interquartile range, almost symmetric about the median, is 0.019.

[56] The method requires initially, operator intervention to set up α and Δfpr as in (A4) and (A5) before the calculation can begin. For example, in (A4) the lower limit Δfpr is set to 0.1 and will attain a maximum value of 0.3 when **R** = 1. Practically, there is no significant issue connected with the finite number of combinations of data points in the subintervals (see *Press et al.* [1992] for a discussion on bootstrapping via sampling with replacement). The quality of the fpr operating point is indicated by the separation of the 10% and 90% percentiles.

[57] Although quartics generally have four roots, it was observed that, for the ROCs calculated in this paper, the roots were never complex. Frequently, doublets were produced outside $[\alpha,\beta]$, the third root was usually negative, and the last root lay inside the $[\alpha,\beta]$. Several other possibilities exist, but to minimize ambiguity, the algorithm is forced to utilize solutions containing just one root in $[\alpha,\beta]$.

[58] For the special unit slope case, the algorithm has an optional switch that allows testing of the roots using the minimum Euclidean distance test, described in section 6, when solutions with more than one root in $[\alpha,\beta]$ are allowed. For the cases considered here, such cases existed for about 10% to 30% of the simulations. Table A1 lists results for several different ROCs. Note that the minimum

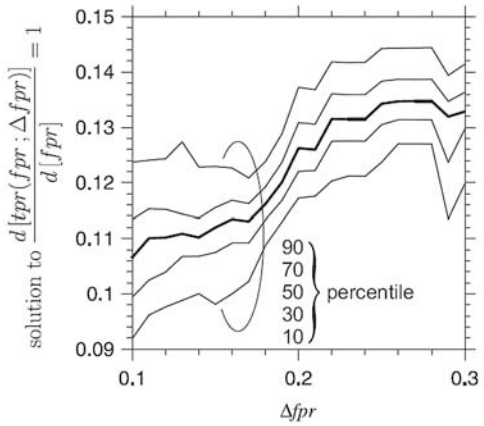

Figure A1. Distribution of solutions showing the sensitivity to Δfpr.

Table A1. Medians and Interquartile Ranges for the Analytic Method Presented in Appendix A Assuming (A4), (A5), and s = 1 for Five Different Illuminations and Viewings of the Field Shown in Figure 1[a]

Solar Zenith Angle	Viewing Zenith Angle	Relative Azimuth	Analytic Estimate (Median; Range)	Euclidean Distance
0	0	0	0.123; 0.020	0.123
0	30	0	0.270; 0.009	0.214
0	60	0	0.315; 0.013	0.335
30	30	0	0.271; 0.020	0.263
60	30	0	0.442; 0.019	0.414

[a]The corresponding minimum Euclidean distances from (0,1) to the ROC are shown for comparison.

Euclidean distance from (0,1) to the ROC appears generally to approximate the analytic routine to within about 10%.

[59] Finally, filtering the ROC curves provided minimal benefits for the ROCs used in this study. However, in a separate study of this method, using data from *Zweig and Campbell* [1993], the algorithm benefited notably as it reduced the difference between the 10% and 90% percentiles.

[60] **Acknowledgments.** This work was performed under NASA grant NRA-02-OES-06 OCO grant 1280999, and the Atmospheric Radiation Measurement (ARM) grant DE-FG02-03ER63521. We wish to thank Giovanni Leoncini at Reading University, U. K., for his helpful criticisms and for the input of professor Joaquim Marques de Sá at the Universidade do Porto, Portugal, for his help with ROCs. Finally, we also gratefully acknowledge the comments and suggestions of our external reviewers that helped improve the presentation and clarity of this paper.

References

Arduini, R. F., P. Minnis, J. K. Ayers, M. M. Khaiyer, and P. Heck (2005), Sensitivity of satellite-retrieved cloud properties to the effective variance of cloud droplet distribution, paper presented at Fifteenth ARM Science Team Meeting, Atmos. Radiat. Measure. Program, Dep. of Energy, Washington, D. C.

Barker, H. W., R. K. Goldstein, and D. E. Stevens (2003), Monte Carlo Simulation of solar reflectances for cloudy atmospheres, *J. Atmos. Sci.*, *60*, 1881–1894.

Brezinski, C. (1973), *Accélération de la Convergence en Analyse Numerique, Lecture Notes Math*, vol. 584, 313 pp., Springer, Berlin.

Cooper, S. J., T. L'Ecuyer, and P. Gabriel (2006), Objective assessment of the information content of visible and infrared radiance measurements for cloud microphysical property retrievals over the global oceans. Part II: Ice clouds, *J. Appl. Meteorol. Climatol.*, *45*(1), 42–62, doi:10.1175/JAM2327.1.

Cornet, C., H. Isaka, B. Guillemet, and F. Szczap (2004), Neural network retrieval of cloud parameters of inhomogeneous clouds from multispectral and multiscale radiance data: Feasibility study, *J. Geophys. Res.*, *109*, D12203, doi:10.1029/2003JD004186.

Cornet, C., J.-C. Buriez, J. Riedi, H. Isaka, and B. Guillemet (2005), Case study of inhomogeneous cloud parameter retrieval from MODIS data, *Geophys. Res. Lett.*, *32*, L13807, doi:10.1029/2005GL022791.

Davis, A., A. Marshak, R. Cahalan, and W. Wiscombe (1997), The LANDSAT scale-break in stratocumulus as a three-dimensional radiative transfer effect, implications for cloud remote sensing, *J. Atmos. Sci.*, *54*, 241–260, doi:10.1175/1520-0469(1997)054<0241:TLSBIS>2.0.CO;2.

Doswell, C. A., III, R. Davies-Jones, and D. L. Keller (1990), On summary measures of skill in rare event forecasting based on contingency tables, *Weather Forecast.*, *5*(4), 576–585, doi:10.1175/1520-0434(1990)005<0576:OSMOSI>2.0.CO;2.

Duda, R. O., and P. Hart (1973), *Pattern Classification and Scene Analysis*, Wiley-Intersci., Hoboken, N. J.

Duda, R. O., P. Hart, and D. Stork (2001), *Pattern Classification*, 2nd ed., Wiley-Intersci., Hoboken, N. J.

Green, D. M., and J. A. Swets (1966), *Signal Detection Theory and Psychophysics*, John Wiley, Hoboken, N. J.

Hand, D. J., and R. J. Till (2001), A simple generalisation of the area under the ROC curve for multiple class classification problems, *Machine Learning*, *45*(2), 171–186, doi:10.1023/A:1010920819831.

Hanley, J. A., and B. J. McNeil (1982), The meaning and use of the area under a receiver operating characteristic (ROC) curve, *Radiology*, *143*, 29–36.

Harvey, L. O., K. R. Hammond, C. M. Lusk, and E. F. Mross (1992), The application of signal detection theory to weather forecasting behavior, *Mon. Weather Rev.*, *120*(5), 863–883, doi:10.1175/1520-0493(1992)120<0863:TAOSDT>2.0.CO;2.

Heidinger, A., and G. L. Stephens (2000), Molecular line absorption in a scattering atmosphere II. Applications to remote sensing in the O_2 A-band, *J. Atmos. Sci.*, *57*, 1615–1634, doi:10.1175/1520-0469(2000)057<1615:MLAIAS>2.0.CO;2.

Iwabuchi, H., and T. Hayasaka (2002), Effects of cloud horizontal inhomogeneity on the optical thickness retrieved from moderate-resolution satellite data, *J. Atmos. Sci.*, *59*, 2227–2242.

Loeb, N. G., and J. A. Coakley Jr. (1998), Inference of marine stratus cloud optical depth from satellite measurements: Does 1D theory apply?, *J. Clim.*, *11*(2), 215–233, doi:10.1175/1520-0442(1998)011<0215:IOMSCO>2.0.CO;2.

Loeb, N. G., and R. Davies (1996), Observational evidence of plane parallel model biases: Apparent dependence of cloud optical depth on solar zenith angle, *J. Geophys. Res.*, *101*, 1621–1634, doi:10.1029/95JD03298.

Loeb, N. G., and R. Davies (1997), Angular dependence of observed reflectances: a comparison with plane-parallel theory, *J. Geophys. Res.*, *102*, 6865–6881, doi:10.1029/96JD03586.

Mann, H. B., and D. R. Whitney (1947), On a test whether one of two random variables is stochastically larger than the other, *Ann. Math. Stat.*, *18*(1), 50–60, doi:10.1214/aoms/1177730491.

Marques de Sá, J. P. (2007), *Applied Statistics Using SPSS, STATISTICA, MATLAB and R*, 2nd ed., 506 pp., Springer, New York.

Mason, I. (1979), On reducing probability forecasts to yes/no forecasts, *Mon. Weather Rev.*, *107*(2), 207–211, doi:10.1175/1520-0493(1979)107<0207:ORPFTY>2.0.CO;2.

Mason, I. (1982), A model for assessment of weather forecasts, *Aust. Meteorol. Mag.*, *30*, 291–303.

Mason, S. J., and N. E. Graham (1999), Conditional probabilities, relative operating characteristics and relative operating levels, *Weather Forecast.*, *14*(5), 713–725, doi:10.1175/1520-0434(1999)014<0713:CPROCA>2.0.CO;2.

Murphy, A. H., and R. L. Winkler (1987), A general framework for forecast verification, *Mon. Weather Rev.*, *115*(7), 1330–1338, doi:10.1175/1520-0493(1987)115<1330:AGFFFV>2.0.CO;2.

Olson, R. H. (1965), On the use of Baye's theorem in estimating false alarm rates, *Mon. Weather Rev.*, *93*(9), 557–558, doi:10.1175/1520-0493(1965)093<0557:OTUOBT>2.3.CO;2.

Peterson, W. W., T. G. Birdsall, and W. C. Fox (1954), The theory of signal detectability, *IRE Trans. Prof. Group Inf. Theory*, *4*, 171–212, doi:10.1109/TIT.1954.1057460.

Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery (1992), *Numerical Recipes in FORTRAN 77: The Art of Scientific Computing*, 2nd ed., 975 pp., Cambridge Univ. Press, Cambridge, U. K.

Provost, F., T. Fawcett, and R. Kohavi (1998), The case against accuracy estimation for comparing induction algorithms, in *Proceedings of the 15th International Conference on Machine Learning*, p. 445, Elsevier, New York.

Rogers, C. D. (2004), *Inverse Methods for Atmospheric Sounding: Theory and Practice*, World Sci, Hackensack, N. J.

Stephens, G. L., and A. Heidinger (2000), Molecular line absorption in a scattering atmosphere: Part I: Theory, *J. Atmos. Sci.*, *57*, 1599–1614.

Stevens, D. E., A. S. Ackerman, and C. S. Bretherton (2002), Effects of domain size and numerical resolution of the simulations of shallow cumulus convection, *J. Atmos. Sci.*, *59*, 3285–3301, doi:10.1175/1520-0469(2002)059<3285:EODSAN>2.0.CO;2.

Swets, J. A. (1973), The relative operating characteristic in psychology, *Science*, *182*(4116), 990–1000, doi:10.1126/science.182.4116.990.

Swets, J. A. (1988), Measuring the accuracy of diagnostic systems, *Science*, *240*(4857), 1285–1293, doi:10.1126/science.3287615.

Tsay, S. C., P. M. Gabriel, M. D. King, and G. L. Stephens (1996), Spectral reflectance and atmospheric energetics in cirrus-like clouds. Part II: Applications of A Fourier-Riccati approach to radiative transfer, *J. Atmos. Sci.*, *53*, 3450–3467, doi:10.1175/1520-0469(1996)053<3450:SRAAEI>2.0.CO;2.

Van Trees, H. L. (2001), *Detection, Estimation and Modulation Theory: Part 1*, 634 pp., John Wiley, Hoboken, N. J.

Varnai, T., and A. Marshak (2002), Observations of three-dimensional radiative effects that influence MODIS cloud optical thickness retrievals, *J. Atmos. Sci.*, *59*, 1607–1618, doi:10.1175/1520-0469(2002)059<1607:OOTDRE>2.0.CO;2.

Wilks, D. S. (1995), *Statistical Methods in the Atmospheric Sciences*, 467 pp., Academic Press, San Diego, Calif.

Woltring, H. J. (1985), On optimal smoothing and derivative estimation from noisy displacement data in biomechanics, *Hum. Movement Sci.*, *4*, 229–245.

Woltring, H. J. (1986), A Fortran package for generalized cross-validatory spline smoothing and differentiation, *Adv. Eng. Software*, *8*(2), 104–113, doi:10.1016/0141-1195(86)90098-7.

Wood, G. A. (1982), Data smoothing and differentiation procedures in biomechanics, *Exercise Sport Sci. Rev.*, *10*, 308–362, doi:10.1249/00003677-198201000-00010.

Zweig, M. H., and G. Campbell (1993), Receiver-operating characteristic (ROC) plots: A fundamental evaluation tool in clinical medicine, *Clin. Chem.*, *39*(4), 561–577.

H. W. Barker, Cloud Physics and Severe Weather Section, Environment Canada, Toronto, Ontario M3H 5T4, Canada.

N. Ferlay, Laboratoire d'Optique Atmospherique, Université des Sciences et Technologies de Lille, F-59655 Villeneuve d'Ascq, France.

P. Gabriel, D. O'Brien, and G. L. Stephens, Department of Atmospheric Science, Colorado State University, 1371 Campus Delivery, Fort Collins, CO 80523, USA. (gabriel@atmos.colostate.edu)