Alternating Decision Trees for Cloud Masking in MODIS and VIIRS NASA Sea Surface Temperature Products

KATHERINE A. KILPATRICK, GUILLERMO PODESTÁ, ELIZABETH WILLIAMS,
SUSAN WALSH, AND PETER J. MINNETT

Rosenstiel School of Marine and Atmospheric Science, University of Miami, Miami, Florida

(Manuscript received 5 July 2018, in final form 9 November 2018)

ABSTRACT

Identification and exclusion of clouds from satellite-based infrared fields is critical to achieve accurate retrievals of sea surface temperature (SST). Historically, identification of clouds has been driven primarily by a few uniformity tests involving a small number of pixels, brightness temperature range tests, and comparisons to low-resolution gap-free reference fields. Collectively these tests are adequate at identifying large, upper-level, very cold cumulus clouds, and uniformity tests identify moderately sized patchy cumulus clouds. But the efficacy of cloud identification often decreases at cloud edges, for small or thin cirrus clouds, and for the lower, more uniform stratus clouds, for which cloud-top temperature can be comparable to that of the sea surface, particularly at high latitudes. The heavy reliance on stringent uniformity thresholds often also has the unintended consequence of eliminating strong SST frontal regions from the pool of best-quality retrievals. This paper presents results for an ensemble cloud classifier based on a machine-learning approach, boosted alternating decision trees (ADtrees), applied to NASA MODIS and VIIRS SST imagery. The ADtree algorithm relies on the use of a majority vote from a collection of both “weak” and “strong” classifiers. This approach offers the potential to identify more cloud types and improve the retention of SST gradients in best-quality SST retrievals and also provides a per pixel confidence estimate in the classification.

1. Introduction

Sea surface temperature (SST) is a key variable for climate and weather study and forecasting; it has been identified by the World Meteorological Organization (WMO) as an essential climate variable (ECV), being “fundamental to advancing scientific understanding of climate” (Bojinski et al. 2014). The synoptic, repeated coverage of satellite remote sensing is important for the production of global SST fields on daily to decadal time scales. Nevertheless, appropriate interpretation of trends and anomalies requires that we characterize the accuracy of satellite-derived SSTs under a broad range of environmental and observational conditions. The accuracy of SST derived from infrared bands of MODIS (Moderate Resolution Imaging Spectroradiometer; Esaias et al. 1998) and VIIRS (Visible Infrared Imaging Radiometer Suite; Hillger et al. 2013) is limited not only by the accuracy of the on-orbit measurements and by an imperfect atmospheric correction, but also by the effects

of undetected clouds and aerosols. As SST retrievals from satellites in the infrared (IR) require clear-sky conditions, the detection of clouds and detection of atmospheric aerosols are important aspects in limiting the accuracy of retrievals (Kilpatrick et al. 2015). Contamination by even thin clouds can cause biases in SST retrievals, so consequently the identification of clouds in infrared imagery has received considerable attention in the literature (Bulgin et al. 2018; Hollstein et al. 2016; Ackerman et al. 2008, 1998; Barnes and Hu 2013; Frey et al. 2008; Merchant et al. 2005; Saunders and Kriebel 1988).

Clouds are generally characterized by higher reflectance and lower temperature than the underlying ocean surface. For this reason, simple threshold tests for visible and longwave infrared (LWIR) channels offer considerable skill in cloud detection (Ackerman et al. 1998). Other useful tests involve differences in brightness temperatures (BTs) between LWIR channels, and spatial uniformity or continuity tests. Unfortunately, many surface conditions reduce cloud–surface contrast in certain spectral regions (e.g., bright clouds over sea ice that may be covered by snow). Similarly, cloud types such as thin cirrus, low-level stratus at night, and small

Corresponding author: Katherine A. Kilpatrick, kkilpatrick@miami.edu

cumulus typically have low contrast with the underlying background (Ackerman et al. 1998). In the daytime portion of an orbit, reflected sunlight provides additional information, allowing tests to distinguish high-reflecting clouds from the low reflectance of the sea surface. However, tests based on visible bands can be compromised by the high reflectance from the sea surface that occurs in regions of sun glitter (direct specular reflection; Cox and Munk 1954).

The cloud mask for NASA MODIS SST fields relies on multiple tests to indicate a level of confidence that a given pixel corresponds to clear sky. For versions R.2014 and earlier of MODIS SST products and AVHRR Pathfinder SST, the cloud mask was developed using recursive binary decision trees (BDtrees) (Kilpatrick et al. 2015, 2001). These trees were based on the classification algorithm of Breiman et al. (1984), a statistical classifier that uses known characteristics of an object (referred to as “attributes” or “features”) to classify that object into one or more classes. In the MODIS cloud mask, several observed or derived variables were used to estimate the probability of a pixel belonging to one of two possible classes: “clear” or “potentially cloud/aerosol contaminated.” For brevity, henceforth we will refer to those pixels potentially contaminated by cloud or aerosols as “cloud-contaminated,” irrespective of the test resulting in that classification.

The performance of classification algorithms such as a BDtree often is assessed through a confusion matrix. For a binary (two class) classification as is the case for cloud contamination, there are four possible outcomes represented by what is referred to as a confusion matrix.

The four cells in the matrix contain the number of records for each combination of observed and predicted classes. Two cells in the matrix list the numbers of records correctly classified as cloud-contaminated or clear (denoted as “true cloud” and “true clear,” respectively). The remaining two cells correspond to classification errors: 1) clear pixels identified as cloudy (“false cloud”) and 2) cloud-contaminated pixels identified as clear (“false clear”). A confusion matrix allows the calculation of various performance metrics such as sensitivity (the proportion of true clouds correctly identified as a cloud) and specificity (the proportion of truly clear pixels that are correctly identified as clear) (Altman and Bland 1994). These metrics allow a more detailed characterization of a classifier’s performance than considering a single metric such as accuracy (i.e., the overall proportion of records correctly classified).

The two classification errors described above have different implications for satellite-derived SST fields. Misclassification of a cloud-contaminated pixel as clear introduces errors in SST retrievals: unidentified

cloud presence often leads to a negative bias in SST (Ackerman et al. 1998). In contrast, an overly conservative cloud mask—where considerable numbers of clear pixels are misclassified as cloud-contaminated—can introduce significant sampling errors as many truly cloud-free pixels are excluded from spatially binned SST fields, leading to incomplete SST coverage. More importantly, the excessive censoring of lower (yet cloud-free) SST values leads to a failure in capturing the true geophysical variability of SST.

False masking of valid yet anomalously cold sea surface pixels is a pervasive problem for cloud detection algorithms (Merchant et al. 2005). Recent research on the proportion of data missing flagged as cloudy in Level-3 global SST fields revealed significant differences in the persistence of clouds between day and night (Liu and Minnett 2016; Liu et al. 2017). For some regions and seasons, these authors found a puzzling behavior of cloud persistence: normally, in the tropics during summer one would expect a higher proportion of cloud-contaminated pixels during daytime. In contrast, Liu and Minnett observed the opposite pattern—a higher proportion of cloudy pixels at night—and tied this counterintuitive pattern to problems with the cloud screening procedure. Such findings induced us to reassess the performance of BDtrees for cloud screening and to explore newer alternative tree classification methods.

In this paper, we compare the ability of two machine-learning (ML) classification algorithms used to distinguish cloudy and clear-sky pixels in MODIS and VIIRS SST images. We compare the performance of BDtrees, the method historically used for cloud masking of NASA infrared SST products, to that of alternating decision trees (ADtree) classifiers (Freund and Mason 1999). Sequential decisions based on a series of radiometric thresholds and uniformity tests have been used in cloud masks, over both land and ocean, since the beginning of the satellite era of environmental measurements. Both BDtrees and ADtrees represent an ensemble of Boolean functions that are easy to interpret and understand, often with minimal computational costs, and mimic human logic. When decision trees are used in classification problems, each node in a tree represents a feature (attribute) and each branch a decision point (test), ending in a terminal leaf with the classification outcome.

A large number of ML classification algorithms exist, and many of them are implemented in widely available, open-source software. Some methods like neural networks are “black boxes” for which classification criteria cannot be accessed or interpreted. Other approaches such as those based on information from nearest neighbors can be extremely computationally intensive and would be difficult to implement efficiently for processing and reprocessing large volumes of satellite data.

The ADtree classification algorithm (Freund and Mason 1999) was selected for this work because of three advantages. First, like BDtrees they are easy to interpret, straightforward to implement, and both have low computational costs. Second, an ADtree not only predicts class membership of a pixel, but also provides an estimate of the confidence in that prediction (called a margin). Finally, an ADtree prediction represents a collective vote from an ensemble of both strong and weak classifiers, rather than a decision from a single terminal node. It is this collective weighted majority vote that is the key advantage and power of an ADtree.

2. Data and methods

a. The data

The data used to build the classification models discussed below must include numerous records for which a broad range of attributes or features are available. The data must also list the known category of each record (i.e., the value that will subsequently be predicted for other records). We used a subset of records randomly selected from the SST matchup databases (MUDBs) described by Kilpatrick et al. (2001, 2015), now publicly available for VIIRS and MODIS from the NASA Ocean Biology distributed archive system (OB.DAAC) SeaBASS validation system (<https://seabass.gsfc.nasa.gov/archive/SSTVAL>). The VIIRS and MODIS MUDBs each contain several million records with temporally (± 30 min) and spatially (within 10 km) coincident in situ and satellite-based variables. For these MUDBs, the in situ SST measurements in the records were obtained from the NOAA/NCEI/STAR (Center for Satellite Applications and Research) in situ Quality Control Monitor (Xu and Ignatov 2014). MUDB records eligible for inclusion in the training dataset were required to be from either drifting or moored buoys and located within the L2 VIIRS or MODIS pixel.

Subsets of records for MODIS and VIIRS sensors were randomly selected from the MUDB for use in training and tenfold cross-validation. Selected records were assigned to the “cloud-contaminated” class if the difference between retrieved skin SST and in situ SST was < -1.5 K, after correcting for the median skin–subsurface SST temperature difference of -0.17 K, to produce a zero bias with respect to the subsurface buoy SST. The cloud threshold of -1.5 K was chosen because it is ~ 3 standard deviations from the typically reported uncertainty of 0.5 K for clear-sky, best-quality IR SST retrievals. Any SST retrieval that is more than -1.5 K cooler than the subsurface buoy SST has an almost certain likelihood of being a bad retrieval that is either cloud or aerosol contaminated.

One of the current important challenges in machine learning research is classification under an imbalanced data distribution, that is, where one of the classes considered has many more records than the minority class (es). Oceans are significantly cloudier than clear: the cloud fraction over the ocean is about 72%, with small seasonal variation (Eastman et al. 2011; King et al. 2013). Consequently, the initial ratio of cloudy versus cloud-free records in the selected subset from the MUDB was roughly 3 to 1. This class imbalance has a negative effect on the performance of traditional classification algorithms: predictive accuracy is biased toward the majority class and is also seen to be highly sensitive to data distribution (Gosain and Sardana 2017). There have been many approaches proposed to address the imbalance issue. Some of them involve adding minority records through oversampling or synthetic generation of minority instances, as in the SMOTE oversampling approach (Chawla et al. 2002). Given the very large number of available MUDB records, we addressed the imbalance by resampling the training subset with undersampling, that is, randomly removing cloudy (majority) instances to produce smaller datasets with approximately equal numbers of cloud-contaminated and clear-sky instances. This class balanced dataset was further split into four subsets for classifier training as described below.

Classification models were built for four different conditions: 1) nighttime; 2) daytime and no sun glint contamination, Cox–Munk (1954) glint coefficient ≤ 0.005 ; 3) daytime moderate glint, Cox–Munk glint coefficient between 0.005 and 0.01; and 4) daytime severe glint. The severe glint condition was defined to occur when red ($\lambda = 678$ nm) reflectance > 0.065 and Cox–Munk glint coefficient > 0.01 . The Cox–Munk glint coefficient is the normalized glint radiance for a solar irradiance of unity modeled as a function of solar and satellite viewing geometry and wind speed. The MODIS and VIIRS attributes used to build the classifier for each condition are given in Table 1. They include satellite-retrieved SST, various statistics (e.g., range, standard deviation) for visible and infrared bands within a 5×5 pixel array centered at the pixel being classified, and ancillary information such as time of year, geographic location, satellite viewing angle, and column water vapor content from the four times daily NCEP–DOE Reanalysis 2.

The number of instances in the training set for each of the four conditions is given in Table 2. The probability density functions and the bar charts in Fig. 1 show the spatial and temporal characteristics of the buoy SST distribution, latitude, and month of year of instances in each of the training sets. Note that the instances used during tree fitting for moderate and high glint condition were identical, but the attributes in the models are

TABLE 1. Matchup database attributes evaluated during classifier training for potential significance. An asterisk (*) indicates that the attribute was used during training for each of the four classification models: night, day no glint, day moderate glint, and day high glint.

Attribute label	Attribute description	Night	Day, no glint	Day, moderate glint	Day, high glint
SST2b	Satellite SST		*	*	*
SST – SST4 (MODIS only)	Satellite longwave IR SST minus midwave IR SST	*	—	—	—
dm5.rho678	Uniformity of reflectance at $\lambda = 678$ nm, 5×5 box max – min	—	*	*	—
dm6.rho748	Uniformity of reflectance at $\lambda = 748$ nm, 5×5 box max-min	—	*	*	—
dm9.rho1380	Uniformity of reflectance at $\lambda = 1380$ nm, 5×5 box max – min	—	*	*	—
dm10.rho1610	Uniformity of reflectance at $\lambda = 1610$ nm, 5×5 box max – min	—	*	*	—
dm12	Uniformity of BT at $\lambda = 3.7 \mu\text{m}$, 5×5 box max – min	*	*	*	*
dm14	Uniformity of BT at $\lambda = 8.5 \mu\text{m}$, 5×5 box max – min	*	*	*	*
dm15	Uniformity of BT at $\lambda = 11 \mu\text{m}$, 5×5 box max – min	*	*	*	*
dm16	Uniformity of BT at $\lambda = 12 \mu\text{m}$, 5×5 box max – min	*	*	*	*
dm12m16	Central pixel BT difference, $\lambda = 3.7\text{--}12 \mu\text{m}$	*	*	*	*
dm12m15	Central pixel BT difference, $\lambda = 3.7\text{--}11 \mu\text{m}$	*	*	*	*
dm15m16	Central pixel BT difference, $\lambda = 11\text{--}12 \mu\text{m}$	*	*	*	*
m5.rho678	Central pixel reflectance, $\lambda = 678$ nm	—	*	*	—
m6.rho748	Central pixel reflectance, $\lambda = 768$ nm	—	*	*	—
m9.rho1380	Central pixel reflectance, $\lambda = 1380$ nm	—	*	*	—
m10.rho1610	Central pixel reflectance, $\lambda = 1610$ nm	—	*	*	—
m14	Central pixel BT, $\lambda = 8.5 \mu\text{m}$	*	*	*	*
glint	Glint coefficient	—	*	*	—
satzen	Satellite zenith angle	*	*	*	*
lat	Latitude	*	*	*	*
lon	Longitude	*	*	*	*
mon	Month (1:12)	*	*	*	*
anc.vv	NCEP–DOE Reanalysis 2 water vapor column content	*	*	*	*
sd.m12	Standard deviation $3.7 \mu\text{m}$ BT 5×5 box	*	*	*	*
sd.m15	Standard deviation $11 \mu\text{m}$ BT 5×5 box	*	*	*	*
sd.m16	Standard deviation $12 \mu\text{m}$ BT 5×5 box	*	*	*	*
sd.sst	Standard deviation SST 5×5 box	*	*	*	*
Tdeflong	$(11 \mu\text{m} - 12 \mu\text{m})/11 \mu\text{m}$ BT	*	*	*	*
maxm14	Max BT at $\lambda = 8.5 \mu\text{m}$ 5×5 box	*	*	*	*
maxm10.rho1610	Max reflectance at $\lambda = 1610$ nm 5×5 box	—	*	*	—
maxm9.rho.1380	Max reflectance at $\lambda = 1380$ nm 5×5 box	—	*	*	—
maxm6.rho748	Max reflectance at $\lambda = 748$ nm 5×5 box	—	*	*	—
maxm5.rho678	Max reflectance at $\lambda = 678$ nm 5×5 box	—	*	*	*
minm14	Minimum BT at $\lambda = 8.5 \mu\text{m}$ 5×5 box	*	*	*	*
minm10.rho610	Minimum reflectance at $\lambda = 1610$ nm 5×5 box	—	*	*	—
minm9.rho1380	Minimum reflectance at $\lambda = 1380$ nm 5×5 box	—	*	*	—
minm6.rho748	Minimum reflectance at $\lambda = 748$ nm 5×5 box	—	*	*	—
minm5.rho678	Minimum reflectance at $\lambda = 778$ nm 5×5 box	—	*	*	—

TABLE 2. Number of records in the training datasets for each illumination condition. Records were randomly selected from the L2 SST MUDBs and down sampled to provide training sets with comparable numbers of cloudy and cloud-free records. Classifications models were trained and validated using tenfold cross-validation.

Sensor	Night	Day, no glint	Day, moderate glint	Day, high glint
MODIS-T	250 000	269 976	27 530	27 530
MODIS-A	250 000	694 780	28 366	28 366
VIIRS—SNPP	118 732	49 006	2346	2346

different. In the case of high glint, many of the reflective bands are saturated and cannot be used; therefore, the attributes used during training for the high glint daytime conditions are the same as those attributes used at night.

b. Machine learning software and decision tree classification algorithms

A variety of open source software packages are available for data mining and machine learning. For our study we used WEKA3.7.11 (Hall et al. 2009) a widely used ML software “workbench” developed and maintained by the University of Waikato Environment for Knowledge and Analysis, New Zealand. The two classifier algorithms, BDtrees and ADtrees, were evaluated for each sensor (MODIS on *Terra*, MODIS on *Aqua*, and VIIRS on *SNPP*) and all four conditions. All classifiers were built and validated using tenfold cross-validation. The decision tree algorithms used here selected attributes based on ranking the information gain ratio (Kullback 1959). This ratio provides a measure of the relative entropy or homogeneity of the class outcome for randomly selected variables, relative to observing another variable.

During training the split that results in the most homogeneous daughter node, lowest relative entropy, is selected until the node is either pure or the information gain is 0.

For the BDtree classifiers we used the WEKA BFDTree package, which implements the algorithm of Shi (2007), in what is referred to as a “best first tree approach.” A best-first tree approach selects attributes to split the datasets based on their contribution to the global entropy loss and not just the loss on a particular split. This approach has been shown to be an improvement over the original BDTree algorithm of Breiman et al. (1984), which used a top-down method to evaluate splits based on the loss only along a particular branch.

The ADtree algorithm in the WEKA ADTree package implements the algorithm of Pfahringer et al. (2001) based on the work of Freund and Mason (1999) optimized to include adaptive boosting. Unlike BDtrees with a single starting root node, ADtrees are built as an ensemble of decision nodes and predictions. Each boosting iteration focuses on misclassified prior instances, adding a new layer of decision and prediction nodes to the ADtree at each iteration. In each boosting iteration the data have

FIG. 1. Spatial and temporal characteristics of buoy SST in training sets: (a) PDF of buoy SST 0°C, (b) count of instances by month for each 2°C in temperature, (c) PDF by latitude, and (d) count by month and latitude.

FIG. 2. Schematic of boosted alternating decision tree classifier. Boxes represent the decision rule in each layer and ovals are the weighted score for the prediction. The sign of the score represents the class. The magnitude of the sum of true rules represents the overall confidence in the predicted outcome.

a different distribution, enabling attributes with weaker correlations to the true classification to be reevaluated. These weaker learners are said to be “boosted in importance” and those associations with the greatest entropy loss on the misclassified priors are added to the ensemble of prior stronger classifiers. This method produces a decision tree with alternating layers of predictions and splitters. The base rule at each branch consists of a condition or set of conditions and two prediction nodes, each with a score. The score is a signed real value number that is updated after each iteration by an amount that is proportional to the gradient descent in the mean squared error of the predictions in the layer. A pixel’s final classification outcome is determined by the sign of the weighted majority vote from the cumulative sum of all the layers of true prediction nodes (Fig. 2). The optimal number of boosting iterations was determined to be 15, by evaluating the decrease in the log loss function (de Boer et al. 2005) after

each iteration. Above 15 iterations there was little gain in classifier accuracy, for either class, compared to the expense in model complexity. In this study a negative vote indicates cloud and a positive indicates clear-sky condition and the absolute magnitude of the majority vote provides an estimate of the overall confidence in the prediction.

c. Level 2 and 3 NASA SST products

Standard NASA SST L2 and L3 products, MODIS R2014.0.1 and VIIRS R2016.0, were obtained from the NASA OB.DAAC. These versions of the SST fields relied on BDtrees to identify clouds, and were used as the baseline against which ADtrees Level-3 performance was assessed. To produce ADtree-based cloud masks, we reprocessed MODIS and VIIRS L1b files (the same input used to generate the existing products). We modified the SeaDAS (<https://seadas.gsfc.nasa.gov/>) L2 gen routine by replacing the operational MODIS

TABLE 3. VIIRS: Classifier tenfold cross-validation statistics for ADtree and BDtree methods applied to subsets of VIIRS matchups. Table provides both the overall percent of correctly and incorrectly classified instances, regardless of class, and the conditional rates for TP = true positive, FP = false positive, PRC = precision/recall curve based on the confusion matrix.

VIIRS Classifier	Night		Day, nonglint		Day, moderate glint		Day, high glint	
	ADtree	BDtree	ADtree	BDtree	ADtree	BDtree	ADtree	BDtree
% correctly classified	89.83	89.56	91.74	91.49	93.31	91.20	88.40	86.57
% misclassified	10.10	10.40	8.26	8.50	6.69	8.70	11.50	13.42
TP cloud	0.87	0.87	0.90	0.92	0.93	0.96	0.85	0.92
TP clear	0.96	0.92	0.93	0.91	0.94	0.87	0.92	0.81
FP cloud	0.04	0.08	0.07	0.09	0.06	0.13	0.07	0.19
FP clear	0.13	0.13	0.10	0.08	0.07	0.04	0.16	0.81
PRC cloud	0.96	0.93	0.97	0.95	0.98	0.87	0.96	0.87
PRC clear	0.95	0.89	0.97	0.95	0.97	0.93	0.93	0.92

TABLE 4. MODIS on *Aqua*: Classifier tenfold cross validation statistics for ADtree and BDTree MODIS *Aqua*. Table provides both the overall percent of correctly and incorrectly classified instances, regardless of class, and the conditional rates for TP = true positive, FP = false positive, PRC = precision/recall curve based on the confusion matrix.

MODIS <i>Aqua</i>	Night		Day, nonglint		Day, moderate glint		Day, high glint	
Classifier	ADtree	BDtree	ADtree	BDtree	ADtree	BDtree	ADtree	BDtree
% correctly classified	89.94	88.24	92.07	88.24	91.43	88.24	89.14	88.24
% misclassified	10.70	11.75	7.94	11.75	8.58	11.75	10.10	11.75
TP cloud	0.89	0.86	0.93	0.86	0.91	0.86	0.87	0.86
TP clear	0.89	0.90	0.92	0.90	0.91	0.90	0.93	0.90
FP cloud	0.10	0.10	0.09	0.10	0.09	0.10	0.07	0.10
FP clear	0.11	0.14	0.07	0.14	0.09	0.14	0.13	0.14
PRC cloud	0.96	0.95	0.97	0.95	0.97	0.95	0.96	0.95
PRC clear	0.96	0.95	0.97	0.95	0.97	0.95	0.95	0.95

BDtrees with the ADtree classifiers shown in the [appendix](#) for each sensor.

3. Results

a. Classifier performance comparisons on training datasets

For each classification model, performance on the training data was evaluated using tenfold cross-validation.

The ADtree classifiers outperformed BDtree classifiers when applied to VIIRS and MODIS SST matchups. Model performance metrics for the training data are shown in [Tables 3](#) and [4](#) for VIIRS and MODIS *Aqua*, respectively. Under all conditions, the ADtrees showed a slightly higher percentage of overall correctly classified records and a reduction in the rate of false positives (clear pixels identified as cloudy), particularly in glint regions. For VIIRS there was a 6–10 percentage point reduction in the rate of false positives for cloud. A good metric for assessing classifier success is the precision/recall curve (PRC) ([Saito and Rehmsmeier 2015](#)). Precision in our study represents the probability that an instance classified as cloud-contaminated really is cloudy, and recall (also known as sensitivity) is a measure of the classifier's ability to actually detect a cloud-contaminated instance. An ideal classifier has

a PRC ratio of 1, meaning a clear-sky pixel is never misclassified as cloud and every cloud-contaminated pixel is detected. The cross-validated PRC values for the ADtrees are several percentage points higher than for BDtrees, with the greatest improvement occurring in daylight under moderate to high glint condition. The PRC values for MODIS ADtrees are also very high and similar to those obtained for VIIRS, but there is a higher rate of false positives for clouds compared to the ADtrees for VIIRS. Values obtained for MODIS on *Terra* (not shown) are very similar to those of MODIS on *Aqua*. This difference in the false positive rate between the sensors may be due to the higher spatial resolution of VIIRS: 750 m at nadir compared to 1 km for MODIS, and reduction in the cross-scan pixel size growth as the scan moves away from nadir through pixel aggregation ([Schueler et al. 2013](#)). A complete listing of the individual tests for each of the four ADtrees for both VIIRS and MODIS SST products is presented in an [appendix](#) and documented on the NASA OB-DAAC SST Algorithm Theoretical Basis (ATBD) web pages (<https://oceancolor.gsfc.nasa.gov/atbd/sst/>).

b. Classifiers impact on L2 retrieval quality and retrieved SST distributions

Any decrease in a cloud classifier's sensitivity has the potential to negatively impact the overall quality of the

TABLE 5. Impact of ADtree and BDtree cloud classification on nighttime SST (NSST) accuracy and uncertainty at L2. SST validation statistics are based on the 2014 MUDB of MODIS and VIIRS sensors. For VIIRS BDtree validation the MUDB was extracted from a nonoperations Miami system with significantly fewer overall targets than the current operational GSFC produced MUDB for MODIS and VIIRS. RSD: robust standard deviation.

2014 night NSST matchups sensor SST skin–buoy SST subsurface						
Sensor	<i>Aqua</i>		<i>Terra</i>		VIIRS	
Method	BDtree	ADtree	BDtree	ADtree	BDtree	ADtree
Median	−0.280	−0.280	−0.275	−0.265	−0.214	−0.165
RSD	0.294	0.294	0.327	0.326	0.372	0.245
Count	172 138	173 230	186 010	187 576	90 403*	176 659

TABLE 6. Impact of ADtree and BDtree cloud classification on daytime SST product quality. SST validation statistics based on the 2014 MUDB of MODIS and VIIRS sensors. During the day only MUDB records with wind speeds $> 6 \text{ m s}^{-1}$ are included due to the potential for significant solar heating of the skin layer at lower wind speeds. For VIIRS BDtree validation the MUDB was extracted from an earlier Miami system with significantly fewer overall targets than the currently available operational GSFC produced MUDB for MODIS and VIIRS.

2014 daytime SST matchups sensor SST skin–buoy SST subsurface						
Sensor	<i>Aqua</i>		<i>Terra</i>		VIIRS	
Method	BDtree	ADtree	BDtree	ADtree	BDtree*	ADtree
Median	−0.170	−0.210	−0.250	0.210	−0.163	−0.125
RSD	0.297	0.313	0.280	0.256	0.315	0.250
Count	114 257	142 107	128 477	178 544	32 031*	151 390

SST products. Comparisons of L2 retrieval accuracy and uncertainty relative to subsurface buoys for the two methods are given for night SST (NSST) in Table 5 and daytime SST in Table 6, for best-quality records in the 2014 MUDB. These records were not used during training and come from a period when all three sensors were on orbit. No significant difference in the bias or the uncertainty of retrievals, day or night, is present between the two methods. The distribution of L2 SST retrievals in the MUDB for best quality (Fig. 3) shows a significant increase in the proportion of daytime SSTs $< 10^\circ\text{C}$ for the ADtree method. In addition, the relative shape of the SST distributions between day and night are more similar to each other for the ADtree classifier, and more representative of the actual distribution of the in situ buoy SSTs (Fig. 1).

c. Comparisons of classifier performance on L2 and L3 images

Inspection of L2 images indicates that the ADtrees improved discrimination of clouds near ocean thermal fronts such as the edge of the Gulf Stream (Fig. 4). The bottom panels used the operational BDtree, and the strong thermal gradients along the Gulf Stream's edge are flagged as cloudy, due to the dominant influence of strict spatial homogeneity tests. In contrast, the L2 images in the upper panels using ADtrees show a reduction in the misclassification (false positive) of SST fronts as cloud, particularly at night, although the misclassification is not entirely eliminated. In general clouds identified by the ADtree method appear to be slightly smaller, and the ADtrees may be better at discerning the location of a cloud's edge, although this is hard to quantify. Note that the horizontal white lines on the sides of the VIIRS images result from on-board along-scan pixel deletion, done to avoid transmitting redundant data caused by the along-track growth in pixel size resulting from an increased path length bow-tie effect (Gladkova et al. 2016). Shown in Fig. 5 is the ADtree cumulative per-pixel vote for the ensemble of classifiers applied to the same daytime VIIRS granule shown in the top left panel of Fig. 4. The edges of clouds, seen as red pixels outlining green and blue

regions, and the brighter red lines collocating along the strong SST front the Gulf Stream are only slightly positive, indicating lower confidence in the ADtree clear sky prediction. These lower confidence but positive pixels are examples where a group of the weaker nodes, voting together as a block, was able to identify clear-sky conditions that the operational BDtree misclassified as cloud. Inspection of global L3 images classified using the operational BDtrees (Fig. 6, left panel) and ADtrees (Fig. 6, right panel) indicate that the better performance of the ADtree, as well as the gain in daytime clear-sky SSTs for VIIRS and MODIS (Fig. 7), occurs primarily in higher latitudes for both day and night (not shown). Globally, individual cloud sizes (“data gaps”) appear slightly smaller, and the overall clear pixel density is greater in many areas using the ADtrees. For daytime images, there is a significant ($\sim 35\%$) increase in the count of cloud free grid cells with a quality level of “good” (NASA QL = 1) and “best” (NASA QL = 0). Retrievals with a quality less than 2 in NASA products are considered valid. At night, the gain in coverage is much more modest, and at $\sim 6\%$ is not easy to discern based solely on visual inspection.

To evaluate any seasonal differences in the number of retrievals between the two methods several days

FIG. 3. Comparison density functions L2 sensor SST and NSST retrievals for BDtree and ADtree classified best-quality retrievals for MODIS and VIIRS present in the MUDB. The distribution of best-quality ADtree classified retrievals has a similar SST distribution between day and night.

FIG. 4. Comparison of cloud mask methods for the Gulf Stream on 19 Jun 2014, Level 2 (left) daytime VIIRS SST_{skin} and (right) nighttime $NSST_{triple}$. White areas indicate pixels identified as cloudy; black indicates land. (top) The ADtree classification clouds are more compact and there is improved retention of clear pixels at the high gradient edges of the Gulf Stream compared to (bottom) the BDTree approach. The white horizontal lines show truncated scan lines resulting from bow-tie effect; when the image is mapped to a geographic projection, these data are taken from adjacent lines (Gladkova et al. 2016).

each year near the annual solstices and equinoxes were processed for every fourth year over the life of the missions, (2000, 2004, 2008, 2012, 2016) and visually examined. Examples for MODIS on *Aqua* L3 4-km daytime maps of the number of good- and best-quality L2 SST retrievals per bin, and the histograms of the SST distributions for both classification methods are shown in Figs. 7 and 8. The nighttime NSST distributions are shown in Fig. 9 and global statistics for the number of valid L3 bins, and total number of 1-km L2 retrievals in the L3 maps are presented in Tables 7 and 8.

4. Discussion

Long-term environmental monitoring and input to weather and climate forecasting models require that both modes of classification error [false positive (FP) cloud and FP clear] be low in derived SST fields. A

highly conservative L2 cloud mask with a low rate of FP clear, while providing highly accurate low uncertainty retrievals, may at the same time greatly undersample the geophysical SST field. Most of the validation efforts and even the benchmark requirements for SST climate data records (CDR) have focused primarily on L2 fields. Some of the sampling errors and differences in cloud persistence reported for MODIS L3 SST fields (Liu 2016; Liu and Minnett 2016) are supported by the results in this study as being a result of BDtrees having a higher rate of FP cloud when applied to NASA IR SST products.

The increases in daytime global coverage at L3 using ADtree classification models is supported by the cross-validation results from the training datasets in Tables 3 and 4. The ADtree classifier outperformed the BDtree on the same training set by several percentage points in glint conditions compared to nonglint or at night. Glint regions can encompass up to 1/3 of any daytime granule

FIG. 5. VIIRS daytime L2 1-km pixel ensemble ADtree majority vote. (top left) The per pixel vote for full daytime granule shown in the top-left panel of Fig. 4. The white box marks a region of the highly dynamic Gulf Stream. (top right) That region shown enlarged. The magnitude of the vote indicates the confidence in a pixel's classification while the sign of the vote indicated cloud/negative or clear/positive. The horizontal lines are truncated scan lines resulting from the bow-tie effect; when the image is mapped to a geographic projection, these data are taken from adjacent lines (Gladkova et al. 2016). (bottom) The red-green-blue (RGB) true color image for the same granule. The red spot is where one of the RGB bands is saturated.

depending on the granule's geographic distance from the center of the sun glitter. The BDtree with a higher rate of false positive for clouds would significantly impact the daytime persistence of “false” clouds in long-term SST trend analysis. The ADtree classifiers were able to increase the number of clear-sky valid L2 retrievals without sacrificing retrieval SST accuracy or uncertainty relative to in situ buoy SST measurements (Table 6).

Based on ADtree reprocessed L2 and L3 imagery, for dates near the solstices and equinoxes, the total increase in the number of L2 valid retrievals during the day ranged between 30% and 60%, relative to BDtree standard NASA products. The largest gains in L2 counts occurred in June and September when the day length grows and the Arctic becomes more ice free (Table 7 and Figs. 7 and 8). This large increase in the total global daytime valid L2 count is, however,

FIG. 6. Comparison of VIIRS SST_{skin} data coverage for daytime retrievals, 4-km maps for 19 Jun 2014 good or better quality, using different cloud mask classification models: (left) cloud identification based on a BDtree and (right) cloud identification based on an ensemble of ADtree models. White indicates areas identified as cloudy. The use of an ADtree algorithm significantly increases the number of valid daytime retrievals everywhere, with the largest gains occurring in the mid- to high latitudes between 30° to 60°.

biased high as a result of the Arctic's repeated sampling view from overlapping orbits, combined with the low sun angles during Northern Hemisphere summer and early fall. For the nonoverlapping portions of the orbit, the L2 count per bin is not uniform across the scanline for either method. The highest counts are seen in the center of the swath, particularly in glint regions, where the ADtree approaches 30 counts per bin compared to 20 for the BDtree. For MODIS there is little difference in the counts per bin at the higher scan angles between the two methods. At the swath edges both classifiers have ~4 to 6 retrievals per bin, which we assume is primarily due to the larger geographical pixel size at larger satellite zenith angle, but cannot rule out the increased path length resulting in potentially more clouds in the field of view. Irrespective of the L2 counts, the number of valid populated 4-km grid cells increased by ~35%, with most of the coverage increase occurring at latitudes higher than 40° in both hemispheres in opposite seasons. Distributions of daytime retrieved SST in L3 mapped products (Fig. 8) show ADtree classifiers identify a higher proportion of temperatures < 15°C, in addition to increasing the frequency at all temperatures. Furthermore, as was seen with the training data the shape of the L3 daytime ADtree SST distribution is more similar to the distribution present in the MUDB (Fig. 1). In contrast, the L3 daytime BDtree distribution is skewed toward higher temperatures relative to in situ observations. These distribution differences indicate that the BDtree daytime classification error is often regional. At night the ADtree classifier only modestly increased the number of valid NSST L2 retrievals and populated L3

grid cells (Table 8) ranging between 2% and 8%. The large daytime regional increase in the number of retrievals in the Arctic would not be expected in the ADtree NSST products, as the sun is often above the horizon during the ice-free periods.

At L2, ADtrees, by determining not only a prediction but also a measure of the confidence in the prediction (Fig. 5), provide valuable information not available with BDtrees. NASA IR SST products have a variety of end users, analyzing nearly two decades of historical data and real-time data for many of purposes, and applications may have different tolerances in regard to the absolute accuracy and uncertainty of the SST product. For applications such as front detection or change in the frequency of fronts, absolute accuracy may be less of a concern than confidently detecting horizontal temperature gradients. Having a per-pixel measure of prediction confidence would provide users with the ability to customize the cloud mask for a particular application.

5. Summary

A new, improved cloud classification methodology, based on an ensemble of ADtree classifiers, was developed for application to NASA IR SST products from MODIS and VIIRS sensors. ADtree classifiers represent an ensemble of both strong and weak classifiers. In some situations, a block of weaker nodes, when voting together for the same outcome, have the ability to modify or override the vote of a single strong prediction node. A collective weighted vote approach is a very different strategy than the BDtree used in NASA operational SST products for MODIS, and

FIG. 7. Seasonal daytime L3 maps of the number of clear good quality L2 pixels in L3 daily SST 4-km bins for MODIS on *Aqua* for (left) ADtree and (right) BDtree. Dates are near the equinoxes and solstices in 2008. The ADtree increases both the total number of populated L3 4-km bins and the total number of L2 retrievals.

TABLE 7. Daytime seasonal comparisons of the number of SST good-quality retrievals in L2 and L3 global products. Number of observations of good quality populated 4-km L3 grid cells and L2 1-km pixels for MODIS on *Aqua* 2008 for dates near the equinoxes and solstices for the BDtree and ADtree methods.

Day SST MODIS-A 2008 date	Classifier	No. of clear 4-km L3 bins	% global increase in populated L3 4-km bins ADtree	Mean L2 pixel count per 4-km bin	Sum global 1-km L2 count	% global increase daily clear L2 1-km pixels ADtree
20 Mar	BDtree	3.88×10^6		9.1	3.82×10^7	
20 Mar	ADtree	5.27×10^6	35.7	9.8	4.98×10^7	30.6
19 Jun	BDtree	3.57×10^6		8.7	3.10×10^7	
19 Jun	ADtree	5.59×10^6	56.8	9.2	5.15×10^7	66.3
23 Sep	BDtree	3.83×10^6		8.5	3.24×10^7	
23 Sep	ADtree	5.72×10^6	49.4	9.3	5.31×10^7	64.0
21 Dec	BDtree	3.83×10^6		8.8	3.36×10^7	
21 Dec	ADtree	5.20×10^6	35.7	9.1	4.75×10^7	41.5

historically was used for NOAA/NASA AVHRR Pathfinder SST, where a pixel can only follow a single path to a terminal node (Kilpatrick et al. 2001). The path along dominant strong nodes often relies on strict uniformity tests, and there is the underlying assumption that the rule is valid everywhere seasonally and geographically.

Validation of ADtree performance based on tenfold cross-validation demonstrated improved PRC values between 3 and 6 percentage points, and reduction in the misclassification rate for false positive for cloud compared to BDtrees. ADtrees applied to L2 granules, around solstices and equinox dates, increased global coverage of MODIS and VIIRS SSTs in L3 4-km daily global maps by $\sim 2\%$ – 10% at night and 35% in the daytime, depending on the location and season. The most significant increases occur at higher latitudes, during summer and fall, and in the glint-contaminated region of the swath. The ADtree classifiers increased global coverage and daily counts of L2 retrievals while still maintaining the same accuracy and uncertainty of retrievals at L2 relative to in situ SST observations in the MUDb. This indicates that the BDtrees are overly conservative, and there are significant numbers of FP for clouds in

the current R2014.1 NASA MODIS IR SST products. The BDtree daytime FP cloud rate impacts the temperature distribution of globally retrieved values, which are skewed to warmer temperatures compared to in situ buoy observations, indicating significant regional differences in the misclassification errors. At night the ADtree slightly increase the number of retrievals across the entire SST domain; however, unlike in the daytime the shape of the temperature distribution for both ADtree and the BDtree are similar to the in situ observations, indicating that the night misclassification errors are more randomly distributed than those in the daytime. Qualitatively the ADtrees also appear to improve the retention of valid L2 pixels in high-gradient SST frontal regions such as the Gulf Stream and pixels near cloud edges, although the confidence in the weighted vote for the classification outcome along these edges is lower.

The ADtree methodology presented here is scheduled to be implemented into operational NASA MODIS SST standard processing in 2019, and reprocessing is planned for continuity in both the cloud mask methodology and the SST retrieval algorithm for NASA VIIRS and MODIS IR SST products.

TABLE 8. Nighttime seasonal comparisons of the number of SST (NSST) good-quality retrievals in L2 and L3 global products. Number of observations of good quality populated 4-km L3 grid cells and L2 1-km pixels for MODIS on *Aqua* 2008 for dates near the equinoxes and solstices for the BDtree and ADtree methods.

Night NSST MODIS-A 2008 date	Classifier	No. of clear 4-km L3 bins	% global increase in populated L3 4-km bins ADtree	Mean L2 pixel count per 4-km bin	Sum global 1-km L2 count	% global increase in daily 1-km L2 pixels ADtree
20 Mar	BDtree	3.94×10^6		7.5	3.14×10^7	
20 Mar	ADtree	4.16×10^6	5.7	7.8	3.23×10^7	3.1
19 Jun	BDtree	3.94×10^6		6.9	2.87×10^7	
19 Jun	ADtree	4.04×10^6	2.3	7.5	3.08×10^7	7.2
23 Sep	BDtree	3.94×10^6		7.3	2.87×10^7	
23 Sep	ADtree	4.16×10^6	5.5	7.7	3.41×10^7	18.9
21 Dec	BDtree	3.71×10^6		7.1	2.59×10^7	
21 Dec	ADtree	3.80×10^6	2.4	7.7	2.84×10^7	9.5

FIG. 8. Seasonal daytime frequencies of SST values in an L3 daily SST 4-km product from MODIS on *Aqua* for (left) ADtree and (right) BDtree. Dates are near the equinoxes and solstices in 2008. The ADtree method increases both the total number of populated bins and the proportion of cooler SST values.

FIG. 9. Seasonal nighttime frequency of NSST values in an L3 daily NSST 4-km product from MODIS on *Aqua* for (left) ADtree and (right) BDtree. Dates are near the equinoxes and solstices in 2008. The ADtree method increases both the total number of populated bins and the proportion of cooler SST values.

Acknowledgments. This study was supported by the NASA Earth Science Physical Oceanography Program, (NNX11AK88G and NNX14AP79A).

APPENDIX

Listing of ADtree Classifiers for VIIRS and MODIS

Below is a listing of each of the four classification ADtree models for VIIRS and both MODIS sensors. Tests and nested pairs are numbered and “|” indicates the hierarchical level of a test.

If a test is true, the value after the colon is added to the sum and forms the cumulative vote for a pixel. The sign of the vote indicates the predicted class, negative is cloud and positive is clear, and the magnitude indicates the confidence of the prediction. See Table 1 for a description of test attribute names and meanings.

a. VIIRS

1) VIIRS DAY NO GLINT ALTERNATING DECISION TREE

```
: 0
| (1)m10.rho1610 < 0.16: 0.805
| | (2)m6.rho748 < 0.062: 0.393
| | | (3)m9.rho1380 < 0.004: 0.287
| | | | (9)Tdeflong < 0.002: -0.681
| | | | (9)Tdeflong >= 0.002: 0.026
| | | | | (13)m6.rho748 < 0.039: 0.364
| | | | | (13)m6.rho748 >= 0.039: -0.21
| | | (3)m9.rho1380 >= 0.004: -1.244
| | (2)m6.rho748 >= 0.062: -0.572
| | | (5)minm10.rho610 < 0.032: 0.455
| | | (5)minm10.rho610 >= 0.032: -0.395
| | (4)satz < 64.994: 0.216
| | | (8)m9.rho1380 < 0.007: 0.065
| | | (8)m9.rho1380 >= 0.007: -1.077
| | (4)satz >= 64.994: -0.708
| (1)m10.rho1610 >= 0.16: -1.755
| | (6)m10.rho1610 < 0.266: 0.642
| | (6)m10.rho1610 >= 0.266: -0.19
| | | (14)dm5.rho678 < 0.103: 0.425
| | | (14)dm5.rho678 >= 0.103: -0.195
| | (10)dm15m16 < 0.235: -0.189
| | (10)dm15m16 >= 0.235: 0.411
| | (15)anc.wv < 2.946: 0.038
| | (15)anc.wv >= 2.946: -1.137
| (7)dm15 < 0.762: 0.156
| (7)dm15 >= 0.762: -0.188
| (11)anc.wv < 1.315: 0.327
| (11)anc.wv >= 1.315: -0.054
```

```
|| (12)sst2b < 278.171: -0.679
|| (12)sst2b >= 278.171: 0.05
Tree size (total number of nodes): 46
Leaves (number of predictor nodes): 31
```

2) VIIRS DAY MODERATE GLINT ALTERNATING DECISION TREE

```
: 0
| (1)minm6.rho748 < 0.104: 0.91
| | (2)m5.rho678 < 0.086: 0.518
| | | (7)minm5.rho678 < 0.067: 0.558
| | | (7)minm5.rho678 >= 0.067: -0.263
| | | | (9)m10.rho1610 < 0.06: -0.231
| | | | (9)m10.rho1610 >= 0.06: 1.712
| | | | (13)m10.rho1610 < 0.046: -1.353
| | | | (13)m10.rho1610 >= 0.046: 0.352
| | (2)m5.rho678 >= 0.086: -0.585
| | | (6)dm12m16 < 12.951: -0.905
| | | (6)dm12m16 >= 12.951: 0.187
| | | | (8)m5.rho678 < 0.098: 0.549
| | | | (8)m5.rho678 >= 0.098: -0.484
| (1)minm6.rho748 >= 0.104: -1.819
| | (4)minm5.rho678 < 0.206: 0.467
| | (4)minm5.rho678 >= 0.206: -1.18
| | | (5)dm5.rho678 < 0.037: 1.747
| | | (5)dm5.rho678 >= 0.037: -1.79
| | (14)anc.wv < 1.705: 0.434
| | (14)anc.wv >= 1.705: -0.645
| (3)m9.rho1380 < 0.002: 0.23
| | (10)lat < 32.13: -0.067
| | | (11)sst2b < 300.034: -0.146
| | | (11)sst2b >= 300.034: 0.824
| | (10)lat >= 32.13: 0.758
| (3)m9.rho1380 >= 0.002: -1.153
| | (12)minm9.rho1380 < 0.005: 0.28
| | (12)minm9.rho1380 >= 0.005: -0.939
| (15)satz < 33.204: 0.2
| (15)satz >= 33.204: -0.331
Tree size (total number of nodes): 46
Leaves (number of predictor nodes): 31
```

3) VIIRS DAY HIGH GLINT ALTERNATING DECISION TREE

```
: 0
| (1)minm14 < 287.451: -0.812
| | (2)lat < 32.315: -0.296
| | | (3)Tdeflong < 0.001: -1.109
| | | | (9)lat < -30: 0.525
| | | | (9)lat >= -30: -1.827
| | | (3)Tdeflong >= 0.001: 0.44
| | | (4)anc.wv < 2.065: 0.49
```

```

|| (4) anc.wv >= 2.065: -1.104
| (2) lat >= 32.315: 0.669
|| (5) dm12m15 < 17.594: 0.452
|| (5) dm12m15 >= 17.594: -0.76
| (1) minm14 >= 287.451: 0.858
|| (7) lon < -69.475: 0.512
|| (7) lon >= -69.475: -0.176
||| (10) lat < 1.01: 0.64
||| (10) lat >= 1.01: -0.176
||| (12) mon < 5.5: 0.69
||| (12) mon >= 5.5: -0.305
||| (15) lat < 22.465: -0.345
||| (15) lat >= 22.465: 0.515
|| (8) dm12m15 < 9.655: 0.562
|| (8) dm12m15 >= 9.655: -0.173
|| (11) m14 < 290.343: -0.39
|| (11) m14 >= 290.343: 0.243
||| (13) Tdeflong < 0.003: -0.817
||| (13) Tdeflong >= 0.003: 0.267
| (6) dm15 < 1.189: 0.059
| (6) dm15 >= 1.189: -1.22
| (14) sd.sst < 0.342: 0.119
| (14) sd.sst >= 0.342: -0.509
Tree size (total number of nodes): 46
Leaves (number of predictor nodes): 31

```

4) VIIRS NIGHT ALTERNATING DECISION TREE

```

: 0.385
| (1) 3.7um - 11um BT < 0.111: -1.942
| (1) 3.7um - 11um BT >= 0.111: 0.469
|| (2) sst2b - sst3b < -0.486: -1.029
||| (7) sst2b - sst3b < -1.013: -0.653
||| (10) 3.7um - 11um BT < 1.477: 0.54
||| (10) 3.7um - 11um BT >= 1.477: -0.789
||| (7) sst2b - sst3b >= -1.013: 0.298
|| (2) sst2b - sst3b >= -0.486: 0.286
||| (4) sst3b < 292.19: -0.316
||| (4) sst3b >= 292.19: 0.245
||| (5) sst2b - sst3b < 0.985: 0.082
||| (6) sst - sst3b < -0.191: -0.43
||| (6) sst - sst3b >= -0.191: 0.197
||| (9) 3.7um - 11um BT < 0.313: -0.701
||| (9) 3.7um - 11um BT >= 0.313: 0.087
|| (5) sst2b - sst3b >= 0.985: -1.084
| (3) sst2b < 270.294K: -3.823
| (3) sst2b >= 270.294K: 0.048
| (8) sd 3.7um BT 5 X 5 box < 0.247: 0.165
| (8) sd 3.7um BT 5 X 5 box >= 0.247: -0.177
Tree size (total number of nodes): 31
Leaves (number of predictor nodes): 21

```

b. MODIS Aqua

1) MODIS-A DAY NO GLINT ALTERNATING DECISION TREE

```

: 0
| (1) cen.rho678 < 0.204: 0.982
|| (4) cen.rho678 < 0.056: 0.328
||| (7) max.rho1380 < 0.006: 0.151
||| (10) satz < 59.043: 0.029
||| (10) satz >= 59.043: -0.618
||| (7) max.rho1380 >= 0.006: -1.071
|| (4) cen.rho678 >= 0.056: -0.565
| (1) cen.rho678 >= 0.204: -1.325
|| (5) min.rho678 < 0.251: 0.81
|| (5) min.rho678 >= 0.251: -0.021
||| (11) lat < -48.685: 0.605
||| (11) lat >= -48.685: -0.088
||| (13) lat < 37.755: -0.311
||| (13) lat >= 37.755: 0.189
||| (14) d.11000m12000 < 0.752: -0.091
||| (14) d.11000m12000 >= 0.752: 0.581
|| (12) d.6715m11000 < -38.023: 0.221
|| (12) d.6715m11000 >= -38.023: -0.201
| (2) min.rho1380 < 0.003: 0.22
|| (8) max.rho1380 < 0.003: 0.069
|| (8) max.rho1380 >= 0.003: -0.422
| (2) min.rho1380 >= 0.003: -0.891
| (3) d.3750m11000 < 3.577: 0.719
| (3) d.3750m11000 >= 3.577: -0.227
|| (9) sst2b < 27.668: -0.017
|| (9) sst2b >= 27.668: 0.472
|| (15) sd.sst < 0.481: 0.094
|| (15) sd.sst >= 0.481: -0.322
| (6) sd.sst < 0.263: 0.193
| (6) sd.sst >= 0.263: -0.276
Tree size (total number of nodes): 46
Leaves (number of predictor nodes): 31

```

2) MODIS-A DAY MODERATE GLINT ALTERNATING DECISION TREE

```

: 0
| (1) min.12000 < 15.119: -0.759
|| (3) d.3750m11000 < 12.733: 0.652
||| (12) min.rho1380 < -0.001: -1.096
||| (12) min.rho1380 >= -0.001: 0.13
|| (3) d.3750m11000 >= 12.733: -0.288
|| (9) dmm.3750 < 0.529: 0.926
|| (9) dmm.3750 >= 0.529: 0.026
| (1) min.12000 >= 15.119: 0.794

```

```

| | (8)d.3750m11000 < 8.37: 0.522
| | (8)d.3750m11000 >= 8.37: -0.268
| (2)dmm.rho1380 < 0.001: 0.381
| | (13)max.rho1380 < 0.002: 0.006
| | (13)max.rho1380 >= 0.002: -0.584
| (2)dmm.rho1380 >= 0.001: -0.619
| (4)sd.sst < 0.397: 0.101
| (4)sd.sst >= 0.397: -0.684
| (5)Tdeflong < 0.036: -0.471
| | (7)lat < 32.335: -0.481
| | (7)lat >= 32.335: 0.684
| (5)Tdeflong >= 0.036: 0.193
| (6)min.rho1380 < 0.003: 0.108
| | (11)lat < -25.425: 0.517
| | (11)lat >= -25.425: -0.082
| | | (14)lat < 27.215: -0.077
| | | | (15)cen.11000 < 14.734: -2.012
| | | | (15)cen.11000 >= 14.734: 0.089
| | | (14)lat >= 27.215: 0.383
| (6)min.rho1380 >= 0.003: -0.913
| (10)sst2b < 27.583: -0.074
| (10)sst2b >= 27.583: 0.691
Tree size (total number of nodes): 46
Leaves (number of predictor nodes): 31

```

3) MODIS-A HIGH GLINT ALTERNATING DECISION TREE

```

: 0
| (1)min.12000 < 15.119: -0.759
| (1)min.12000 >= 15.119: 0.794
| (2)sd.sst < 0.254: 0.304
| | (8)sd.sst < 0.14: 0.237
| | (8)sd.sst >= 0.14: -0.276
| | (10)lat < -4.225: 0.401
| | (10)lat >= -4.225: -0.144
| (2)sd.sst >= 0.254: -0.508
| (3)d.11000m12000 < 0.303: -0.715
| (3)d.11000m12000 >= 0.303: 0.184
| | (5)lat < -23.905: 0.662
| | (5)lat >= -23.905: -0.087
| | (7)sd.sst < 0.494: 0.114
| | (7)sd.sst >= 0.494: -0.55
| (4)lat < 32.415: -0.137
| | (6)sst2b < 26.408: -0.245
| | (6)sst2b >= 26.408: 0.499
| | (9)Tdeflong < 0.038: -0.415
| | (9)Tdeflong >= 0.038: 0.128
| (4)lat >= 32.415: 0.506
Tree size (total number of nodes): 31
Leaves (number of predictor nodes): 21

```

4) MODIS-A NIGHT ALTERNATING DECISION TREE

```

: 0
| (1)d.3750m12000 < 0.117: -1.279
| | (10)lat < 33.035: -0.289
| | | (11)lat < -42.235: 0.747
| | | (11)lat >= -42.235: -0.564
| | | | (14)cen.3750 < 13.117: -0.58
| | | | (14)cen.3750 >= 13.117: 0.638
| | (10)lat >= 33.035: 0.355
| | (12)cen.3750 < 9.447: -0.307
| | (12)cen.3750 >= 9.447: 0.747
| (1)d.3750m12000 >= 0.117: 0.47
| | (2)sd.11000 < 0.155: 0.69
| | | (7)dmm.3750 < 0.524: 0.15
| | | (7)dmm.3750 >= 0.524: -0.794
| | (2)sd.11000 >= 0.155: -0.43
| | (3)sstmsst4 < -0.787: -1.404
| | | (13)sstmsst4 < -1.253: -1.086
| | | (13)sstmsst4 >= -1.253: 0.388
| | (3)sstmsst4 >= -0.787: 0.197
| | | (4)dmm.3750 < 1.229: 0.287
| | | | (8)sstmsst4 < -0.383: -0.531
| | | | (8)sstmsst4 >= -0.383: 0.279
| | | (4)dmm.3750 >= 1.229: -0.497
| | | (5)cen.12000 < 16.353: -0.3
| | | | (6)d.4050m11000 < -2.171: 0.516
| | | | (6)d.4050m11000 >= -2.171: -0.395
| | | | | (9)sst4 < 16.212: -0.694
| | | | | (9)sst4 >= 16.212: 0.392
| | | (5)cen.12000 >= 16.353: 0.451
| | | (15)sstmsst4 < 1.033: 0.049
| | | (15)sstmsst4 >= 1.033: -0.718
Tree size (total number of nodes): 46
Leaves (number of predictor nodes): 31

```

c. MODIS Terra

1) MODIS-T DAY NONGLINT ALTERNATING DECISION TREE

```

0:
| (1)cen.rho678 < 0.206: 0.894
| | (2)cen.rho678 < 0.051: 0.54
| | | (5)satz < 64.915: -0.057
| | | | (9)max.rho1380 < 0.003: 0.087
| | | | | (14)satz < 54.529: 0.03
| | | | | (14)satz >= 54.529: -0.573
| | | | (9)max.rho1380 >= 0.003: -0.828
| | | (5)satz >= 64.915: -2.7

```



```

|| (2) cen.rho678 >= 0.051: -0.49
| | (3) min.rho1380 < 0.004: 0.21
| | | (10) cen.rho678 < 0.073: 0.074
| | | (10) cen.rho678 >= 0.073: -0.467
| | (3) min.rho1380 >= 0.004: -1.008
| (1) cen.rho678 >= 0.206: -1.402
| | (8) min.rho678 < 0.239: 0.821
| | (8) min.rho678 >= 0.239: -0.033
| | | (15) d.6715m11000 < -42.423: 0.283
| | | (15) d.6715m11000 >= -42.423: -0.184
| (4) d.3750m11000 < 3.493: 0.53
| | (11) d.11000m12000 < 0.178: -0.589
| | (11) d.11000m12000 >= 0.178: 0.174
| (4) d.3750m11000 >= 3.493: -0.169
| | (7) min.rho1380 < 0.008: 0.097
| | (7) min.rho1380 >= 0.008: -0.761
| (6) sd.sst < 0.273: 0.258
| (6) sd.sst >= 0.273: -0.194
| (12) sd.sst < 0.119: 0.459
| (12) sd.sst >= 0.119: -0.037
| | (13) sst2b < 27.293: -0.054
| | (13) sst2b >= 27.293: 0.411
Tree size (total number of nodes): 46
Leaves (number of predictor nodes): 31

```

2) MODIS-T DAY MODERATE GLINT ALTERNATING DECISION TREE

```

: 0
| (1) cen.rho678 < 0.098: -1.026
| | (8) cen.rho678 < 0.069: -0.29
| | (8) cen.rho678 >= 0.069: 0.312
| | (10) cen.rho1380 < 0.002: -0.154
| | (10) cen.rho1380 >= 0.002: 0.574
| | (15) max.7325 < -10.303: 0.15
| | (15) max.7325 >= -10.303: -0.393
| (1) cen.rho678 >= 0.098: 0.761
| | (3) cen.11000 < 16.122: 0.538
| | (3) cen.11000 >= 16.122: -0.707
| | (2) max.rho1380 < 0.005: -0.357
| | (5) Tdeflong < 0.037: 0.45
| | | (7) lat < 28.27: 0.383
| | | (7) lat >= 28.27: -0.56
| | (5) Tdeflong >= 0.037: -0.134
| | (12) satz < 34.756: -0.098
| | (12) satz >= 34.756: 0.424
| (2) max.rho1380 >= 0.005: 1.399
| | (9) min.rho1380 < 0.01: -0.326
| | (9) min.rho1380 >= 0.01: 1.404
| (4) sd.sst < 0.382: -0.13
| | (13) lat < -5.005: -0.4
| | (13) lat >= -5.005: 0.107
| | (14) sd.sst < 0.224: -0.182

```

```

|| (14) sd.sst >= 0.224: 0.216
| (4) sd.sst >= 0.382: 0.632
| (6) d.8550m11000 < -1.641: -0.268
| (6) d.8550m11000 >= -1.641: 0.201
| | (11) sst2b < 27.723: 0.069
| | (11) sst2b >= 27.723: -0.903
Legend: -ve = Good, +ve = Bad
Tree size (total number of nodes): 46
Leaves (number of predictor nodes): 31

```

3) MODIS-T DAY HIGH GLINT ALTERNATING DECISION TREE

```

:0
| (1) min.12000 < 15.376: 0.908
| | (4) d.11000m12000 < 0.27: 0.446
| | (4) d.11000m12000 >= 0.27: -0.568
| | (9) min.rho1380 < -0.001: 0.981
| | (9) min.rho1380 >= -0.001: -0.137
| (1) min.12000 >= 15.376: -0.819
| | (5) Tdeflong < 0.031: 0.969
| | (5) Tdeflong >= 0.031: -0.025
| | (8) max.7325 < -9.485: 0.208
| | (8) max.7325 >= -9.485: -0.428
| (2) max.rho1380 < 0.005: -0.235
| | (11) cen.rho1380 < 0.002: -0.1
| | (11) cen.rho1380 >= 0.002: 0.506
| (2) max.rho1380 >= 0.005: 1.176
| (3) sd.sst < 0.36: -0.222
| (3) sd.sst >= 0.36: 0.657
| (6) sst2b < 27.443: 0.077
| | (7) sd.sst < 0.19: -0.355
| | (7) sd.sst >= 0.19: 0.179
| | (10) min.rho1380 < 0.01: -0.022
| | | (12) lat < 29.775: 0.128
| | | | (13) lat < -15.305: -0.431
| | | | (13) lat >= -15.305: 0.202
| | | | (14) max.11000 < 14.722: 1.906
| | | | (14) max.11000 >= 14.722: -0.087
| | | | (15) Tdeflong < 0.04: 0.336
| | | | (15) Tdeflong >= 0.04: -0.156
| | | (12) lat >= 29.775: -0.379
| | (10) min.rho1380 >= 0.01: 1.948
| (6) sst2b >= 27.443: -0.688
Tree size (total number of nodes): 46
Leaves (number of predictor nodes): 31

```

4) MODIS-T NIGHT ALTERNATING DECISION TREE

```

:0
| (1) d.3750m12000 < -0.053: -1.257
| | (11) lat < 33.195: -0.278

```

```

||| (12) lat < -40.815: 0.619
||| (12) lat >= -40.815: -0.711
|||| (15) cen.3750 < 6.477: -3.733
|||| (15) cen.3750 >= 6.477: -0.111
|| (11) lat >= 33.195: 0.333
|| (13) cen.3750 < 9.372: -0.292
|| (13) cen.3750 >= 9.372: 0.764
| (1) d.3750m12000 >= -0.053: 0.43
|| (2) dmm.11000 < 0.486: 0.628
||| (7) sd.4050 < 0.146: 0.177
||| (7) sd.4050 >= 0.146: -0.723
|| (2) dmm.11000 >= 0.486: -0.45
|| (3) sstmsst4 < -0.878: -1.353
||| (10) sstmsst4 < -1.533: -1.439
||| (10) sstmsst4 >= -1.533: 0.346
|| (3) sstmsst4 >= -0.878: 0.219
||| (4) sd.3750 < 0.448: 0.29
|||| (9) sstmsst4 < -0.422: -0.504
|||| (9) sstmsst4 >= -0.422: 0.268
||| (4) sd.3750 >= 0.448: -0.484
||| (5) cen.12000 < 16.736: -0.285
|||| (6) d.4050m11000 < -2.199: 0.518
|||| (6) d.4050m11000 >= -2.199: -0.316
||||| (8) cen.12000 < 11.896: -0.527
||||| (8) cen.12000 >= 11.896: 0.4
||| (5) cen.12000 >= 16.736: 0.5
||| (14) sstmsst4 < 1.183: 0.051
||| (14) sstmsst4 >= 1.183: -0.898
Tree size (total number of nodes): 46
Leaves (number of predictor nodes): 31

```

REFERENCES

- Ackerman, S. A., K. I. Strabala, W. P. Menzel, R. A. Frey, C. C. Moeller, and L. E. Gumley, 1998: Discriminating clear sky from clouds with MODIS. *J. Geophys. Res.*, **103**, 32 141–32 157, <https://doi.org/10.1029/1998JD200032>.
- , R. E. Holz, R. Frey, E. W. Eloranta, B. C. Maddux, and M. McGill, 2008: Cloud detection with MODIS. Part II: Validation. *J. Atmos. Oceanic Technol.*, **25**, 1073–1086, <https://doi.org/10.1175/2007JTECHA1053.1>.
- Altman, D. G., and J. M. Bland, 1994: Diagnostic tests. 1: Sensitivity and specificity. *BMJ*, **308**, 1552, <https://doi.org/10.1136/bmj.308.6943.1552>.
- Barnes, B. B., and C. Hu, 2013: A hybrid cloud detection algorithm to improve MODIS sea surface temperature data quality and coverage over the eastern Gulf of Mexico. *IEEE Trans. Geosci. Remote Sens.*, **51**, 3273–3285, <https://doi.org/10.1109/TGRS.2012.2223217>.
- Bojinski, S., M. Verstraete, T. C. Peterson, C. Richter, A. Simmons, and M. Zemp, 2014: The concept of essential climate variables in support of climate research, applications, and policy. *Bull. Amer. Meteor. Soc.*, **95**, 1431–1443, <https://doi.org/10.1175/BAMS-D-13-00047.1>.
- Breiman, L., J. H. Friedman, R. A. Olshen, and C. J. Stone, 1984: *Classification and Regression Trees*. Chapman & Hall, 368 pp.
- Bulgin, C. E., J. P. D. Mittaz, O. Embury, S. Eastwood, and C. J. Merchant, 2018: Bayesian cloud detection for 37 years of Advanced Very High Resolution Radiometer (AVHRR) global area coverage (GAC) data. *Remote Sens.*, **10**, 97, <https://doi.org/10.3390/rs10010097>.
- Chawla, N. V., K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, 2002: SMOTE: Synthetic Minority Over-Sampling Technique. *J. Artif. Intell. Res.*, **16**, 321–357, <https://doi.org/10.1613/jair.953>.
- Cox, C., and W. Munk, 1954: Measurements of the roughness of the sea surface from photographs of the sun's glitter. *J. Opt. Soc. Amer.*, **44**, 838–850, <https://doi.org/10.1364/JOSA.44.000838>.
- de Boer, P. T., D. P. Kroese, S. Mannor, and R. Y. Rubinstein, 2005: A tutorial on the cross-entropy method. *Ann. Oper. Res.*, **134**, 19–67, <https://doi.org/10.1007/s10479-005-5724-z>.
- Eastman, R., S. G. Warren, and C. J. Hahn, 2011: Variations in cloud cover and cloud types over the ocean from surface observations, 1954–2008. *J. Climate*, **24**, 5914–5934, <https://doi.org/10.1175/2011JCLI3972.1>.
- Esaias, W. E., and Coauthors, 1998: An overview of MODIS capabilities for ocean science observations. *IEEE Trans. Geosci. Remote Sens.*, **36**, 1250–1265, <https://doi.org/10.1109/36.701076>.
- Freund, Y., and L. Mason, 1999: The alternating decision tree learning algorithm. *Proc. 16th Int. Conf. on Machine Learning*, Bled, Slovenia, International Machine Learning Society, 124–133, <http://perun.pmf.uns.ac.rs/radovanovic/dmsem/cd/install/Weka/doc/classifiers-papers/trees/ADTree/atrees.pdf>.
- Frey, R. A., and Coauthors, 2008: Cloud detection with MODIS. Part I: Improvements in the MODIS cloud mask for Collection 5. *J. Atmos. Oceanic Technol.*, **25**, 1057–1072, <https://doi.org/10.1175/2008JTECHA1052.1>.
- Gladkova I., A. Ignatov, F. Shahriar, Y. Kihai, D. Hillger, and B. Petrenko, 2016: Improved VIIRS and MODIS SST imagery. *Remote Sens.*, **8**, 79, <https://doi.org/10.3390/rs8010079>.
- Gosain, A., and S. Sardana, 2017: Handling class imbalance problem using oversampling techniques: A review. *2017 Int. Conf. on Advances in Computing, Communications and Informatics (ICACCI)*, Udupi, India, IEEE, 79–85.
- Hall, M., E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, 2009: The WEKA data mining software: An update. *ACM SIGKDD Explorations Newsletter*, No. 11, Association for Computing Machinery, New York, NY, 10–18, <https://doi.org/10.1145/1656274.1656278>.
- Hillger, D., and Coauthors, 2013: First-light imagery from *Suomi NPP* VIIRS. *Bull. Amer. Meteor. Soc.*, **94**, 1019–1029, <https://doi.org/10.1175/BAMS-D-12-00097.1>.
- Hollstein, A., K. Segl, L. Guanter, M. Brell, and M. Enesco, 2016: Ready-to-use methods for the detection of clouds, cirrus, snow, shadow, water and clear sky pixels in Sentinel-2 MSI images. *Remote Sens.*, **8**, 666, <https://doi.org/10.3390/rs8080666>.
- Kilpatrick, K. A., G. P. Podestá, and R. H. Evans, 2001: Overview of the NOAA/NASA Pathfinder algorithm for sea surface temperature and associated matchup database. *J. Geophys. Res.*, **106**, 9179–9198, <https://doi.org/10.1029/1999JC000065>.
- , and Coauthors, 2015: A decade of sea surface temperature from MODIS. *Remote Sens. Environ.*, **165**, 27–41, <https://doi.org/10.1016/j.rse.2015.04.023>.
- King, M. D., S. Platnick, W. P. Menzel, S. A. Ackerman, and P. A. Hubanks, 2013: Spatial and temporal distribution of clouds observed by MODIS onboard the Terra and Aqua satellites. *IEEE Trans. Geosci. Remote Sens.*, **51**, 3826–3852, <https://doi.org/10.1109/TGRS.2012.2227333>.

- Kullback, S., 1959: *Information Theory and Statistics*. John Wiley & Sons, 395 pp.
- Liu, Y., 2016: Sampling errors in satellite-derived infrared sea surface temperatures. Ph.D. thesis, Meteorology and Physical Oceanography Program, Rosenstiel School of Marine and Atmospheric Science, University of Miami, 84 pp., https://scholarlyrepository.miami.edu/oa_dissertations/1659/.
- , and P. J. Minnett, 2016: Sampling errors in satellite-derived infrared sea-surface temperatures. Part I: Global and regional MODIS fields. *Remote Sens. Environ.*, **177**, 48–64, <https://doi.org/10.1016/j.rse.2016.02.026>.
- , T. M. Chin, and P. J. Minnett, 2017: Sampling errors in satellite-derived infrared sea-surface temperatures. Part II: Sensitivity and parameterization. *Remote Sens. Environ.*, **198**, 297–309, <https://doi.org/10.1016/j.rse.2017.06.011>.
- Merchant, C. J., A. R. Harris, E. Maturi, and S. MacCallum, 2005: Probabilistic physically based cloud screening of satellite infrared imagery for operational sea surface temperature retrieval. *Quart. J. Roy. Meteor. Soc.*, **131**, 2735–2755, <https://doi.org/10.1256/qj.05.15>.
- Pfahring, B., G. Holmes, and R. Kirkby, 2001: Optimizing the induction of alternating decision trees. *Fifth Pacific-Asia Conf. on Advances in Knowledge Discovery and Data Mining*, Hong Kong, China, PAKDD, 477–487, <http://www.cs.waikato.ac.nz/ml/publications/2001/pakdd2001.pdf>.
- Saito, T., and M. Rehmsmeier, 2015: The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PLOS ONE*, **10**, e0118432, <https://doi.org/10.1371/journal.pone.0118432>.
- Saunders, R. W., and K. T. Kriebel, 1988: An improved method for detecting clear sky and cloudy radiances from AVHRR data. *Int. J. Remote Sens.*, **9**, 123–150, <https://doi.org/10.1080/01431168808954841>.
- Schueler, C. F., T. F. Lee, and S. D. Miller, 2013: VIIRS constant spatial-resolution advantages. *Int. J. Remote Sens.*, **34**, 5761–5777, <https://doi.org/10.1080/01431161.2013.796102>.
- Shi, H., 2007: Best-first decision tree learning. M.S. thesis, Dept. of Computer Science, University of Waikato, 104 pp.
- Xu, F., and A. Ignatov, 2014: In situ SST Quality Monitor (*iQuam*). *J. Atmos. Oceanic Technol.*, **31**, 164–180, <https://doi.org/10.1175/JTECH-D-13-00121.1>.