Improved cloud mask for NASA sea-surface temperature products from MODIS and VIIRS

Katherine Kilpatrick, Susan Walsh, Peter Minnett
University of Miami Rosenstiel School for Marine and Atmospheric Science
4600 Rickenbacker Causeway, Miami, FL 33149 USA email: kkilpatrick@miami.edu

GHRSST XIX EUMETSAT, Darmstadt, Germany June 2018

ABSTRACT

Deriving accurate retrievals of sea surface temperature (SST) from satellite-based measurements in the infrared requires the identification and exclusion of cloud contaminated pixels. The cloud mask historically used for both NOAA Pathfinder SST V5.3 and NASA MODIS SST R2014.0 use binary decision tree (BDtree) classifiers, and have been shown to be overly conservative, rejecting a significant number of clear pixel as cloudy in some conditions. A highly conservative cloud classifier, with a low false positive rate for clear sky conditions, while selecting very accurate retrievals may lead to errors in aggregated data, due to a potential of under sampling the true geophysical variability of the SST field (Liu and Minnett, 2016; Liu et al, 2017).

The cloud mask for NASA VIIRS SST R2016 uses a classifier based on an ensemble of alternating decision tree models (ADtree) trained with boosting (Kilpatrick et. al. 2017). The ADtree model provides a majority vote from an ensemble of classifiers scaled to the predictive power and confidence in the predicted class.

This a very different strategy than a BDtree, where during classification a pixel is limited to transit a single path to a terminal cloud/no cloud node. The use of the ADtree models for VIIRS significantly improved the number of valid clear sky retrievals in areas of sun glint, around cloud edges, at high latitudes, and in oceanic frontal regions. We report on our recent work extending the ADtree methodology to the cloud masks for MODIS on Aqua and Terra based on the experience gained with VIIRS.

METHODS and DATASETS

The two classifier algorithms (BDtrees and ADtrees) were trained and applied to matchups from each sensor (MODIS on Terra, MODIS on Aqua, and VIIRS on S-NPP). For our study we used WEKA3.7.11 (Hall et al., 2009) widely used Machine Learning software developed and maintained by the University of Waikato Environment for Knowledge and Analysis. For the BDtree classifiers we used the BFDtree package of Shi (2007) and for the ADtree we used the algorithm package of Pfahringer et al. (2001).

For each sensor classification models were built for four different illumination conditions: (i) nighttime, (ii) daytime, no glint (glint coefficient ≤ 0.005), (iii) daytime moderate glint coefficient between 0.005 to 0.01, and (iv) daytime severe glint. The severe glint condition was defined to occur when red (λ = 678 nm) reflectance > 0.065 and glint coefficient > 0.01. The glint coefficients were derivted using the Cox and Munk (1954) expression. The MODIS and VIIRS attributes used to build the classifier for each condition are given in Table 1.

Oceans are significantly cloudier than continents: the cloud fraction over the ocean is about 72%, with small seasonal variation. Consequently, the ratio of cloudy versus cloud-free records in the MUDB is roughly 3 to 1. This class imbalance can have a negative effect on the performance of traditional classification algorithms (Gosain and Sardana, 2017): predictive accuracy is biased towards the majority class and is also seen to be highly sensitive to data distribution. Given the very large number of available MUDB records, we addressed the imbalance by under-sampling, that is, randomly removing cloudy (majority) instances to produce subsets with approximately equal numbers of cloud-contaminated and clear-sky instances. For each condition we used a subset of records randomly selected from the SST matchup databases (MUDBs) described by Kilpatrick et al. (2001 and 2015), now publically available for VIIRS and MODIS through the NASA Ocean Biology distributed archive system (OB.DAAC) SeaBASS validation system. https://seabass.gsfc.nasa.gov/archive/SSTVAL.

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	294,780	28,366	28,366
VIIRS – S-NPP	118,732	49,006	2346	2346

Table 1. Number of records used in the training and validation for each illumination condition. Records were randomly selected from the L2 SST MUDBs and sub-sampled to provide training sets with comparable numbers of cloudy and cloud-free records. Classifications models were trained and validated using 10-fold cross-validation.

CLASSIFICATION MODEL CROSS-VALIDATION FROM MATCHUPS

Tables 2-4. MODIS and VIIRS: Classifier 10-fold cross validation statistics for Alternating (ADtree) and Binary (BDtree) Decision Tree methods. Under all conditions, the ADtree showed a slightly higher percentage of overall correctly classified records, and a reduction in the percentage of false positives (clear pixels identified as cloudy), particularly in glint regions where there was a 6-10% reduction in false positives. A good metric for assessing classifier success is the Precision/Recall (PRC) curve. 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.

Table 2 VIIRS ADtree and BDTree classifier validation statistics

VIIRS	Night		Day, non-glint		Day, mod glint		Day, high glint	
Model	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

TP = true positive, FP = false positive, PRC = Precision/Recall.

In all cases the ADtree PRC value were several points higher than BDtrees indicating improved performance.

Table 3 MODIS ADtree and BDtree classifier validation statistics

MODIS- A	Night		Day, non-glint		Day, mod glint		Day, high glint	
Model	ADtree	BDtree	ADtree	BDtree	ADtree	BDtree	ADtree	BDtree
% correctly classified	89.90	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

Table 4 MODIS ADtree classifier validation statistics

ADtree	Night		Day, non-glint		Day, mod glint		Day, high glint	
sensor	MODIS-A	MODIS-T	MODIS-A	MODIS-T	MODIS-A	MODIS-T	MODIS-A	MODIS-T
% correctly classified	89.90	88.61	92.07	92.08	91.43	91.84	89.14	90.15
% misclassified	10.70	11.30	7.94	7.91	8.58	8.10	10.10	9.85
TP cloud	0.89	0.88	0.93	0.93	0.91	0.93	0.87	0.92
TP clear	0.89	0.89	0.92	0.91	0.91	0.91	0.93	0.88
FP cloud	0.10	0.11	0.09	0.09	0.09	0.09	0.07	0.12
FP clear	0.11	0.12	0.07	0.07	0.09	0.07	0.13	0.08
PRC cloud	0.96	0.96	0.97	0.97	0.97	0.97	0.96	0.96
PRC clear	0.96	0.96	0.97	0.97	0.97	0.97	0.95	0.97

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. *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. Furthermore, VIIRS uses a pixel aggregation scheme to reduce the growth in the across-track pixel size as the satellite zenith angle increases, resulting in a three-fold reduction in the size of a VIIRS pixel at the edge of the swath (Schueler et al, 2013).*

Coverage Impact on L2 and L3 products

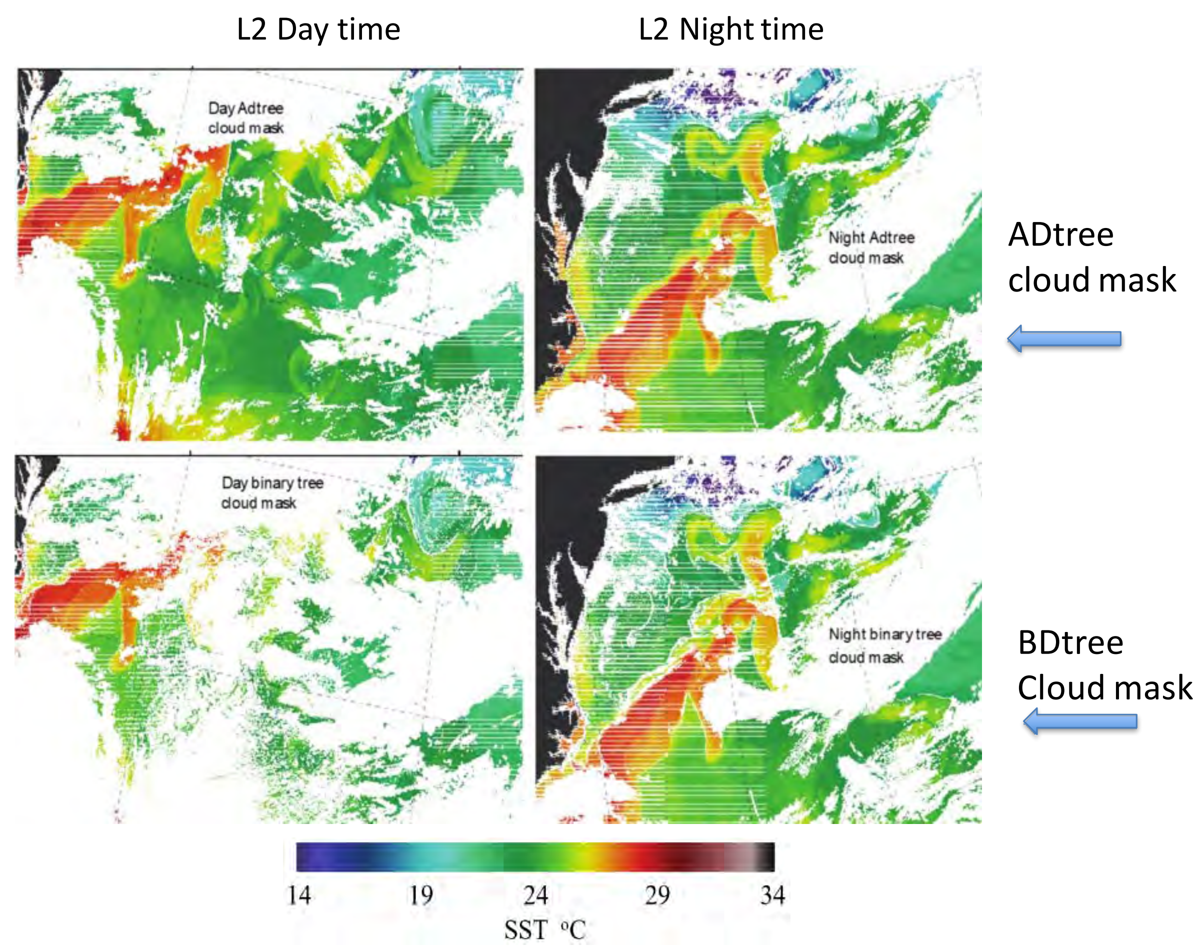

Figure 1. *Comparison of cloud mask methods for the Gulf Stream area on June 19th 2014, Level 2 daytime VIIRS SST_{skin} (left) and nighttime $NSST_{triple}$ (right). White areas indicate pixels identified as cloudy; black indicates land.* ***The Alternating Decision Tree classification (top panels) clouds are more compact and there is improved retention of clear pixels at the high gradient edges of the Gulf Stream*** *compared to the Binary Tree approach (bottom panels). The white horizontal lines show truncated scan lines resulting from bow-tie effect; when the images are mapped to a geographic projection, these data are taken from adjacent lines (Gladkova et al., 2016).*

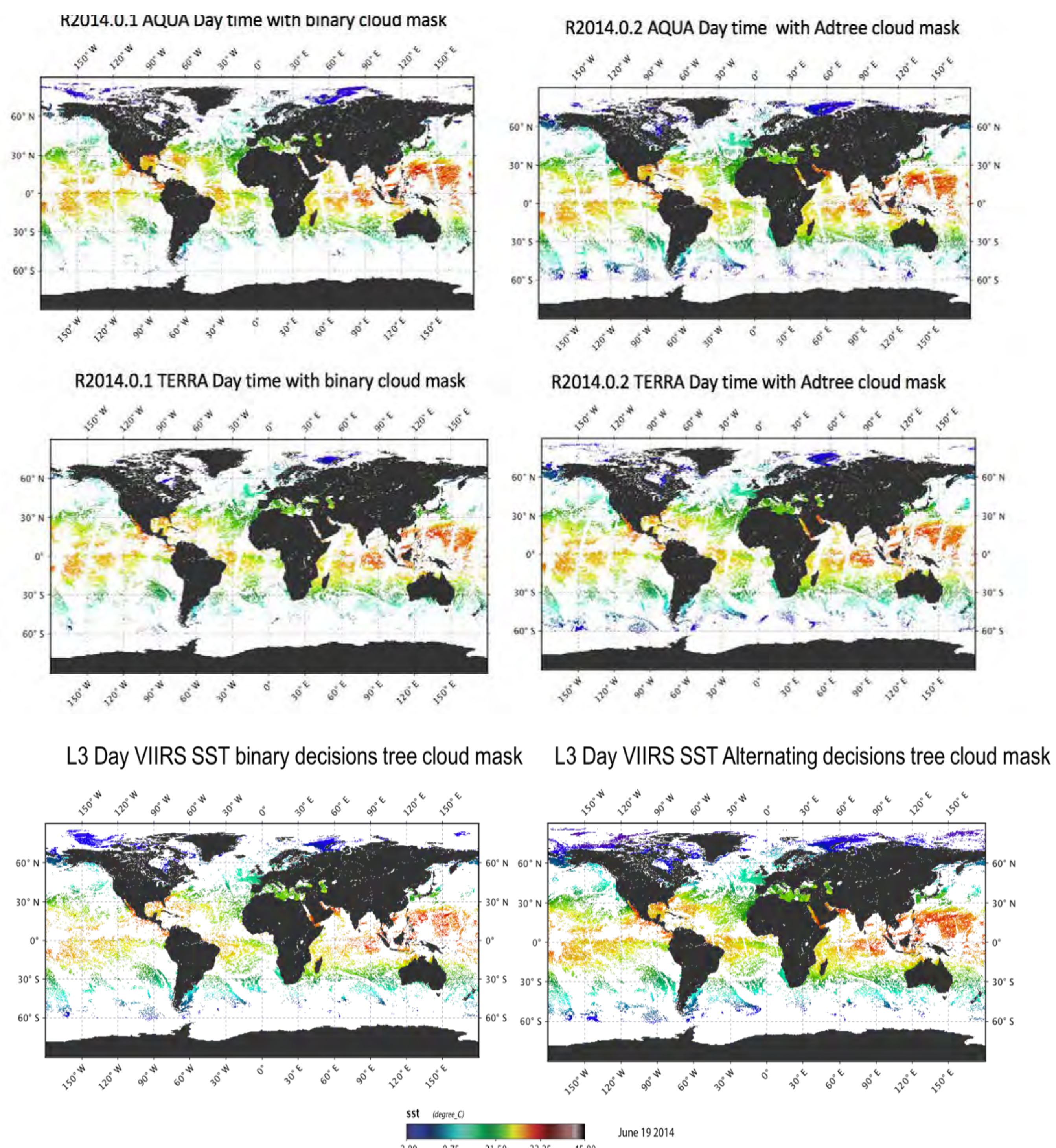

Figure 2. *Comparison of BDTree and ADtree MODIS and VIIRS SST_{skin} data coverage for daytime retrievals, L3 4km maps for June 19th 2014 good or better quality (NASA SST QL ≥ 1), using different cloud classification models. Left panels: cloud identification based on the BDtree. Right panel: cloud identification based on an ensemble of ADtree models.* ***Use of ADtree models increased global coverage of VIIRS and MODIS SST in L3 4km daily global files by ~ 5-10% at night and up to 35% in the daytime, depending on the location and season compared to BDtrees. The largest gains are seen in the mid to high latitudes poleward of 30°.***

References

Cox, C., & Munk, W. (1954). Statistics of the sea surface derived from sun glitter. Journal of Marine Research, 13, 198-227
Freund, Y., Mason, L.: The alternating decision tree learning algorithm. In: Proceeding of the Sixteenth International Conference on Machine Learning, Bled, Slovenia, 124-133, 1999.
Gosain, A., & Sardana, S. (2017). Handling class imbalance problem using oversampling techniques: A review. In, 2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI) (pp. 79-85)
Gladkova, I., Ignatov, A., Shahriar, F., Kihai, Y., Hillger, D., & Petrenko, B. (2016). Improved VIIRS and MODIS SST Imagery. Remote Sensing, 8, 79
Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I.H. (2009). The WEKA data mining software: an update. SIGKDD Explor. Newsl., 11, 10-18.
Kilpatrick, K.A., Podestá, G., Walsh, S., Williams, E., Halliwell, V., Szczodrak, M., Brown, O.B., Minnett, P.J., & Evans, R. (2015). A decade of sea surface temperature from MODIS. *Remote Sensing of Environment 165*, 27-41.
Kilpatrick K.A., Walsh S., Williams E. , Minnett P.. Classification of SST Quality Using a Combined Forest of Weak and Strong Classifiers. (2017). Poster. GHRSST meeting Washington DC.
Liu, Y., & Minnett, P.J. (2016). Sampling errors in satellite-derived infrared sea-surface temperatures. Part I: Global and regional MODIS fields. Remote Sensing of Environment, 177, 48-64.
Liu, Y., Chin, T.M., & Minnett, P.J. (2017). Sampling errors in satellite-derived infrared sea-surface temperatures. Part II: Sensitivity and parameterization. Remote Sensing of Environment, 198, 297-309
Pfahringer, B., Holmes, G., & Kirkby, R. (2001). Optimizing the Induction of Alternating Decision Trees. Fifth Pacific-Asia Conference on Advances in Knowledge Discovery and Data Mining. (pp. 477-487).
Schueler, C.F., Lee, T.F., & Miller, S.D. (2013). VIIRS constant spatial-resolution advantages. International Journal of Remote Sensing, 34,5761-577.
Shi, H. (2007). Best-first Decision Tree Learning (Thesis, Master of Science (MSc)). The University of Waikato, Hamilton, New Zealand

Acknowledgement
This study was funded by the NASA Physical Oceanography Program.

UNIVERSITY OF MIAMI
ROSENSTIEL SCHOOL of MARINE & ATMOSPHERIC SCIENCE