Ice-Over-Water Cloud Identification in an Artificial Neural Network Approach

Sunny Sun-Mack[1], Patrick Minnis[1], Yan Chen[1], Gang Hong[1] , William L. Smith, Jr.[2]

[1]Analytical Mechanics Associates, Inc. Hampton, VA 23666, [2]Atmospheric Sciences, NASA Langley Research Center, Hampton, VA, 23681

Corresponding author email address: szedung.sun-mack-1@nasa.gov

Introduction

An artificial neural network (ANN) algorithm, employing several Aqua MODIS channels, the retrieved cloud phase and total cloud visible optical depth, and temperature and humidity vertical profiles is trained to detect multilayer (ML) ice-over-water cloud systems identified by matched 2008 CloudSat and CALIPSO (CC) data. The trained MLANN was applied to 2009 MODIS data resulting in combined ML and single layer detection accuracies of 87% (89%) and 86% (89%) for snow-free (snow-covered) regions during the day and night, respectively. When corrected for the viewing-zenith-angle dependence of each parameter, the ML fraction detected is relatively invariant across the swath. Compared to the CC ML variability, the MLANN is robust seasonally and interannually, and produces similar distribution patterns over the globe, except in the polar regions. Additional research is needed to conclusively evaluate the VZA dependence and further improve the MLANN accuracy. This approach should greatly improve the monitoring of cloud vertical structure using operational passive sensors.

Approach

- **Train artificial neural network to detect multilayer (ML) clouds:** MLANN using 2008 C3M data and GEOS-5 reanalyses
  - **C3M:** Aqua MODIS radiances, Ed4+ retrievals from VISST/SIST, CloudSat-CALIPSO cloud profiles, 2/3 of matched cloudy MODIS pixel must be cloudy, use CALIPSO HA ≤ 5 km
  - **GEOS-5:** humidity profiles, skin temperature
  - **Auxiliary:** Lat, lon, elevation, SZA
  - **Breakdown:** Day and night, ice/snow and snow-free surfaces, ice and water phase => 8 categories
- **Assumptions and Constraints**
  - **Ice-over-water ML systems only**, ice = all cloud layers containing ice tops, water = all cloud layers containing liquid tops => 2 effective layers
    - Phase determined by CALIPSO or CloudSat (T > 253 K = liquid)
    - Top of liquid cloud must be 1 km or more below the base of the ice layer, lower layer can be mixed ice/water, but predominantly water
    - Assume pixels with MODIS COD < 0.5 are single layer (SL)
- **Validation:** apply to 2009 data and compare to CC
- **Application to full swath:** adjust input parameters for view zenith angle (VZA) dependence
- **Combine with NN CTH/CBH results for SL and upper ML clouds, and lower ML**

MLANN Input Parameters

Regional Parameters	GEOS-5.4	MODIS Thermal Data (K)	MODIS Solar Data†
Latitude (°), Longitude (°)	Surface skin temp (K)	T_{37}, T_{67}, T_{85}, T_{11}, T_{12}, T_{13}@	R_{CM}, τ_{CM}
Surface type, elevation	Relative humidity (%) at 8 levels*	BTD_{3711}, BTD_{6711}, BTD_{8511}, BTD_{1112}, BTD_{1113}@	$\rho_{1.38}$, $\rho_{1.61}$, $\rho_{2.13}$
Solar zenith angle (°)†			$\rho_{1.61}$ - $\rho_{2.13}$

* levels: surface, 850, 700, 500, 400, 300, 200, 100 hPa
† day only
@, snow/ice free only

Training Results

Confusion Matrix

	CC Single	CC Multi	Total
MLANN Single	SS	SM	SS + SM
MLANN Multi	MS	MM	MS + MM
Total	SS + MS	SM + MM	SS + MM

SS = True SL
SM = False SL
MS = False ML
MM = True ML

Assessment Parameters

Fraction correct (FC) or Accuracy: ACC = FC = SS +MM.
Real Risk: RR = SM +MS = 1 - ACC.
False ML rate: FM = MS / (MS + MM).
Precision: PR = 1 - FM, fraction of ML detections that are true
Recall: RC = MM / (SM +MM).
Single-layer Confidence: CoS = 1 - SS/(SS + SM).
False SL Rate: FSL = MS / (MS +SS)
Net Gain of Accuracy: NGA = MM - MS.

CloudSat and CALIPSO

	Snow-free, Day			Snow-free, Night			Snow-cover, Day *			Snow-cover, Night*		
MLANN	SL	ML	Total	SL	ML	Total	SL	ML	Total	SL	ML	Total
Ice SL,%	73.2	11.3	84.5	65.9	12.7	78.6	86.6	7.8	69.5	86.4	9.4	95.8
Ice ML,%	4.6	10.9	15.5	6.0	15.4	21.4	1.8	3.9	30.5	1.5	2.7	4.2
Total, %	77.8	22.2	**84.1**	71.9	28.1	**81.3**	88.4	11.7	**90.4**	87.9	12.1	**89.1**
# pixels x 10^3	3,748	1,070	4,818	4,097	1,599	5,696	2,549	390	2,884	11,41	1,500	12,91
Liquid SL, %	76.3	7.4	83.7	79.5	7.9	87.4	82.7	8.2	71.3	87.3	9.0	96.3
Liquid ML, %	3.9	12.4	16.3	3.3	9.3	12.6	2.5	6.6	28.7	1.3	2.4	3.7
Total, %	80.2	19.8	**88.7**	82.8	17.2	**88.8**	85.2	14.8	**89.3**	88.6	11.4	**89.7**
# pixels x 10^3	4,502	1,112	5,614	5,297	1,100	6,397	5,647	996	6,643	844	319	3,941

- Accuracy ranges from 81.3% - 88.8% at night over SF surfaces
  - Lower accuracy for ice clouds compared to liquid
    - More large COD ice clouds in ice category
- Accuracy higher for snow-cover (SC) surfaces, but fewer available ice-over-water clouds and lower recall than over SF
  - Fewer ML detected for ice clouds
- False alarm rate (FM) relatively low

Validation Results, 2009 C3M Data

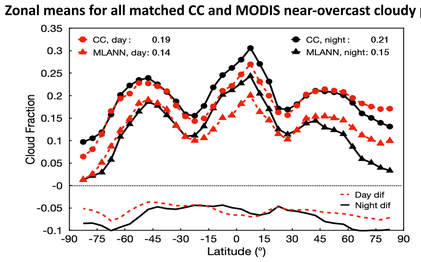
Combined Ice and Water Clouds

CloudSat and CALIPSO

	Snow-free, Day			Snow-free, Night			Snow-cover, Day			Snow-cover, Night		
MLANN	SL	ML	Total	SL	ML	Total	SL	ML	Total	SL	ML	Total
SL,%	75.3	8.9	84.2	74.0	10.1	84.1	83.6	8.4	92.0	86.2	9.8	96.0
ML,%	4.1	11.7	15.8	4.3	11.6	15.9	2.3	5.7	8.0	1.5	2.5	4.0
Total, %	79.4	20.6	**87.0**	77.3	21.7	**85.6**	85.9	14.1	**89.3**	87.7	12.3	**88.7**
# pixels x 10^3	28,883	7,493	36,376	33,739	9,908	43,647	8,235	1,352	9,587	14,277	2,002	16,279

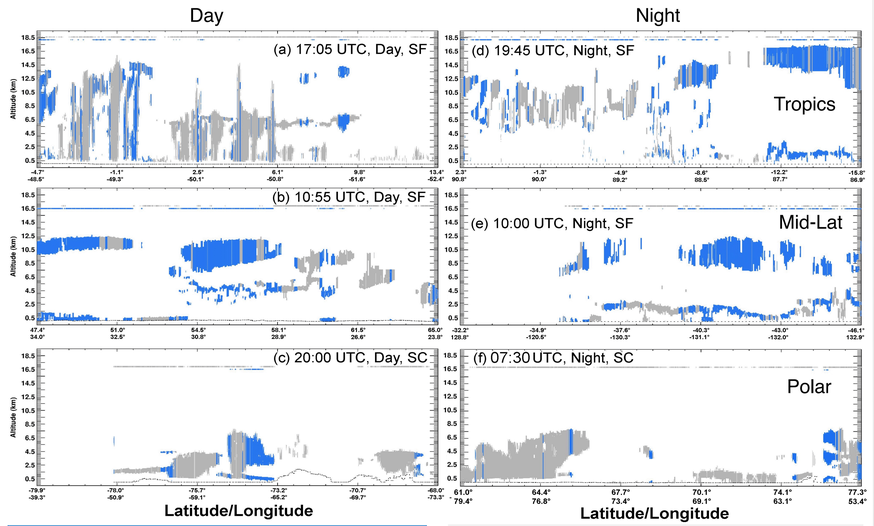
- Results very similar to training output
  - Minimal degradation in accuracy from training
- NGA = 7.5% for SF, day/night
- NGA = 4.4% for SC day, 1.0% for SC night
- Difference of CC-MLANN ~5% of total detected ML for SF, ~7% of total detected ML for SC
- MLANN correctly identifies ~55% of ice-over-water systems for SF, ~20-40% over SC

Zonal means for all matched CC and MODIS near-overcast cloudy pixels



- Full sampling diminishes relative fraction of ML
  - Differences between CC and MLANN remain about the same
- MLANN results track well with CC zonally and regionally (not shown)
  - Less so for polar regions at night

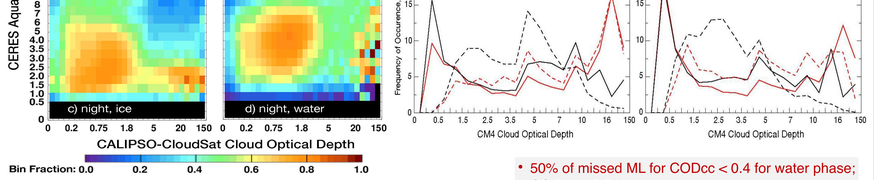
Examples of Results, CALIPSO-CloudSat Profiles, 25 December 2009



- Ice over water ML shown in blue profiles (1 km separation)
- SL, single phase, and ML with separation < 1 km shown in gray
- MLANN ML shown as blue dots at top

- Performance appears to be quite reasonable over snow-free surfaces
- Missed ML clouds more common over SC

Sensitivity of MLANN Detection to Cloud Optical Depths, Snow-Free Surfaces,2009



- Daytime:
  - RC greatest for COD_{CCice} = 1.7, COD_{CM4} = 11, RC < 50% for COD_{CCice} < 0.3
  - RC < 50% when $COD_{CCice} \approx COD_{CM4}$ or when COD_{CCice} > 4-5
  - Few ML detected when COD_{CM4} < 1.5
- Nighttime:
  - RC greatest for COD_{CCice} = 0.9, COD_{CM4} = 2 for ice
  - RC greatest for COD_{CCice} = 0.9, COD_{CM4} = 4 for water
  - RC < 50% when COD_{CCice} < 0.1, >5.0

- 50% of missed ML for CODcc < 0.4 for water phase; CODcc > 2-3 for ice phase
- 50% false ML for CODcm < 3.5 at night, < 7 during day for water phase
- 50% false ML for CODcm < 2.5 at night,< 5 during day for ice phase

Comparison with Other Methods

Algorithm	FC/ACC	PR	RC	CoS	FSL	NGA
MYD06 C6.1, τ > 5	67	54	46	73	22	2.2
POLDER 95, τ > 5	70	58	47	74	*18*	4.4
VIIRS, $\tau_{\lambda\lambda}$ > 1 ($\tau_{\lambda\lambda}$ < 1)	-	-	*65* (53)	79 (64)	23 - 36	-
Himawari Training Day	85	**81**	**72**	87	-	**18.3**
Himawari Validation Day	-	*70*	-	*89*	-	-
MLANN day	**87**	74	55	**90**	**5**	7.6
Himawari Training All	79	**73**	**64**	82	-	**14.3**
Himawari Validation All	-	*64*	-	*85*	-	-
MLANN Night	**86**	72	52	**88**	**6**	6.5

MYD06 - Platnick et al., 2017
• global
POLDER - Desmons et al., JAMC, 2017
• overcast global
VIIRS - Wang et al., RSE, 2019
• Global, uses CALIPSO only*, combine probable ML & ML
Himawari - Tan et al., TGARS, 2022
• 60N-60S, 80E - 160W
• **All** ML types, ice/liq, liq/liq, ice/ice, wat/ice

Published methods all vary in their constraints, sampling, domain, and reporting methods exacerbating comparisons

Definitions:
Fraction correct (FC) or Accuracy: ACC = FC = SS +MM
Precision: PR = 1 - FM, fraction of ML detections that are true
Recall: RC = MM / (SM +MM).
Single-layer Confidence: CoS = 1 - SS/(SS + SM).
False SL Rate: FSL = MS / (MS +SS)
Net Gain of Accuracy: NGA = MM - MS

Day
- MLANN outperforms MODIS and POLDER algorithms on all measures
- VIIRS day-only method has better recall than MLANN, but FSL very high
  - NGA likely to be negative or near zero, no false alarm data reported
- Himawari training set best of bunch, but not robust
- Independent validation drops values by 10 points or more
  - *Expect NGA to be considerably smaller as CoS increased*
  - *Precision drop indicates false ML increase*

Night
- Only Himawari and MLANN available
- Relative results similar to daytime, Himawari stats do not hold up in validation

Fractional Net Gain in Accuracy (NGA) in Cloud Layering Identification

NGA = trueML fraction - falseML fraction, in % of total cloud cover

- trueML fraction = ML pixels from NN that match ice-over-water ML pixels from CloudSat-CALIPSO data
- falseML fraction = ML pixels from NN that match SL pixels from CloudSat-CALIPSO data
- ML = ice-over-water cloud system with separation distance, SD = ice base CTH - water top CTH
- **bold** denotes higher value in one-to-one comparisons, *parentheses* indicate MLtrue

†April 2009, Aqua MODIS, CALIPSO Horizontal averaging ≤ 5 km

**July 2009, Aqua MODIS, CALIPSO Horizontal averaging ≤ 80 km*

Time, Ed4 phase	†Ed4	†MLANN SD > 1 km	*MLANN SD > 3 km	*MLANN SD > 1 km
Day, ice	-6.2 (10.5)	**6.2 (12.1)**	4.3 (7.6)	**7.8 (13.5)**
Day, water	1.6 (5.3)	**8.2 (12.6)**	4.7 (9.6)	**8.0 (13.5)**
Night, ice	0.2 (11.7)	**10.2 (17.1)**	8.3 (15.5)	**12.9 (23.2)**
Night, water	-0.6 **(2.8)**	**5.2 (8.1)**	4.6 (9.4)	**6.3 (12.4)**

- MLANN trained using SD > 1 km produces a net gain in information allowing the confident deconstruction of the MODIS radiances into 2 separate cloud layers for more ML clouds than either Ed4 CO2-based ML detection or MLANN trained with SD > 3 km
- MLANN trained using SD is as accurate or more accurate than other currently available methods.
- Positive NGA indicates a reduction in uncertainty

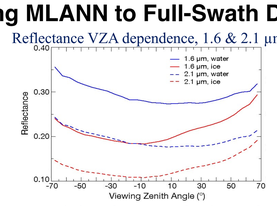
Applying MLANN to Full-Swath Data

Problem
- MLANN trained at near-nadir view
- All input parameters vary with VZA

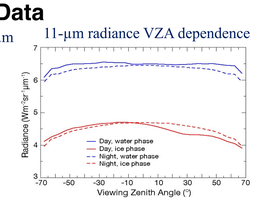
Solution
- Compute VZA dependence of all input
  - JAJO 2019 data, SF & SC sfcs
- Normalize to the near-nadir view
  - Similar to limb-darkening function
- Apply normalization factors to adjust all input to near-nadir view
- Use as input to MLANN for all views

Results
- Adjustment factors flatten dependence for ice phase, but not liquid

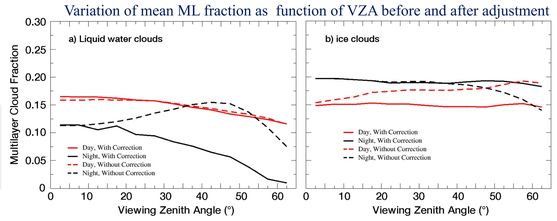
Reflectance VZA dependence, 1.6 & 2.1 μm



11-μm radiance VZA dependence

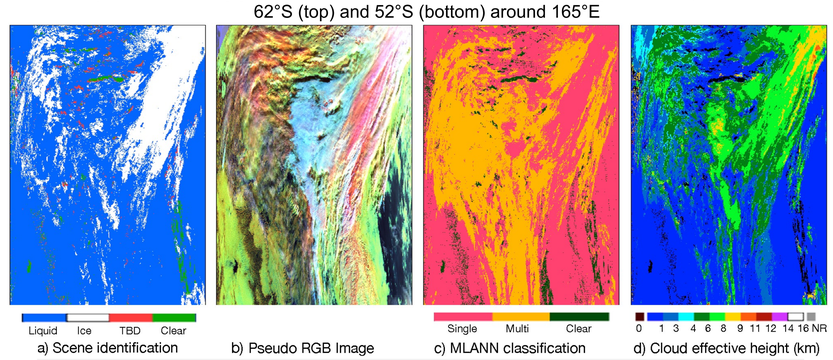


Variation of mean ML fraction as function of VZA before and after adjustment



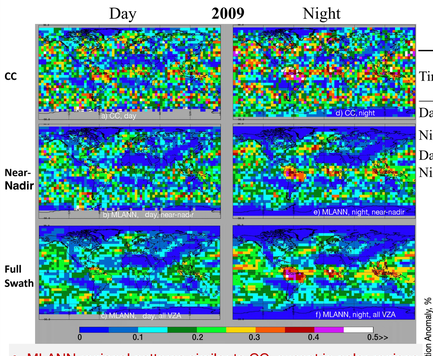
Example Swath Application, Aqua MODIS data, 16 April 2009, ~3:50 UTC

62°S (top) and 52°S (bottom) around 165°E



- Thin cirrus revealed as ice in some cases, blurry part of the RGB image, greater CTH in upper left and center
- MLANN catches much of the ML out to the edge and leaves many of the optically thick high clouds as SL

Swath Comparisons of Mean ML from CALIPSO-CloudSat & Aqua MODIS data

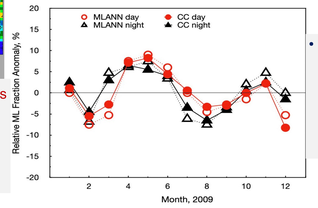


Summary of SF Results, JAJO

L/O = land/ocean ratio

Time	CC 2009	MLANN near-nadir 2009	MLANN all VZA 2009	MLANN near-nadir 2013	MLANN all VZA 2013
Day, ML (%)	15.4	12.1	11.8	12.6	12.3
Night, ML (%)	17.7	12.6	13.5	12.8	14.1
Day, L/O	0.77	0.64	0.65	0.63	0.64
Night, L/O	0.84	0.74	0.75	0.76	0.77

- MLANN 3 - 5% less than CC, smaller L/O values
- MLANN nadir slightly > swath during day, < swath at night
- L/O consistent for all MLANN cases, day & night
- MLANN yields fewer ML over land than ocean



- MLANN swath captures seasonal variability to within 1-2% of CC both day and night

- MLANN regional patterns similar to CC, except in polar regions
- ML-minima more defined in MLANN results, maybe missing thin Ci from nearby areas
- MLANN swath and near-nadir results comparable, swath smoother results

Conclusion

- Multilayer detection with neural network is among the best to date
  - Minimizes false detections
  - Yields net gain in accuracy of 6-8%, detects more than half of all ice/water ML clouds
    - Should improve characterization of radiation budget in atmosphere, at SFC, maybe TOA
    - Not perfect, still many ML clouds to detect
- Appears to be robust regionally, inter annually, and seasonally

Submitted to Atmospheric Measurement Techniques, November 2023