EUMETSAT NWC SAF

SUPPORT TO NOWCASTING AND VERY SHORT RANGE FORECASTING

Scientific and Validation Report for the Cloud Product Processors of the NWC/PPS

NWC/CDOP3/PPS/SMHI/SCI/VR/Cloud, Issue 3, Rev. 0

12 October 2021

Applicable to SAFNWC/PPS version 2021

Applicable to the following PGEs:

Acronym	**Product ID**	**Product name**	**Version number**
CMa	NWC-064	Cloud Mask	5.1
CMa-Prob	NWC-154	Cloud Probability	1.1
CT	NWC-067	Cloud Type	3.1
CTTH	NWC-070	Cloud Top Temperature and Height	5.1
CMIC	NWC-082	Cloud Micro Physics	2.1

Prepared by Swedish Meteorological and Hydrological Institute (SMHI)

REPORT SIGNATURE TABLE

Function	Name	Signature	Date
Prepared by	SMHI		12 October 2021
Reviewed by	SAFNWC Project Team EUMETSAT		28 September 2021
Authorised by	Anke Thoss, SMHI *SAFNWC PPS Manager*		12 October 2021

DOCUMENT CHANGE RECORD

Version	Date	Pages	Changes
3.0d	22 March 2021	50	First original version for PPS v2021.
3.0d	1 September 2021	75	Changes after PPS v2021 RR: RID-029: removed a table that by mistake was doubled Updated version numbers in 1.2. (Those on p1 were already correct.) RID-032: Clarified version of CPR data RID-048: Clarified how bias and STD are translated to MAE RID-025: Clarified that we are using 23 granules from Metop-B RID-030: Clarify description Table 3 RID-031: Make sure tables can be read as stand-alone RID-041: Include polar statistics in VR report (Action 003) RID-022: Included information on SLSTR validation plans (Action 001) For completeness included MODIS-C6.1 as a point of comparison also for phase RID-023: Included position of SLSTR granule in Figure 6 (Now 10). As a consequence also Figure 7 (Now11) was updated. RID-049: Updated caption for figure 7. RID-033: replace PGE-number with product acronym RID-019, 024, 037: editorial Excluding bad quality low-energy CALIOP (MERSI-2) Extended cloud probability validation with some more figures. Included section with validation and investigation of errors. Specially for cloud phase. Related to action 004 but focused on showing that product errors caused by errors in the RTMs are small for cloud phase. And as the errors are small modelling them in the ATBD is of limited use for the users. Added Cloud probability validation for SYNOP as that is in the requirements. Included CTTH validation separated both for semi-transparent and opaque clouds and for low-level, medium-level, high-level clouds. Requested by CLARA-A3 simulator developers. Changed to use correct CMa-Prob acronym. Added validation table in ANNEX-B for combination of CMa and CMa-Prob recommended in scientific user manual for high quality clear retrievals. Added NWP-data resolutions. Updated Metop-C results, for some tables data within 5minutes was used.
3.0	12 October 2021	75	Changes related to DRR v2021: RID012 Fixed three numbers marked with wrong color RID034 Improved caption Figure 4 RID036 Adding missing accuracies in Table 3 and Table 19. RID037 Clarified CMa validation strategy RID038 Acronyms POD and FAR are now consistently in upper case. RID039 editorial Some smaller editorials.

Table of Contents

1 INTRODUCTION 9

1.1 PURPOSE 9
1.2 SCOPE 9
1.3 DEFINITIONS AND ACRONYMS 9
1.4 REFERENCES 10
1.4.1 Applicable documents 10
1.4.2 Reference documents 11
1.5 DOCUMENT OVERVIEW 11

2 DEFINITION OF VERIFICATION MEASURES USED 12

3 DATA USED 14

3.1 IMAGER SATELLITE DATA USED 14
3.1.1 AVHRR GAC data (NOAA-18) 14
3.1.2 VIIRS data (Suomi-NPP) 14
3.1.3 MODIS data (EOS-Aqua) 14
3.1.4 MERSI-2 15
3.1.5 EPS-SG 15
3.1.6 AVHRR (Metop-B) 15
3.1.7 SLSTR (Sentinel 3a) 15
3.2 NWP DATA 15
3.3 THE SYNOP DATA 15
3.4 THE CALIPSO DATASET 16
3.5 THE CPR (CLOUDSAT) DATASET 20
3.6 THE AMSR-E DATASET 20
3.7 THE MODIS COLLECTION 6.1 L2 DATASET 21

4 RESULTS AND DISCUSSION 22

4.1 CLOUD MASK 22
4.1.1 SYNOP validation 22
4.1.2 CALIOP validation 23
4.1.3 Preliminary EPS-SG validation 28
4.2 CLOUD TYPE 29
4.3 CLOUD TOP HEIGHT 31
4.3.1 CALIOP validation 33
4.3.2 CPR (CloudSat) validation 38
4.4 CMIC CLOUD PHASE (CPH) 41
4.5 CMIC LIQUID WATER PATH (LWP) 42
4.5.1 Global validation of LWP over Sea with AMSR-E 42
4.6 CLOUD PROBABILITY 46
4.6.1 Cloud mask and cloud probability comparison 50
4.7 SLSTR EXAMPLE 52
4.7.1 SLSTR future plans 52
4.8 POLAR STATISTCS 54
4.8.1 Cloud mask and cloud probability 54
4.8.2 Cloud type 60
4.8.3 CTTH 61
4.8.4 CMIC Cloud phase 62

5 VALIDATION OF ERROR ESTIMATES 63

5.1 CLOUDMASK AND CLOUD PROBABILITY ERRORS 63
5.2 CTTH ERRORS 63
5.3 CLOUD PHASE ERRORS 64
5.3.1 When do we expect large errors for CMIC cloud phase? 67

6 SUMMARY AND CONCLUSIONS 70

7 REFERENCES 72

ANNEX A. LIST OF TBC, TBD, OPEN POINTS AND COMMENTS 74

ANNEX B. VALIDATION RESULTS FOR COMBINED USE OF CMA AND CMA-PROB FOR HIGH QUALITY CLEAR DETECTION. 75

List of Tables and Figures

Table 1: List of Applicable Documents 11
Table 2: List of Referenced Documents 11
Table 3: Cloud mask validation scores for 15 S-NPP orbits against global SYNOP. Results are global, results for Europe only are included in a separat line. Europe here is defined as latitude between 35 and 72 and longitude between -25 and 60 degrees. Green: within target accuracy, red: outside threshold accuracy. Yellow marks measures with requirements. 22
Table 4: Cloud Mask global validation for PPS-v2021. Green: within target accuracy, red: outside threshold accuracy. Yellow marks measures with requirements. Measures are intended for SYNOP data, we do not expect measures to be met for very thin clouds. For the "no polar night" category all data where sun satellite angle is above 95 and absolute latitude is above 70 is excluded. 23
Table 5: Cloud Mask global validation for PPS-v2021. Green: within target accuracy, red: outside threshold accuracy. Yellow marks measures with requirements. We do not expect good results for polar night and CALIOP 5km thin clouds (pink). 25
Table 6: Cloud mask validation for PPS-v2021 over Europe compared to CALIOP. Europe here is defined as latitude between 35 and 72 and longitude between -25 and 60 degrees. Green: within target accuracy, red: outside threshold accuracy. Yellow marks measures with requirements. 27
Table 7: Cloud Mask snow validation for PPS-v2021 compared to NSIDC flag in CALIOP data. 28
Table 8: Validation results for a preliminary version of PPS v2021 for the V1 of the EPS-SG test data. The truth is the cloud mask used as input for the simulation. 28
Table 9: Overview of PPS and CALIOP cloud type matches that are considered to be correct marked with (X). 29
Table 10: Cloud type validation compared to CALIOP. See Table 9 to see which classifications are considered successful/unsuccessful. Green: within target accuracy, red: outside threshold accuracy. Yellow marks measures with requirements. 30
Table 11: Confusion table of PPS v2021 cloud type version 2021 compared to CALIOP for VIIRS. 31
Table 12: PPS-v2021 cloud top height validation. Results outside threshold accuracy (semi-transparent) are marked red, results within target accuracy (opaque) are marked green. Yellow: measures with requirement. We use MAE and median used as error distributions are non-Gaussian. 35
Table 13: Validation results separated by opaque semi-transparent. For comparison results for MYDO6_L2 are included, PPSv2021 has better scores. 36
Table 14: Validation results separated by opaque semi-transparent filtered to get the cases where CALIOP and the imager really should agree. Filtering is done as similar as possible to GEO validation, considering different validation software. Only pixels where 9 neighbouring measurements have the same cloud type and the variation in CALIOP pressure are less than 200hPa are included. Pixels where the 5km CALIOP top-layer are thinner than 0.2 are excluded. This leaves 1/4 of the data. 37
Table 15 Validation results separated by opaque semi-transparent and separated for different heights. For comparison results for MYDO6_L2 are included, PPSv2021 has better scores. 38
Table 16: PPS-v2021 cloud top height validation with CPR (CloudSat). Given are total results as well as results separated for low, medium and high for MODIS. Results outside threshold accuracy (semi-transparent) are marked red, results within target accuracy (opaque) are marked green. Yellow: measures with requirement. However median and IQR are more informative. 40
Table 17: Success measures for CMIC cloud phase compared to CALIOP. All datasets are global except for Metop-B that has only high latitude data. Green: within target accuracy, red: outside threshold accuracy. Yellow marks measures with requirements. 42
Table 18: Required and achieved accuracies for the global validation over sea of liquid water path compared to AMSR-E, CPR (CloudSat) or intercompared with PPS CMIC-lwp for MODIS. Note that requirements are not defined for comparison with CPR (CloudSat) or for inter comparisons between sensors. And CPR (CloudSat) RVOD is based much on MODIS data which can't be assumed to have higher quality than CMIC-lwp. Green: within target accuracy, red: outside threshold accuracy. Yellow marks measures with requirements. 43
Table 19: Cloud probability validation scores for 15 S-NPP orbits against global SYNOP. Results are global, result for Europe only are included in a separat line. Europe here is defined as latitude between 35 and 72 and longitude between -25 and 60 degrees. Green: within target accuracy, red: outside threshold accuracy. Yellow marks measures with requirements. 46

TABLE 20: SCORES FOR CMA-PROB FOR THE NPP-CALIOP MATCHUP DATA. AS A REFERENCE CMA HAVE KUIPERS 0.710 AND HITRATE 0.863 FOR THE SAME DATASET. POD-CLOUDY AND FAR-CLEAR ARE NOT EXPECTED TO MEET THE REQUIREMENT FOR ALL LIMITS. INSTEAD IT IS IMPORTANT THAT THEY MEET THE REQUIREMENTS FOR AT LEAST ONE PROBABILITY CLOUD MASK THRESHOLD. IT IS ALSO IMPORTANT THAT SOME PROBABILITY CLOUD MASK THRESHOLDS GIVE LOW VALUES ON FAR-CLEAR AND FAR-CLOUDY RESPECTIVELY. GREEN: WITHIN TARGET ACCURACY, RED: OUTSIDE THRESHOLD ACCURACY. YELLOW MARKS MEASURES WITH REQUIREMENTS....47
TABLE 21: VALIDATION OF CMA-PROB WITH LIMIT 50 AGAINST CALIOP-DATA. GREEN: WITHIN TARGET ACCURACY, RED: OUTSIDE THRESHOLD ACCURACY. YELLOW MARKS MEASURES WITH REQUIREMENTS MEASURES ARE INTENDED FOR SYNOP DATA, WE DO NOT EXPECT MEASURES TO BE MET FOR VERY THIN CLOUDS. FOR THE "NO POLAR NIGHT" CATEGORY ALL DATA WHERE SUN SATELLITE ANGLE IS ABOVE 95 AND ABSOLUTE LATITUDE IS ABOVE 70 IS EXCLUDED. 48
TABLE 22: OVERVIEW OF POLAR STATISTICS OVER SEA (INCLUDING ICE AND OPEN WATER) FOR CLOUD MASK COMBINED WITH CLOUD PROBABILITY WHERE PPS IS CONSIDERING IT CLOUDY IF A CLOUD IS DETECTED BY CLOUD PROBABILITY ABOVE 50% OR BY CLOUD MASK. LINES WITH POD-CLOUDY>85% AND FAR-CLOUDY<20% ARE MARKED GREEN. LINES WITH HIT RATE BELOW 60% ARE MARKED RED....54
TABLE 23: POLAR STATISTICS FOR CMA AND CMA-PROB FOR VIIRS DATA DIVIDED IN ILLUMINATION CATEGORIES AND SURFACE TYPE (LAND/WATER/ICE). LINES WITH POD-CLOUDY>85% AND FAR-CLOUDY<20% ARE MARKED GREEN. LINES WITH HIT RATE BELOW 70% ARE MARKED RED. NOTE CMA IN THE TOP AND CMA-PROB RESULTS IN THE BOTTOM. 55
TABLE 24: POLAR STATISTICS FOR CMA AND CMA-PROB FOR AVHRR NOAA-18 GAC DATA DIVIDED IN ILLUMINATION CATEGORIES AND SURFACE TYPE (LAND/WATER/ICE). LINES WITH POD-CLOUDY>85% AND FAR-CLOUDY<20% ARE MARKED GREEN. LINES WITH HIT RATE BELOW 70% ARE MARKED RED....57
TABLE 25: POLAR STATISTICS FOR CMA AND CMA-PROB FOR MODIS DIVIDED IN ILLUMINATION CATEGORIES AND SURFACE TYPE (LAND/WATER/ICE). LINES WITH POD-CLOUDY>85% AND FAR-CLOUDY<20% ARE MARKED GREEN. LINES WITH HIT RATE BELOW 70% ARE MARKED RED. 58
TABLE 26: POLAR STATISTICS FOR CMA AND CMA-PROB FOR MERSI2 DIVIDED IN ILLUMINATION CATEGORIES AND SURFACE TYPE (LAND/WATER/ICE). LINES WITH POD-CLOUDY>85% AND FAR-CLOUDY<20% ARE MARKED GREEN. LINES WITH HITRATE BELOW 70% ARE MARKED RED. 59
TABLE 27: POLAR STATISTICS FOR CMA AND CMA-PROB FOR AVHRR (METOP-B) DIVIDED IN ILLUMINATION CATEGORIES AND SURFACE TYPE (LAND/WATER/ICE). 60
TABLE 28: POLAR CLOUD TYPE VALIDATION COMPARED TO CALIOP. SEE TABLE 9 TO SEE WHICH CLASSIFICATIONS ARE CONSIDERED SUCCESSFUL/UNSUCCESSFUL. DATA WITHIN THRESHOLD ACCURACY ARE MARKED GREEN. THERE ARE NO REQUIREMENTS FOR POLAR DATA SPECIFICALLY. 61
TABLE 29: PPS-V2021 POLAR CLOUD TOP HEIGHT VALIDATION. RESULTS INSIDE THRESHOLD ACCURACY (MEDIAN AND MAE) ARE MARKED GREEN. THERE ARE NO REQUIREMENTS FOR POLAR DATA SPECIFICALLY. 62
TABLE 30: POLAR SUCCESS MEASURES FOR CMIC CLOUD PHASE COMPARED TO CALIOP. 62
TABLE 31: VALIDATION OF ERROR ESTIMATES FOR CMA AND CMA-PROB. FOR CMA THERE IS A LOW-QUALITY FLAG AND FOR CMA-PROB VALUES BETWEEN 10 AND 90 ARE TREATED AS LOW QUALITY IN THIS INVESTIGATION. 63
TABLE 32: TABLE SHOWING HIT RATE FOR CLOUD PHASE AND HOW THAT IMPROVES WHEN DATA THAT ARE MORE LIKELY TO CONTAIN ERRORS ARE EXCLUDED. FOR THE GEO STYLE VALIDATION, WE KEEP ONLY RETRIEVALS THAT HAVE CALIOP OPTICAL DEPTH OF THE TOP LAYER LARGER THAN 0.2 AND VARIATION LESS THAN 200 HPA AND THE SAME PPS CLOUD TYPE FOR THE 9 CLOSEST NEIGHBOURS ALONG THE CALIOP TRACK. 67
TABLE 33: POD-CLEAR AND FAR CLEAR FOR CMA AND CMA-PROB AND FOR THEM COMBINED. RESULTS ARE SEPARATED IN LAND AND SEA CATEGORIES. 75

FIGURE 1: GLOBAL MAP SHOWING THE LOCATION AND FREQUENCY OF ALL VIIRS AND SYNOP CO-LOCATIONS, BASED ON THE 15 S-NPP ORBITS FROM 2015. 16
FIGURE 2: EQUATORIAL CROSSING TIMES OF THE NOAA, METOP AND SNPP/JPSS SPACECRAFTS FROM NOAA-7 TILL TODAYS NOAA-20 AND METOP-B. 17
FIGURE 3: DISTRIBUTION IN TERMS OF SATELLITE ZENITH (UPPER PANEL), SUN ZENITH ANGLE (MIDDLE PANEL) AND TIME DIFFERENCE (LOWER PANEL) OF CALIPSO-AVHRR COLOCATIONS FROM THE 2006-2009 NOAA-18 GAC DATA RECORD. OBSERVATIONS SPAN NIGHT AND DAY WITH AN ALMOST FLAT DISTRIBUTION FOR NIGHT, TWILIGHT AND DAY, AND WITH THE MAJORITY OF OBSERVATIONS HAVING A NEAR NADIR AVHRR VIEW....18
FIGURE 4: THE NON-GAUSSIAN ERROR DISTRIBUTION FOR 15 ORBITS S-NPP IN RED. THE CTTH DOES NOT REACH THE GAUSSIAN THRESHOLD ACCURACY FOR STD OF 2000M. HOWEVER IT REACHES (WITH A LARGE MARGIN)

THE MAE AND MEDIAN THRESHOLD ACCURACY MORE SUITABLE ALSO FOR NON-GAUSSIAN DISTRIBUTIONS. THE NON-GAUSSIAN ERROR DISTRIBUTION IS COMPARED TO TWO FICTIONAL ERROR DISTRIBUTIONS THAT DO REACH THRESHOLD OR EVEN TARGET ACCURACY BUT OVERALL HAS MORE LARGE ERRORS. IT IS CLEAR THAT BECAUSE THE CTTH ERROR DISTRIBUTION IS NON-GAUSSIAN THE STD AND BIAS ARE NOT THE MOST APPROPRIATE MEASURES TO USE TO DESCRIBE THE ERRORS OF THE ALGORITHM. NOTE THAT WE DO EXPECT SOME LARGE ERRORS DUE TO DIFFERENCES IN FOV, AND SENSITIVITY BETWEEN THE IMAGER AND THE CALIOP INSTRUMENT. 32

FIGURE 5: COMPARISON OF CTTH V2014 (LEFT) WITH CTTH V2018 (RIGHT). WE CAN SEE THAT THE RETRIEVALS ARE MORE CONSISTENT FOR PPS-V2018 AND THE SQUARED GAPS WITH MISSING DATA IN VERSION 2014 ARE FILLED FOR VERSION 2018. NOTE THAT CTTH V2014 AND V2018 ARE TWO COMPLETELY DIFFERENT ALGORITHMS. THE PPS V2021 CTTH ALGORITHM IS IDENTICAL TO V2018. MINOR DIFFERENCES IN CTTH BETWEEN V2018 AND V2021 IS LIKELY CAUSED BY UPDATES TO THE CLOUD MASK. 33

FIGURE 6 FROM THE TOP AVHRR-GAC COMPARED TO AMSR-E, MODIS COMPARED TO AMSR-E AND VIIRS COMPARED TO CPR (CLOUDSAT). FROM LEFT TO RIGHT ERROR DENSITY FUNCTION, SCATTER PLOT AND LWP DENSITY FUNCTION. THE ERROR DENSITY FUNCTION (RED CURVE) IS COMPARED TO A GAUSSIAN DISTRIBUTION WITH THE SAME BIAS AND RMS. 45

FIGURE 7:. INTER COMPARISON BETWEEN PPS CMIC-LWP FOR MODIS AND MERSI-2. FROM LEFT TO RIGHT ERROR DENSITY FUNCTION, SCATTER PLOT AND LWP DENSITY FUNCTION. THE ERROR DENSITY FUNCTION (RED CURVE) IS COMPARED TO A GAUSSIAN DISTRIBUTION WITH THE SAME BIAS AND RMS. NOTE THAT BOTH MEDIAN AND IQR ARE SMALL. SHOWING THAT THE PRODUCTS ARE VERY SIMILAR. DATA POINTS ARE LESS THAN 20 MINUTES APART. AND FOR EACH SENSOR A GAUSSIAN WEIGHTED (DEPENDING ON DISTANCE) AVERAGE IS CALCULATED FOR THE 25 CLOSEST PIXELS WITHIN 5KM. DATA POINTS WITH VALUES ABOVE 500 KG/M2 ARE EXCLUDED SIMILAR TO THE VALIDATIONS WITH AMSR-E AND CLOUDSAT RVOD PRODUCT. ..46

FIGURE 8: SHOWING THE ACTUAL PROBABILITY FOR CLOUDY FOR THE CLOUD PROBABILITY CLASSES. 50

FIGURE 9: CONFUSION HISTOGRAMS FOR VIIRS, MODIS AND MERSI-2 CLOUD PROBABILITY AND CMA. 51

FIGURE 10: SHOWING THE SLSTR GRANULE ON A MAP. 52

FIGURE 11: TO THE LEFT A GRANULE OF MODIS FROM EOS-AQUA REMAPPED TO THE SAME PROJECTION AS THE SLSTR GRANULE. TO THE RIGHT THE SAME GRANULE FOR SLSTR SENTINEL-3A. THE GRANULES ARE ONLY 5 MINUTES APART JUST OUTSIDE THE COAST OF ANTARCTICA 20200312 01:40 UTC FOR EXACT LOCATION SEE FIGURE 10. CHANNEL 8.7 IS TURNED OFF FOR MODIS. THE PRODUCTS DISPLAYED ARE, FROM TOP: CLOUD TYPE, CLOUD PROBABILITY, CTTH, AND RGB-IMAGE. THE RGB-IMAGES ARE MADE WITH CHANNELS 1.6µM, 37µM, 11µM AND 12µM. 53

FIGURE 12: ERROR ESTIMATES FOR CTTH VIIRS DATA. THE ERROR ESTIMATE FOR CTTH COMES AS A LOWER AND UPPER LIMIT. FOR THE FIGURE THE LARGEST (IN ABSOLUTE VALUE) OF THE LIMITS IS USED TO SORT DATA ALONG X-AXIS. WE CAN SEE THAT MOST OF THE ERROR ESTIMATES ARE SMALL (HIGH BARS BETWEEN -125HPA AND 125HPA). AND MOST OF THE ACTUAL ERRORS ARE WITHIN THE ERROR ESTIMATE +/- AN ADDITIONAL 20HPA (ORANGE BARS ALMOST AS HIGH AS BLUE). 64

FIGURE 13: CLOUD PHASE ERROR DEPENDENCE ON CTTH ERROR. BLUE IS FRACTION OF WRONG PHASE AND READ IS THE PERCENT OF CTTH RETRIEVALS. NOTE THAT THE OVERLAP YIELDS A DARK RED COLOR. NOTE THAT WHEN THE CTTH ABSOLUTE ERROR IS SMALLER THAN 50HPA THE CLOUD PHASE HAS FEW (7.6%) BAD RETRIEVALS. AND WHEN THE ERROR FOR CTTH IS LARGE THE CLOUD PHASE WILL ALMOST ALWAYS BE WRONG. NOTE THAT THE LARGE ERRORS FOR CTTH ALSO INCLUDE CASES WHEN THE IMAGER AND CALIOP SEE DIFFERENT CLOUDS, FOR EXAMPLE AT A CLOUD EDGE. THE RED CURVE SHOWS THE AMOUNT OF CTTH RETRIEVALS WHICH HAS MOST OF THE DATA IN THE MIDDLE (SMALL ERRORS). 66

FIGURE 14: CLOUD PHASE RETRIEVALS ARE BETTER WHERE THE ESTIMATED CTTH ERROR IS SMALLER (BELOW 50HPA IN ABSOLUTE VALUE). BLUE IS FRACTION OF WRONG PHASE AND READ IS THE PERCENT OF CTTH RETRIEVALS. NOTE THAT THE OVERLAP YIELDS A DARK RED COLOR NOTE THAT COMPARED TO FIGURE 13 (WHICH SHOWED CLOUD PHASE ERROR COMPARED TO THE ACTUAL CTTH ERROR) THIS FIGURE SHOWS THE CLOUD PHASE ERROR AS A FUNCTION OF THE CTTH ERROR ESTIMATES AVAILABLE IN THE PPS PRODUCTS.68

FIGURE 15: FRACTION OF WRONG CLOUD PHASE RETRIEVALS AS A FUNCTION OF CLOUD TYPE FOR FOUR DIFFERENT SENSORS. FOR ALL SENSORS, ERRORS ARE LARGEST FOR THE VERY THIN CIRRUS CLOUDS (11) AND THE MID-LEVEL CLOUDS (7) CATEGORY. VERY HIGH CLOUD (9) HAS ALMOST ALL RETRIEVALS CORRECT. METOP-B DATA HAVE NOT BEEN INCLUDED IN THIS FIGURE BECAUSE OF THE SMALL AMOUNT OF PIXELS. 69

1 INTRODUCTION

The EUMETSAT "Satellite Application Facilities" (SAF) are dedicated centres of excellence for processing satellite data, and form an integral part of the distributed EUMETSAT Application Ground Segment (http://www.eumetsat.int). This documentation is provided by the SAF on Support to Nowcasting and Very Short Range Forecasting, SAFNWC. The main objective of SAFNWC is to provide, further develop and maintain software packages to be used for Nowcasting applications of operational meteorological satellite data by National Meteorological Services. More information can be found at the SAFNWC webpage, http://nwc-saf.eumetsat.int. This document is applicable to the SAFNWC processing package for polar orbiting meteorological satellites, SAFNWC/PPS, developed and maintained by SMHI (http://nwcsaf.smhi.se).

1.1 Purpose

This document is a report presenting validations results of the cloud products from NWC/SAF. The threshold, target and optimal accuracies validated against are described in the Product Requirement Document [AD.4.].

Note that PPS version 2021 contains mostly technical updates and minor changes compared to version 2018 and results are expected to be very similar to version 2018. The two products developed by CMSAF: CMa-Prob and CMIC have larger updates. A minor preliminary validation of the V1 EPS-SG test-data for CMa is included to verify that the quality for EPS-SG is still expected to meet the requirements.

1.2 Scope

This document presents the validation result of the NWC/SAF cloud products: CMa version 5.1, CMa-Prob version 1.1, CT version 3.1, CTTH version 5.1 and CMIC version 2.1 - all applicable to PPS version 2021.

1.3 Definitions and Acronyms

Acronym	Explanation
AMSR-E	Advanced Microwave Scanning Radiometer for EOS
AEMET	Agencia Estatal de Meteorología (Spain)
AVHRR	Advanced Very High Resolution Radiometer
CALIOP	Cloud-Aerosol Lidar with Orthogonal Polarisation
CDOP3	Third Continuous Development and Operational Phase
CMa	Cloud Mask (also PGE01)
CMa-Prob	Cloud probability (also PGE01c)
CMIC	Cloud MIcro Physics (also CPP)
CPP	Cloud Physical Properties (also CMIC/PGE05)
CPR (CloudSat)	Cloud Profiling Radar for CloudSat
CT	Cloud Type (also PGE02)
CTTH	Cloud Top Temperature, Height and Pressure (also PGE03)
EOS	Earth Observation System
EPS	EUMETSAT Polar System
EPS-SG	EUMETSAT Polar System Second Generation
EUMETSAT	European Organisation for the Exploitation of Meteorological Satellites
FOV	Field of View
GAC	Global Area Coverage
LWP	Liquid Water Path
MERSI	Medium Resolution Spectral Imager
METimage	Meteorological Imager
MODIS	Moderate-resolution Imaging Spectro-radiometer
NOAA	National Oceanic and Atmospheric Administration
PC	Precipitating Cloud (also PGE04) (discontinued product)
PGE	Process Generating Element
PPS	Polar Platform System
SAF	Satellite Application Facility
SAFNWC	Satellite Application Facility for support to NoWCasting
SLSTR	Sea and Land Surface Temperature Radiometer
SMHI	Swedish Meteorological and Hydrological Institute
SYNOP	Surface synoptic observations
TBC	To Be Confirmed
TBD	To Be Defined
VIIRS	Visible Infrared Imaging Radiometer Suite

See [RD.1.] for a complete list of acronyms for the SAFNWC project.

1.4 REFERENCES

1.4.1 Applicable documents

The following documents, of the exact issue shown, form part of this document to the extent specified herein. Applicable documents are those referenced in the Contract or approved by the Approval Authority. They are referenced in this document in the form [AD.X]

For dated references, subsequent amendments to, or revisions of, any of these publications do not apply. For undated references, the current edition of the document referred applies.

Current documentation can be found at SAFNWC Helpdesk web: http://www.nwcsaf.org

Ref	Title	Code	Vers	Date
[AD.1.]	NWCSAF Project Plan	NWC/CDOP3/SAF/AEMET/MGT/PP	1.5	15/04/21
[AD.3.]	Software Verification and Validation Plan for the SAFNWC/PPS	NWC/CDOP3/PPS/SMHI/MGT/SVVP NWC/CDOP3/EPSSG/SMHI/MGT/SVVP	1.1 1.0	17/10/18 15/04/21
[AD.4.]	NWCSAF Product Requirements Document	NWC/CDOP3/SAF/AEMET/MGT/PRD	1.4	02/06/21
[AD.5.]	System and Components Requirements Document for the SAFNWC/PPS	NWC/CDOP3/PPS/SMHI/SW/SCRD	2.3	12/10/21

Table 1: List of Applicable Documents

1.4.2 Reference documents

The reference documents contain useful information related to the subject of the project. These reference documents complement the applicable ones, and can be looked up to enhance the information included in this document if it is desired. They are referenced in this document in the form [RD.X]

For dated references, subsequent amendments to, or revisions of, any of these publications do not apply. For undated references, the current edition of the document referred applies

Current documentation can be found at SAFNWC Helpdesk web: http://www.nwcsaf.org

Ref	Title	Code	Vers	Date
[RD.1.]	The Nowcasting SAF Glossary	NWC/CDOP3/SAF/AEMET/MGT/GLO	1.0	20/10/20
[RD.2]	Algorithm Theoretical Basis Document for the Cloud Top Temperature, Pressure and Height of the NWC/PPS	NWC/CDOP3/PPS/SMHI/SCI/ATBD/CTTH	3.0	26/04/21
[RD.3.]	Product Validation report for the SAFNWC/PPS version 2008 (and 2.0)	SAF/NWC/CDOP3/SMHI-PPS/SCI/VR/1	2.1.1	19/03/08
[RD.5]	Algorithm Theoretical Basis Document for the Cloud Mask of the NWC/PPS	NWC/CDOP3/PPS/SMHI/SCI/ATBD/CloudMask	3.0	26/04/21
[RD.6]	Algorithm Theoretical Basis Document for Cloud Micro Physics of the NWC/PPS	NWC/CDOP3/PPS/SMHI/SCI/ATBD/CMIC	3.0	12/10/21

Table 2: List of Referenced Documents

1.5 DOCUMENT OVERVIEW

This document contains the scientific validation results for NWCSAF PPS v2021. After this introduction follows section 2 which lists and defines the verification measures used throughout the data analysis. Section 3 describes the satellite datasets and validation datasets (also happens to be satellite based) used and section 4 presents and discuss the results. Section 6 summarises and concludes, and a few scientific references cited are given in section 7. ANNEX A contains a list of still open TBCs and TBDs.

2 DEFINITION OF VERIFICATION MEASURES USED

In the following chapters we present the validation results using standard verification measures. Below we provide a short definition of each validation measure utilized, where 'validating truth' is the result of another product we are validating against. In the definitions below, we use cloud mask as an example, but the verification measures are applicable to additional PPS products.

We define N as the total number of observations, whereas A, B, C, and D are assigned numerical values based on statistical estimates described below.

	Validating truth Cloudy	Validating truth Cloud-free
PPS Cloudy	A	B
PPS Cloud-free	C	D

Bias:

$$\frac{1}{N}\sum_{k}(y_k - o_k)$$

Where the sum is over all N data pairs of PPS cloud cover (y) and the validating truth cloud cover (o).

Root Mean Square Deviation (RMS):

$$\sqrt{\frac{1}{N}\sum_{k}(y_k - o_k)^2}$$

Standard deviation (STD):

$$\sqrt{\frac{1}{N-1}\sum_{k}(y_k - o_k - Bias)^2}$$

Mean Absolut Error (MAE):

$$\frac{1}{N}\sum_{k}|y_k - o_k|$$

Inter Quartile Range (IQR):

$$\boldsymbol{Q_3 - Q_1}$$

where Q is quartile.

Hit rate (HR) also sometimes denoted PC (percent correct):

(A+D)/N

POD-*cloudy*:

A/(A+C)

FAR-*cloudy*:

B/(A+B)

POD-*clear*:

D/(B+D)

FAR-*clear*:

C/(C+D)

Hanssen-Kuiper Skill Score:

(AD - BC) / ((A+C) (B+D))

PEX:

Part of error larger than x (km).

It has to be emphasized here that the set of statistical scores presented above are obviously not all independent. For instance, in the binary case (cloud mask validation) the hit rate (HR) and the RMS are directly related through:

$$HR = 1 - RMS^2$$

This is because:

$$N * RMS^2 = \sum_k (y_k - o_k)^2$$

$$= C + B$$

$$= N - (A + D)$$

Even though there usually is such a tight interconnection between the different statistical measures we will be using all of them in the following to emphasize different aspects of the validation results.

3 DATA USED

3.1 IMAGER SATELLITE DATA USED

For the validation 15 orbits of S-NPP VIIRS, 99 orbits with AVHRR GAC data, 12 days of MODIS data was used, 920 granules of MERSI-2 data and 23 granules of AVHRR (Metop-B) was used.

3.1.1 AVHRR GAC data (NOAA-18)

We are using NOAA-18 GAC (Global Area Coverage) data from CLARA-A2 (The CM SAF Cloud, Albedo And Surface Radiation dataset from AVHRR data, second edition) see Karlsson et al. 2017 for more information on the CLARA-A2 dataset. The CLARA-A2 was produced using PPS-v2014. GAC data are averaged from 1km resolution data in a creative way. For each GAC pixel three lines times 5 columns of 1km data are used. The geolocation of the central pixels is saved as geolocation. Only data from the middle line of these three lines in original resolution are used. In this line the average of 4 pixels of a 5-pixel group is used to generate the new GAC pixel. In practice this means that when matched to a truth, it will happen that the truth and the GAC-pixel have FOV that does not overlap at all, even if matched very close in both time and space. Generally clouds features are much larger than a few kilometres, which makes GAC matchup datasets possible to use for validation despite the limitations. For the validation we have included 99 orbits of AVHRR GAC from NOAA-18 from 2006 to 2009.

3.1.2 VIIRS data (Suomi-NPP)

The Suomi National Polar-orbiting Partnership (S-NPP) spacecraft was launched successfully in late October 2011. The largest of its five payloads is the Visible Infrared Imaging Radiometer Suite (VIIRS). This visible/infrared radiometer features 22 spectral bands with 0.371 km nadir resolution for the five imager resolution bands and 0.742 km nadir resolution for the Day/Night band and the moderate resolution bands.

To account for along-track distortions, overlapping pixels are removed (so called 'Bowtie Removal'). Problems related to across-track distortions are avoided by pixel aggregation. The pixels are at nadir angle rather oblong and more quadratic at swath edges. Three oblong pixels close to nadir are aggregated into one more quadratic pixel. At the swath edge no pixels are aggregated, and in between two pixels are aggregated together. This makes the pixels more equally sized across the swath. As a positive side effect, the aggregation increases the signal to noise ratio near nadir.

For the processing of PPS cloud products, the AVHRR-heritage channels (0.6, 0.8, 1.6, 3.7, 11 and 12 microns) and the 8.5 and 1.3 micron channel are used. For the validation we have included 15 orbits of Suomi-NPP data from 2015.

3.1.3 MODIS data (EOS-Aqua)

For the processing of PPS cloud products, the AVHRR-heritage channels (0.6, 0.8, 1.6, 3.7, 11 and 12 microns) and the 8.5- and 1.3-micron channel are used. Note that due to increased noise for channel 8.5-micron channel this channel is switched off in default configuration to avoid stripy products.

We have also included 12 days of MODIS from EOS-Aqua the 1st every month 2010.

3.1.4 MERSI-2

For the processing of PPS cloud products, the AVHRR-heritage channels (0.6, 0.8, 1.6, 3.7, 11 and 12 microns) and the 8.5- and 1.3-micron channel are used. The validation data for MERSI-2 consist of 920 granules from 2020 February, April, May and June.

MERSI-2 has problems with noise for channel 12micron and therefore we use 11micron and 3.7micron for the CTTH.

3.1.5 EPS-SG

For the EPS-SG comparison 20 granules from version 1 of the test data was used.

3.1.6 AVHRR (Metop-B)

For AVHRR Metop-B 23 granules from 2015 December was used. Note that Metop-B only has colocations with CALIOP and CPR (Cloudsat) for high latitudes and cannot be used for global validation.

3.1.7 SLSTR (Sentinel 3a)

A granule of sentinel is included as an example showing the demonstrational functionality for SLSTR.

3.2 NWP DATA

For the validation forecasted NWP data from ECMWF is used. Data for all pressure levels available were included in the data: 91 levels up to 2013-06-30 and 137 levels after that. The 30 to 48 top most layers are not used.

For validation using MODIS, MERSI-2, AVHRR (Metop-B) and VIIRS data forecast lengths of 6-15h were used. Forecast valid times differed at most 2.0h from the time of the first scan line in the swath was used. Sea-ice information was not used in the validation as it was shown in the v2018 validation that it does not affect the overall results. Spatial resolution of 0.5 degrees was used except for VIIRS data for which a 0.25 degree grid was used.

For GAC validation NWP from ERA5 was used with at all hours.

For GAC validation 2018 NWP analysis data from ERA-interim at 6, 12 18 and 24h where used. Forecast valid times differed at most 3h, from the time of the middle scan line in the swath.

NWP snow data was also used for all cases. It is necessary for the cloud mask probability product.

3.3 THE SYNOP DATA

The PPS cloudmask has been validated against global SYNOP (surface synoptic observations) reports, using global S-NPP VIIRS data. The SYNOP data used have been acquired from DWD, and kindly provided by Martin Stengel and Anke Kniffka.

For the SYNOP data we use a number of pixels closer than 10km to the SYNOP station. For GAC at most 16 and for VIIRS/MODIS at most 250 pixels where used. The time difference between the VIIRS pixel and the SYNOP report time is allowed to deviate by up to 30 minutes.

The geographical distribution of Satellite-SYNOP collocations based on the S-NPP VIIRS dataset is shown in Figure 1. The distribution is global but there is an obvious concentration of SYNOP matchups in central and northern Europe, including Germany.

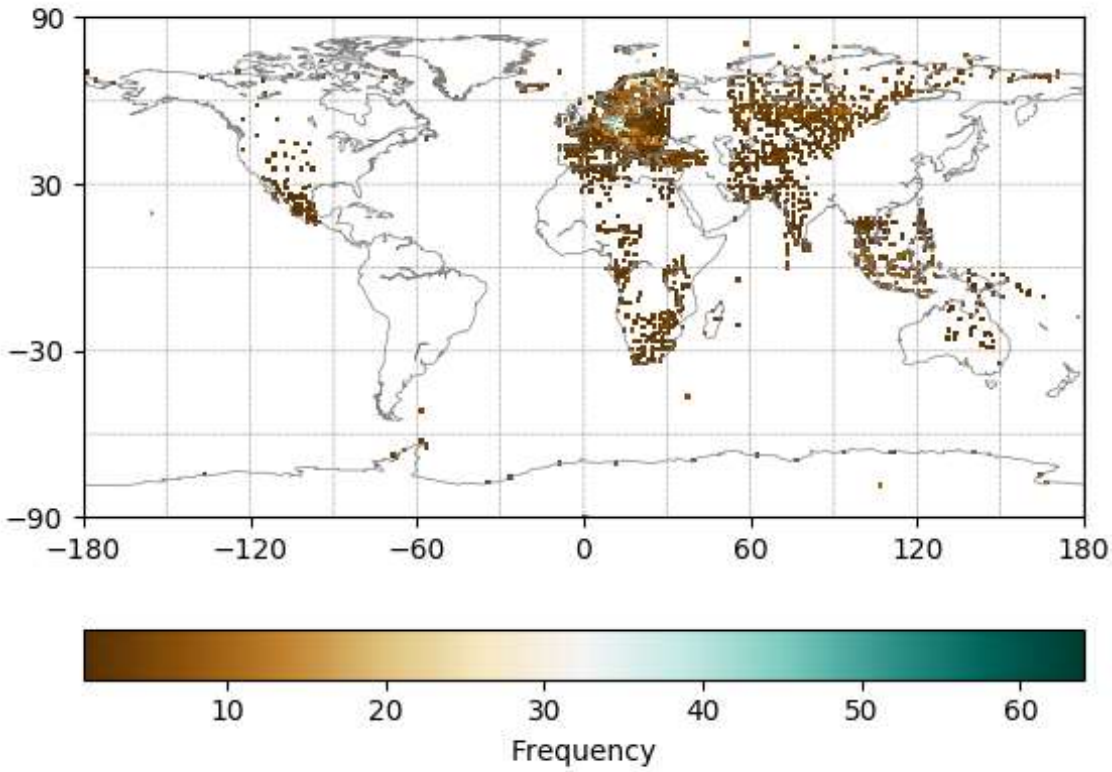

Figure 1: Global map showing the location and frequency of all VIIRS and SYNOP co-locations, based on the 15 S-NPP orbits from 2015.

3.4 THE CALIPSO DATASET

Several active and passive satellite sensors are flying in a formation called the A-train. One satellite flying in the A-train is the Cloud-Aerosol Lidar and Infrared Pathfinder Satellite (CALIPSO) that was launched in April 2006. The CALIPSO payload consists of three nadir-viewing instruments: Cloud-Aerosol Lidar with Orthogonal Polarization (CALIOP), the imaging infrared radiometer (IIR), and the wide field camera (WFC). We have used data from the CALIOP instrument for the MODIS/AVHRR/VIIRS-CALIPSO comparison presented in the study.

MODIS on EOS-Aqua is also in the A-train and all orbits have matches with CALIOP. Fortunately, there is an overlap of both the VIIRS data record and the AVHRR data record with the data from CALIPSO. Though their orbits differ, the orbits of the three afternoon satellites Suomi NPP and NOAA-18 do align periodically with the A-train formation. Where the Suomi NPP orbital plane is well maintained and stable over the lifetime of the satellite this is not the case with the NOAA satellites. However, until around 2010 the NOAA-18 satellite orbit was rather well aligned with CALIPSO, see plot of the equatorial crossing times in Figure 2.

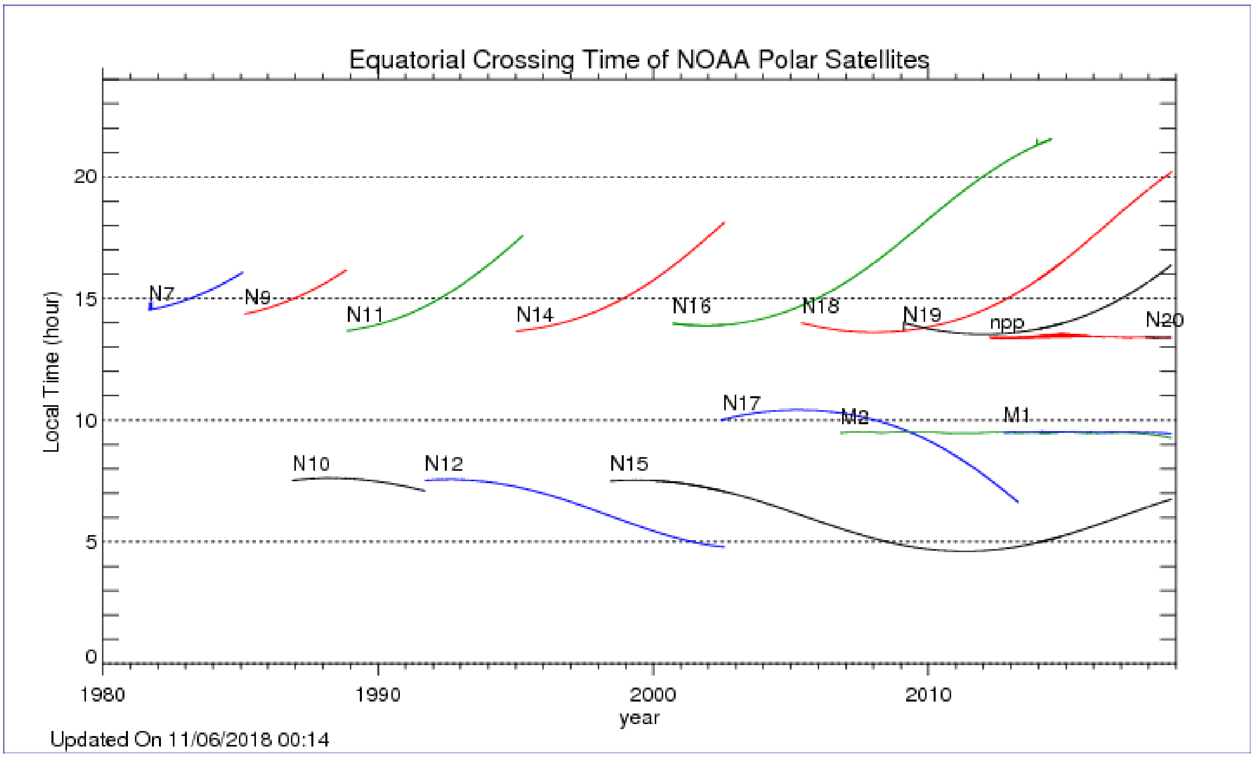

Figure 2: Equatorial crossing times of the NOAA, Metop and SNPP/JPSS spacecrafts from NOAA-7 till todays NOAA-20 and Metop-B.

In this report we have investigated 15 orbits of matching Suomi NPP overpasses with CALIPSO observations from 2015. We have also included 24 days scenes of matching MODIS and CALIPSO data. A match here is defined successful if observations at one position by both the Suomi NPP and CALIPSO satellite were performed within a +/- 3 minutes time-window. No correction for the parallax has been attempted, as the error introduced by ignoring the parallax effect is assumed small over this dataset. As can be seen from Figure 3 the AVHRR/VIIRS observations are if not close to nadir, then at least with rather low zenith angles, indicating that parallax effects should be relatively small on average. See also discussion below under co-location criteria.

We have also matched and analyzed 99 GAC orbits of AVHRR for NOAA-18, year 2006-2009, with CALIPSO data.

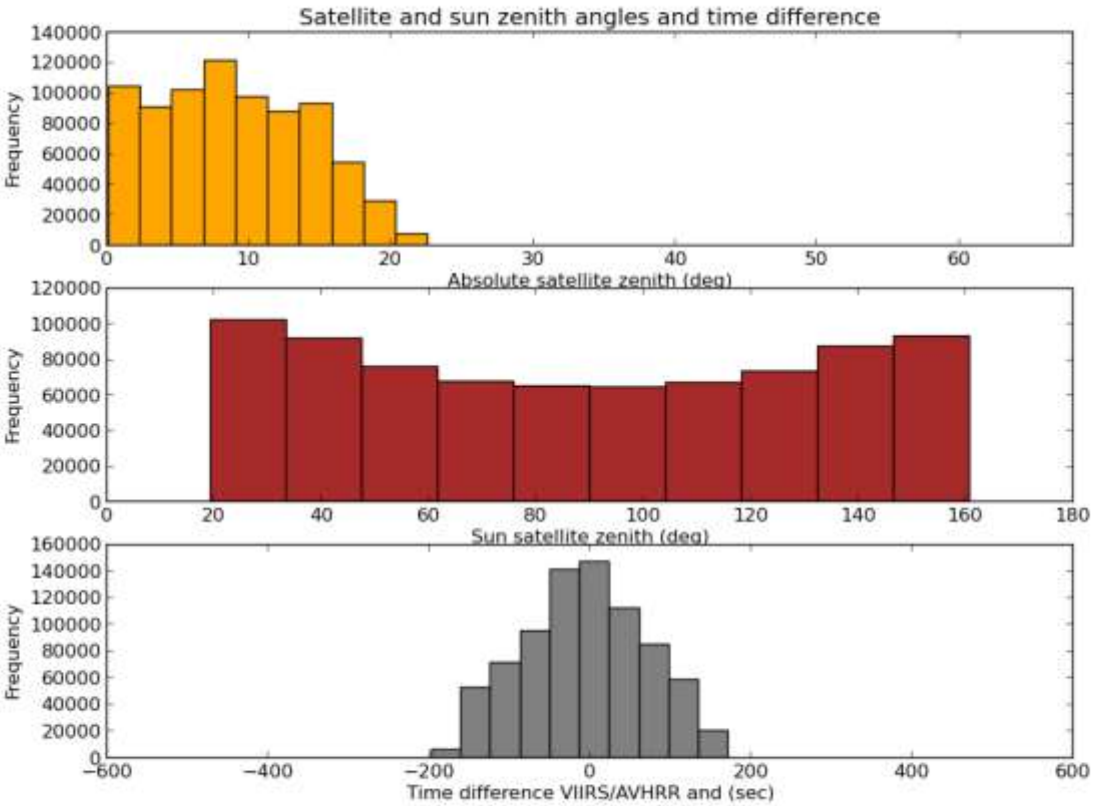

Figure 3: Distribution in terms of satellite zenith (upper panel), sun zenith angle (middle panel) and time difference (lower panel) of CALIPSO-AVHRR colocations from the 2006-2009 NOAA-18 GAC data record. Observations span night and day with an almost flat distribution for night, twilight and day, and with the majority of observations having a near nadir AVHRR view

We use the CALIOP (Winker et. al 2016) version 4 cloud layer data product from NASA Langley produced at 1km horizontal and 30m vertical resolution to quantify cloud fraction and cloud height (and some additional information from the 5km product for data filtering). For the GAC orbits we use the 5km data, combined with information from the single shot data. The single shot data is also included in the 5km files, see Karlsson et. Håkansson 2018 for more details. These data were obtained from the NASA Langley Research Center Atmospheric Science Data Center. More information on CALIPSO can be found at http://www-CALIPSO.larc.nasa.gov/. In the following we use the term CALIPSO synonymous with the CALIOP instrument on CALIPSO.

Technical details about the CALIOP datasets:

Archive Center:

Atmospheric Science Data Center archive center details

Distributing Center:

* NASA Langley Research Center Atmospheric Science Data Center
* http://eosweb.larc.nasa.gov
* Contacts: larc@eos.nasa.gov

ShortNames:

1. CAL_LID_L2_01kmCLay-Standard-V4-10
2. CAL_LID_L2_05kmCLay-Standard-V4-10
3. CAL_LID_L2_05kmALay-Standard-V4-10

Version:

V4-10 for data from 2020 we use version V4-20

Descriptions:

1. CALIPSO Lidar Level 2 1 km cloud layer data
2. CALIPSO Lidar Level 2 5 km cloud layer data

Co-location criteria with VIIRS/AVHRR/MODIS/MERSI2:

The CALIOP pixel is co-located with the nearest VIIRS/AVHR/MODIS pixel. Since CALIOP is a nadir viewing instrument with 70x333 m wide sampling, the MODIS and VIIRS pixel in moderate resolution (which is used in general) covers an area most comparable to the aggregated 1 km CALIOP product. However, note that the instruments have different FOVs and CALIOP is only seeing part of the MODIS/VIIRS pixel. Additional uncertainties are navigation uncertainties, typically of less than a pixel, and parallax effects for VIIRS/AVHRR/MODIS, which are not being corrected for. The theoretical maximum displacement due to a VIIRS pixel for a 10 km high cloud at the outer part of the swath (which corresponds to a scan-angle as large as 56°) can be on the order of 15 km. However since the ground track of CALIPSO approximately coincide with those of S-NPP, EOS-Aqua, and NOAA-18 (at least during the time period used herein[1]) a more typical displacement would be for a viewing angle of less than 10 degrees, and thus be below 2 km. All co-locations between VIIRS/AVHRR and CALIOP are made within a 3-min time window. Thus, any VIIRS/MODIS/AVHRR pixels with less than +/- 3-min separation from CALIPSO are retained and analyzed in the statistics below. For MERSI.2 +/- 10 minutes are allowed.

Adapt for detectability differences:

Active (like CALIOP) and passive instruments (like VIIRS) feature different detectability of hydro-meteors. This means that results are expected to differ dependent on which instruments data is considered regardless of the performance of the algorithm used. In short: differences between the validation truth and the PPS results are expected because there is systematic difference between the instruments.

The aim of this report is to validate the applied algorithm, not the instrument. Therefore, along with unfiltered results for cloudmask, results where all clouds optically thinner than 0.2 are excluded are presented. Karlsson and Håkansson (2018) found that an optical thickness of 0.225 can be considered the average optical detection limit for PPS on AVHRR. In lack of other similar studies, we regard this as the optical detection limit also for VIIRS and MODIS data.

Variables used:

For cloud mask and cloud probability pixels with a layer_top_altitude that is not nodata is used as cloudy. For GAC (5km) also pixels where more than half of the corresponding single-shot pixels are cloudy are treated as cloudy even if the 5km number_layers_found is zero. For cloud top height validation layer_top_altitude is used. The feature_classification_flags are used for cloud phase and cloud type. DEM_elevation is used to transform PPS-height to meter over sea-level.

[1] As seen from Figure 2 NOAA-18 is subject to severe orbital drift after around 2010

3.5 The CPR (CloudSat) dataset

Also CloudSat (CLOUD SATellite) flies in the A-train. CloudSat has a Cloud Profiling Radar, with acronym CPR (CloudSat), which gives a vertical profile of clouds. The 2B-GEOPROF dataset (Marchand et al., 2008) can be used for cloud height validation and the 2B-CWC-RVOD (Austin et al., 2009) dataset can be used for a cautious LWP validation. The horizontal resolution of CPR is 1.4km x 3.5km. Because of ground clutter the CPR misses a lot of low clouds and cannot be used for cloudmask validation. Especially FAR-cloudy would be miss-leading.

Technical details about the CloudSat 2B-GEOPROF dataset:
Archive Center:
NSIDC archive center details
Distributing Center:
* the CloudSat Data Processing Center
* http://www.cloudsat.cira.colostate.edu/order-data
* Contacts: http://www.cloudsat.cira.colostate.edu/contact

2B-CWC-RVOD

- Version 5 for VIIRS
- Description: Contains LWP retrievals.

2B-GEOPROF

- Version 5
- Description: Contain profile and height information on clouds.

Co-location criteria with VIIRS/AVHRR/MODIS:

The CPR pixel is co-located with the nearest VIIRS/AVHR/MODIS pixel a time difference of +/- 3 minutes is allowed.

3.6 The AMSR-E dataset

For this study the LWP product (Wentz et al, 2004) of the Advanced Microwave Scanning Radiometer for EOS (AMSR-E) onboard the AQUA platform has been used to provide an independent dataset to compare the CMIC LWP against. The passive microwave measurements provided by AMSR-E provide a somewhat more direct means of estimating the LWP compared to what can be achieved by the VIIRS/AVHRR/MODIS based CMIC products. The coarse spatial resolution of the AMSR-E channels and other obvious limitations mentioned later (see 4.5.1) are of course important limiting factors when searching for a method to validate the CMIC LWP product. The AMSR-E data does not provide any ground truth. AMSR-E data for 2010 (January, April, July and October) and AMSR-E data to match the AVHRR GAC dataset was used.

Technical details about the AMSR-E dataset:
Archive Center:
NSIDC archive center details
Distributing Center:
* NSIDC National Snow and Ice Data Center
* http://nsidc.org
* Contacts: nsidc@nsidc.org
ShortName:
AE_Ocean
Version:
2

Description:

The AMSR-E/Aqua Level-2B ocean product includes Sea Surface Temperature at 56 and 38 km, near-surface wind speed at 38 and 21 km, column water vapor at 21 km, and columnar cloud liquid water at 12 km, generated by the Wentz algorithm using Level-2A TBs.

Co-location criteria with VIIRS/AVHRR/MODIS:

The average of the MODIS or AVHRR pixels within the AMSR-E footprint is compared to the AMSR-E estimate, which has a resolution of approximately $12km^2$. All time differences are less than 3minutes. Possible parallax effects are not considered and can be up to the order of 1 AMSR-E pixel displacement. The up to eight closest pixels are matched to the AMSR-E footprint, the centre of all included pixels are closer than 5.4km from the centre of the AMSR-E footprint. For GAC data only the five closest neighbours are considered. An average CMIC LWP is calculated for pixels that full fill: sea, sun-zenith angle less than 72 degrees, liquid phase 0<=CMIC LWP<3000 g/m^2. If the AMSR-E pixel LWP is between 0 and 170 g/m^2 the matching result is included in the analysis.

3.7 The MODIS Collection 6.1 L2 DATASET

We have chosen to make also an inter-comparison of the CTTH, CMa and Cloud phase with the corresponding official product MYD06_L2 (Ackerman et al., 2015) from the MODIS team at NASA Langley.

Technical details about the MODIS LWP dataset:

Archive Center:

LAADS archive center details

Distributing Center:

* LAADS Level 1 and Atmosphere Archive and Distribution System
* http://ladsweb.nascom.nasa.gov
* Contacts: modapsuso@sigmaspace.com

ShortName:

MYD06_L2

Version:

Collection 6.1

Description:

Aqua Atmosphere Level 2 Cloud product.

Co-location criteria:

MODIS-C6.1 data are collocated in the same way as the PPS-MODIS data.

4 RESULTS AND DISCUSSION

In this section, the comparison results are shown separated by cloud product (cloud mask, cloud type and cloud top temperature and height, cloud physical properties, and cloud probability). Throughout this section yellow marks indicate where we have requirements, green marks indicate within target accuracy and red marks indicate outside threshold accuracy. This means that yellow boxes with measures marked red means that the respective requirements is not met by PPS-v2021. PPSv2021 meets all requirements and there are no such boxes. Datasets expected to have bad scores due to sensor differences or difficult conditions (polar night) for which the requirements are not intended, are marked with pink. In some cases, results for MODIS-C6.1 validated with the same methods are included as a reference point (these reference data are marked with orange).

4.1 Cloud Mask

4.1.1 SYNOP validation

For the SYNOP validation a mean cloud fraction is calculated for PPS for the pixels closest to the SYNOP station. Only pixels within 10 km are used. For VIIRS the maximum number of pixels used is limited to the 250 nearest pixels. All cases where SYNOP or PPS have cloud fraction cover close to 50% (between 25% and 75%) where excluded. The results of the global validation using the archive of 15 S-NPP orbits are presented below in Table 3. For all sun illuminations (first row) the data corresponds to the colocations shown in Figure 1. The data are also stratified according to sun illumination. See [RD.5] for the definition of day, night and twilight.

Table 3: Cloud mask validation scores for 15 S-NPP orbits against global SYNOP. Results are global, results for Europe only are included in a separat line. Europe here is defined as latitude between 35 and 72 and longitude between -25 and 60 degrees. Green: within target accuracy, red: outside threshold accuracy. Yellow marks measures with requirements.

	Bias (%)	Hit rate	POD cloudy	FAR cloudy	POD clear	FAR clear	N
All	1.3	0.94	97.2	4.4	83.9	10.7	8614
Day	1.3	0.95	98.0	3.5	83.9	9.8	6388
Night	2.9	0.87	91.8	12.4	80.9	12.9	1279
Twilight	-0.4	0.96	97.0	2.4	91.1	10.6	947
Europe	1.9	0.96	98.6	3.7	82.8	7.3	6319
Threshold Accuracy			> 85	< 20			
Target Accuracy (Europe)			> 90 (95)	< 15 (10)			
Optimal Accuracy (Europe)			> 95 (98)	< 10 (5)			

The target accuracy is reached for the global NPP data validation for all conditions. Note that these results are not valid on pixel level. To investigate pixel level performance a truth with finer horizontal resolution (like CALIOP) is needed. For the AVHRR, MODIS and MERSI-2 sensor we will do CALIOP validation only. CALIOP validations have better global coverage and give more information on pixel level. Generally worse scores are expected for validation with CALIOP as there are no averaging done and CALIOP detects thinner clouds. This means that if requirements are met in CALIOP validation, it is reasonable to assume that they are met also compared to SYNOP.

4.1.2 CALIOP validation

It must be stated that accuracy requirements for cloudmask are defined for comparison against European SYNOP stations however comparison with CALIOP allows performance evaluation on pixel level. In Table 4 global accuracy measures for the cloudmask as compared to CALIOP are presented. Note for the GAC data orbits from 2009 were excluded as these where used in the tuning process. Clouds with thickness below 0.2 are excluded when evaluating POD-cloudy. For the FAR-cloudy target accuracy is reached for all cases. POD-cloudy threshold or target accuracy is reached for all cases (AVHRR, VIIRS, MODIS, MERSI-2). The “no polar night” category is defined as all data where sun satellite angle is above 95 and absolute latitude above 70 is excluded. The requirements are specified to hold globally except for polar night. Compared to PPSv2018 POD-cloudy is increased, and POD-clear is decreased, HR is not affected. In Table 5 results presented for day, night and twilight are shown. Results are generally best for day time. In twilight for MERSI-2 POD-clear is a bit lower and FAR-cloudy is outside threshold accuracy. This is likely caused by the noise for the 12micron channel. Results for Metop-B night are also lower but this is expected due to polar-night conditions. In Table 6 results are shown for an area covering most of Europe. The area was defined as latitude between 35 and 72 and longitude between -25 and 60 degrees. Here target accuracy is met for FAR-cloudy and threshold accuracy for POD-cloudy. In Table 7 validation scores for the snow detection of PPS cloud mask are shown. There are no requirements on the snow detection, but good snow detection increase the accuracy of the cloudmask. Note that for VIIRS the SYNOP validation has better scores than the CALIOP validation. This means that as MERSI-2, AVHRR and MODIS meet requirements compared to CALIOP they also meet requirements compared to SYNOP where better scores are expected.

Table 4: Cloud Mask global validation for PPS-v2021. Green: within target accuracy, red: outside threshold accuracy. Yellow marks measures with requirements. Measures are intended for SYNOP

data, we do not expect measures to be met for very thin clouds. For the "no polar night" category all data where sun satellite angle is above 95 and absolute latitude is above 70 is excluded.

CALIOP comparison	BIAS %	HR	K	POD-cloudy %	POD-cloudy% Filt 0.2	FAR-cloudy %	POD-clear %	FAR-clear %	N
S-NPP VIIRS data global 15-16 orbits									
PPS-NPP 2018 (all)	-2.5	0.86	0.71	87.6	89.4	9	83.3	22.3	623052
PPS-NPP 2021 (all)	-2.1	0.86	0.71	88.0	89.8	9.2	83.1	21.7	643585
2021 no polar night	-0.6	0.87	0.72	90.0	91.6	9.3	82.3	19.1	590023
EOS-Aqua global data 12 days									
PPS-MODIS v2018	-4.5	0.87	0.74	87	89	6.8	87	23.5	6557753
PPS-MODIS v2021	-3.9	0.87	0.72	87.1	89.2	7.6	85.4	23.6	6504273
2021 no polar night	-2.3	0.88	0.74	89.4	91.3	7.5	84.6	20.9	5962319
MODIS-C6.1 L2	-3.5	0.87	0.73	87.7	89.7	7.6	85.3	22.8	6504358
NOAA-18 AVHRR GAC data 66 orbits (2006-2008)									
GAC 2018	-9.4	0.82	0.66	81	88.2	6.9	85	36	504627
PPS-GAC 2021	-8.5	0.82	0.66	81.7	89	7.2	84.2	35.1	497759
2021 no polar night	-6.5	0.84	0.68	84.3	91.1	7.2	83.8	31.8	456690
Feng-Yun–3 MERSI-2 data 920 granules 2020									
PPS v2021	3.3	0.82	0.57	88.8	90.3	15.4	68.4	24.2	1358616
2021 no polar night	4.6	0.84	0.6	91.5	92.8	14.3	68.4	20.4	1219232
Metop–B AVHRR global metop 23 granules 2015 (mostly twilight polar)									
PPS v2021	-5.1	0.78	0.55	77.8	80.4	15.2	77.7	31.4	21821
2021 no polar night	-3.6	0.85	0.69	85.1	87.2	9.8	84	23.6	13529
Requirement Accuracy (global)									
Threshold accuracy				85 %		20 %			
Target accuracy				90 %		15 %			
Optimal Accuracy				95 %		10 %			

Table 5: Cloud Mask global validation for PPS-v2021. Green: within target accuracy, red: outside threshold accuracy. Yellow marks measures with requirements. We do not expect good results for polar night and CALIOP 5km thin clouds (pink).

Observed Accuracy --- Global data									
	BIAS %	HR	K	POD-cloudy%	POD-cloudy% Filt 0.2	FAR-cloudy %	POD-clear %	FAR-clear %	N
S-NPP VIIRS data global 15-16 orbits									
NPP day	2.5	0.88	0.75	92.6	93.9	11	82	12.4	291239
NPP night	-6.5	0.85	0.7	83.9	86.3	7.2	85.9	28.9	294164
NPP twilight	-2.5	0.84	0.61	87.8	89.2	9.1	73.5	33.2	58182
EOS-Aqua global data 12 days									
MODIS day	-0.2	0.9	0.78	92.1	93.4	7.6	85.4	15	2850359
MODIS night	-7	0.85	0.7	83.6	86.5	6.9	86.5	29.4	3056085
MODIS twilight	-5.5	0.81	0.62	81.6	83.3	11.0	80.3	30.9	597829
NOAA-18 AVHRR GAC data 66 orbits (2006-2008)									
GAC day	-4.8	0.85	0.7	85.5	92.9	7.9	84.7	26.3	217910
GAC night	-11.8	0.8	0.64	78.8	86.5	6.2	85	41.6	235480
GAC twilight	-9.4	0.79	0.57	80.2	85.3	8.5	76.5	45.1	44369
	Feng-Yun – 3 MERSI-2 data 920 granules 2020								
MERSI-2 day	8.3	0.85	0.62	94.9	95.8	16	67.3	12.1	575349
MERSI-2 night	-2.6	0.8	0.58	83.2	85.3	13.4	74.4	31	638648
MERSI-2 twilight	9.7	0.76	0.34	90.1	90.7	20.8	44.1	34.6	144619
	Metop – B AVHRR global metop 23 granules 2015 (mostly twilight polar)								
AVHRR day	-1.9	0.94	0.82	85.5	94.2	8.5	96.9	5.6	269
AVHRR night	-9	0.66	0.32	63.2	66.0	25.5	68.9	43.5	8535
AVHRR twilight	-2.6	0.85	0.7	86.6	88.7	9.8	83.4	22.1	13017
Requirement Accuracy (global)									
Threshold accuracy				85 %		20 %			
Target accuracy				90 %		15 %			
Optimal Accuracy				95 %		10 %			

Table 6: Cloud mask validation for PPS-v2021 over Europe compared to CALIOP. Europe here is defined as latitude between 35 and 72 and longitude between -25 and 60 degrees. Green: within target accuracy, red: outside threshold accuracy. Yellow marks measures with requirements.

Observed Accuracy --- Europe									
	BIAS %	HR	K	POD-cloudy %	POD-cloudy% Filt 0.2	FAR-cloudy %	POD-clear %	FAR-clear %	N
	S-NPP VIIRS data 15-16 orbits								
PPS 2018 (all)	-1.5	0.91	0.75	93.9	94.4	4.4	81.2	24.8	22521
PPS 2021 (all)	-0.6	0.92	0.76	94.9	95.5	4.4	80.9	21.4	22520
	EOS-Aqua data 12 days								
PPS-MODIS v2018	-1.1	0.91	0.80	93.1	94.7	5.4	86.9	16.3	323275
PPS-MODIS v2021	-0.2	0.90	0.77	93.0	94.7	6.7	83.7	17	319107
	NOAA-18 AVHRR GAC data 66 orbits (2006-2008)								
GAC 2018	-4.1	0.86	0.70	88.0	94.2	6.8	82.4	28.6	25741
GAC 2018	-3.8	0.86	0.69	87.8	94.3	7.4	80.8	29.2	25316
	Feng-Yun – 3 MERSI-2 data 920 granules 2020								
PPS 2021 (all)	1.7	0.89	0.75	92.9	93.9	9.4	81.7	14.1	68590
	Metop-B AVHRR								
PPS 2021 (all)	3.6	0.94	0.27	98.7	99.2	4.9	28.1	39	1352
Requirement Accuracy (global)									
Threshold accuracy				85 %		20 %			
Target accuracy				95 %		10 %			
Optimal Accuracy				98 %		5 %			

Table 7: Cloud Mask snow validation for PPS-v2021 compared to NSIDC flag in CALIOP data.

	POD-snow %	FAR-snow%	FAR-snow not clouds %	N
PPS-NPP (day)	71.7	14.9	1.6	217910
PPS-NPP (twilight)	36.5	32.1	1.7	44369
PPS-MODIS (day)	70.0	18.0	5.0	2850359
PPS-MODIS (twilight)	46.5	26.8	5.2	597829
GAC (day)	76.7	34.2	4.2	217910
GAC (twilight)	46.8	45.3	7.2	44369
PPS-MERSI-2 (day)	50.0	10.7	2.2	575349
PPS-MERSI-2 (twilight)	22.3	18.5	0.9	144619
PPS-Metop-B (day)	95.3	5.6	0.0	269
PPS-Metop-B (twilight)	61.6	14.4	0.0	61.6

4.1.3 Preliminary EPS-SG validation

During the development of PPSv2021 20 granules of the EPS-SG testdata (V1) were processed and validated. The validation was performed by Loredana Spezzi. The cloud mask used as input to the simulation of the test data was used as truth. Results for PPS were well within target accuracy and PPSv2021 gives a CMa product that is identical for 99% of the pixels to the CMa product used in the comparison. This means that PPSv2021 is on track to meet requirements for EPS-SG.

Table 8: Validation results for a preliminary version of PPS v2021 for the V1 of the EPS-SG test data. The truth is the cloud mask used as input for the simulation.

	Hit rate	POD cloudy	FAR cloudy	POD clear	FAR clear	N
All	93	95	7	86	13	13204800
Threshold Accuracy		> 85	< 20			
Target Accuracy		> 90	< 15			
Optimal Accuracy		> 95	< 10			

4.2 Cloud Type

For a validation of the cloud type, we use the classification provided in the CALIPSO data (at 30 m vertical resolution). CALIPSO cloud types can be condensed into three height classes, which are low- (pressure > 680 hPa), medium- (pressure 440-680 hPa) and high-level (pressure < 440 hPa) clouds. These levels correspond more or less (not exactly because of slightly differing bins) to the PPS cloud classes: Low level clouds (including fractional), medium level clouds and high + very high-level clouds. Note that in PPS-v2021 the fractional class is only used for low level fractional clouds. However, the PPS cirrus class corresponds to CALIOP medium- and high-level semi-transparent clouds. In Table 9 what matches of classes are treated as successful are shown and marked with (X). We consider it ok that some high transparent cirrus clouds are classified as high opaque. But we don't want the deep convective or the altostratus (op) classed as semi-transparent (cirrus). If CALIPSO can't see through the top layer neither should the Imager. To treat CALIPSO-altocumulus/PPS-cirrus as correct matches may be questionable. But on the other hand, we can have thin altocumulus. Convective high clouds can have a cirrus layer on top. This means that some of the PPS-cirrus/CALIPSO-convective is not wrongly classified however they will be treated as misclassifications.

Table 9: Overview of PPS and CALIOP cloud type matches that are considered to be correct marked with (X).

		PPS cloud types				
CALIOP cloud types		*Low*	*Low (frac)*	*Medium*	*High*	*Cirrus*
		Low		*Medium*	*High*	*Cirrus*
CALIOP low	*low overcast (tp)*	*X*	*X*			
	low overcast (op)	*X*	*X*			
	transition stratocumulus	*X*	*X*			
	low broken cumulus	*X*	*X*			
CALIOP medium	*altocumulus (tp)*			*X*		*X*
	altostratus (op)			*X*		
CALIOP high	*cirrus (tp)*				*X*	*X*
	convective (op)				*X*	

In Table 10 results for the cloud type validation is shown. Note that the GAC data are not included as PPS cloud type is traditionally not used by GAC users. Note that performance for AVHRR is expected to be the same as for MODIS/VIIRS for global data. For PPS-v2021 threshold accuracy are reached for all cases (VIIRS, MODIS, AVHRR, MERSI2) and all illumination conditions day, night and twilight. For all cases except FAR Medium (MODIS, MERSI-2) also target accuracies are met. Note that for low clouds both POD and FAR even reach optimal accuracies, except for MERSI-2. One needs to keep in mind that evaluation was originally planned against interactive targets, not CALIOP data. However, it is not certain that comparison with interactive targets will show better or fairer results. In Table 11 it is shown, for each CALIOP cloud type, the percent of classifications for each PPS class. We have chosen to not separate the low and very low classes for PPS, despite that the very low class is aiming at detecting clouds less than 500m above ground. This means that some clouds in this class could be medium or high clouds for CALIOP and still in reality be correctly classified. However, they should not be a large percentage of clouds. The very low cloud category has a POD of around 8% and a FAR of 1%. This means that almost all clouds detected as very low truly are very low. But many clouds with height below 500 m are also classed as low or fractional by PPS, which is ok. PPS fractional is considered a true match with all CALIOP low classes. This makes sense because what is fractional for the PPS FOV (1km x 1km) does not need to be fractional for the CALIOP FOV which is much smaller (70x330m).

Table 10: Cloud type validation compared to CALIOP. See Table 9 to see which classifications are considered successful/unsuccessful. Green: within target accuracy, red: outside threshold accuracy. Yellow marks measures with requirements.

	POD Low (%)	**POD** Medium (%)	**POD** High (%)	**FAR** Low (%)	**FAR** Medium (%)	**FAR** High (%)	**FAR** Cirrus (%)	N
PPS VIIRS v2018	90.8	66.0	76.2	9.6	39.2	8.5	30.1	346199
PPS VIIRS v2021	90.7	66.4	76.3	9.6	38.6	8.5	29.6	371660
PPS VIIRS (day)	89.3	62.5	75.9	6.3	37.9	11.0	36.4	164658
PPS VIIRS (night)	93.0	72.4	78.6	12.7	39.4	5.4	19.9	168686
PPS VIIRS (twilight)	88.0	61.8	66.1	11.9	38.3	13.0	42.9	38316
PPS MODIS v2018	92.4	67.8	77.0	10.3	42.1	6.7	25.7	1958146
PPS MODIS v2021	92.4	67.8	77.0	10.2	42.1	6.7	25.2	2025206
PPS MODIS (day)	92.4	68.1	77.1	7.6	37.5	8.6	33.3	925450
PPS MODIS (night)	91.1	64.9	77.0	12.6	48.4	3.7	15.4	930426
PPS MODIS (twilight)	94.4	73.2	78.2	12.3	35.8	14.4	38.1	169330
MERSI-2 2021	62.4	64.7	84.9	7.3	57.9	20.3	29.3	956562
AVHRR Metop-B v2021	88.2	70.6	79.9	22.0	27.3	17.0	18.6	12047
Required Accuracy			Threshold			Target		Optimal
POD			50%			70%		80%
FAR			60%			40%		20%

Table 11: Confusion table of PPS v2021 cloud type version 2021 compared to CALIOP for VIIRS.

		PPS type					
		N	Low	Frac	Medium	High	Cirrus
	CALIOP type:						
Low	low overcast (tp)	18311	**70.9**	**23.4**	3.0	0.1	2.5
	low overcast (oq)	37562	**86.3**	**7.5**	3.4	0.2	2.6
	transition stratocumulus	66864	**47.3**	**43.6**	5.7	0.3	3.1
	low broken cumulus	16355	**33.6**	**45.3**	10.3	0.9	9.9
Medium	altocumulus (tp)	31320	8.8	6.2	**34.7**	5.3	**45.0**
	altostratus (oq)	33629	3.5	0.5	**54.0**	7.9	34.1
High	cirrus (tp)	110016	3.7	3.1	9.0	**15.2**	**69.1**
	deep convective (op)	57603	0.0	0.0	1.8	**61.1**	37.1

4.3 Cloud Top Height

In this section, the performance of the cloud top height algorithm is investigated. PPS-v2018 had a completely new neural network cloud top temperature, pressure and height algorithm (see Figure 5 and more details in Håkansson et al. 2018 and RD.2). Cloud top height is traditionally validated with bias and STD. As stated in the PRD for non-Gaussian error distributions we use median and MAE to evaluate the performance, as the error distributions are inadequately described by STD and bias (Håkansson et al. 2018). We will include also IQR, PE0.25, PE0.5 and PE1.0. The STD and bias requirements are translated to these new measures. We calculated threshold, target and optimal accuracy for PE0.25, PE0.5, PE1.0, and IQR using a normal distribution with zero bias and corresponding threshold, target or optimal STD (these measures will therefore be stricter than the bias/STD). For median a reasonable performance is to match the requirements for bias. For MAE limits are calculated using the corresponding bias and (STD) for a Gaussian distribution. So for MAE we create Gaussian distributed data with optimal bias and STD. We measure MAE for that data and that will be the optimal accuracy for MAE. This is repeated for target and threshold accuracy. This ensures that if we do have Gaussian data (which we don't) we have the same requirements even if we use MAE/median instead of bias/STD. In Figure 4 it is illustrated why we cannot use bias and STD.

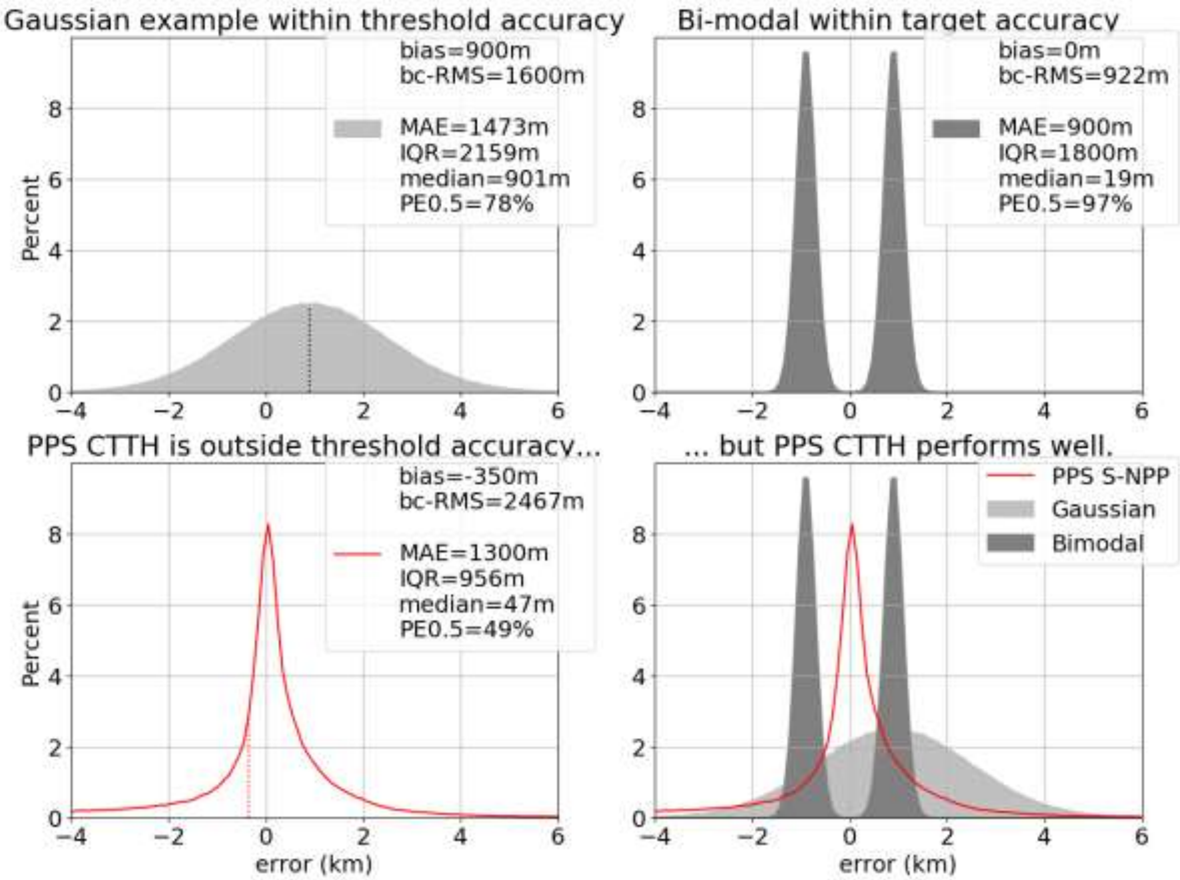

Figure 4: The non-Gaussian error distribution for 15 orbits S-NPP in red. The CTTH does not reach the gaussian threshold accuracy for STD of 2000m. However it reaches (with a large margin) the MAE and median threshold accuracy more suitable also for non-Gaussian distributions. The non-Gaussian error distribution is compared to two fictional error distributions that do reach threshold or even target accuracy but overall has more large errors. It is clear that because the CTTH error distribution is non-Gaussian the STD and bias are not the most appropriate measures to use to describe the errors of the algorithm. Note that we do expect some large errors due to differences in FOV, and sensitivity between the imager and the CALIOP instrument.

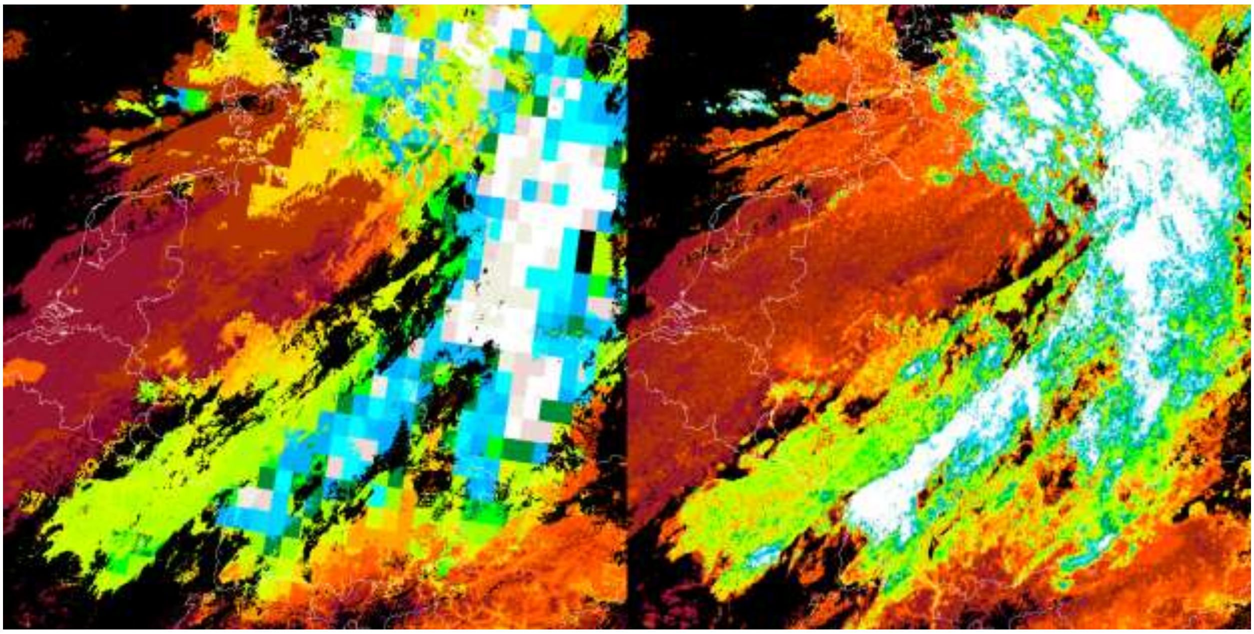

Figure 5: Comparison of CTTH v2014 (left) with CTTH v2018 (right). We can see that the retrievals are more consistent for PPS-v2018 and the squared gaps with missing data in version 2014 are filled for version 2018. Note that CTTH v2014 and v2018 are two completely different algorithms. The PPS v2021 CTTH algorithm is identical to v2018. Minor differences in CTTH between v2018 and v2021 is likely caused by updates to the cloud mask.

4.3.1 CALIOP validation

Table 12 presents and overview on the general quality for the height determination at several cloud classes and for VIIRS, MODIS, AVHRR (GAC), MERSI-2 and AVHRR (Metop-B), and also presenting results separated in semi-transparent and opaque (as determined by CALIOP). The retrieval rate (amount of clouds with cloud mask that also get cloud top height) is 100%. For the GAC data orbits from 2007 and 2008 were excluded as these were used in training a GAC network. For the same reason for MODIS only the respective 1st day of February, April, June, August, October and December included in the validation.

It is clear in Table 12 that PPS-v2021 reaches threshold accuracy for all the measures suitable for non-Gaussian distributions and for many also target or even optimal accuracy is met.

Note that in the matchup data we do have some very thin cloud layers which we do not expect the imager to detect, and for these we expect severe underestimations (not because of algorithm problems but because of differences in instrument sensitivity). As the imager and lidar FOV are different, we will also have cases where CALIOP detects mostly a low-level cloud and the imager mostly a high-level cloud and vice versa. For these cases we also expect very large differences which will contribute to a large STD as the errors are squared for that measure. Note that for PPS-v2021, for NPP (Table 12) the PE0.5 of 49% means that 49% of the errors are larger than 0.5km in magnitude! A Gaussian error distribution with zero bias and target STD would have PE0.5=75% meaning that 75% of the error would be larger than 0.5km in magnitude. Only 32% of the errors are larger than 1km in magnitude for PPS-v2021 (NPP). A Gaussian zero-bias threshold STD distribution on the other hand would have 60% of the errors above 1km. This means that we really cannot use bias and STD to evaluate the results due to the non-Gaussian error distributions.

As mentioned before, the restrictions due to the sensor differences are also valid for the cloud height inter-comparison. However, the CALIPSO dataset comes closest to what a trustworthy and continuous dataset should look like but more efforts to assure a fair comparison need to be taken.

In Table 13 results are divided into semi-transparent and opaque according to the CALIOP cloud classes. We can see that PPS-v2021 has the same accuracy as PPSv2018 and passes the target requirements for most datasets and threshold accuracy for the rest. Note that semi-transparent clouds for GAC-data are also inside target accuracy when thin cirrus detected only at 20km or 80km resolution for CALIOP are excluded. Clouds so thin that the lidar can't detect them on 5km scale are not expected to be even seen by the imager and correct height retrieval is not expected. For the CloudSat validation Table 16, GAC results are similar to the other datasets. If CTTH is produced with CMa-Prob as input with a low probability cloud mask threshold (i.e. 10%), more thin cirrus is likely to be included and overall a larger negative bias is expected. However, the median is not expected to be affected. The median is for most cases within optimal accuracy except for MODIS-semi-transparent and FY3D opaque. The MAE is well within target accuracy for all opaque clouds except for FYD where it is within threshold accuracy. For MAE semi-transparent clouds all datasets are within threshold accuracy.

To handle different FOV and instrument differences data can be filtered to get the results where the lidar and imager should agree. Filtered data are show in Table 14. Filtering is done as similar as possible to GEO validation, considering different validation software. Only pixels where 9 neighbouring measurements have the same cloud type and the variation in CALIOP pressure are less than 200hPa are included. Pixels where the 5km CALIOP top-layer are thinner than 0.2 are excluded. This leaves around 25% of the data. With this type of filtering results are within target accuracy for all datasets and for many cases even within optimal accuracy.

For MERSI-2 where we use channel 3.7micron instead of 12micron, results are not as good as for the sensors where we can use 11 and 12micron. But as seen in Table 12 results are matching those of MODIS collection 6 (MYD06_L2) and are within requirements which means they are still good. When looking at MERSI-2 separated for semi-transparent and opaque clouds we can see that results are worse for opaque clouds compared to the other datasets. One part of the explanation for this is that the 3.7-11micron difference is more sensitive to thin cirrus clouds. And high thin clouds are more likely to get correct height but with the drawback that some clouds will be misclassified as thin high cirrus. The differences could also be caused by differences between the 3.7micron channel from MODIS used for training and the 3.7micron channel for MERSI-2. This could be tested in future experiments by validating also MOIDS and VIIRS with the 11- and 3.7micron CTTH.

Table 12: PPS-v2021 cloud top height validation. Results outside threshold accuracy (semi-transparent) are marked red, results within target accuracy (opaque) are marked green. Yellow: measures with requirement. We use MAE and median used as error distributions are non-Gaussian.

Comparison to CALIOP data.	Bias (m)	Median (m)	IQR (m)	PE0.5 (%)	PE>1.0 (%)	MAE (m)	std (m)	N
	S-NPP VIIRS 15 orbits							
NPP v 2018	-350	47	956	49	32	1300	2467	359306
NPP v2021	-345	43	960	49	32	1298	2462	371174
NPP Low	354	82	426	21	9	528	1344	139071
NPP Medium	188	75	880	46	26	851	1432	64937
NPP High	-1133	-158	2706	73	52	2113	3166	167166
Low & inversion	400	184	770	40	16	652	1114	4562
	NOAA-18 GAC 39 orbits (2006 and 2009)							
GAC v2018	-1091	-158	1640	56	39	1768	3125	163693
GAC v2021	-995	-95	1498	55	38	1730	3128	163212
	MODIS data 6 days from 2010							
MODIS v2018	-586	-61	1078	48	31	1264	2373	2024112
MODIS v2021	-587	-65	1076	48	31	1261	2364	2028230
MODIS C6.1	-1414	-691	2134	71	49	1928	2911	2042013
	Feng-Yun – 3 MERSI-2 data 920 granules 2020							
MERSI2 v2021	781	515	1745	71	49	1774	2734	796006
	Metop–B AVHRR global metop 23 granules 2015 (mostly twilight polar)							
AVHRR v2021	-231	105	1469	59	41	1227	1794	12047
						?		
Threshold	2000/1000	2000/1000	2700	80	60	2350/1800	2000	
Target	1500/500	1500/500	2200	75	50	1750/1250	1500	
Optimal	200	200	670	30	5	430	500	

Table 13: Validation results separated by opaque semi-transparent. For comparison results for MYDO6_L2 are included, PPSv2021 has better scores.

Comparison to CALIOP data.	Bias (m)	Median (m)	IQR (m)	PE0.5 (%)	PE>1.0 (%)	MAE (m)	std (m)	N
	S-NPP VIIRS 15 orbits							
Opaque	309	110	624	33	15	599	1130	195642
Semi	-1074	-147	2476	67	50	2077	3223	175532
	NOAA-18 GAC 28 orbits (2009)							
Opaque	252	56	529	29	13	531	1021	63970
Semi	-1799	-594	2998	71	54	2503	3711	99242
Total OD > 0.225	-1588	-519	2647	70	52	2281	3459	88635
Top OD > 0.225	-420	-54	1472	60	39	1369	2284	52669
No CALIOP 20-80km	-861	-182	1859	64	44	1718	2759	65633
	MODIS data 6 days							
Opaque	161	39	538	31	14	558	1081	1073067
Semi	-1429	-534	2429	67	50	2051	3036	955163
Opaque MY06_L2	-283	-250	1329	60	33	937	1360	1081572
Semi MY06_L2	-2687	-1679	3542	82	67	3043	3589	960441
	MERSI-2 (uses channel 3.7 and 11)							
Opaque	1386	704	1467	66	42	1583	2275	419845
Semi	106	97	2468	76	57	1988	3030	376161
	Metop-B AVHRR (mostly polar night and twilight)							
Opaque	483	252	846	38	22	668	920	4323
Semi	-631	-84	2508	70	52	1540	2025	7724
Threshold	2000/1000	2000/1000	2700	80	60	2350/1800	2000	
Target	1500/500	15000/500	2200	75	50	1750/1250	1500	
Optimal	200	200	670	30	5	430	500	

Table 14: Validation results separated by opaque semi-transparent filtered to get the cases where CALIOP and the imager really should agree. Filtering is done as similar as possible to GEO validation, considering different validation software. Only pixels where 9 neighbouring measurements have the same cloud type and the variation in CALIOP pressure are less than 200hPa are included. Pixels where the 5km CALIOP top-layer are thinner than 0.2 are excluded. This leaves 1/4 of the data.

Comparison to filtered CALIOP.	Bias (m)	Median (m)	IQR (m)	PE0.5 (%)	PE>1.0 (%)	MAE (m)	std (m)	N
	S-NPP VIIRS 15 orbits							
Opaque	115	52	449	24	8	381	618	69067
Semi	-247	20	1037	51	32	1070	1893	27563
	NOAA-18 GAC 28 orbits (2009)							
Opaque	30	1	313	18	6	309	531	21461
Semi	-183	-11	731	42	24	843	1658	8029
	MODIS data 6 days							
Opaque	-36	-6	424	23	8	368	590	426509
Semi	-568	-163	1223	52	32	1043	1741	172517
	MERSI-2 (uses channel 3.7 and 11)							
Opaque	682	463	751	53	25	889	1420	162225
Semi	223	234	1668	68	44	1289	1945	64233
	Metop-B AVHRR (mostly polar night and twilight)							
Opaque	367	211	652	31	17	537	707	1709
Semi	-140	212	1083	56	34	1061	1632	1649
Threshold	2000/1000	2000/1000	2700	80	60	2350/1800	2000	
Target	1500/500	1500/500	2200	75	50	1750/1250	1500	
Optimal	200	200	670	30	5	430	500	

Table 15 Validation results separated by opaque semi-transparent and separated for different heights. For comparison results for MYDO6_L2 are included, PPSv2021 has better scores.

Comparison to CALIOP.	Median Low (m)	Median Medium (m)	Median High (m)		MAE Low (m)	MAE Medium (m)	MAE High (m)	N
	S-NPP VIIRS 15 orbits							
Opaque	74	113	301		448	637	850	195642
Semi	111	-3	-847		767	1081	2776	175532
	NOAA-18 GAC 28 orbits (2009)							
Opaque	325	17	7		393	558	801	63970
Semi	51	-59	-1032		593	111	2953	99242
	MODIS data 6 days							
Opaque	47	111	-108		427	607	811	1073067
Semi	77	-147	-1317		586	1113	2710	955163
Opaque MY06_L2	-101	-23	-879		711	1186	1292	1081572
Semi MY06_L2	-391	-2126	-2395		830	2407	3860	960441
	MERSI-2 (uses channel 3.7 and 11)							
Opaque	742	841	492		1799	1467	1137	419845
Semi	1226	664	-541		2707	1789	1857	376161
	Metop-B AVHRR (mostly polar night and twilight)							
Opaque	77	358	556		490	741	813	4323
Semi	215	-108	-386		735	1156	1801	7742

4.3.2 CPR (CloudSat) validation

In Table 16 the cloud top height results are also compared to CPR (CloudSat) this gives an independent validation as CPR (CloudSat) was not the instrument used for training. Measures IQR, MAE, PE0.5, PE1.0 and median show good scores well within corresponding target accuracy for most cases. MAE is between target accuracy for semi-transparent and opaque which is reasonable as table show opaque and semi-transparent clouds mixed. Note that compared to CloudSat AVHRR GAC have very similar statistics as MODIS and VIIRS. There is only a 100m difference in MAE compared to 500m when validated with CALIPSO. And this is probably because the very thin clouds available only in in CALIOP 5km data are now not included (as this is CloudSat data).

Table 16: PPS-v2021 cloud top height validation with CPR (CloudSat). Given are total results as well as results separated for low, medium and high for MODIS. Results outside threshold accuracy (semi-transparent) are marked red, results within target accuracy (opaque) are marked green. Yellow: measures with requirement. However median and IQR are more informative.

Comparison to CPR (CloudSat) data.	Bias (m)	Median (m)	IQR (m)	PE0.5 (%)	PE>1.0 (%)	MAE (m)	std (m)	N
	NOAA-18 GAC 28 orbits (2009)							
GAC v2018	573	263	1717	62	43	1433	2177	402266
GAC v2021	697	368	1812	63	45	1483	2191	418711
	S-NPP VIRRS orbits							
NPP v2018	-19	103	1526	63	42	1382	2191	98540
NPP v2021	26	117	1543	63	43	1381	2172	139009
	MODIS 6 days data							
MODIS-C 6.1	-811	-435	1929	69	47	1675	2497	1324989
PPS-v2018	6	-30	1309	57	37	1249	2020	1321846
PPS-v2021	7	-32	1309	57	37	1247	2017	1324738
V2021 Low	271	-126	523	28	13	689	1596	355072
V2021 Medim	159	-43	1238	56	35	1167	1856	276188
V2021 High	-189	159	2039	72	50	1565	2239	693478
Threshold	2000/1000	2000/1000	2700	80	60	2350/1800	2000	
Target	1500/500	1500/500	2200	75	50	1750/1250	1500	
Optimal	200	200	670	30	5	430	500	

4.4 CMIC CLOUD PHASE (CPH)

Cloud phase, as it is seen from a satellite, is a measure that describes whether the dominant number of observed photons is reflected/emitted by solid or liquid water particles. The penetration depth, the position and size of the probed layer is dependent on the observing wavelength as well as on the cloud composition. This implies that an earthbound observer is likely to probe different volumes, i.e. there is no ground truth for the cloud phase.

The most reliable source for a cloud phase determination from the current A-Train instruments is represented by the active laser probe CALIOP on board the CALIPSO platform. The decision, liquid or ice phase, is made on basis of the depolarization ratio of the backscattered signal (see Hu et al., 2009 for details). A known problem that reduces the quality of the cloud phase product is the detection of multiple scattered radiances. Another problem, that of horizontally oriented ice crystals, has been taken into account by tilting the instrument (to 3º off nadir) and enhancing the viewing zenith angle in 2007 (Hu et al., 2009). One other problem is that if we have a multilayer cloud the different layers can have different cloud phase. This could be a problem if the top layer is too thin for AVHRR to detect. We then compare the cloud phase calculated for PPS on one cloud layer with the CALIPSO cloud phase of a different layer. This is even a problem in case of no physical separation of these layers, i.e. a water cloud with a thin iced top may (for a passive instrument) still reflect/emit the radiative pattern of a water cloud. We tried to tackle this challenge by using only pixels where the upper three CALIOP levels give a concordant phase.

The CMIC Cloud Phase gives either liquid water or ice, while the CALIOP Cloud Ice/Water Phase Discrimination uses the classes: ice, water and oriented plates. The classes liquid and water are considered as a match. The CMIC class ice is considered a match with the two CALIOP classes: ice and oriented plates.

The cloud phase is validated for all five datasets NPP, AVHRR GAC, MERSI-.2, AVHRR (Metop-B) and MODIS. For MODIS only the part of the data, (1st every even month), which was not used for the cloud top height training was considered. The cloud phase algorithm is using the cloud top height as input. Comparing the required scores (Table 17) with the algorithm performance shows that all accuracies are within threshold accuracy. For POD target accuracy is reached for all datasets except for POD-liquid MERSI-2. For FAR target accuracy is reached for all datasets except for FAR-ice (MERSI-2) and FAR-liquid (AVHRR Metop-B). The higher FAR liquid for Metop-B is explained by the fact that the Metop-B contains mostly ice-clouds (for METOP-B only high latitude colocations are available, and for those only a third of the clouds are water clouds). For MERSI-2 the differences are most likely due to the differences in CTTH (see CTTH validation). For MERSI-2 it is more likely to get high ice clouds correct than low water clouds.

Table 17: Success measures for CMIC cloud phase compared to CALIOP. All datasets are global except for Metop-B that has only high latitude data. Green: within target accuracy, red: outside threshold accuracy. Yellow marks measures with requirements.

	HR	POD liquid	FAR liquid	POD solid	FAR solid	N
GAC v2018	0.82	0.83	0.23	0.82	0.13	131899
GAC v2021	0.83	0.85	0.23	0.82	0.12	131700
MODIS v2018	0.84	0.85	0.15	0.83	0.18	1966656
MODIS v2021	0.85	0.87	0.15	0.82	0.16	1970926
MODIS-C6.1	0.84	0.77	0.10	0.90	0.22	1681303
NPP v2018	0.83	0.82	0.14	0.84	0.20	349921
NPP v2021	0.84	0.84	0.14	0.84	0.18	361557
MERSI-2 2021	0.81	0.75	0.12	0.88	0.25	769576
AVHRR Metop-B 2021	0.86	0.82	0.25	0.87	0.09	11640
NPP v2018 (CTTH not input)	0.77	0.87	0.26	0.65	0.19	349921
Threshold accuracy		≥ 0.70	≤ 0.35	≥ 0.60	≤ 0.35	
Target accuracy		≥ 0.80	≤ 0.20	≥ 0.80	≤ 0.20	
Optimal accuracy		≥ 0.90	≤ 0.10	≥ 0.90	≤ 0.10	

4.5 CMIC LIQUID WATER PATH (LWP)

4.5.1 Global validation of LWP over Sea with AMSR-E

Up to now, one of the most feasible methods to determine the vertical integrated liquid water content of the atmosphere is the observation of emissions in the microwave spectral range. Over sea we validated the CMIC liquid water path retrieval with AMSR-E (Advanced Microwave Scanning Radiometer for EOS) estimates of LWP.

The most important advantage of microwave LWP retrievals is that they are dependent on fewer assumptions, than is possible for retrievals based on observations in VIS and IR channels. Still, this method is far from being perfect, i.e. Greenwald (2009) identified strong dependencies of the AMSR-E LWP product on surface wind speed but over open water it is still superior to all other methods with comparable temporal and spatial coverage.

To keep the absorption coefficient within the valid range, rain contaminated pixels have to be excluded. Unfortunately, the naturally emitted energy is very low in this wavelength band (typically between 0.9 and 1.3 cm). This requires large antennas, which leads on the other hand to large FOVs and thus a rather low spatial resolution. Due to complex contributions from land surfaces, LWP results based on microwave observations are only applicable if the footprint is not 'land-contaminated'. For this study the LWP product of the AMSR-E onboard the AQUA platform is used to provide the reference data. More information on the AMSR-E dataset used can be found in section 3.6 and at http://nsidc.org/data/amsre/.

To avoid contamination by rain, the validation is restricted to a AMSR-E LWP between 0 g/m² and 170 g/m² the AMSR-E pixel is matched to the mean LWP of the 5-8 nearest neighbours within the AMSR-E footprint. The 6 days of MODIS data from 2010 from February, April, June, August, October and December 99 orbits of GAC data was used for the LWP validation. For the GAC data unfortunately only AMSR-E granules from the northern hemisphere were used. Results are shown in Table 18 and threshold or target accuracy are reached for all datasets. For AVHRR (GAC) bias is even within optimal accuracy. See also Figure 6 for scatter plots and error density functions.

For VIIRS we have no matches with AMSR-E (AMSR-E's utilization period ended before launch of Suomi-NPP) so for VIIRS we have co-located CMIC-lwp with CPR (Cloudsat) data (the CWC_RVOD product). Note that this data uses the MODIS lwp which comes from a similar sensor as VIIRS and should not be assumed to be superior. Only data where both algorithms had values between 0 and 500 where included and where both algorithms agreed that it was cloudy and there was no ice water path. In Table 18 we can see that the scores are within threshold accuracy. See also Figure 6 for scatter plots and error density functions.

For MERSI-2 we have no matches with AMSR-E. And for the time period of MERSI-2 data selected (2020) there are no CPR (CloudSat) data released yet. (Also, RVOD data set will never be released for this period as EOS-Aqua is no longer in the same orbit as CloudSat and CALIPSO). So for MERSI-2 we have 22 matching granule pairs with MODIS for 1st of January 2020 08:25 to 10:15. As CMIC-lwp for MODIS meet the requirements making sure that retrievals are very similar between MODIS and MERSI-2 indicates that also CMIC-lwp for MERSI-2 meets the requirements. And as the median (-0.7 g/m2) and inter quartile range (20 g/m2) are small we can confirm that the two products are very similar. For this inter-comparison data points are less than 20 minutes apart. And for each sensor a Gaussian weighted (depending on distance) average is calculated for the 25 closest pixels within 5km. Data points with values above 500 g/m2 are excluded similar to the validations with AMSR-E and CloudSat RVOD product. For difference distributions and scatterplots see Figure 7.

Table 18: Required and achieved accuracies for the global validation over sea of liquid water path compared to AMSR-E, CPR (CloudSat) or intercompared with PPS CMIC-lwp for MODIS. Note that requirements are not defined for comparison with CPR (CloudSat) or for inter comparisons between sensors. And CPR (CloudSat) RVOD is based much on MODIS data which can't be assumed to have higher quality than CMIC-lwp. Green: within target accuracy, red: outside threshold accuracy. Yellow marks measures with requirements.

	RMS [g/m²]	Bias [g/m²]	IQR [g/m²]	Median [g/m²]	N AMSR-E	N CPR RVOD	N MODIS - PPS CMIC-lwp
PPS v2018 MODIS (different days)	57.0	-5.1	42.2	-11.1	6830430		
PPS v2021 MODIS	59.0	-6.2	40.2	-11.5	4990233		
Accuracy v2018 (GAC)	44.6	-3.3	43.7	-7.3	815057		
Accuracy v2021 (GAC)	43.4	-4.1	39.5	-6.8	793625		

NPP- VIIRS 2021	78.4	1.9	67.1	-4.3		4002	
MERSI-2 2021 (inter comparison)	48	-2.0	21.1	-0.7			8789900
Threshold accuracy	≤ 100	≤ 20	≤ 135	≤ 20			
Target accuracy	≤ 50	≤ 10	≤ 65	≤ 10			
Optimal accuracy	≤ 20	≤ 5	≤ 25	≤ 5			

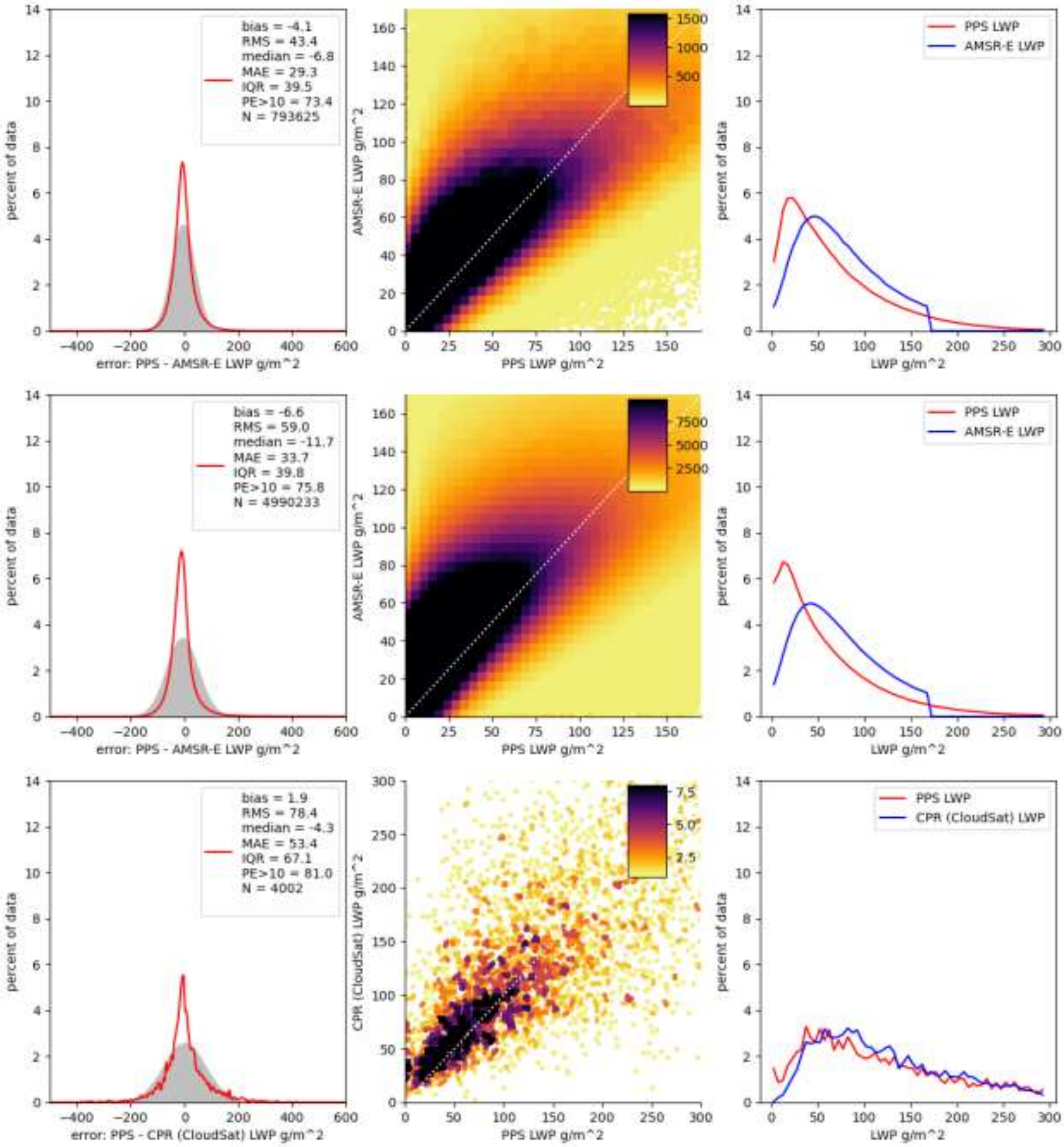

Figure 6 From the top AVHRR-GAC compared to AMSR-E, MODIS compared to AMSR-E and VIIRS compared to CPR (CloudSat). From left to right error density function, scatter plot and lwp density function. The error density function (Red curve) is compared to a gaussian distribution with the same bias and RMS.

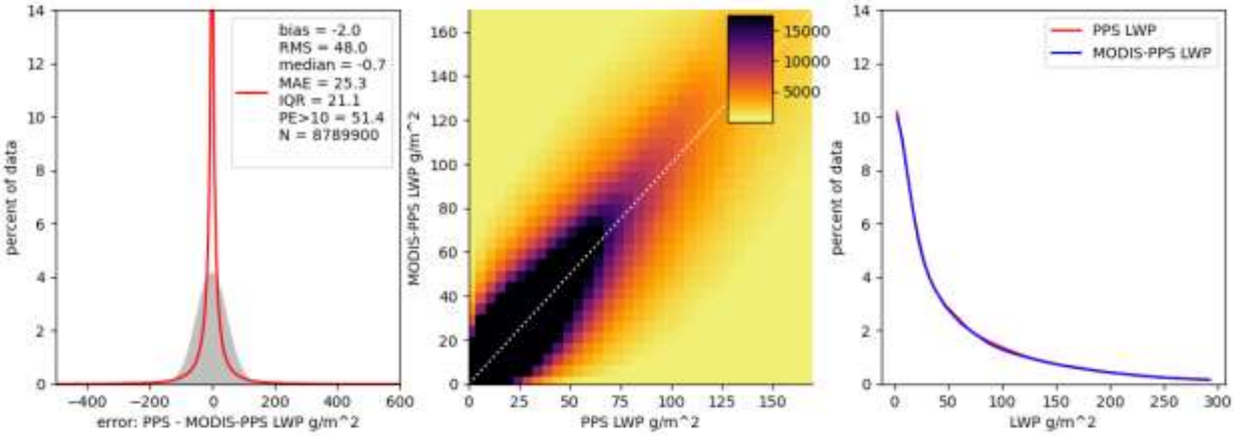

Figure 7:. Inter comparison between PPS CMIC-lwp for MODIS and MERSI-2. From left to right error density function, scatter plot and lwp density function. The error density function (Red curve) is compared to a gaussian distribution with the same bias and RMS. Note that both median and IQR are small. Showing that the products are very similar. Data points are less than 20 minutes apart. And for each sensor a Gaussian weighted (depending on distance) average is calculated for the 25 closest pixels within 5km. Data points with values above 500 kg/m2 are excluded similar to the validations with AMSR-E and CloudSat RVOD product.

4.6 Cloud probability

The cloud probability product is validated with CALIOP co-located with VIIRS/AVHRR/MODIS and MERSI2. The cloud probability is not a binary cloud mask. For each pixel the probability of it being cloudy is presented. Users can use the product as a binary cloud mask by first applying a probability cloud mask threshold suitable for their application. By selecting a low probability cloud mask threshold of for example 5% a clear conservative cloud mask is achieved. If a high probability cloud mask threshold for example 95% is used the result is a binary cloud mask with very low FAR-cloudy. All probability cloud mask thresholds between 5% and 30% will give a binary cloud mask that meets the requirements on POD-cloudy and FAR-cloudy, but other probability cloud mask thresholds might be more suitable for some applications. Note that a probability cloud mask threshold on 5% - 20% will give a cloud mask with lower values on FAR-clear (compared to CMa FAR-clear 22%). This can be useful for applications that need to screen out clouds before making retrievals for clear pixels. Similarly using a probability cloud mask threshold above 50% will decrease the FAR-cloudy to below 5% (compared to CMa FAR-cloudy 9%) which can be useful for applications that need to screen out clear pixels. The requirement for CMa-Prob are for the 50% limit compared to SYNOP. This is evaluated in Table 19 for SYNOP and in Table 21 using CALIOP data. Cloudy pixels with CALIOP total optical thickness below 0.2 are excluded. All datasets pass the threshold requirements. POD-cloudy passes target requirements for several datasets and FAR-cloudy passes optimal requirements for several datasets.

Table 19: Cloud probability validation scores for 15 S-NPP orbits against global SYNOP. Results are global, result for Europe only are included in a separat line. Europe here is defined as latitude

between 35 and 72 and longitude between -25 and 60 degrees. Green: within target accuracy, red: outside threshold accuracy. Yellow marks measures with requirements.

	Bias (%)	Hit rate	POD cloudy	FAR cloudy	POD clear	FAR clear	N
All	-1.4	0.93	94.8	3.4	88.7	16.6	8422
Day	-2.0	0.94	95.0	2.6	89.2	19.0	6337
Night	0.6	0.89	91.0	10.1	86.4	12.2	1239
Twilight	-0.7	0.96	97.2	1.9	92.4	10.7	846
Europe	0.4	0.95	97.4	3.2	86.4	11.6	6145
Threshold Accuracy			> 85	< 20			
Target Accuracy (Europe)			> 90 (95)	< 15 (10)			
Optimal Accuracy (Europe)			> 95 (98)	< 10 (5)			

Table 20: Scores for CMa-Prob for the NPP-CALIOP matchup data. As a reference CMa have Kuipers 0.710 and Hitrate 0.863 for the same dataset. POD-cloudy and FAR-clear are not expected to meet the requirement for all Limits. Instead it is important that they meet the requirements for at least one Probability cloud mask threshold. It is also important that some probability cloud mask thresholds give low values on FAR-clear and FAR-cloudy respectively. Green: within target accuracy, red: outside threshold accuracy. Yellow marks measures with requirements.

Probability cloud mask threshold (%)	POD-cloudy (%)	POD-Clear (%)	FAR-cloudy (%)	FAR-Clear (%)	Hit rate	Kuipers	Hit rate (v2018)	Kuipers V(2018)
0	100	0.00	34.34	x	0.657	0.000	0.672	0.000
5	93.71	68.50	14.95	14.93	0.851	0.622	0.811	0.489
10	91.33	76.93	11.67	17.72	0.864	0.683	0.837	0.590
15	89.41	82.07	9.49	19.78	0.869	0.715	0.847	0.637
20	87.91	85.03	8.18	21.38	0.869	0.729	0.852	0.667
25	86.58	87.11	7.23	22.75	0.868	0.737	0.854	0.688
30	85.48	88.58	6.53	23.86	0.865	0.741	0.855	0.702

35	84.47	89.78	5.95	24.85	0.863	0.743	0.854	0.709
40	83.55	90.76	5.47	25.74	0.86	0.743	0.852	0.713
45	82.66	91.54	5.08	26.59	0.857	0.742	0.850	0.716
50	81.76	92.27	4.71	27.42	0.854	0.74	0.847	0.717
55	80.9	92.92	4.38	28.21	0.85	0.738	0.844	0.716
60	80.01	93.5	4.07	29.01	0.846	0.735	0.840	0.715
65	79.08	94.03	3.8	29.84	0.842	0.731	0.836	0.713
70	78.08	94.55	3.52	30.71	0.837	0.726	0.832	0.710
75	77.02	95.03	3.27	31.62	0.832	0.72	0.826	0.705
80	75.79	95.52	3	32.64	0.826	0.713	0.820	0.698
85	74.31	96.02	2.72	33.84	0.818	0.703	0.812	0.690
90	72.21	96.57	2.43	35.49	0.806	0.688	0.802	0.677
95	68.65	97.25	2.05	38.13	0.785	0.659	0.782	0.652
100	0	100.00	x	65.66	0.343	0.000	0.341	0.000

Table 21: Validation of CMa-Prob with limit 50 against CALIOP-data. Green: within target accuracy, red: outside threshold accuracy. Yellow marks measures with requirements Measures are intended for SYNOP data, we do not expect measures to be met for very thin clouds. For the "no

polar night" category all data where sun satellite angle is above 95 and absolute latitude is above 70 is excluded.

CALIOP against CMa-Prob (50%)	BIAS %	HR	K	POD-cloudy %	POD-cloudy% Filt 0.2	FAR-cloudy %	POD-clear %	FAR-clear %	N
S-NPP VIIRS data global 15-16 orbits									
PPS-NPP 2021 (all)	-9.3	0.85	0.74	81.7	84.8	4.7	92.3	27.4	643586
2021 no polar night	-8.2	0.87	0.76	83.6	86.5	4.5	92.4	25.5	590024
EOS-Aqua global data 12 days									
PPS-MODIS v2021	-6.3	0.85	0.72	84.5	86.8	6.7	87.5	26.6	6504358
2021 no polar night	-5.2	0.87	0.74	86.5	88.7	6.2	87.9	24.4	5962404
NOAA-18 AVHRR GAC data 66 orbits (2006-2008)									
PPS-GAC 2021	-10.5	0.82	0.66	79.9	88.1	6.3	86.6	36.7	497760
2021 no polar night	-8.5	0.84	0.69	82.6	90.4	6.2	86.6	33.3	456691
Feng-Yun–3 MERSI-2 data 920 granules 2020									
PPS v2021	5.3	0.8	0.52	88.7	89.9	17.8	62.8	25.7	1639263
2021 no polar night	5.8	0.83	0.56	91.4	92.5	15.9	64.7	21.3	1467311
Metop–B AVHRR global metop 23 granules 2015 (mostly twilight polar)									
PPS v2021	6.3	0.7	0.34	80.9	83.7	26.6	53.2	36.4	21821
2021 no polar night	5.8	0.82	0.57	90.2	92.8	17.3	67.1	20.2	13529
Requirement Accuracy (global)									
Threshold accuracy				85 %		20 %			
Target accuracy				90 %		15 %			
Optimal Accuracy				95 %		10 %			

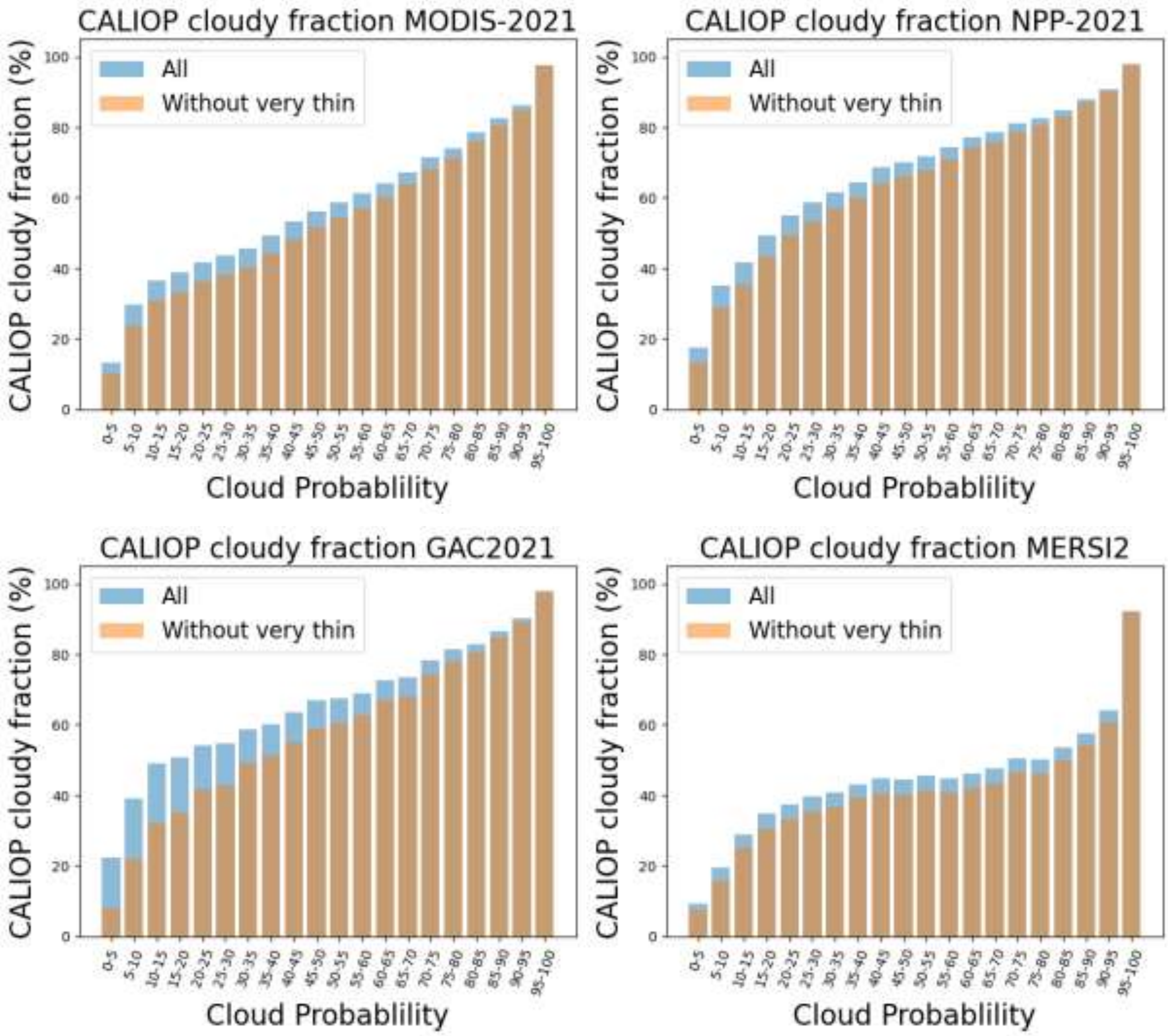

Figure 8: Showing the actual probability for cloudy for the cloud probability classes.

4.6.1 Cloud mask and cloud probability comparison

Cloud mask and cloud probability are based on the same thresholds and most of the time agree on classification (cloud or clear) see Figure 9.

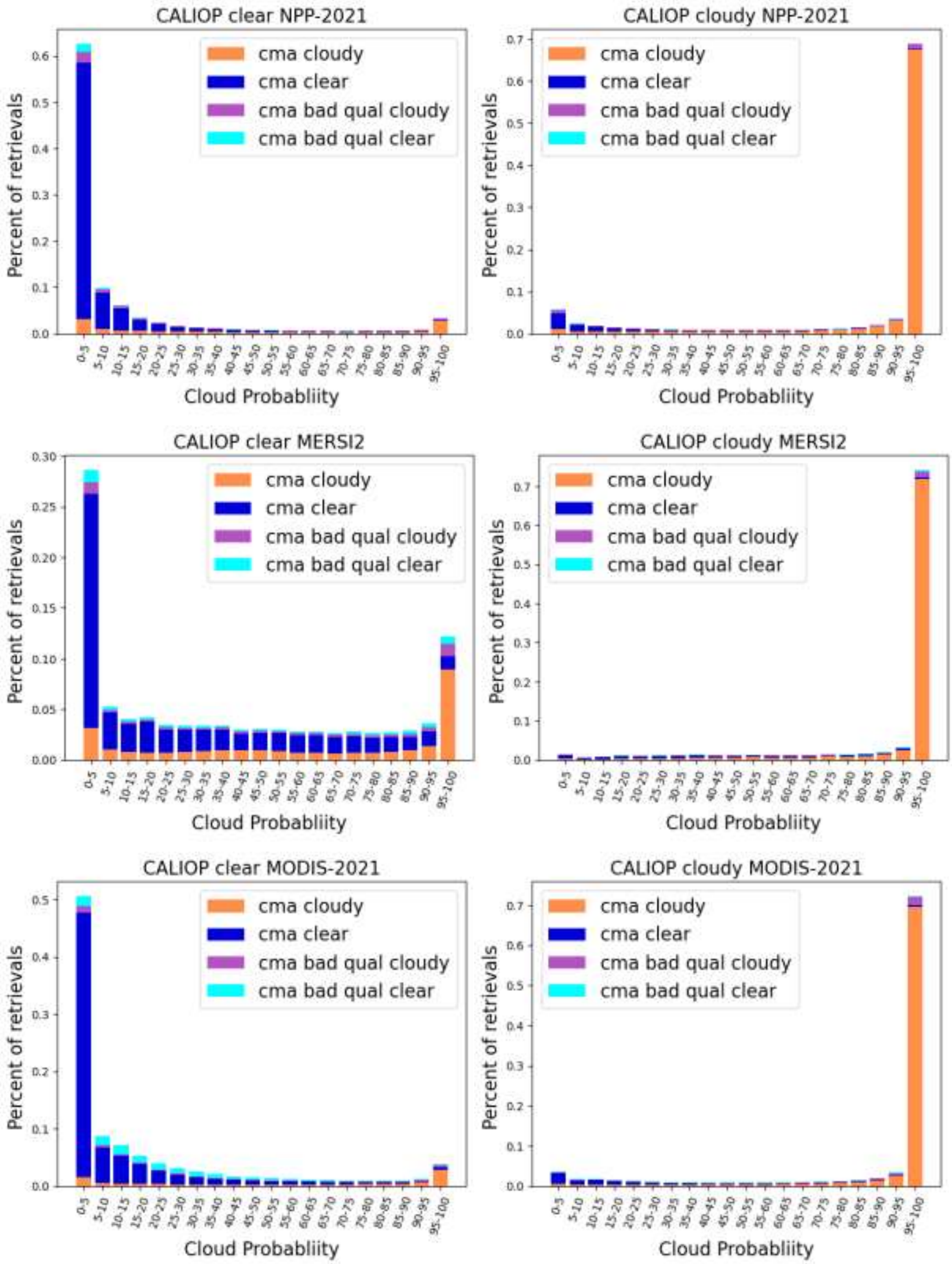

Figure 9: Confusion histograms for VIIRS, MODIS and MERSI-2 cloud probability and CMa.

4.7 SLSTR EXAMPLE

As seen in Figure 11 SLSTR gives a product that is very similar to the results for MODIS. Looking at more cases there seem to be a bit larger problem with false clouds over cold ground from the water cloud test for SLSTR compared to MODIS. CTTH, CT and CMa-Prob are very similar for MODIS and SLSTR for this case.

4.7.1 SLSTR future plans

Currently SLSTR is included as a demonstrational product in PPS. This means it is not quantitatively validated but visually intercompared with MODIS for a handful of scenes. A proper NWCSAF validation is currently not planned and not expected before earliest with the release of PPS-NGI-2 planned for 2026. User feedback and validation activities by users are encouraged.

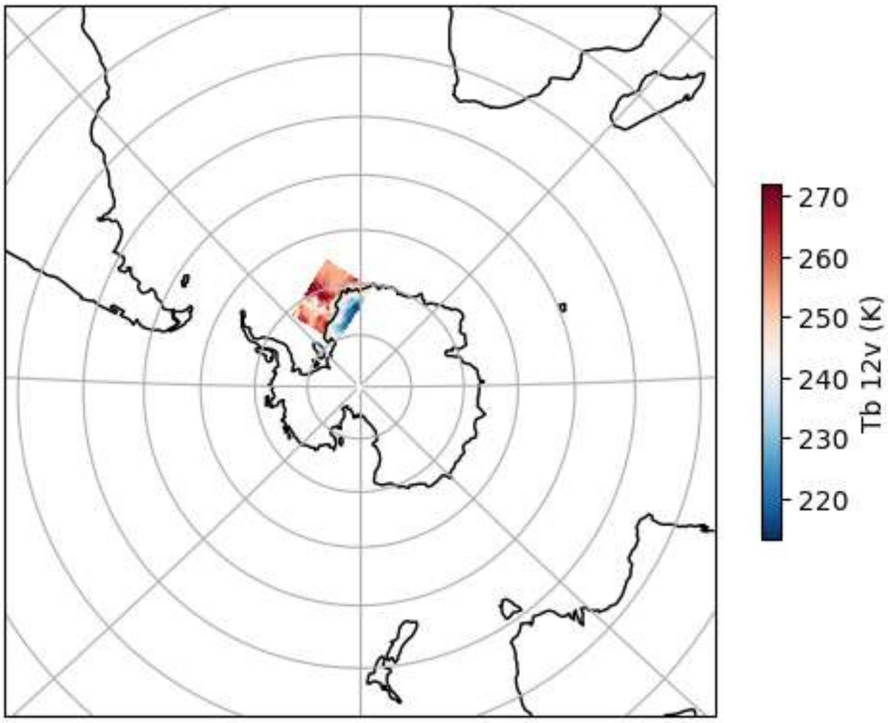

Figure 10: Showing the SLSTR granule on a map.

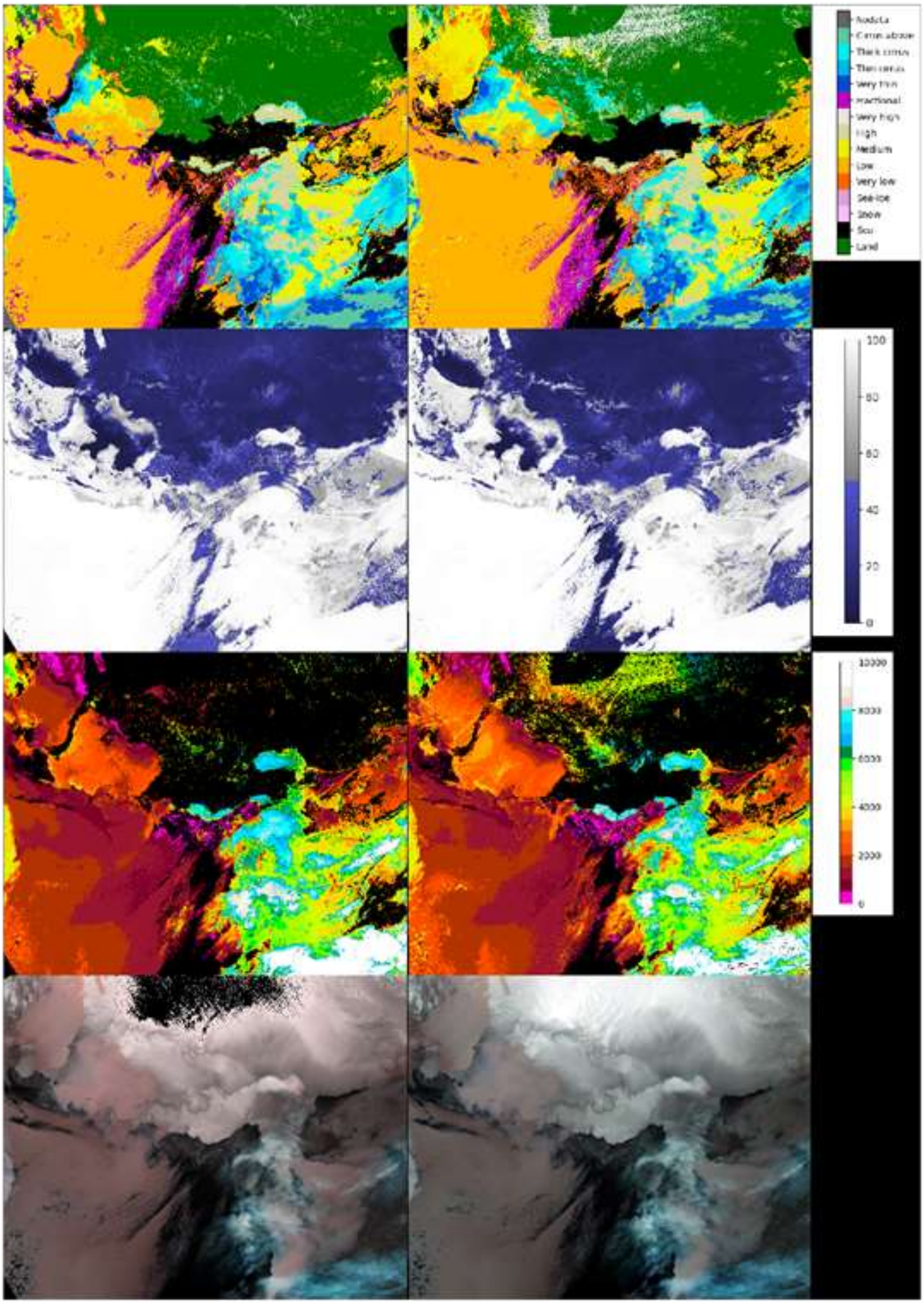

Figure 11: To the left a granule of MODIS from EOS-Aqua remapped to the same projection as the SLSTR granule. To the right the same granule for SLSTR Sentinel-3A. The granules are only 5 minutes apart just outside the coast of Antarctica 20200312 01:40 UTC for exact location see Figure 10. Channel 8.7 is turned off for MODIS. The products displayed are, from top: Cloud Type, Cloud Probability, CTTH, and RGB-image. The RGB-images are made with channels 1.6µm, 37µm, 11µm and 12µm.

4.8 POLAR STATISTCS

4.8.1 Cloud mask and cloud probability

In this section we present statistics for the polar regions (defined as latitude above 70 or below -70). In Table 22 we show statistics for CMa combined with CMa-Prob over sea. The idea is that this could be useful for retrieving sea ice or water temperatures. In the table PPS is considered cloudy if indicated by CMa or if the cloud probability is above 50%. Note that the cloud detection is quite high between 77% and 98% specially for day and twilight. FAR-clear is better for day than for night and twilight as expected and especially FAR-cloudy for MERSI-2 and AVHRR (Metop-B) polar night is even higher than 20%. In Table 23 to Table 27 there are results separated per sensor. We can see that results are best for MODIS and VIIRS and that results are generally worst over polar land in night and twilight with hit rates for many sensors below 70%. For MERSI-2 and AVHRR (Metop-B) also results over polar ice during night look bad.

Table 22: Overview of polar statistics over sea (including ice and open water) for cloud mask combined with cloud probability where PPS is considering it cloudy if a cloud is detected by cloud probability above 50% or by cloud mask. Lines with POD-cloudy>85% and FAR-cloudy<20% are marked green. Lines with hit rate below 60% are marked red

CALIOP against CMa combined with CMa-Prob	BIAS %	HR	K	POD-cloudy %	POD-cloudy% Filt 0.2	FAR-cloudy %	POD-clear %	FAR-clear %	N
S-NPP VIIRS data global 15-16 orbits									
sea_polar_day	1.1	0.94	0.81	97.0	97.6	4.3	84.3	11.1	22598
sea_polar_night	-4.6	0.78	0.45	83.0	84.8	11.7	61.5	49.1	23913
sea_polar_twilight	2.3	0.86	0.50	92.4	93.1	10.1	57.7	34.8	19761
EOS-Aqua global data 12 days									
sea_polar_day	1.0	0.94	0.79	96.5	97.1	4.7	82.5	13.3	243575
sea_polar_night	-8.5	0.76	0.50	77.4	79.6	12.2	72.4	44.5	261585
sea_polar_twilight	-0.7	0.83	0.58	88.0	89.0	11.1	69.8	32.0	183339
MODIS C-6.1									
sea_polar_day	-3.8	0.91	0.80	91.9	92.9	3.5	87.9	25.3	243575
sea_polar_night	-18.9	0.75	0.57	69.2	71.9	6.2	88.2	47.3	261585
sea_polar_twilight	-14.3	0.79	0.63	75.8	77.1	5.9	86.9	43.4	183339
NOAA-18 AVHRR GAC data 66 orbits (2006-2008)									
sea_polar_day	-5.8	0.87	0.71	87.9	93.0	5.0	83.4	34.1	16866
sea_polar_night	-10.3	0.76	0.49	77.3	82.8	10.5	71.8	49.5	19990
sea_polar_twilight	-3.0	0.83	0.48	87.6	91.1	9.0	60.8	47.9	11427
Feng-Yun–3 MERSI-2 data 920 granules 2020									
sea_polar_day	16.7	0.82	0.36	99.0	99.2	19.6	37.5	6.7	86383
sea_polar_night	13.5	0.65	0.13	83.8	84.3	30.5	28.8	52.1	52044
sea_polar_twilight	18.5	0.74	0.15	95.1	95.1	24.3	19.8	39.6	33461
Metop–B AVHRR global metop 23 granules 2015 (mostly twilight polar)									
sea_polar_night	22.6	0.60	0.15	84.5	85.7	40.0	30.3	38.8	7783
sea_polar_twilight	9.2	0.88	0.35	98.5	98.9	11.3	36.4	17.2	2001

Table 23: Polar statistics for CMa and CMa-Prob for VIIRS data divided in illumination categories and surface type (land/water/ice). Lines with POD-cloudy>85% and FAR-cloudy<20% are marked

green. Lines with hit rate below 70% are marked red. Note CMa in the top and CMa-Prob results in the bottom.

CALIOP against CMa and CMa-Prob	BIAS %	HR	K	POD-cloudy %	POD-cloudy% Filt 0.2	FAR-cloudy %	POD-clear %	FAR-clear %	N
CMa S-NPP VIIRS data global 15-16 orbits									
ice_polar_day	-0.9	0.94	0.83	95.2	95.8	3.7	88.2	15.1	19984
ice_polar_night	-20.1	0.71	0.48	66.6	69.1	8.7	81.8	53.9	20065
ice_polar_twilight	-2.7	0.77	0.51	81.1	82.3	15.4	70.0	35.3	11287
water_polar_day	0	0.93	0.63	96.1	96.9	3.8	66.8	33.5	2614
water_polar_night	-9.4	0.89	0.68	89.5	90.7	0.8	78.6	79.1	3848
water_polar_twilight	-2.3	0.97	0.71	97.1	97.4	0.5	73.8	65.9	8474
land_polar_day	5.0	0.88	0.77	91.9	92.7	17.6	85.2	6.7	18712
land_polar_night	-19.5	0.75	0.52	57.5	63.0	8.0	94.6	32.7	29649
land_polar_twilight	2.4	0.75	0.49	78.6	81.1	24.8	70.1	26.0	15143
CMa-Prob > 50% S-NPP VIIRS data global 15-16 orbits									
ice_polar_day	1.6	0.94	0.82	97.5	98.1	4.6	84.8	8.9	19984
ice_polar_night	-2.1	0.80	0.51	85.3	87.7	12.3	65.8	39.1	20065
ice_polar_twilight	3.1	0.81	0.54	87.9	89.1	15.9	66.2	27.0	11287
water_polar_day	-1.6	0.94	0.72	95.5	96.6	2.7	76.8	33.8	2614
water_polar_night	-2.7	0.94	0.52	95.6	96.6	1.6	56.5	68.6	3848
water_polar_twilight	-0.6	0.98	0.63	98.7	99.1	0.7	64.3	50.2	8474
land_polar_day	-3.4	0.92	0.83	87.0	89.1	5.4	96.2	9.2	18712
land_polar_night	-8.4	0.84	0.68	76.1	82.6	9.2	91.7	22.0	29649
land_polar_twilight	0.8	0.76	0.52	78.4	80.6	22.7	73.4	25.3	15143

Table 24: Polar statistics for CMa and CMa-Prob for AVHRR NOAA-18 GAC data divided in illumination categories and surface type (land/water/ice). Lines with POD-cloudy>85% and FAR-cloudy<20% are marked green. Lines with hit rate below 70% are marked red.

CALIOP against CMa and CMa-Prob	BIAS %	HR	K	POD-cloudy %	POD-cloudy% Filt 0.2	FAR-cloudy %	POD-clear %	FAR-clear %	N
CMa NOAA-18 AVHRR GAC data 66 orbits (2006-2008)									
ice_polar_day	-10.4	0.84	0.72	82.9	89.1	4.0	88.8	38.1	14641
ice_polar_night	-21.7	0.71	0.53	66.1	72.6	6.6	86.7	52.8	18502
ice_polar_twilight	-13.0	0.79	0.60	78.5	83.6	6.0	81.3	49.8	8956
water_polar_day	-4.7	0.91	0.65	92.2	94.6	2.8	73.2	52.1	2225
water_polar_night	1.9	0.95	0.40	98.5	98.7	3.4	41.7	37.5	1488
water_polar_twilight	-10.3	0.85	0.58	86.5	88.4	2.7	71.4	69.6	2471
land_polar_day	-13.2	0.78	0.58	66.9	77.0	10.8	90.9	29.1	15132
land_polar_night	-40.5	0.54	0.29	37.8	49.6	9.6	90.8	61.0	21079
land_polar_twilight	-8.4	0.69	0.38	67.4	74.2	21.7	70.8	41.7	11532
CMa-Prob > 50% NOAA-18 AVHRR GAC data 66 orbits (2006-2008)									
ice_polar_day	-7.5	0.85	0.71	85.6	91.3	5.1	85.2	35.1	14641
ice_polar_night	-21.3	0.68	0.43	63.9	70.0	10.4	78.7	56.8	18502
ice_polar_twilight	-11.9	0.75	0.45	76.5	80.8	10.0	68.2	56.4	8956
water_polar_day	-2.6	0.93	0.68	94.6	97.1	2.6	73.7	43.0	2225
water_polar_night	0.7	0.95	0.49	97.8	98.4	2.9	51.2	41.9	1488
water_polar_twilight	0.4	0.93	0.48	96.5	97.6	4.0	51.9	45.3	2471
land_polar_day	-20.1	0.75	0.52	57.3	69.6	7.6	94.7	33.7	15132
land_polar_night	-44.4	0.51	0.26	33.2	44.5	8.3	93.1	62.1	21079
land_polar_twilight	-18.5	0.64	0.34	55.5	62.5	20.2	78.1	47.0	11532

Table 25: Polar statistics for CMa and CMa-Prob for MODIS divided in illumination categories and surface type (land/water/ice). Lines with POD-cloudy>85% and FAR-cloudy<20% are marked green. Lines with hit rate below 70% are marked red.

CALIOP against CMa and CMa-Prob	BIAS %	HR	K	POD-cloudy %		POD-cloudy% Filt 0.2	FAR-cloudy %	POD-clear %	FAR-clear %	N
	CMa EOS-Aqua global data 12 days									
ice_polar_day	-1.7	0.93	0.82	94.2		94.9	3.7	88.2	17.6	201464
ice_polar_night	-21.2	0.72	0.53	64.5		67.4	7.2	88.7	47.6	232304
ice_polar_twilight	-9.9	0.81	0.63	79.0		80.4	7.9	84.5	36.2	148229
water_polar_day	-3.6	0.93	0.81		94.2	95.0	1.8	86.8	34.0	42111
water_polar_night	-3.4	0.94	0.81		95.2	96.3	1.2	86.0	40.5	29281
water_polar_twilight	-3.7	0.93	0.78		93.9	94.9	2.0	84.3	37.0	35110
land_polar_day	0.5	0.85	0.69		83.0	84.3	18.0	86.3	12.9	192802
land_polar_night	-24.4	0.70	0.38		43.8	49.3	12.2	94.2	36.2	280369
land_polar_twilight	-0.3	0.72	0.43		68.6	70.7	31.0	74.3	26.1	171880
	CMa-Prob > 50% EOS-Aqua global data 12 days									
ice_polar_day	-0.3	0.93	0.80		95.1	95.8	4.6	85.1	15.9	201464
ice_polar_night	-15.8	0.70	0.43		66.9	69.3	13.4	76.5	49.6	232304
ice_polar_twilight	-8.4	0.77	0.54		77.4	78.7	11.9	76.2	40.4	148229
water_polar_day	-2.8	0.95	0.87		95.7	96.9	1.2	90.8	26.8	42111
water_polar_night	-3.0	0.95	0.81		95.6	96.8	1.2	85.6	38.4	29281
water_polar_twilight	-1.0	0.94	0.75		96.2	97.0	2.7	78.5	28.1	35110
land_polar_day	1.7	0.85	0.70		84.3	85.9	18.9	85.3	12.2	192802
land_polar_night	-21.4	0.67	0.32		43.7	48.7	22.0	88.3	37.7	280369
land_polar_twilight	-0.1	0.65	0.29		61.2	62.6	38.7	67.7	32.4	171880

Table 26: Polar statistics for CMa and CMa-Prob for MERSI2 divided in illumination categories and surface type (land/water/ice). Lines with POD-cloudy>85% and FAR-cloudy<20% are marked green. Lines with hitrate below 70% are marked red.

CALIOP against CMa and CMa-Prob	BIAS %	HR	K	POD-cloudy %	POD-cloudy% Filt 0.2	FAR-cloudy %	POD-clear %	FAR-clear %	N
	CMa	Feng-Yun–3 MERSI-2 data 920 granules 2020							
ice_polar_day	17.5	0.80	0.35	98.4	98.7	21.1	36.9	9.3	92980
ice_polar_night	10.0	0.64	0.12	80.1	80.7	30.5	32.3	54.3	63813
ice_polar_twilight	17.0	0.74	0.20	93.8	93.9	24.3	26.4	36.4	37608
water_polar_day	3.9	0.94	0.63	98.9	99.3	5.4	63.9	9.8	13868
water_polar_night	-0.6	0.94	0.81	95.6	95.8	3.7	85.9	16.4	2252
water_polar_twilight	12.4	0.86	0.29	98.8	98.9	14.4	30.2	14.8	4161
land_polar_day	14.4	0.77	0.57	89.7	90.7	32.9	67.2	10.2	46334
land_polar_night	-18.2	0.63	0.24	42.0	45.4	32.2	81.7	39.5	105867
land_polar_twilight	3.2	0.64	0.27	65.9	66.5	38.2	61.4	34.5	35988
	CMa-Prob > 50%	Feng-Yun–3 MERSI-2 data 920 granules 2020							
ice_polar_day	-0.6	0.92	0.80	93.6	94.4	5.6	86.7	15.0	92980
ice_polar_night	-8.8	0.63	0.24	65.2	66.5	24.7	58.7	53.3	63823
ice_polar_twilight	1.0	0.76	0.41	83.8	84.3	17.4	56.8	41.1	37608
water_polar_day	2.8	0.95	0.72	99.0	99.4	4.1	72.9	8.1	13868
water_polar_night	17.9	0.80	0.08	98.7	98.8	19.5	9.2	34.8	2252
water_polar_twilight	17.3	0.82	0.06	99.3	99.5	18.2	7.1	27.8	4161
land_polar_day	-0.9	0.84	0.67	80.1	81.8	18.1	86.8	14.6	46334
land_polar_night	6.0	0.52	0.04	56.1	57.2	50.1	48.3	45.4	105877
land_polar_twilight	3.2	0.59	0.19	61.4	62.1	42.4	57.2	39.0	35988

Table 27: Polar statistics for CMa and CMa-Prob for AVHRR (Metop-B) divided in illumination categories and surface type (land/water/ice).

CALIOP against CMa and CMa-Prob	BIAS %	HR	K	POD-cloudy %	POD-cloudy% Filt 0.2	FAR-cloudy %	POD-clear %	FAR-clear %	N
CMa Metop–B AVHRR global metop 23 granules 2015									
ice_polar_night	-1.0	0.65	0.30	63.5	66.5	35.2	66.8	34.5	6248
ice_polar_twilight	-7.8	0.86	0.60	87.6	88.2	3.9	72.7	56.7	1205
water_polar_night	-1.8	0.90	0.72	92.9	94.0	5.0	79.5	27.1	1535
water_polar_twilight	9.7	0.87	0.51	98.0	98.1	13.0	53.4	10.5	796
land_polar_day	-2.0	0.90	0.67	72.2	76.6	20.2	95.2	7.1	556
land_polar_night	-8.8	0.64	0.29	58.5	61.6	30.1	70.8	40.5	3714
land_polar_twilight	-3.2	0.88	0.72	78.7	85.8	13.5	93.1	11.4	4721
CMa-Prob > 50% Metop–B AVHRR global metop 23 granules 2015									
ice_polar_night	20.4	0.50	0.00	69.5	70.6	50.9	30.4	49.2	6248
ice_polar_twilight	4.5	0.87	0.20	95.2	95.7	9.4	24.5	60.0	1205
water_polar_night	0.7	0.91	0.69	94.6	96.5	6.2	74.2	23.3	1535
water_polar_twilight	5.4	0.89	0.62	96.4	97.6	10.0	66.0	14.9	796
land_polar_day	3.4	0.90	0.76	84.3	89.7	27.6	91.6	4.3	556
land_polar_night	-5.3	0.52	0.05	50.7	53.8	43.7	54.2	51.3	3714
land_polar_twilight	10.6	0.76	0.54	81.0	87.3	37.4	72.8	12.8	4721

4.8.2 Cloud type

Cloud type statistic for polar data are shown in Table 28.

Table 28: Polar cloud type validation compared to CALIOP. See Table 9 to see which classifications are considered successful/unsuccessful. Data within threshold accuracy are marked green. There are no requirements for polar data specifically.

	POD Low (%)	POD Medium (%)	POD High (%)	FAR Low (%)	FAR Medium (%)	FAR High (%)	FAR Cirrus (%)	N
PPS VIIRS v2021	86.4	59.3	73.9	12.6	39.8	18.4	33.2	66885
PPS VIIRS (day)	89.2	53.7	79.2	9.4	26.5	25.1	44.6	24224
PPS VIIRS (night)	82.7	68.9	72.8	15.6	55.9	9.2	23.9	22092
PPS VIIRS (twilight)	86.1	62.0	71.7	14.0	39.8	19.7	32.9	20569
PPS MODIS v2021	90.6	62.9	73.1	11.6	3.08	14.3	32.7	330950
PPS MODIS (day)	90.8	60.8	70.4	8.2	33.1	21.8	43.3	141413
PPS MODIS (night)	91.6	66.0	76.1	15.9	50.1	4.1	17.7	111223
PPS MODIS (twilight)	88.8	64.3	69.9	12.7	33.3	20.3	38.8	78314
MERSI-2 2021	74.6	68.2	76.7	9.9	47.2	21.8	29.3	152847
AVHRR Metop-B v2021	88.3	72.3	81.2	30.5	32.6	15.8	9.1	7214

4.8.3 CTTH

Polar validation results for CTTH are shown in Table 29.

Table 29: PPS-v2021 POLAR cloud top height validation. Results inside threshold accuracy (median and MAE) are marked green. There are no requirements for polar data specifically.

Comparison to CALIOP data.	Bias (m)	Median (m)	IQR (m)	PE0.5 (%)	PE>1.0 (%)	MAE (m)	std (m)	N
	S-NPP VIIRS 15 orbits							
NPP v2021	-359	75	1177	54	35	1211	2053	66883
	NOAA-18 GAC 39 orbits (2006 and 2009)							
GAC v2021	-1155	-164	1919	60	43	1840	3271	25010
	MODIS data 6 days from 2010							
MODIS v2021	-382	-21	1116	51	30	1084	1869	330721
MODIS C6.1	-1667	-1134	2554	76	59	2069	2515	331074
	Feng-Yun – 3 MERSI-2 data 920 granules 2020							
MERSI2 v2021	219	358	1527	68	43	1309	1965	152761
	Metop–B AVHRR global metop 23 granules 2015 (mostly twilight polar)							
AVHRR v2021	-336	29	1834	62	45	1311	1843	7214

4.8.4 CMIC Cloud phase

Table 30: Polar success measures for CMIC cloud phase compared to CALIOP.

	HR	POD liquid	FAR liquid	POD solid	FAR solid	N
GAC v2021	0.84	0.79	0.18	0.87	0.16	19711
MODIS v2021	0.84	0.83	0.11	0.86	0.21	319368
MODIS-C6.1	0.77	0.56	0.05	0.97	0.31	222878
NPP v2021	0.82	0.76	0.12	0.89	0.22	64877
MERSI-2 2021	0.81	0.78	0.15	0.84	0.22	145636
AVHRR Metop-B 2021	0.86	0.85	0.32	0.86	0.06	6957

5 VALIDATION OF ERROR ESTIMATES

5.1 CLOUDMASK AND CLOUD PROBABILITY ERRORS

Cloud mask comes with a quality flag indicating low or high quality. The goal with this flag is to flag the uncertain retrievals and we expect the high-quality retrievals to be more accurate. In Table 31 we can see that most (88-92%) of the data is flagged high-quality by CMa and the hit rate for high-quality data is 17 to 26 percentage points higher than for the low-quality data. The quality flag is working as intended. For the cloud probability the uncertainty is instead directly included in the product. According to the theory data with cloud probability close to 50% is the most uncertain and should have higher risk of being incorrect. In Table 31 we can see that 65% to 88% of the data has either a cloud probability above 80 % or below 20% (defined in this investigation as high-quality). And the high-quality retrievals for cloud probability have a hit rate that is 29 to 34 percentage points better than the low-quality (cloud probability between 20 and 80%) retrievals. The built-in error estimates of CMa-Prob are working as intended.

Table 31: Validation of error estimates for CMa and CMa-Prob. For CMa there is a low-quality flag and for CMa-Prob values between 10 and 90 are treated as low quality in this investigation.

SENSOR	Hit rate (low-quality)	Hit rate (high-quality)	Part of data that is high-quality.	N
CMa VIIRS	63.9	88.1	92.4	643025
CMa-Prob VIIRS	55.4	88.9	87.8	606931
CMa MODIS	70.4	88.7	89.2	3453079
CMa-Prob MODIS	58.4	90.5	85.0	3363627
CMa AVHRR GAC	63.7	84.6	90.1	293513
CMa-Prob ACHRR GAC	54.8	87.1	83.1	291279
CMa MERIS-2	57.3	84.1	91.8	1358616
CMa-Prob- MERSI-2	53.1	88.2	77.8	1347004
CMa AVHRR Metop-B	58.5	80.1	85.3	27171
CMa-Prob AVHRR Metop-B	52.1	79.4	63.0	27167

5.2 CTTH ERRORS

For CTTH an upper and a lower limit for CTTH can be optionally produced. For users not interested in error estimates it does not need to be processed. The error estimates for CTTH are constructed using quantile regression neural networks (Pfreundschuh et. al 2018) and are computed for the 16% and 84%

percentile. This means that according to theory 68% of the errors should fall within the error estimates. Compare with including standard deviation for Gaussian data. For Gaussian data 68% of the retrievals fall within +/- one standard deviation of the mean. For the VIIRS data when checking against CALIOP 63% of the data falls within the error estimates (compare with expected 68%). If we include an additional 5hPa error 70% of the data fall within upper -5hPa and lower limit + 5hPa limit. In Figure 12 we can see that most of the error estimates are in absolute value between 25hPa and 125hPa. Only 10% of the errors are expected to be smaller than 25hPa in absolute value. And most of the data (>80%) fall within 20hPa from the error estimate if not already inside the error estimate limits.

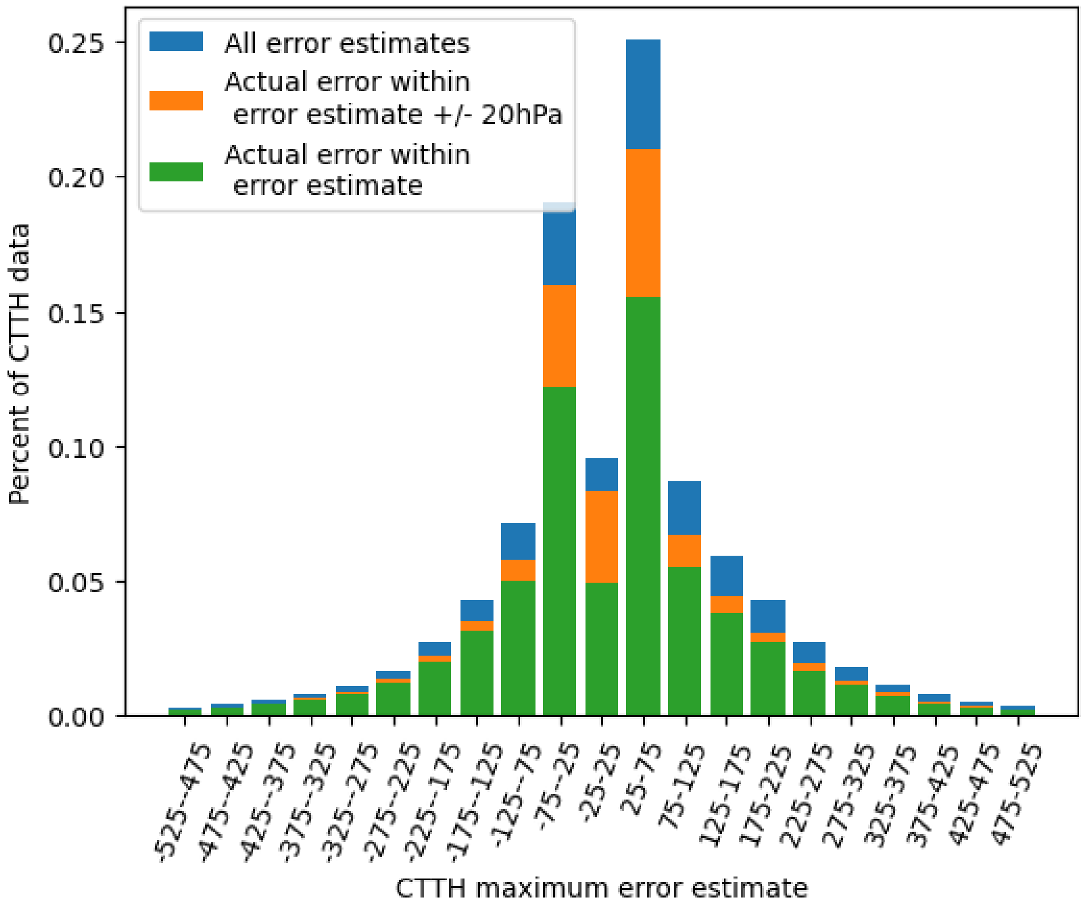

Figure 12: Error estimates for CTTH VIIRS data. The error estimate for CTTH comes as a lower and upper limit. For the figure the largest (in absolute value) of the limits is used to sort data along x-axis. We can see that most of the error estimates are small (high bars between -125hPa and 125hPa). And most of the actual errors are within the error estimate +/- an additional 20hPa (orange bars almost as high as blue).

5.3 Cloud phase errors

The cloud phase flag does not have a pixel level error estimate. This is normally the case also for other cloud phase algorithms that use thresholding or Optimal Estimation. For example, for OCA the phase that gives the smallest retrieval cost/error are chosen but the phase has no error estimate (Loredana Spezzi personal communication May 2021). However, the cloud CCI project has now switched to a cloud phase algorithm using a neural network approach that also gives estimation of the errors (Poulsen et al. 2019). Note that the PPS cloud phase algorithm is developed in CMSAF and it might during CDOP-4 be replaced with a different algorithm that gives also error estimates, for example a neural network approach. However, for version 2021 of PPS the cloud phase algorithm in PPS consists of several threshold tests using channels (11, 12, 1.6 and 3.7micron). The phase is also reset to a different phase if results are inconsistent with cloudy LUT entries during day time. For more details see the ATBD (RD.6). After the phase is retrieved, a consistency check is performed against cloud top temperature. Very cold clouds are not allowed to be liquid, and warm clouds are not allowed to be ice. This means that the cloud phase algorithm is very dependent on the cloud top temperature quality. This can also be seen in Table 17 where including cloud top temperature improves the hit rate with 6 percentage points. The cloud phase algorithm has a hit rate of around 83%. In reality it might be a bit higher as the hit rate is calculated using CALIOP data and CALIOP has much smaller field of view meaning that the imager and CALIOP are not always looking at the same cloud. The 17% of non-successful retrievals comes from several different sources of errors:

1. FOV differences. CALIOP and the imager do not always see the same cloud, especially at cloud edges.
2. Errors in inputs to the algorithm:
   a) Errors in cloud top temperature. The cloud phase algorithm in the last step uses the cloud top temperature to reclassify phase values if they are not consistent with the cloud top temperature.
   b) Errors in surface type that is input to the cloud phase retrieval algorithm. For example, the parameter RNIR for the *first reflectance test* need to be smaller than NIR_PHASE_THRES which depends on the type of surface water/snow/other. And the type of surface is not always correctly known.
3. Limitations of and errors in the algorithm itself, including:
   a) Uncertainties in the initial cloud phase algorithm offsets and limits. These tests contain thresholds and offsets that could be non-optimal for a specific pixel. For example, the cirrus test is only successful for clouds with T11<295, can there be a cirrus cloud with T11 = 295.1K? This cirrus cloud risks ending up with the wrong phase. However changing limits and offsets risk misclassifying other and more pixels.
   b) Some of the thresholds in the cloud phase algorithm are indirectly based on RTM simulations. In these simulations there are likely some errors.
4. Limitations of the instrument. Do we expect a HR of 100% to be possible with the spectral and spatial resolution we have for the imager? In this context, it should also be kept in mind that the reference instrument (CALIOP) does not perfectly represent the truth.

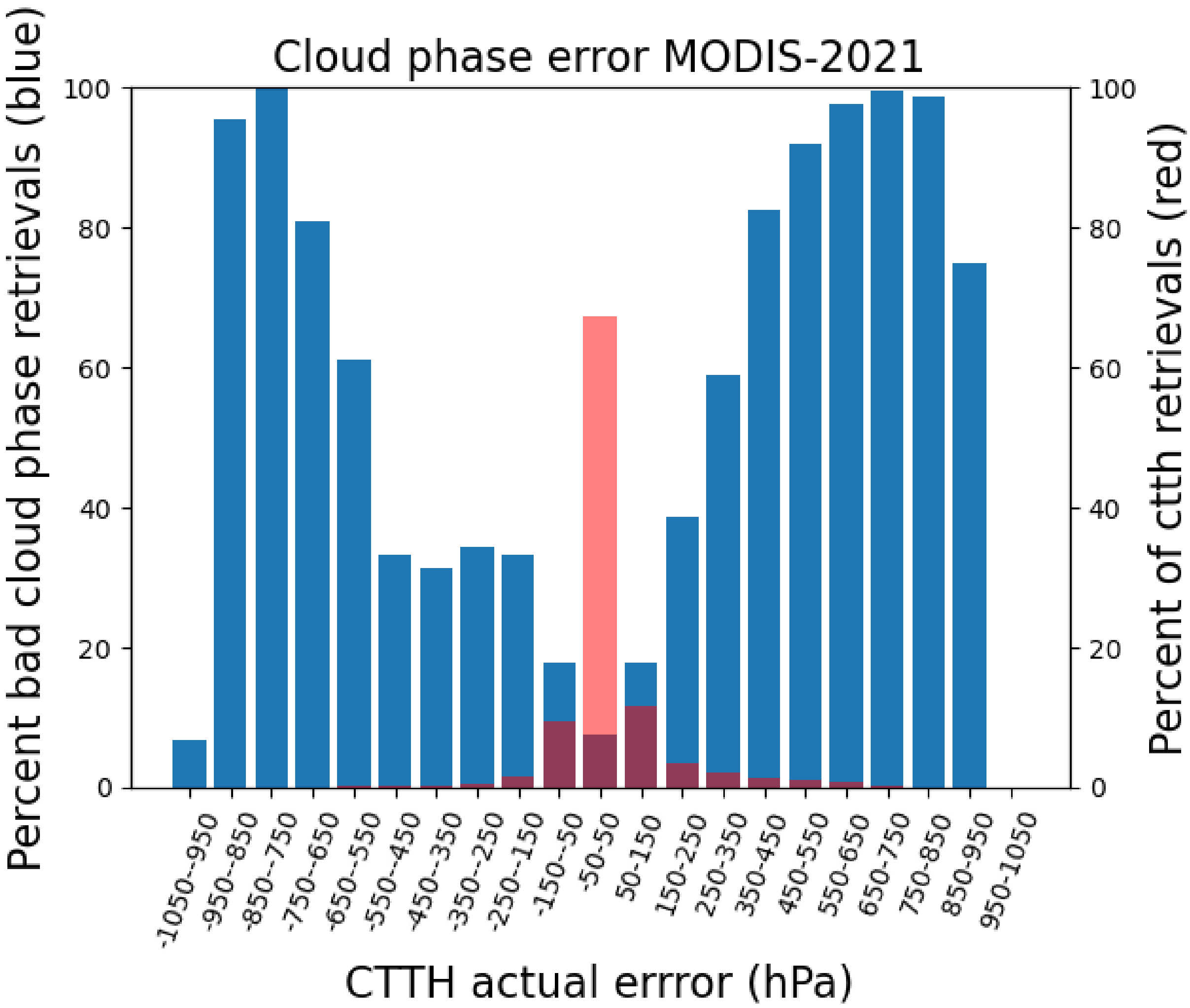

Figure 13: Cloud phase error dependence on CTTH error. Blue is fraction of wrong phase and read is the percent of CTTH retrievals. Note that the overlap yields a dark red color. Note that when the CTTH absolute error is smaller than 50hPa the cloud phase has few (7.6%) bad retrievals. And when the error for CTTH is large the cloud phase will almost always be wrong. Note that the large errors for CTTH also include cases when the imager and CALIOP see different clouds, for example at a cloud edge. The red curve shows the amount of CTTH retrievals which has most of the data in the middle (small errors).

CMIC does not contain cloud phase error estimates on pixel level. Phase estimates are correct for 83% of the data. For comparison the CTTH error estimates correctly captures around 68% of the errors. From a more detailed study of error characteristics, conclusions can be drawn on the most important error sources as well as for upper limits for remaining error sources of less importance. Looking in Table 32 we can see that hit rate spans from 81% to 86%. In Figure 13 we can see that the quality of the cloud phase depends very much on the CTTH error. If we exclude data where the error in CTTH is larger than 50hPa, by which most of the errors in points 1 (since significant FOV differences normally lead to large CTTH errors) and 2a are removed, this gives hit rates over 90% and optimal performance for the remaining pixels. If we further refine the selection to be only over surface type water we exclude errors from point 2b. And we now have retrievals with hit rates between 92%-95%. The around 5%-8% remaining errors are likely due to the combined effect of the remaining points 3) limitations of the algorithm, including non-optimal threshold settings and errors in underlying RTM simulations, and 4) the separability between ice and water clouds even possible for the instrument. And considering only daytime data the remaining error is between 4%-5%. There is also a column included in Table 32 with validation done as similar as possible to the validation for the GEO package. This shows that around 10 percentage points of the original 17% errors are caused either by clouds too thin for the imager to detect or in homogeneous clouds where the risk is high that the lidar and imager are looking at different clouds.

Table 32: Table showing hit rate for cloud phase and how that improves when data that are more likely to contain errors are excluded. For the geo style validation, we keep only retrievals that have CALIOP optical depth of the top layer larger than 0.2 and variation less than 200 hPa and the same PPS cloud type for the 9 closest neighbours along the CALIOP track.

	HR (%)	HR (%) As for GEO	Data left (%)	HR (%) CTTH error <50hPa	Data left (%)	HR (%) CTTH error <50hPa Over water	Data left (%)	HR (%) CTTH error <50hPa Over water day	Data left (%)
VIIRS	84.0	92.7	26.6	90.8	66.5	92.8	41.3	95.7	20.3
MODIS	84.6	93.5	30.2	92.4	67.2	94.2	43.2	95.7	22.3
GAC	83.0	93.1	21.4	92.7	67.7	94.6	42.5	94.7	20.0
MERSI-2	80.8	89.6	29.2	93.4	44.3	95.2	24.5	96.0	14.1
For completeness including also VIIRS without CTTH as input and Metop-B (few pixels)									
VIIRS-2018 no ctth input	76.6	88.1	26.4	84.2	66.4	86.9	41.2	88.1	20.4
Metop-B	85.8	96.5	20.9	90.6	55.1	91.8	11.2	85.1	3.2

5.3.1 When do we expect large errors for CMIC cloud phase?

For practical use of the product it is of less interest what is exactly causing the errors and more important to have a quality flag available to indicate when there is a large risk for errors. The cloud phase does not have a quality flag for PPS version 2021 but the cloud type product and the CTTH error estimates can be used to identify pixels which are more likely have an incorrect phase. As seen in Figure 15 the error is largest for the cloud categories *mid-level* and *very thin cirrus*. And in Figure 14 we can see that if the estimated error for CTTH is smaller than 50hPa the hit rate for cloud phase is higher.

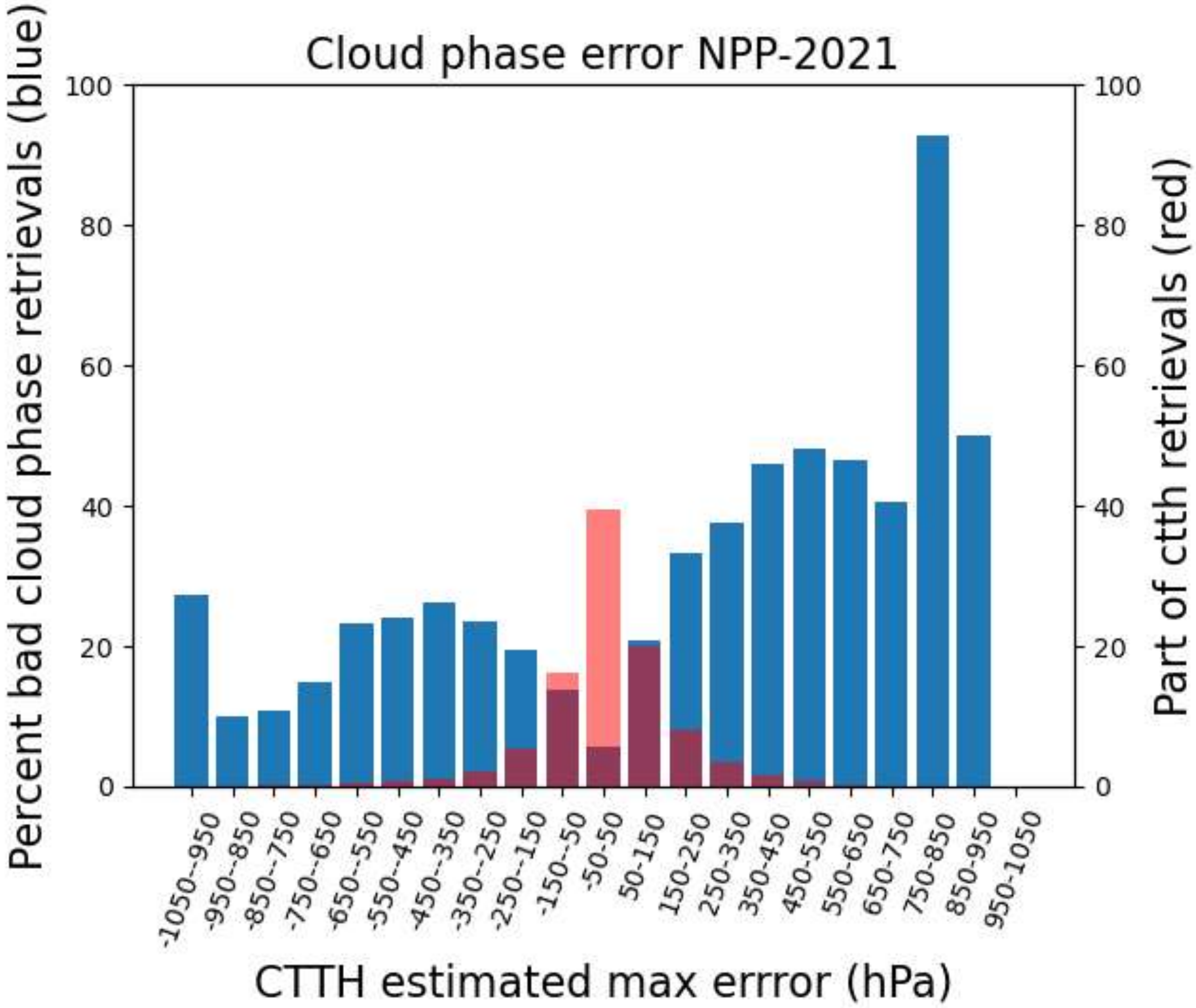

Figure 14: Cloud phase retrievals are better where the estimated CTTH error is smaller (below 50hPa in absolute value). Blue is fraction of wrong phase and read is the percent of CTTH retrievals. Note that the overlap yields a dark red color Note that compared to Figure 13 (which showed cloud phase error compared to the actual CTTH error) this figure shows the cloud phase error as a function of the CTTH error estimates available in the PPS products.

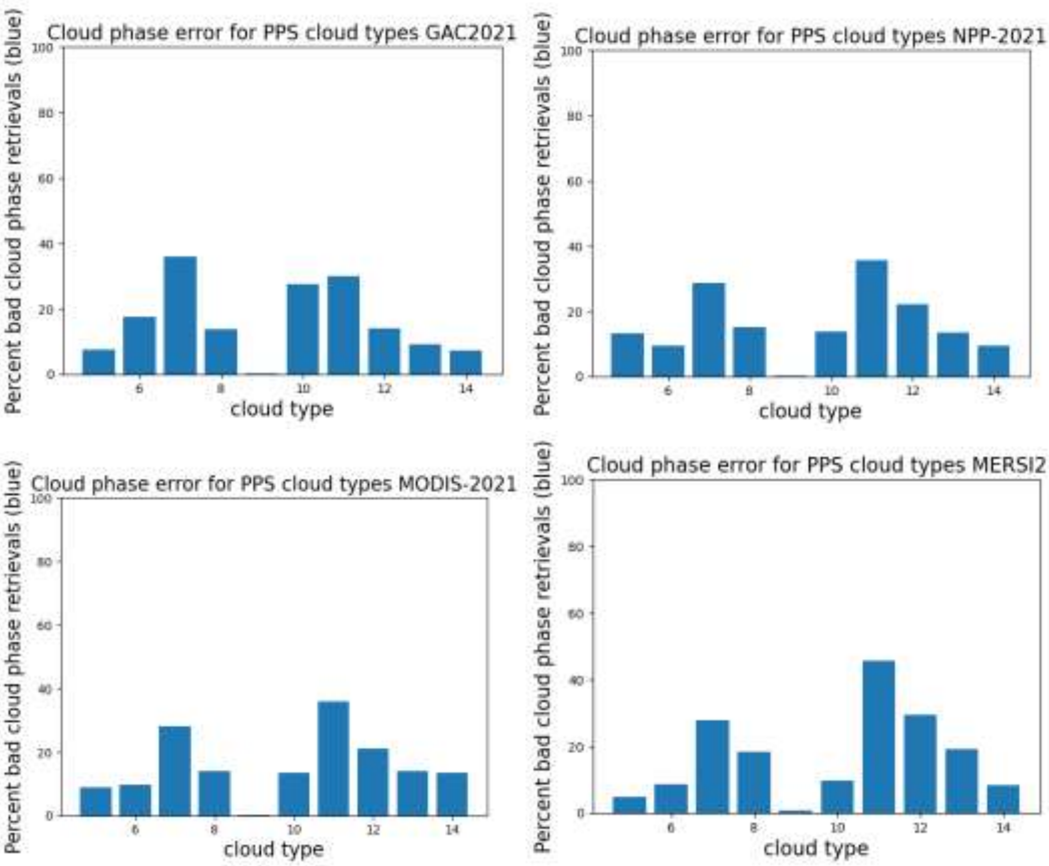

Figure 15: Fraction of wrong cloud phase retrievals as a function of cloud type for four different sensors. For all sensors, errors are largest for the very thin cirrus clouds (11) and the mid-level clouds (7) category. Very high cloud (9) has almost all retrievals correct. Metop-B data have not been included in this figure because of the small amount of pixels.

6 SUMMARY AND CONCLUSIONS

In this report we validate the 2021 version of PPS for AVHRR, VIIRS and MODIS, and MERSI2 data. The validation is made for the products: CMa, CMa-Prob, CT, CTTH, CMIC phase and CMIC LWP over sea. For CMa validation is performed against a dataset of global SYNOP reports (kindly provided by DWD). For validations also a set of CALIPSO lidar products, level 2 MODIS-collection 6.1 products, AMSR-E and CPR (CloudSat) radar products are used. Five imager datasets have been used for collocation: one with 17 orbits of S-NPP VIIRS data from 2015, one with 6 days (1st of each month 2010) of global MODIS data and another using a global dataset of 99 AVHRR GAC orbits from 2006 to 2009. For MERSI-2 we have 920 granules from 4 months (February, April, May, June) 2020, and for AVHRR Metop-B 23 global Metop granules from December 2015.

We find that quite a number of target-, and even optimal accuracies are reached, and all scores are within threshold accuracy. Results for PPSv2021 have small differences compared to PPSv2018.

For CMa we see a small increased POD-cloudy and a small decreased in FAR-cloudy while generally HR and Kuipers are unaffected. The cloud mask is meeting the target accuracy globally with best results for day and twilight according to the SYNOP validation with VIIRS. Over the European domain target accuracy is met. Using CALIPSO as the reference truth for the cloud mask globally (excluding polar night) or over Europe all scores (for AVHRR/VIIRS/MODIS/MERSI-2) are within threshold accuracy and for POD-cloudy often within or close to target accuracy and for FAR-cloudy often within optimal accuracy.

The EPS-SG validation (kindly carried out by Loredana Spezzi) of the EPS-SG test-data meet optimal accuracy, and the PPSv2021 results are very similar (99% identical) to the ones used for this validation. Meaning that PPSv2021 is on track to meet requirements for EPS-SG day1.

For the cloud type and the separation in low, medium level and high clouds, all scores match either target or optimal accuracy except FAR-medium-level clouds that match threshold accuracy.

The results for version PPS-v2021 CTTH are similar to the ones for PPS-v2018. As the error distributions are non-Gaussian, and considering that some very large errors are expected due to sensitivity differences and differences in FOV, bias and STD are not the measures that best describe the error distributions, and according to the PRD performance are evaluated for MAE and median. The CTTH meets threshold accuracy for all datasets both compared to CALIOP and CloudSat. In many cases target accuracy are reached and for median even optimal accuracy. Results for opaque clouds are generally better and for AVHRR/VIIRS and MODIS target accuracy are reached for opaque clouds. Note that for MODIS performance is clearly better than the performance of MODIS collection 6.1 (MYD06_L2), meaning the PPSv2021 truly have a very good CTTH product.

For the phase validation MODIS/VIIRS/AVHRR and MERSI2 all meet threshold and for most cases also target accuracy.

The LWP validation for MODIS and AVHRR-GAC with AMSR-E over sea has scores within target accuracy. The LWP validation for VIIRS with CloudSat RVOD data is also within target accuracy as is the inter-comparison of MERI-2 CMIC LWP with MODIS CMIC LWP even though scores are intended for validation with AMSR-E.

For the cloud mask probabilities validated with VIIRS, it is seen that a probability cloud mask threshold between 5% and 30% will give a binary cloud mask that meets the requirements on POD-cloudy and FAR-cloudy. It is possible to select the probability cloud mask threshold best suited for the application to get a cloud conservative or clear conservative cloud mask. All datasets pass the threshold requirements for Limit 50% and POD-cloudy passes target requirements for several datasets and FAR-cloudy passes optimal requirements for several datasets.

The error estimates of cloud mask, cloud probability and cloud top pressure have been validated to show that they work as intended. For cloud phase an investigation of the errors shows that the CTTH error estimate and the cloud type are useful to identify pixels with larger risk for misclassifications and that the product errors originating from RTM simulations are small.

7 REFERENCES

Austin, R. T., A. J. Heymsfield, and G. L. Stephens, (2009) Retrieval of ice cloud microphysical parameters using the CloudSat millimeter-wave radar and temperature, J. Geophys. Res., 114, D00A23, doi:10.1029/2008JD010049

Ackerman, S., Menzel, P., and Frey, R.: MODIS Atmosphere L2 Cloud Product (06_L2), 2015, https://doi.org/10.5067/MODIS/MYD06_L2.006

Greenwald, T.J., 2009: A 2 year comparison of AMSR-E and MODIS cloud liquid water path observations, Geophys. Res. Lett., 36, L20805, doi:10.1029/2009GL040394.

Håkansson, N., Adok, C., Thoss, A., Scheirer, R., and Hörnquist, S. 2018:Neural network cloud top pressure and height for MODIS, Atmos. Meas. Tech., 11, 3177–3196, https://doi.org/10.5194/amt-11-3177-2018.

Karlsson, K.G.; Håkansson, N. 2018: Characterization of AVHRR global cloud detection sensitivity based on CALIPSO-CALIOP cloud optical thickness information: Demonstration of results based on the CM SAF CLARA-A2 climate data record. Atmos. Meas. Tech. 2018, 11, 633–649

Karlsson, K.-G., Anttila, K., Trentmann, J., Stengel, M., Meirink, J.F., Devasthale, A., Hanschmann, T., Kothe, S., Jääskeläinen, E., Sedlar, J., Benas, N., van Zadelhoff, G.-J., Schlundt, C., Stein, D., Finkensieper, S., Håkansson, N.,and Hollmann, R. 2017: CLARA-A2: the second edition of the CM SAF cloud and radiation data record from 34 years of global AVHRR data, Atmos. Chem. Phys., 17, 5809–5828, https://doi.org/10.5194/acp-17-5809-2017.

Kato, S. and A. Marshak, 2009: Solar zenith and viewing geometry-dependent errors in satellite retrieved cloud optical thickness: Marine stratocumulus case, J. Geophys. Res., 114, doi: 10.1029/2008JD010579.

Marchand, R., G.G. Mace, T. Ackerman, and G. Stephens (2008), Hydrometeor Detection Using Cloudsat—An Earth-Orbiting 94-GHz Cloud Radar. J. Atmos. Oceanic Technol., 25, 519–533.

Minnis, P., C. R. Yost, S. Sun-Mack, and Y. Chen (2008), Estimating the top altitude of optically thick ice clouds from thermal infrared satellite observations using CALIPSO data, Geophys. Res. Lett., 35, L12801, doi:10.1029/2008GL033947.

Pfreundschuh, S. and Eriksson, P. and Duncan, D. and Rydberg, B. and Håkansson, N. and Thoss, A. (2018) A neural network approach to estimating a posteriori distributions of Bayesian retrieval problems. Atmospheric Measurement Techniques. doi: 10.5194/amt-11-4627-2018. url: https://www.atmos-meas-tech.net/11/4627/2018/

Poulsen C. A, Thomas G. E, Siddans R., Povey A.,, McGarragh G., Schlundt C., Stapelberg S., Stengel M. and Grainger R. G. (2019) Algorithm Theoretical Baseline Document v6.2Community Cloud retrieval for Climate (CC4CL) https://climate.esa.int/media/documents/Cloud_Algorithm-Theoretical-Baseline-Document-ATBD-CC4CL_v6.2.pdf (last access: 12 May 2021)

Wentz, F. J. and T. Meissner. 2004. AMSR-E/Aqua L2B Global Swath Ocean Products derived from Wentz Algorithm, Version 2. [Liquid water path 2006-2010]. Boulder, Colorado USA. NASA National Snow and Ice Data Center Distributed Active Archive Center. doi: https://doi.org/10.5067/AMSR-E/AE_OCEAN.002. [20191012]

Winker, D. (2016). CALIPSO LID L1 Standard HDF File - Version 4.10 [Data set]. NASA Langley Research Center Atmospheric Science Data Center DAAC. https://doi.org/10.5067/caliop/calipso/lid_l1-standard-v4-20

Winker, D. (2016). CALIPSO LID L2 1km Standard HDF File - Version 4.10 [Data set]. NASA Langley Research Center Atmospheric Science Data Center DAAC. https://doi.org/10.5067/CALIOP/CALIPSO/LID_L2_05kmCLay-Standard-V4-20

CALIPSO DOIs

CALIPSO Cloud 1km

https://doi.org/10.5067/CALIOP/CALIPSO/LID_L2_01kmCLay-Standard-V4-10

https://doi.org/10.5067/CALIOP/CALIPSO/LID_L2_01kmCLay-Standard-V4-20

CALIPSO Cloud 5km

https://doi.org/10.5067/CALIOP/CALIPSO/LID_L2_05kmCLay-Standard-V4-10

https://doi.org/10.5067/CALIOP/CALIPSO/LID_L2_05kmCLay-Standard-V4-20

ANNEX A. List of TBC, TBD, Open Points and Comments

TBD/TBC	Section	Resp.	Comment

ANNEX B. Validation results for combined use of CMa and CMa-Prob for high quality clear detection.

When high quality clear retrievals are needed we suggest that CMa and CMa-Prob are combined. And that only pixels which are high quality, clear for CMa and have a cloud probability less than 5% are used. This reduces the false alarm rate clear (see Table 33). Note that over polar land and sea (Metop-B) the proposed 5% threshold is too rigorous, screening out most clear pixels while over polar land still yielding a quite high FAR-clear of 18%. The FAR-clear is also higher for GAC data, one reason is that this is validated with the CALIPSO-5km data that contains very thin clouds.

Table 33: POD-clear and FAR clear for CMa and CMa-Prob and for them combined. Results are separated in land and sea categories.

SENSOR	CMa		CMa-Prob L-5%		CMa and CMa-Prob L-5%		CMa and CMa-Prob L-5% No low-quality	
	POD-clear	FAR-clear	POD-clear	FAR-clear	POD-clear	FAR-clear	POD-clear	FAR-clear
LAND VIIRS	84.4	19.9	63.0	14.1	57.0	10.3	54.5	9.6
SEA VIIRS	82.0	23.2	73.1	15.6	69.4	12.9	68.0	12.7
LAND MODIS	85.0	21.5	42.4	12.9	39.5	7.6	37.5	7.2
SEA MODIS	85.7	25.3	63.3	15.1	61.2	11.8	59.5	11.7
LAND MERIS-2	74.7	24.5	31.8	7.4	27.8	6.0	26.8	5.7
SEA MERSI-2	62.8	23.8	30.9	7.4	26.3	7.0	24.9	6.7
LAND Metop-B	82.9	21.1	10.0	19.5	9.8	18.3	9.6	18.4
SEA Metop-B	66.5	40.3	3.6	5.4	3.5	3.1	3.5	3.1
LAND GAC	84.9	36.6	47.6	21.4	45.8	20.0	44.7	19.5
SEA GAC	83.8	34.2	56.5	22.4	55.6	21.5	54.3	20.8