EUMETSAT

NWC SAF

SUPPORT TO NOWCASTING AND VERY SHORT RANGE FORECASTING

Algorithm Theoretical Basis Document for Cloud Top Temperature, Pressure and Height of the NWC/PPS

NWC/CDOP3/PPS/SMHI/SCI/ATBD/CTTH, Issue 2, Rev. 1

13 December 2018

Applicable to SAFNWC/PPS version 2018

Applicable to the following PGEs:

Acronym	Product ID	Product name	Version number
CTTH	NWC-069	Cloud Top Temperature and Height	5.0

Prepared by Swedish Meteorological and Hydrological Institute (SMHI)

REPORT SIGNATURE TABLE

Function	Name	Signature	Date
Prepared by	SMHI		13 December 2018
Reviewed by	SAFNWC Project Team EUMETSAT		3 February 2017
Authorised by	Anke Thoss, SMHI *SAFNWC PPS Manager*		13 December 2018

DOCUMENT CHANGE RECORD

Version	Date	Pages	Changes
1.0d	22 January 2014	34	Replacing CDOP-document: SAF/NWC/CDOP/SMHI-PPS/SCI/ATBD/3 First version for SAFNWC/PPS v2014 Changes since v2012: -Updated the document according to scientific changes of v2014. -Updated the output format, according to v2014 changes.
1.0	15 September 2014	35	Implemented RIDs from PCR-v2014: -Action4: Provided rationale for using the RTTOV radiative transfer model. -DF2 (rewriting section 1.6; updating to RTTOV 11; option to use actual surface emissivity together with RTTOV) -LSc1 (formal issues) -LSc2 (summary of requirements) -LSc3, PW24-PW25, PW28, PW30-PW33, PW35-PW36 (editorials and clarifications) -PW29 (updating terminology: histogram => scatter plot) General changes: -Added short description about surface emissivity, in scientific updates, and in list of inputs.
2.0d	23 December 2016	39	-Major change of the CTTH algorithm. Now using a neural network. Note: Several sections do not have a proper change log, since they are fully replaced. -Describe channels needed, for each type of instrument, incl. MERSI-2.
2.0	20 February 2017	40	Implemented RIDs from PCR-v2018: -Heinemann-045: Updated scientific reference list -Heinemann-044: Clarifications in section about NN variables. (4.1.2.3) -Heinemann-018: Added a sub-section about processing time, and describe the configuration options related to time and dependencies.
2.1beta	9 May 2018	37	Document code changed from NWC/CDOP2/PPS/SMHI/SCI/ATBD/5 to NWC/CDOP3/PPS/SMHI/SCI/ATBD/5. Changes for v2018: - The statusflag was overhauled, to fit the CTTH-NN algorithm. Some related changes in the quality control as well. - Updates of what networks to train after investigation of performance for higher satellite zenith angles - Changed some of the validation quantities. -Alternative input: CMa-prob.
2.1d	17 October 2018	38	Updates for PPSv2018 ORR: - Updated loss function from MSE to MAE - Added GAC training dataset information - Updated processing time section -Added section 4.2.1 describing the configuration.

2.1	13 December 2018	39	Updates according to PPS v2018 ORR: OBJ2_UM_SCI_Heinemann_039: Removed most PGE-<*number*> notations in this document. OBJ2_UM_SCI_Heinemann_041: editorial Updates according to PPS v2018 PCR: -Added discussion about instrument channel spectral characteristics dependence. (4.2.5.1) from Action: PCR-PPS-v2017/v2018-002 Other changes: -Describe configuration SM_CMAPROB_CLOUD_THRESHOLD.

Table of Contents

1. INTRODUCTION 8

1.1 PURPOSE 8
1.2 SCOPE 8
1.3 DEFINITIONS AND ACRONYMS 8
1.4 REFERENCES 9
1.4.1 Applicable documents 9
1.4.2 Reference documents 10
1.4.3 Scientific references 10
1.4.4 Other references 11
1.5 DOCUMENT OVERVIEW 11
1.6 SCIENTIFIC UPDATES SINCE PPS VERSION 2014 12

2. INTRODUCTION TO THE SAFNWC/PPS CTTH 13

2.1 SATELLITE CHANNELS 13
2.2 REQUIREMENTS 14

3. ALGORITHM OVERVIEW 16

4. ALGORITHM DESCRIPTION 17

4.1 THEORETICAL DESCRIPTION 17
4.1.1 Physics of the Problem 17
4.1.2 Description of the neural network 17
4.1.2.1 Training against calipso pressure. 18
4.1.2.2 MODIS/AVHRR and CALIPSO training data 18
4.1.2.3 Variables 19
4.1.2.4 Separate networks trained for MODIS, AVHRR, VIIRS, MERSI-2 20
4.1.2.5 Technical 21
4.1.2.6 Tuning network parameters. 21
4.1.2.7 Handling of extremes 21
4.2 PRACTICAL CONSIDERATION 22
4.2.1 Configuration 22
4.2.2 Validation 22
4.2.3 Quality control 25
4.2.4 Processing time 25
4.2.5 Assumptions and limitation 26
4.2.5.1 Spectral response function differences 26
4.2.6 List of inputs 26
4.2.6.1 Satellite data 26
4.2.6.2 Cloud Mask 26
4.2.6.3 NWP data 26
4.2.6.4 Parameter files and algorithm configuration files 27
4.2.7 Description of output 27
4.2.8 Visualisation 31

ANNEX A. LIST OF TBC, TBD, OPEN POINTS AND COMMENTS 32

ANNEX B. THE MULTILAYER PERCEPTRON 33

B 1 CONFIGURATION OF THE MULTILAYER PERCEPTRON 35
B 1.1 Activation functions 35
B 1.2 Loss function, Number of epochs and batch size 35
B 1.3 Weight and bias initialization 36
B1.4 Data preprocessing 36
B 1.5 Performance evaluation metrics 36
B 2 TUNING NETWORK PARAMETERS 37

List of Tables and Figures

TABLE 1: LIST OF APPLICABLE DOCUMENTS 10
TABLE 2: LIST OF REFERENCED DOCUMENTS 10
TABLE 3 THE AVHRR/3 CHANNELS AND THEIR APPROXIMATE SPECTRAL POSITIONS ON NOAA 15, 16 AND 17 (SMALL DEVIATIONS IN THE SPECTRAL RESPONSE DO OCCUR BETWEEN DIFFERENT NOAA SATELLITES)....14
TABLE 4: NAMES AND SPECTRAL SPECIFICATIONS OF MODIS CHANNELS, USED SO FAR. 14
TABLE 5: NAMES AND SPECTRAL SPECIFICATIONS OF MERSI-2 CHANNELS, AS PLANNED TO USE 14
TABLE 6: NAMES AND SPECTRAL SPECIFICATIONS OF VIIRS CHANNELS, USED SO FAR. A NEW SPECTRAL RANGE (COMPARED TO THE PREVIOUSLY USED AVHRR CHANNELS) IS INTRODUCED BY THE USE OF CHANNEL M14, BUT THE VIIRS CHANNELS ARE SIMILAR TO THOSE USED FOR MODIS 14
TABLE 7 NAMES AND SPECTRAL SPECIFICATIONS OF VIIRS CHANNELS, WHEN USING A COMBINATION OF I-BAND AND M-BAND CHANNELS 14
TABLE 8 ACCURACY REQUIREMENTS FOR CLOUD TOP HEIGHT 15
TABLE 9 INSTRUMENT CHANNEL ACRONYMS USED BY CTTH 17
TABLE 10 DATA USED FOR TRAINING, VALIDATION AND TESTING THE NETWORK. 18
TABLE 11 DATA USED FOR TRAINING, VALIDATION AND TESTING THE NETWORK GAC. 18
TABLE 12 DESCRIPTION OF THE VARIABLES USED IN THE NEURAL NETWORK TRAINING. 19
TABLE 13 DESCRIPTION OF THE NEURAL NETWORKS TRAINED FOR PPS-V2018 ON MODIS-CALIOP MATCHUPS. 20
TABLE 14 PARAMETERS USED FOR THE NETWORKS. 21
TABLE 15: PERFORMANCE FOR THE CTTH PRESSURE PRELIMINARY V2018 FOR THE DIFFERENT NETWORKS. THE DATA USED IS THE TEST-DATA (NOT USED FOR TRAINING THE NETWORKS 14TH OF JANUARY, JUNE AND SEPTEMBER AND 1ST NOVEMBER). RESULTS FOR OLD PPS ALGORITHM ARE FOR PIXELS WHERE BOTH PPS AND CALIOP CLOUD MASK INDICATE CLOUD AND 14TH OF JANUARY, JUNE AND SEPTEMBER. RESULTS FOR CTTH NEURAL NETWORK ARE FOR ALL PIXELS WHERE CALIOP CLOUD MASK INDICATE CLOUD. 24
TABLE 16 GAIN, INTERCEPT AND NO-DATA VALUES FOR CLOUD TOP TEMPERATURE 28
TABLE 17 GAIN, INTERCEPT AND NO-DATA VALUES FOR CLOUD TOP PRESSURE 28
TABLE 18 GAIN, INTERCEPT AND NO-DATA VALUES FOR CLOUD TOP HEIGHT 28
TABLE 19 CTTH FLAGS 29
TABLE 20 VALIDATION MSE WHEN CHANGING NETWORK PARAMETERS. UNIFORM WEIGHT INITIALIZATION WAS USED AND THE SAME RANDOM SEED. ITS IS THE MERSI-2 NETWORK. THE VARIATION BETWEEN DIFFERENT RANDOM SEEDS IS AROUND 0.002 (SEE TABLE 21). THE SAME RANDOM SEED WAS USED FOR ALL TESTS....37
TABLE 21 VALIDATION MSE WHEN CHANGING WEIGHT INITIALIZATION AND RANDOM SEED. FOR THE OTHER PARAMETERS 30-15 HIDDEN NEURONS WHERE USED, LEARNING RATE WAS 0.01 LEARNING RATE DECAY 1E-06 AND MOMENTUM 0.9. WE CAN SEE HOW MUCH THE MSE VARIES WITH RANDOM SEED AND THAT GLOROT UNIFORM WEIGHT INITIALIZATION IS BETTER FOR THIS VERY SMALL TEST. 38
TABLE 22 MEAN ABSOLUTE ERROR (MAE) IN HPA FOR DIFFERENT HEIGHT CLASSES FOR THREE MODELS (MARKED WITH CIRCLES) FROM TABLE 20. 39

FIGURE 1 COMPARISON OF CTTH PPS-V2018 (LEFT) WITH PPS-V2014(RIGHT). IN THE MIDDLE AN RGB. INSTRUMENT MODIS. NOTICE THE SQUARED ARTEFACTS REMOVED IN VERSION 2018, AND THE CLOUDS MISSING HEIGHT IN PPS-OLD CTTH UPPER RIGHT CORNER. 22
FIGURE 2 SHOWING THE BIAS DISTRIBUTION OF CTTH (V2018 PRELIMINARY) TO THE LEFT AND PPS CTTH-OLD (V2014) TO THE RIGHT. THE RED SQUARE IS THE BIAS MEAN. THE RED LINE IS THE BIAS MEDIAN. THE PINK AREA IS WHERE BIAS MEAN IS OUTSIDE THRESHOLD ACCURACY. THE BOX EDGES ARE THE FIRST AND THIRD QUARTILE. THE WHISKERS INDICATE THE 5TH AND THE 95TH PERCENTILE. NOTICE THE LARGE IMPROVEMENT OF MEAN AND MEDIAN BIAS FOR HIGH CLOUDS. IT IS NOW WITHIN THRESHOLD ACCURACY ALTHOUGH STILL NOT WITHIN. THE DATA IS ONE ORBIT OF SUOMI-NPP (2150711 19:44)....23
FIGURE 3 COMPARISON OF OLD-CTTH (TOP) AND CTTH (BOTTOM) TO CALIOP-HEIGHT. THIS IS A SECTION OF AN AVHRR-GAC ORBIT FROM NOAA-18 20090921, TIME 16:34. NOTICE THE IMPROVED HEIGHT FOR LOW CLOUDS AT TRACK POSITION 3700 AND 5400. ALSO NOTICE THE IMPROVED HEIGHT FOR THE HIGHER CLOUDS AROUND TRACK POSITION 5700. 24
FIGURE 4 EXAMPLE OF CTTH IMAGE DISPLAY USING IMAGE VIEWER DEVELOPED AT SMHI. SCENE: EOS-AQUA, JANUARY 14TH ,2010, 00:05UTC. TO THE RIGHT AN RGB IMAGE USING CHANNELS 20, 31, AND 32 (3.7, 11 AND 12 µM) AND TO THE LEFT THE CORRESPONDING CTTH. THE TWO INFORMATION DIALOGS, TO THE RIGHTMOST, PROVIDE INFORMATION FOR THE PIXEL UNDER THE MOUSE-POINTER (MARKED IN THE RIGHT PICTURE WITH A CROSS). 31
FIGURE 5 ARCHITECTURE OF A MULTILAYER PERCEPTRON 33

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 the Algorithm Theoretical Basis Document for the Cloud Top Temperature and Height (CTTH) of the SAFNWC/PPS software package.

This document contains a description of the algorithm, including scientific aspects and practical considerations.

1.2 SCOPE

This document describes the algorithms implemented in the CTTH version 5.0 of the 2018 SAFNWC/PPS software package delivery.

1.3 DEFINITIONS AND ACRONYMS

Acronym	Explanation
ACPG	AVHRR/AMSU Cloud Product Generation software (A major part of the SAFNWC/PPS s.w., including the PGEs.)
AEMET	Agencia Estatal de Meteorología (Spain)
AHAMAP	AMSU-HIRS-AVHRR Mapping Library (A part of the SAFNWC/PPS s.w.)
AMSU	Advance Microwave Sounding Unit
AVHRR	Advanced Very High Resolution Radiometer
CDOP	Continuous Development and Operational Phase
CDOP-2	Second Continuous Development and Operational Phase
CMa	Cloud Mask (also PGE01)
CMa-prob	Cloud Probability (also PGE01c)
CPP	Cloud Physical Products (also PGE05)
CT	Cloud Type (also PGE02)
CTTH	Cloud Top Temperature, Height and Pressure (also PGE03)
EPS	EUMETSAT Polar System
EUMETSAT	European Organisation for the Exploitation of Meteorological Satellites
GEO	Geosynchronous equatorial orbit
MERSI	Medium Resolution Spectral Imager
MHS	Microwave Humidity Sounding Unit
MODIS	Moderate Resolution Imaging Spectroradiometer
NN	Neural Network
NOAA	National Oceanic and Atmospheric Administration
NWP	Numerical Weather Prediction
PC	Precipitating Cloud (also PGE04)
PGE	Process Generating Element
PPS	Polar Platform System
RGB	Red Green Blue
RTM	Radiative Transfer Model
RTTOV	Radiative Transfer for TOVs
SAF	Satellite Application Facility
SAFNWC	Satellite Application Facility for support to NoWcasting
SMHI	Swedish Meteorological and Hydrological Institute
TBC	To Be Confirmed
TBD	To Be Defined
TOA	Top Of Atmosphere
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	1.2	10/10/18
[AD.2]	NWCSAF Product Requirements Document	NWC/CDOP3/SAF/AEMET/MGT/PRD	1.1	17/12/18
[AD.3]	System and Components Requirements Document for the SAFNWC/PPS	NWC/CDOP3/PPS/SMHI/SW/SCRD	2.1	13/12/18

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/CDOP2/SAF/AEMET/MGT/GLO	2.1	03/02/17
[RD.3]	Output Data Format of the SAFNWC/PPS	NWC/CDOP3/PPS/SMHI/SW/DOF	2.0	13/12/18
[RD.4]	Algorithm Theoretical Basis Document for the Cloud Mask of the NWC/PPS	NWC/CDOP3/PPS/SMHI/SCI/ATBD/CloudMask	2.1	13/12/18
[RD.4b]	Algorithm Theoretical Basis Document for the Cloud Probability of the NWC/PPS	NWC/CDOP3/PPS/SMHI/SCI/ATBD/CloudProbability	1.0	13/12/18
[RD.5]	Algorithm Theoretical Basis Document for the Cloud Type of the NWC/PPS	NWC/CDOP3/PPS/SMHI/SCI/ATBD/CloudType	2.1	13/12/18
[RD.6]	Scientific and Validation Report for the Cloud Product Processors of the NWC/PPS	NWC/CDOP3/OOS/SMHI/SCI/VR/Cloud	2.0	13/12/18

Table 2: List of Referenced Documents

1.4.3 Scientific references

Bengio, Yoshua. (2012) Practical recommendations for gradient-based training of deep architectures. Neural Networks: Tricks of the Trade. Springer Berlin Heidelberg, 437-478

Cotter, A., Shamir, O., Srebro, N., and Sridharan, K. (2011). Better mini-batch algorithms via accelerated gradient methods. Advances in neural information processing systems.

Garcia, Luis A., and Abdalla Shigidi. (2006) Using neural networks for parameter estimation in ground water. Journal of Hydrology 318.1: 215-231.

Gardner, M. W. and Dorling, S. (1998). Artificial neural networks (the multilayer perceptron)-a review of applications in the atmospheric sciences. Atmospheric environment, 32(14):2627–2636

Håkansson, N. and Adok, C. and Thoss, A. and Scheirer, R. and Hörnquist, S. (2018) Neural network cloud top pressure and height for MODIS. Atmospheric Measurement Techniques Discussions. doi: 10.5194/amt-2017-443

Karlik, B. and Olgac, A. V. (2011). Performance analysis of various activation functions in generalized mlp architectures of neural networks. International Journal of Artificial Intelligence and Expert Systems, 1(4):111–122.

Karlsson, K.-G. and Dybbroe, A., 2009. Evaluation of Arctic cloud products from the EUMETSAT Climate Monitoring Satellite Application Facility based on CALIPSO-CALIOP observations. Atmos. Chem. Phys. Discuss., 9, 16755-16810, 2009.

Karlsson, K.-G., Rühelä A., Müller R., Meirink J.F., Sedlar J., Stengel M., Lockhoff M., Trentmann J., Kasper F., Hollmann R. and Wolters E., 2013. CLARA-A1: a cloud, albedo and radiation dataset from 28yr of global AVHRR data.. Atmos. Chem. Phys. 13,5351-5367, 2013 doi:10.5194/acp-13-5351-2013

Karlsson, K.-G. and Johansson E, 2013, On the optimal method for evaluating cloud products from passive satellite imagery using CALIPSO-CALIOP data: example investigating the CM SAF CLARA-A1 dataset. Atmos Meas. Thech. 6, 1271-1286, 2013 doi:105194/amt-6-1271-2013

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.: CLARA-A2: The second edition of the CM SAF cloud and radiation data record from 34 years of global AVHRR data, Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2016-935, in review, 2016.

Leen, Todd K, and Genevieve B. Orr. Optimal stochastic search and adaptive momentum. Advances in Neural Information Processing Systems 6 (NIPS 1993)

Menzel, W. P., l. Smith, W., and Stewart, T. R., 1983. Improved Cloud Motion Wind Vector and Altitude Assignment using VAS. Journal of Climate and Applied Meteorology 22, 377-384.

Theano Development Team (2016). Theano: A Python framework for fast computation of mathematical expressions. arXiv e-prints, abs/1605.02688, http://arxiv.org/abs/1605.02688

1.4.4 Other references

Chollet Français, Keras (2015) Github, https://github.com/fchollet/keras.

1.5 DOCUMENT OVERVIEW

This document contains a theoretical description of the algorithms for cloud top Temperature, Pressure and Height derivations. The document has been structured in the following sections:

- Section 1 contains the current introduction along with the list of used acronyms and applicable and reference documents.
- Section 2 A short introduction to the CTTH product
- Section 3 A short overview of the CTTH algorithm
- Section 4 Algorithm description in more detail

1.6 SCIENTIFIC UPDATES SINCE PPS VERSION 2014

For v2018 CTTH has been totally updated, the algorithm has changed from one algorithm for opaque clouds and another algorithm for semi-transparent clouds, to a neural network used for all types of clouds.

Previously the CTTH was dependent on the Cloud Type product. Now it is only dependent on the Cloud Mask product (which comes earlier in the PPS processing chain).

Compared to the earlier version, the current algorithm with several different networks takes advantage of extra channels when available. Useful channels that will improve results are brightness temperatures 3.7, 8.6, 7.3, 13.3 μm. PPS histogram technique could not benefit from these extra channels.

Current algorithm will also run for AVHRR1. AVHRR1 has no channel 12μm brightness temperature. PPS histogram technique algorithm could not run without it, while one of the neural networks is trained using only 3.7 and 11 μm brightness temperatures, simulated threshold and NWP-data.

The v2018 NN algorithm does not have the squared artefacts common in the PPS histogram technique; this makes the information visually much clearer.

The neural network was trained using MODIS collocated with CALIPSO. Separate networks was also trained using AVHRR-GAC data, these are intended for GAC processing.

The neural network algorithm has even higher retrieval rates and also generally improved bias and RMS.

The neural network algorithm has no use for RTTOV-online simulations. And RTTOV is no longer needed to be installed with PPS..

2. INTRODUCTION TO THE SAFNWC/PPS CTTH

The cloud top temperature, pressure and height (CTTH) retrieval based on polar orbiter data developed within the SAFNWC project aims at Nowcasting applications. The main use of this product is in the analysis and early warning of thunderstorm development and the height assignment for aviation forecasting. The product may also serve as input to mesoscale models for use in Nowcasting in general, or as input to other satellite retrievals used for Nowcasting. The SAFNWC CTTH retrieval, based on imager data from polar orbiter, will also be used to build up cloud climatologies within the CMSAF.

The CTTH product aims at providing information on the cloud top temperature and height for all pixels identified as cloudy in the satellite scene.

Many NMSs of EUMETSAT member states (including SMHI) use still today the uncorrected brightness temperature information from IR imagery as a rough estimation of cloud top temperatures. For the optically thick clouds this estimation is in most cases acceptable. However, for pixels containing semi-transparent or fractional clouds (often representing a large fraction of cloudy pixels) this information is definitely misleading, yielding sometimes to quite a large underestimation of true cloud top heights.

The objective of the SAFNWC CTTH product has been to create a retrieval that as far as possible (considering both computational accuracy and CPU efficiency aspects) compensates for the semi-transparency effect and the effect of an absorbing atmosphere between the cloud top and the satellite sensor.

It must, however, be remembered that neither the NOAA, Metop nor the EOS satellites do provide the most optimal platform for semi-transparency correction and cloud top temperature and height retrieval in general. The derivation of the cloud top height using the instruments on these satellites will naturally be rather indirect requiring a lot of ancillary data like NWP model output. Other more direct techniques exist, e.g. using stereo-scope imagery requiring a setup of two geostationary satellites with overlapping fields of view.

Sounding channels as on the HIRS instrument would provide the possibility for applying the radiance rationing technique, as detailed by Menzel et al. (1983). This technique applies to single layers of high semi-transparent clouds. The HIRS channels do, however, have rather poor horizontal and vertical resolution. The AVHRR and similar instruments provide window channels which may be used to estimate height of semi-transparent clouds. The true temperature of a semi-transparent cloud will effect the difference between the two window channels, and the difference to the surface temperature.

For the SAFNWC CTTH we have chosen the latter technique to be applied to AVHRR data and likewise channel combinations from other instruments.

2.1 SATELLITE CHANNELS

Satellite channels used by the SAFNWC/PPS are from the Metop and NOAA imager instrument AVHRR/3 as well as from MODIS and VIIRS, and planned for MERSI-2. See Table 3 -Table 7 for more details about the different channels.

Table 3 The AVHRR/3 channels and their approximate spectral positions on NOAA 15, 16 and 17 (small deviations in the spectral response do occur between different NOAA satellites).

	Ch 3B	**Ch 4**	**Ch 5**
λ (µm)	**3.55-3.93**	**10.3-11.3**	**11.5-12.5**

Table 4: Names and spectral specifications of MODIS channels, used so far.

	Ch 20	**Ch 27**	**Ch 28**	**Ch 29**	**Ch 31**	**Ch 32**	**Ch 33**	**Ch 34**	**Ch 35**	**Ch 36**
λ (µm)	**3.75**	**6.7**	**7.3**	**8.6**	**11.0**	**12.0**	**13.3**	**13.6**	**13.9**	**14.2**

Table 5: Names and spectral specifications of MERSI-2 channels, as planned to use

λ (µm)	**3.80**	**7.20**	**8.55**	**10.8**	**12.0**

Table 6: Names and spectral specifications of VIIRS channels, used so far. A new spectral range (compared to the previously used AVHRR channels) is introduced by the use of channel M14, but the VIIRS channels are similar to those used for MODIS

	Ch M12	**Ch M14**	**Ch M15**	**Ch M16**
λ (µm)	**3.7**	**8.55**	**10.763**	**12.013**

Table 7 Names and spectral specifications of VIIRS channels, when using a combination of I-band and M-band channels

	Ch I4	**Ch M14**	**Ch M15**	**Ch M16**
λ (µm)	**3.74**	**8.55**	**10.763**	**12.013**

2.2 REQUIREMENTS

The requirements for the SAFNWC/PPS products are described in the Product Requirements Document [AD.2]. In Table 8 is given a summary of the requirement specific for the CTTH product.

Table 8 Accuracy requirements for Cloud Top Height

	Bias (opaque clouds)	**Standard deviation (opaque clouds)**	**Bias (semi-transparent clouds)**	**Standard deviation (semi-transparent clouds)**
Threshold accuracy	1000 m	2000 m	2000 m	2000 m
Target accuracy	500 m	1500 m	1500 m	1500 m
Optimal accuracy	200 m	500 m	200 m	500 m

3. ALGORITHM OVERVIEW

- Neural network trained offline to retrieve cloud top pressure. MODIS data collocated with CALIPSO are used for the training. Several networks with different channels are trained. Networks are also trained with AVHRR-GAC data, these are intended for GAC processing.
- Neural network with coefficients are loaded considering available channels.
- Retrieve the cloud top pressure from the neural network.
- Read NWP-data and retrieve height and temperature corresponding to the pressure retrieved by the neural network.

4. ALGORITHM DESCRIPTION

4.1 THEORETICAL DESCRIPTION

4.1.1 Physics of the Problem

As described in ATBD-01 [RD.4] and ATBD-02 [RD.5] also this algorithm tries to extract information from radiances, observed by the detector. In this case the cloud top pressure (and derived properties) is assigned by relating it to the measured brightness temperature.

There are certain differences for pixels with and without contributions from ground. Cases where information from emissions underneath the cloud do not contribute to the signal (opaque clouds) are relatively easy to handle in this sense. The brightness temperature of channel 11 µm is then close to the cloud temperature. The situation is more complex if this is not true (semi-transparent or fractional clouds). For these clouds a mixture of surface and cloud temperature is measured in channel 11 µm. These clouds also have a larger difference between channel 11µm and 12µm. The texture of channel 11 is also typically larger for fractional clouds compared to semi-transparent clouds. The neural network is trained using several variables that holds information on the type of cloud and there for can also estimate a pressure for these more complex cases.

Radiation is not emitted from a clouds facet (however this is defined) but from the whole cloud. Emission is a volume property. Before a certain beam reaches a detector, it underlies a complex interplay of absorption and emission. This means that the observed brightness temperature is representative for a layer somewhat below the cloud top. How far below depends on the optical properties of the cloud.

Table 9 Instrument channel acronyms used by CTTH

Acronym	Description
T11	**Brightness temperature at 11 µm**
T12	**Brightness temperature at 12 µm**
T3.7	**Brightness temperature at 3.7 µm**
T6.7	**Brightness temperature at 6.7 µm**
T7.3	**Brightness temperature at 7.3 µm**
T8.5	**Brightness temperature at 8.5 µm**
T13.3	**Brightness temperature at 13.3 µm**
T13.6	**Brightness temperature at 13.6 µm**
T13.9	**Brightness temperature at 13.9 µm**
T14.2	**Brightness temperature at 14.2 µm**

4.1.2 Description of the neural network

A multilayer perceptron is used for the neural network. The multilayer perceptron is a supervised learning technique since it learns by example. For a more detailed description of the multilayer

perceptron see ANNEX B. The CTTH algorithm of PPS version 2018 is also well described in Håkansson et al (2018). The final version of PPS-v2018 is trained with CALIOP version 4 data and uses MAE as loss function in the training of the networks.

4.1.2.1 Training against calipso pressure.

We chose to train the network to retrieve cloud top pressure. One reason for this is that the NWP-data in PPS already provide temperatures on pressure levels. Minor testing showed that similar results can be achieved for retrieval of height directly from the networks.

4.1.2.2 MODIS/AVHRR and CALIPSO training data

Global MODIS data co-located with CALIPSO for 18 days during 2010 was used to train the networks. For training the neural network the data needs to be divided into three parts: training, validation and test. The validation set is used to tune the parameters of the neural network. The test set is a dataset not used during training and is used to evaluate the performance of the network. For GAC training 1118 orbits from Noaa-18 and Noaa-19 2007-2010 was used. (Table 10,

Table 11).

Table 10 Data used for training, validation and testing the network.

Training data	1st January Mars July September 14th Mars April May August October December
Validation data	1st May 14th Mars July November
Test data	1st November 14th January June September

Table 11 Data used for training, validation and testing the network GAC.

Training data	Noaa-18 537 orbits from 2007 and 2008 Noaa19 262 oribts from 2010
Validation data	Noaa-18 191 orbits from 2007 and 2008 Noaa19 101 oribts from 2010 and 2009
Test data	Nooa-18 28 orbits from 2009

The data used for training and validation consisted of 50% low clouds, 25 % medium level clouds and 25% high clouds. The observations from each cloud class are sampled from all observations in the data belonging to the respective cloud class.

The training data is shuffled before each epoch to get a random sample of data in each batch. For the training data 1 500 000 observations were used and for the validation data 375 000 observations, resulting in four times more training data. The validation data is used to decide when to quit training, this is needed to avoid over fitting. After choosing the final model, the model is tested on an unseen

test data that has not been used during the training of the network, to evaluate the performance of the network.

The performance of the algorithm is validated using completely different datasets and/or instruments (VIIRS, AVHRR, MODIS). See validation report [RD.6] for more info.

4.1.2.3 Variables

Variables used to train the network are: variables derived from the IR imager channels, simulated cloud free thresholds and NWP-variables (mostly temperature at different pressure levels). The simulated cloud free thresholds are the same as used in the CMa product. RTM simulations of cloud free land and ocean surfaces for different sun and satellite angles are made for about 10000 NWP-profiles. The column integrated water vapour, surface temperature, satellite zenith angle, illumination condition (day or night), and fraction of land is used to read the corresponding threshold. The simulated cloud free thresholds was not included in the networks released with v2018 of PPS. See Table 12 for more information about the variables used to train the neural networks.

Table 12 Description of the variables used in the neural network training.

Variable type	**Variables**	**Why is it needed ?**
Pressure	Surface pressure (psur)	Gives network highest possible pressure for the pixel
NWP temperatures	Temperature at surface 950, 850, 700, 500 and 250 hPa.	Example: If the temperature of the cloud is close to the 250 hPa temperature, the clouds are more likely to be at a height close to 250hPa.
Predicted Water vapour	Column integrated water vapour (ciwv)	Gives for example information on expected difference between t11 and t12.
Texture: standard variation of variable in 5x5 pixels (gac 3x3)	Texture for T11-T12, T11-T3.7, T11, T3.7	Holds information on if cloud is likely to be opaque, semi-transparent, or edge/fractional.
Brightness temperature of 11 or 12 um	T11 or T12	For opaque clouds this should be close to the temperature of the cloud.
Brightness temperature of water vapour channels	T6.7, T7.3	Important to identify high semi transparent clouds.
Brightness temperature of CO_2 channels	T13.3 T13.6, T13.9, T14.2	Important to identify high semi transparent clouds.
Channel temperature differences between pixel and warmest/coldest neighbour for the same channel.	T12-T12warmest T12-T12coldest, T11-T11warmest, T11-T11coldest	Holds information on if cloud is likely opaque, semi-transparent, or edge /fractional.
Channel temperature	T11-T12, T11-T37, T8.5-T12,	Holds information on if cloud

Variable type	Variables	Why is it needed ?
differences between different channels.	T7.3-T12, T13.3-T12, T6.7-T12 Differences calculated for pixel, warmest and coldest neighbour.	is likely opaque, and if so where to place it in the atmosphere, or whether it is semi-transparent, or edge /fractional. The T11-T12 differences for warmest/coldest pixel contains the key information also used by the PPS histogram technique for semi transparent clouds, used by CTTH up to PPS v2014.
Not Used Offline simulated clear-sky temperature differences. Dependent on: • surface (land/sea) • illumination (day/night) • surface temperature • ciwv • satellite zenith angle • fraction of land	Simulated mean T11-T12 (T11T12_sim) Simulated mean T11-T3.7 (T11T3.7_sim) Simulated mean minus two standard deviations of T11-TSUR (T11Tsur_sim).	Holds information of what channel differences could be expected for a cloud free pixel. Should make algorithm better for high satellite zenith angles although it is trained for nadir and helps global applicability. Currently not included due to no apparent performance improvement for higher satellite zenith angles.

4.1.2.4 Separate networks trained for MODIS, AVHRR, VIIRS, MERSI-2

PPS version 2018 takes advantage of additional channels from instruments where they are available. Several networks were trained using channel combination corresponding to the instruments AVHRR, VIIRS, MERSI-2, MODIS, AVHRR1. See Table 13 for information on the different networks trained.

T11w = T11 warmest, T11c = T11 coldest.

Table 13 Description of the neural networks trained for PPS-v2018 on MODIS-CALIOP matchups.

Network	Variables used
NN-T11T12 : Imager channels T11 and T12.	T12, T11-T12, T11w-T12w, T11c-T12c, T12w-T12, T12c-T12, ciwv, tsur, psur, t950, t850, t700, t500, t250, T11T12text, T11text,
VIIRS: Imager channels T11, T12 and T8.5 (not included in release due to bad performance for high satellite zenith angles)	T12, T11-T12, T11w-T12w, T11c-T12c, T12w-T12, T12c-T12, ciwv, tsur, psur, t950, t850, t700, t500, t250, T11T12text, T11text, T8.5-T11,

Network	**Variables used**
MERSI-2: Imager channels T11, T12 and T8.5 and T7.3 (not included in release due to bad performance for high satellite zenith angles)	T12, T11-T12, T11w-T12w, T11c-T12c, T12w-T12, T12c-T12, ciwv, tsur, psur, t950, t850, t700, t500, t250, T11T12text, T11text, T8.5-T11, T7.3,
MetImage: Imager channels T11, T12, T8.5, T7.3, T6.7 and T13.3 (not included in release due to bad performance for high satellite zenith angles)	T12, T11-T12, T11w-T12w, T11c-T12c, T12w-T12, T12c-T12, ciwv, tsur, psur, t950, t850, t700, t500, t250, T11T12text, T11text, T8.5-T11, T7.3, T6.7, T13.3
MODIS: Imager channels T11, T12, T8.5, T7.3, T6.7 and T13.3 13.6, 13.9, 14.2 (not included in release due to bad performance for high satellite zenith angles)	T12, T11-T12, T11w-T12w, T11c-T12c, T12w-T12, T12c-T12, ciwv, tsur, psur, t950, t850, t700, t500, t250, T11T12text, T11text, T8.5-T11, T7.3, T6.7, T13.3, T13.6, T13.9, T14.2, T11T12_sim, T11Tsur_sim
NN-T11T37 : Imager channels T11 and T37 (intended for AVHRR1 processing)	T11, T11-T37, T11w-T37w, T11c-T37c, T11w-T11, T11c-T11, ciwv, tsur, psur, t950, t850, t700, t500, t250,, T11text, T37text

4.1.2.5 Technical

Python modules scikit-learn (see Pedgrosa 2011), keras (see Chollet 1.4.4) and theano (Theano Development Team (2016)) is used for training the network and retrieving the cloud top pressure.

4.1.2.6 Tuning network parameters.

The learning rate, learning rate decay, momentum, number of hidden layers and number of hidden neurons were varied to see what combination would give the best results. Also the weight initialization was tested with both uniform and glorot uniform distribution. The tuning was done for the MERSI-2 network. As there was not enough time to perform tuning for all networks, the same tuning was used for all the networks. Table 14 describes the parameters we choose to use.

Table 14 Parameters used for the networks.

Number of hidden layers	Number of hidden neurons	Learning rate	Learning rate decay	Momentum	Weight initialization
2	30-15	0.01	1E-6	0.9	Glorot uniform

4.1.2.7 Handling of extremes

If a cloud top pressure higher than surface pressure is retrieved, it is set to surface pressure. If a cloud top pressure below 70hPa is retrieved it is set to nodata.

4.2 PRACTICAL CONSIDERATION

4.2.1 Configuration

In pps_basic_configure.py what network to use for different satellites can be configured. Generally this does not need to be altered as the default "nn_t11t12" is the best and most robust performing network delivered with pps-v2018. However when processing on GAC data it is better to use the "nn_avhrr_gac_mae" network. Networks named *_avhrr1_* are suitable for the AVHRR1 instrument.

4.2.2 Validation

Currently CLOUDSAT/CALIPSO data represents the best available reference dataset for quality assessment of cloud height products. Validation on global MODIS, VIIRS, AVHRR (GAC) data against CALIPSO is done for version 2018 of CTTH, see the validation report for results [RD.6]. The CTTH algorithm of PPS version 2018 is also well described and performance is compared to MODIS collection 6 as well as PPS-v2014 in Håkansson et al (2018). See Figure 3 for an example of visual comparison of CTTH and CALIPSO-data. And see Figure 1 for a comparison of CTTH preliminary version 2018 to version 2014. In Figure 2 the difference in error distributions for one orbit of VIIRS data is shown. Notice the 500 meter mean bias improvement for high clouds.

The CM SAF cloud dataset CLARA-A1 and CLARA-A2, used PPS (version 2010 and 2014) to retrieve cloud top temperature and height. Please note that this was a completely different algorithm. Results for the older versions of CTTH can be found in for example (Karlsson and Johansson 2013, Karlsson et al. 2013, Karlsson et al. 2009 and Karlsson et al. 2016).

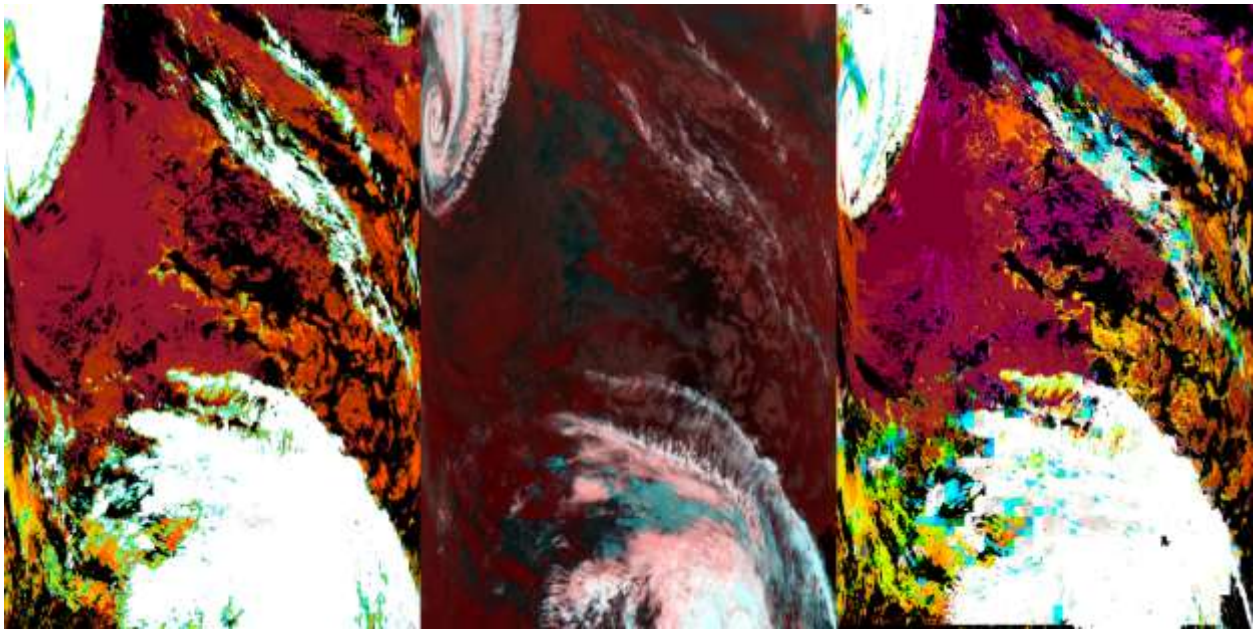

Figure 1 Comparison of CTTH PPS-v2018 (left) with PPS-v2014(right). In the middle an RGB. Instrument MODIS. Notice the squared artefacts removed in version 2018, and the clouds missing height in PPS-old CTTH upper right corner.

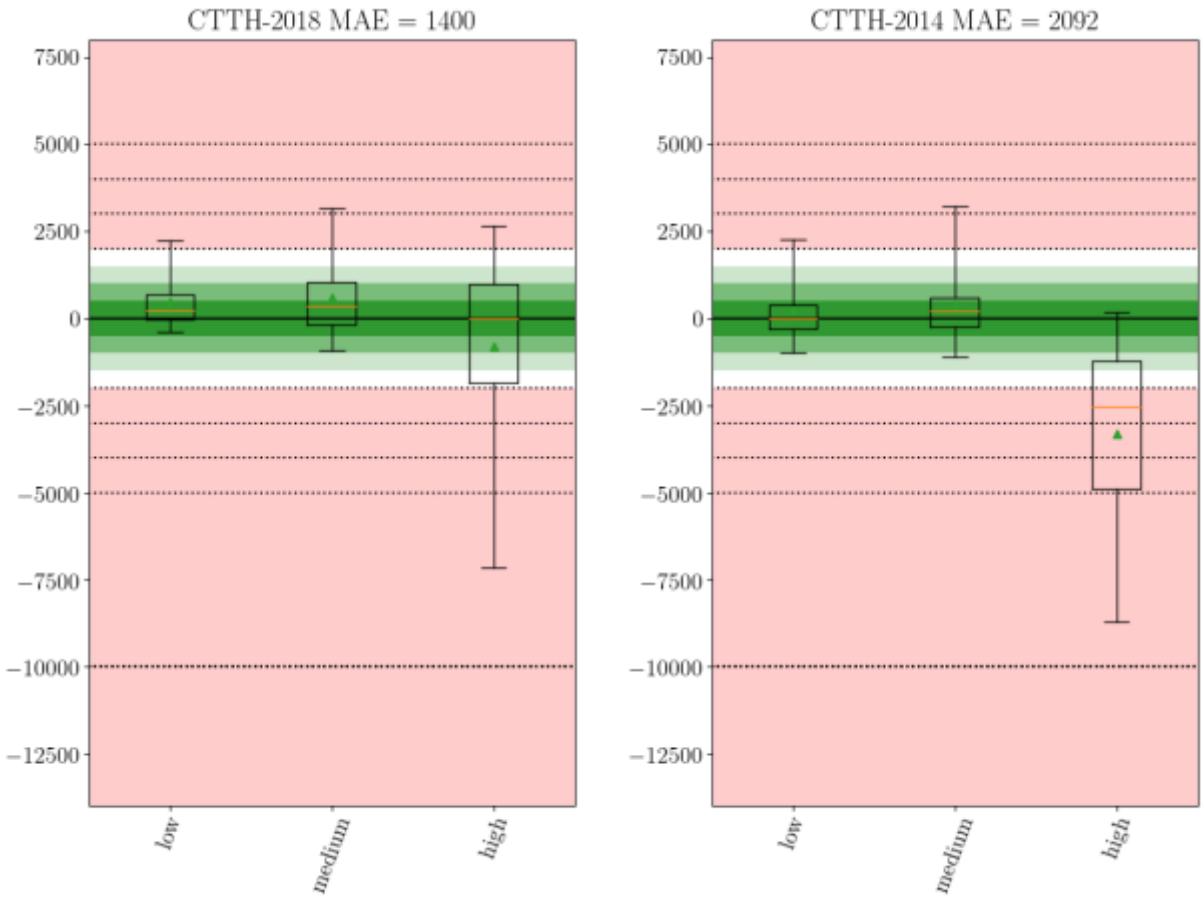

Figure 2 Showing the bias distribution of CTTH (v2018 preliminary) to the left and PPS CTTH-old (v2014) to the right. The red square is the bias mean. The red line is the bias median. The pink area is where bias mean is outside threshold accuracy. The box edges are the first and third quartile. The whiskers indicate the 5th and the 95th percentile. Notice the large improvement of mean and median bias for high clouds. It is now within threshold accuracy although still not within. The data is one orbit of Suomi-NPP (2150711 19:44).

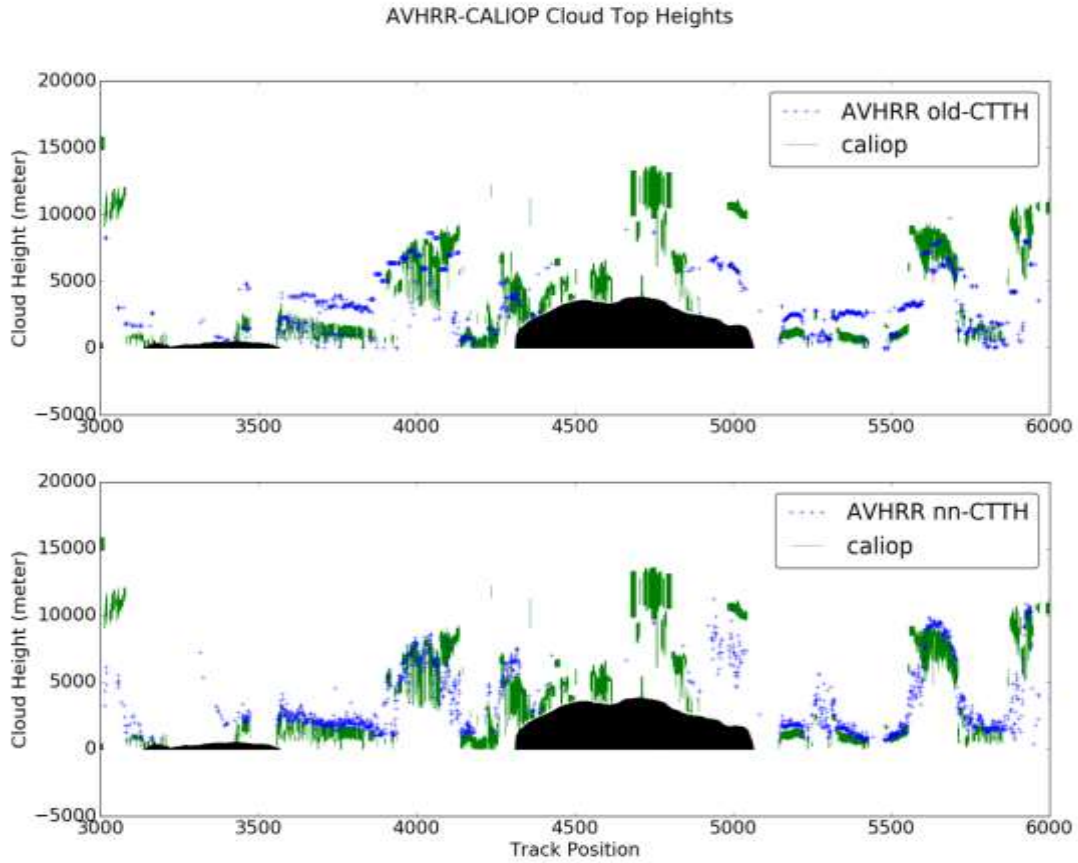

Figure 3 Comparison of old-CTTH (top) and CTTH (bottom) to CALIOP-height. This is a section of an AVHRR-GAC orbit from NOAA-18 20090921, time 16:34. Notice the improved height for low clouds at track position 3700 and 5400. Also notice the improved height for the higher clouds around track position 5700.

Table 15: Performance for the CTTH pressure preliminary v2018 for the different networks. The data used is the test-data (not used for training the networks 14th of January, June and September and 1st November). Results for old PPS algorithm are for pixels where both PPS and CALIOP cloud mask indicate cloud and 14th of January, June and September. Results for CTTH neural network are for all pixels where CALIOP cloud mask indicate cloud.

	MAE (hPa)				Median (hPa)				IQR (hPa)			
	all	low	medium	high	all	low	medium	High	all	low	medium	high

NN-T11T37	74	53	67	95	0	-21	-14	38	87	65	93	119
NN-T11T12	70	53	66	88	36	4	66	54	76	58	85	114
Old algorithm PPS v2014	120	77	86	171	42	3	4	127	149	106	118	186

4.2.3 Quality control

The CTTH processing flags include a flag for low confidence (see 4.2.7). This flag is set to N/A in case of nodata (eg. cloud-free), when CTTH data is present it is normally set to good quality. If there are problems or doubts, the quality can be set to bad or questionable -currently it is only being set to questionable in situations when the cloud top pressure is higher than the surface pressure. This is condition also indicated by the status flag.

If a CTTH value is considered too bad to keep, the CTTH-values are set to nodata, quality flag is set to N/A and one of the status flag is set to tell about the condition. This will be done: if the cloud top pressure originally was set to a value outside allowed limits (below 70hPa or above 1.400hPa), or if the NWP-data (surface temperature) is missing in a pixel.

In the presence of low-level inversion cloud height retrieval is more difficult. However validation divided by CALIOP cloud type show that also cases where low level inversion is present does not consistently show degraded results; varies between validation sets and between CALIOP cloud types. However for low level clouds (as determined by CALIOP) the amount of errors larger than 500m is larger for cases with low level inversion than for cases with out low level inversion. However the overall validation scores for low level clouds with inversion show statistics for most cases with in target accuracy, which makes it misleading to flag all cases with low level inversion as bad quality.

4.2.4 Processing time

If you are in a hurry to get your CTTH product, and only being interested in the CTTH-pressure dataset, you can configure CTTH to produce only cloud top pressure, not height or temperature. That will reduce processing time substantially. Please note that other configurations (eg. only cloud top height) are possible, but they will not save processing time.

But keep in mind that the quality of the down stream products (i.e. CT, PC and CPP) from CTTH will get a bit reduced quality under this configuration.

Note that the processing time for CTTH is much dependent on the resolution of the NWP data used. PPS can be configured to use only a subset of the vertical levels available in a NWP grib file, this is set the pps_nwp.ini and can reduce both processing time and RAM- memory usage significantly. For NWP data with 137 hybrid levels using only a subset of levels is very highly recommended.

See also section 4.2.6.4 for configuration options, as well as details on PGE dependences.

4.2.5 Assumptions and limitation

The data used for the multilayer perceptron is non independent and identically distributed (non-i.i.d). The fact that there is correlation among the observations is not taken into account and the observations are treated as i.i.d. when using the multilayer perceptron.

As the network is trained for near nadir satellite zenith angles, there could be a risk for satellite zenith angle dependence. However it is shown in Håkansson et al (2018) that there are no satellite zenith angle dependence for the neural networks NN-T11T12 and NN-T11T37.

4.2.5.1 Spectral response function differences

As the networks are trained for MODIS but applied on VIIRS and AVHRR data there is a risk for degraded products due to spectral response function differences. However the validation report [RD.6] shows similar validation statistics for both VIIRS and MODIS, despite the spectral response function differences. This means that at least for channel 11 and channel 12 µm the spectral response function differences are not a problem.

4.2.6 List of inputs

4.2.6.1 Satellite data

The same spectral information as needed by the CMa is needed (see section *List of inputs* in ATBD-01 [RD.4]), as the cloud mask is mandatory input to the CTTH.

For CTTH channel 11 µm is mandatory. Additionally either channel 12 or 3.7 µm is needed. Channels 8.5, 6.7, 7.3 and 13.3 µm are optional and will improve the performance.

4.2.6.2 Cloud Mask

A cloud mask product as provided by PPS is a mandatory input for the CTTH. The default input would be the CMa-product ([RD.4]). An option is to use the Cloud Probability (CMa-prob) product instead (see 4.2.6.4) ([RD.4b]).

4.2.6.3 NWP data

Like the CMa and the CT, CTTH uses NWP parameters as either provided by a short range forecast (lead times between 6 and 24 hours) in case of nowcasting, or as provided by a valid analysis in case of off line processing (as e.g. in re-processing for climate applications).

As compared to the cloud mask and cloud type the CTTH is, however, much more dependent on NWP information. The CTTH requires the temperature at quite high vertical resolution for the accuracy of height and temperature. Using full vertical resolution is fine for the quality, but takes unnecessary processing time. Levels above 70 hPa (i.e. lower pressure than 70 hPa) can be omitted without problems. For levels below 70 hPa a selection can be made -the fewer levels the faster processing, but at a cost in quality. I you have 137 levels originally, you would need at least every third level. (Make the settings in pps_nwp.ini -see UM for details.)

CTTH uses the parameters mapped to pixel resolution, and read from a PPS intermediate file, as inputs to the neural network. For the retrieval of height and temperature from pressure 3-dimensional NWP-variables are needed. These are used in higher spatial resolution compared to PPS-V2014. These are not written to file, but instead read and remapped when needed.

4.2.6.4 Parameter files and algorithm configuration files

The CTTH has a one configuration parameter related to what input is wanted. It can be found in the file `pps_config_common.cfg`:

- USE_CMAPROB (default False): Whether to use CMa-prob as input, otherwise CMa is used as input.

The CTTH has a few configuration parameters related to how much is wanted in the final output. These can be found in the file `pps_config_common.cfg` and are listed here:

- GENERATE_PRESSURE (default True): Whether the cloud top pressure output is wanted.
- GENERATE_TEMPERATURE (default True): Whether the cloud top temperature output is wanted.
- GENERATE_HEIGHT (default True): Whether the cloud top height (in meters) output is wanted.
- GENERATE_PROCESSING_FLAG (default True): Whether the processing flags are wanted.

If you want to save time you can chose the configuration: GENERATE_PRESSURE=True, GENERATE_TEMPERATURE=False, GENERATE_HEIGHT=False. But, of course, then you will only get the cloud top pressure, not the other two main datasets. Please note that other configurations (eg. only cloud top height) are possible, but they will not save processing time.

Please observe that changes in the CTTH configuration will affect also the other PGEs processed:

- CMa and CMa-prob: Not affected, as they are not using the CTTH product as input.
- CT: Uses all three CTTH datasets. Can run on only CTTH-pressure, but with reduced quality.
- PC: Gets only indirect effects from using the CT.
- CPP: Uses CTTH-height and CTTH-temperature, but can also run with just one of them, or without using any CTTH data, but with somewhat reduced quality.

Among the environment variables (source_me/.profile) you find SM_CMAPROB_CLOUD_THRESHOLD. This variable is only relevant when Cloud Probability is used as input, instead of the Cloud Mask. The value defines a threshold on cloud probability, for what probability is considered to be cloudy. Default value is 50.0 (i.e. 50% probability).

If you want to configure for using less than full vertical resolution for the NWP-data, you should have a pps_nwp.ini-file setting hybrid_level_selection.

4.2.7 Description of output

The CTTH produces three parameters for the cloud top, namely the temperature, the height in meters and the height in pressure units.

In addition, for each pixel a set of processing flags describe the method applied and provide information on the conditions under which the pixel was processed, and thought to be important for the assessment of the quality of the cloud top estimation.

So in total the content of the CTTH consists of six 16 bit dataset. Three datasets are for temperature, pressure and height. Three datasets contains information in the status, quality and conditions flags. They are all described in more detail below.

The CTTH algorithm tries to produce an output for every pixel classified as cloudy by the Cloud Mask. Cloud free pixels and pixels outside the satellite swath become non-processed. In addition pixels where the calculated cloud top pressure is out of range will be set to nodata, and also pixels where the NWP-data is missing.

Cloud Top Temperature

The cloud top temperature is stored using a linear conversion from 16bit count to temperature, as

T = gain × count + intercept

The gain, intercept and no-data value (for missing data = outside swath - or no data = no result due to failed retrieval or corrupt input data) are listed below:

Table 16 Gain, intercept and no-data values for cloud top temperature

Gain	Intercept	Nodata
0.01K/count	**0.0K**	**65535**

Cloud Top Pressure

The cloud top pressure is stored using a linear conversion from 16bit count to pressure like as done for the temperature. The gain, intercept and no-data value are listed below:

Table 17 Gain, intercept and no-data values for cloud top pressure

Gain	Intercept	Nodata
10Pa/count	**0.0Pa**	**65535**

Cloud Top Height

The cloud top height is stored using a linear conversion from 16bit count to height like as done for the temperature. The gain, intercept and no-data value are listed below:

Table 18 Gain, intercept and no-data values for cloud top height

Gain	Intercept	Nodata
1m/count	**0.0m**	**65535**

Processing flags

There are three flags ctth_quality, ctth_conditions and ctth_status. They contain information about the quality and conditions of each pixel. See

Table 19 for a more detailed description.

Table 19 CTTH flags

Bit number	**Flag: ctth_quality**	**Explanation**
0	Non-processed	Containing no data. Cloudfree pixel or pixel where no reliable cloud pressure could be retrieved.
1	Spare	
2	Spare	
3 to 5	Quality	0: N/A nodata 1: Good 2: Questionable (cloud top pressure above psur) 3: Bad 4: Interpolated
Bit number	**Flag: ctth_status**	**Explanation**
0	Cloud-free	Cloud-free
1	Pressure below lower bound	Calculations gave a cloud top pressure below lower bound (i.e. the clouds were placed too high, but has been replaced by nodata)
2	Pressure above upper bound	Calculations gave a cloud top pressure above upper bound (i.e. the clouds were placed too low, but has been replaced by nodata)
3	Pressure above surface pressure	Cloud top pressure is above surface pressure. (i.e. the clouds can be suspected to be too low, and are flagged as questionable quality)
4	Inversion	Low level thermal inversion in NWP
5	NWP low quality	NWP data suspected low quality
Bit number	**Flag: ctth_conditions**	**Explanation**
0	Outside swath	Pixel is out of swath or points to space
1-2	Illumination	Defines the illumination condition: 0: N/A 1: Night 2: Day

		3: Twilight
3	Sunglint	Sunglint
4-5	Land Sea	Defines whether it is land or sea: 0: N/A 1: Land 2: Sea 3: Coast
6	High Terrain	High terrain
7	Rough Terrain	Rough terrain
8-9	Satellite input data	Satellite input data status: 0: N/A 1: All satellite data are available 2: At least one useful channel is missing 3: At least one mandatory channel is missing
10-11	NWP input data	NWP input data status: 0: N/A 1: Data Available 2: Useful data missing 3: Mandatory data missing
12-13	Product input data	Product input data status: 0: N/A 1: Data Available 2: Useful data missing 3: Mandatory data missing
14-15	Auxilary input data	Auxiliary data status: 0: N/A (not classified pixel or auxiliary data not used) 1: All auxiliary data are available 2: At least one useful auxiliary field is missing 3: At least one mandatory auxiliary is missing

Please note that for CTTH produced by neural network, several of the flags in ctth_status are never used. They are temporarily saved for consistency with the old CTTH algorithm.

4.2.8 Visualisation

The CTTH product is like the other PPS cloud and precipitation products first of all a digital product available in netCDF which should be used together with the appended flags, e.g. as input to an automatic mesoscale analysis or nowcasting scheme. A plain CTTH image showing just the temperature for example using a colour palette, without additional flags or quantitative numbers, is even more difficult to use than a plain cloud mask or cloud type image. Figure 1 shows such a plain image of the height in meters of a CTTH product (together with an RGB and the CTTH of PPS version 2014).

A dedicated PPS viewer developed at SMHI (Imager Viewer) may be rather useful in the case of the CTTH. Figure 4 shows an example with the Image Viewer. With the Imager Viewer it is possible to get the full information available in the netCDF file for the pixel under the mouse-pointer, as illustrated in Figure 4. Image Viewer is distributed from the same site as the PPS software.

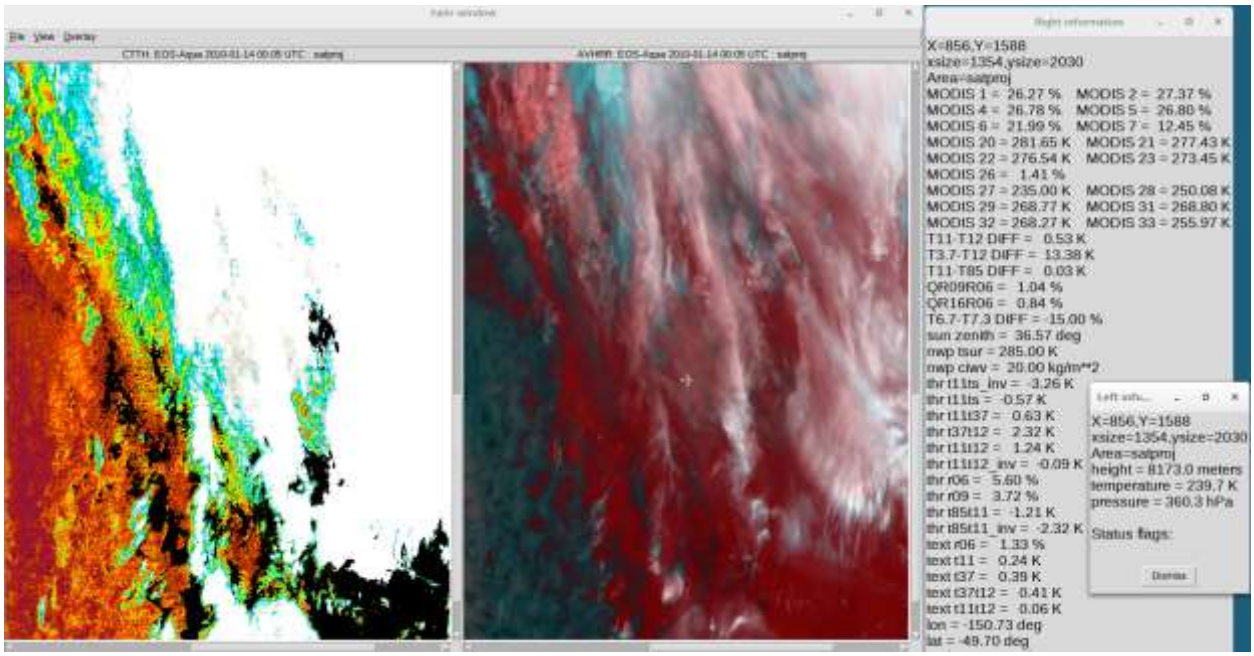

Figure 4 Example of CTTH image display using Image Viewer developed at SMHI. Scene: EOS-Aqua, January 14th ,2010, 00:05UTC. To the right an RGB image using channels 20, 31, and 32 (3.7, 11 and 12 µm) and to the left the corresponding CTTH. The two information dialogs, to the rightmost, provide information for the pixel under the mouse-pointer (marked in the right picture with a cross).

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

TBD/TBC	Section	Resp.	Comment
TBD01	4.1.2.4	SMHI	Neural network variables used for MODIS. The potential to use all MODIS $C0_2$ channels is not yet fully exploited. Network trained for MetImage gives already a good approximation, by using one $C0_2$ channel. MODIS $C0_2$ channels are currently not used at all due to problems for high satellite zenith angels. **Closed!**
TBC01	4.2.2 and 4.2.5	SMHI	Check algorithm performance when inversion is present. If performance turns out to be significantly inferior under those conditions set the low quality flag. Results checked performance is not significantly inferior. **Closed!**

ANNEX B. The multilayer perceptron

Artificial neural networks are widely used for non-linear regression problems. One type of neural network is the multilayer perceptron. It consists of three different types of layers, the input layer, the hidden layer and the output layer. The input layer is used to pass the inputs to the model. A multilayer perceptron can have more than one hidden layer.

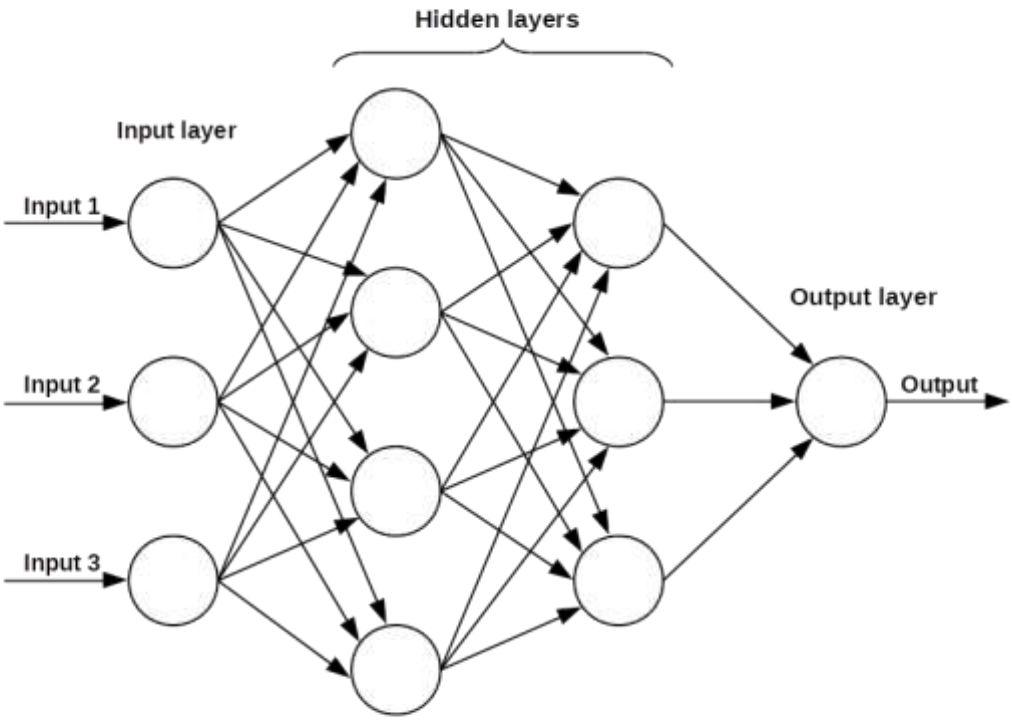

Figure 5 Architecture of a multilayer perceptron

In the different layers there are neurons, also called nodes. In Figure 5 you can see an example of architecture of a multilayer perceptron with two hidden layers. The input layer has as many neurons as there are inputs in the neural network. The number of hidden neurons in the hidden layer determines the complexity of the neural network, by increasing the number of hidden neurons the network becomes more complex. The output layer when using a neural network and retrieving one continuous variable has one neuron. The multilayer perceptron is a supervised learning technique since it learns by example. Weights and output signals connect the neurons in the network. The output signal is a function of the sum of all inputs to a neuron transformed by an activation function. The reason the multilayer perceptron can solve non-linear relationships is because of the non-linear activation functions used in the hidden layer/layers. The multilayer perceptron is part of the class feedforward neural networks since an output from a neuron is scaled by the connecting weight and fed forward as an input to the neurons in the next layer (see for example Gardner and Dorling, 1998).

If the output for an input when training the multilayer perceptron is not equal to the target output an error signal is propagated back in the network and the weights of the network are adjusted resulting in a reduced overall error. This algorithm is called the Backpropagation algorithm.

Formula for the output of the ith neuron of the (k+1) th layer of a multilayer perceptron can be defined as:

$$\mathrm{O_i(k+1) = \sum_{j=1}^{N_k} w_{ij}(k+1)\, a_i(k) + b_i(k+1),} \qquad 1$$

$$a_i(k+1) = \sigma\big(O_i(k+1)\big). \qquad 2$$

For neuron i at layer k $a_i(k)$ is the activation value, $O_i(k)$ is the net input and $b_i(k)$ is the bias. The $w_{ij}(k)$ is the weight value linked between neuron i at layer k and neuron j of layer k-1.Function σ in Equation 2 is the activation function.

The initialization of weights before training the network is important for the neural network to learn faster. Using bias in the neural network allows the activation function to shift which could help the neural network to learn.

To determine the changes in the weights an optimization method is used during the backpropagation algorithm. The optimization method used for the multilayer perceptron is mini-batch stochastic gradient descent which performs mini-batch training. A mini-batch is a sample of observations in the data. Several observations are used to update weight and biases, which is different from the traditional stochastic gradient descent where one observation at a time is used for the updates (Cotter et al., 2011). Having an optimal mini-batch size is important for the training of a neural network since a too large size can lead to the network taking a long time to converge.

During training of the network the mean squared error (MSE) is used as the loss function that is minimized during training. For the final version of v2018 MAE was used instead as this gave slightly better results. There are different learning parameters that need to be tuned for the neural network training to be effective during training. The learning rate is a parameter that determines the size of change in the weights; a too large learning rate will result in large weight changes which can result in the neural network error changing abruptly. If a learning rate on the other hand is too small the training time of the network will be long. The momentum is a parameter which adds a part of the weight change to the current weight change, using momentum can help avoid the network getting trapped in local minimas (Gardener and Dorling, 1998). A high value of momentum speeds up the training of the network. Choosing the number of hidden neurons and hidden layers for the neural network is also important for the training to be effective. Too few hidden neurons will result in underfitting.

There is also an option to use learning rate decay which is used to decrease the learning rate after each update as the training progresses which could improve convergence (Leen and Genevieve, 1993).

The learning rate decay function for decaying the learning rate is:

$$\varepsilon_{m+1} = \frac{\varepsilon_0}{1+\lambda m}. \qquad 3$$

where ε_0 is the initial learning rate, λ is the decay rate chosen and m is the number of iterations. One iteration is one update of the parameters of the neural network using the number of observations specified as batch size.

To avoid the neural network from overfitting a method called "early stopping" can be used. During early stopping the validation error is monitored to prevent the network from overfitting. In the beginning of training the network the training error and validation error generally decreases but after a certain number of epochs the validation error can start increasing while the training error keeps decreasing. At this point the training should be stopped since the network is starting to overfit the data. One epoch is a forward and backward pass of all the training data. Overfitting of the network results in the network not performing well on unseen data. If the validation error does not keep decreasing after a certain number of epochs one should therefore stop training the network. The validation error can fluctuate a bit during training, because of local minimas, a decrease in validation error can be followed by an increase in validation error. Therefore the number of consecutive epochs with no decrease in validation error before stopping the training of the network should be chosen carefully. The model saved after the neural network training has stopped is the model at the epoch where the validation loss is at its lowest.

B 1 CONFIGURATION OF THE MULTILAYER PERCEPTRON

B 1.1 Activation functions

The multilayer perceptron used has two hidden layers. The activation function used for the hidden layers is the activation function tangent hyperbolic (see Karlik and Olgac, 2011 for mote info) and for the output layer a linear activation function is used. When retrieving continuous targets such as cloud top pressure the linear activation function at the output layer is useful since it can output values in the range (-∞,∞).

The linear activation function has the form $f(z) = z$ and tangent hyperbolic has the form:

$$f(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}}. \qquad 4$$

B 1.2 Loss function, Number of epochs and batch size

During training of the network the mean squared error (equation 7) is used as the loss function that is tried to be minimized during training.

The number of epochs with no improvement before the training is stopped is chosen to be large since the validation error can fluctuate; training is stopped if there is no decrease in validation error for 300 consecutive epochs. The mini-batch size used for the neural network is 250.

The neural networks are trained for maximum 2650 epochs. The training is therefore stopped if early stopping occurs or if the network has trained for 2650 epochs. The 2650 was chosen as the maximum number of epochs since at that epoch a counter inside the training algorithm

is quite close to reaching the maximum integer that can be represented accurately when using float32.

B 1.3 Weight and bias initialization

There are many different weight initialization methods. One of the methods tried for initializing the neural network weights is initializing the weights by random sampling from the uniform distribution in the range (-0.05, 0.05). Another type of weight initialization is the glorot uniform weight initialization which is based on the idea that layers with more neurons should have smaller weights (see Bengio 2012 for more info). The weight initialization samples weights from the uniform distribution in the range (-range, range) where

$$range = \sqrt{\frac{6}{fanIn+fanOut}}. \qquad 5$$

The parameter $fanIn$ is the number of inputs to the neuron and $fanOut$ is the number of outputs to which any input contributed to. The biases of the neural network are initialized to 0.

B1.4 Data preprocessing

Missing values were removed from the data before training the network. Standardization of data in neural networks is important and makes the training of the network faster as well as helps prevent the neural network to get stuck in a local minimum and helps ensure convergence. The data is normalized by subtracting the mean and dividing with the standard deviation: $z = (x - \mu)/\sigma$.

B 1.5 Performance evaluation metrics

Some different metrics can be used to evaluate the performance of the network. The MSE is used as loss function. In the following formulas $\hat{y}_t$ is the predicted value and y_t is the observed value.

The mean absolute error (MAE):

$$MAE = \frac{\sum_{t=1}^{n} |\hat{y}_t - y_t|}{n}. \qquad 6$$

The mean squared error MSE:

$$MSE = \frac{\sum_{t=1}^{n} (\hat{y}_t - y_t)^2}{n}. \qquad 7$$

The root mean squared error RMSE:

$$RMSE = \sqrt{MSE}. \qquad 8$$

The bias:

$$bias = \frac{\sum_{t=1}^{n}(\hat{y}_t - y_t)}{n}. \qquad 9$$

The bias corrected root mean squared error can be calculated with bcRMSE = $\sqrt{RMSE^2 - bias^2}$.

B 2 TUNING NETWORK PARAMETERS

There are several existing rule of thumbs how to set the network parameters. The important thing is to check the MSE error to se how different parameter combinations behave. In Table 20 the results from different network configurations are shown. It is the MERSI-2 network that is considered throughout this chapter. In the rightmost column the validation mean squared error (MSE) for the standardized data can be seen. This is the MSE of the validation data at the epoch with the lowest validation MSE from the neural network training. First the different number of hidden neurons using one or two hidden layers was explored while keeping the learning rate, learning rate decay and momentum fixed. Then the number of neurons where fixed and the other parameters where varied. Last some networks with fewer neurons where tested to see how simple model we could chose and still have similar MSE.

From Table 20 one can clearly see that the neural network configurations with one hidden layer performed worse than the ones with two hidden layers. The model marked with * was chosen to continue with and tested with the glorot uniform weight initialization, see Table 21. We chose to use the glorot weight initialization as it was best in the small test (Table 21). In Table 21 we can see that also the random seeds have some impact on the MSE. The three models marked with circles were investigated to see what impact the differences in validation MSE would have on MAE of pressure for low, medium and high clouds. Notice that the difference is below 2hPa (Table 22). We chose to use the float32 data type to keep down memory usage, however we tested to also train some networks with float64 data types and they had similar MSE.

Table 20 Validation MSE when changing network parameters. Uniform weight initialization was used and the same random seed. Its is the MERSI-2 network. The variation between different random seeds is around 0.002 (see Table 21). The same random seed was used for all tests.

Hidden Neurons	Learning rate	Learning rate decay	Momentum	Validation MSE
10	0.001	1E-09	0.9	0.2035
15	0.001	1E-09	0.9	0.1957
20	0.001	1E-09	0.9	0.1915
25	0.001	1E-09	0.9	0.1869
30	0.001	1E-09	0.9	0.1837
35	0.001	1E-09	0.9	0.1843
40	0.001	1E-09	0.9	0.1824

Hidden Neurons	Learning rate	Learning rate decay	Momentum	Validation MSE
30-10	0.001	1E-09	0.9	0.1720
30-15	0.001	1E-09	0.9	0.1737°
30-20	0.001	1E-09	0.9	0.1706°
35-15	0.001	1E-09	0.9	0.1728
40-10	0.001	1E-09	0.9	0.1726
40-20	0.001	1E-09	0.9	0.1716
40-30	0.001	1E-09	0.9	0.1699
40-40	0.001	1E-09	0.9	0.1721°
30-15	0.01	0	0.9	0.1731
30-15	0.01	1E-06	0.9	0.1708 *
30-15	0.01	1E-09	0.9	0.1728
30-15	0.005	0	0.9	0.1722
30-15	0.005	1E-06	0.9	0.1719
30-15	0.005	1E-09	0.9	0.1722
30-15	0.001	0	0.9	0.1737
30-15	0.001	1E-06	0.9	0.1728
30-15	0.01	1E-06	0.8	0.1722
30-10	0.01	1E-06	0.9	0.1732
20-10	0.01	1E-06	0.9	0.1756

Table 21 Validation MSE when changing weight initialization and random seed. For the other parameters 30-15 hidden neurons where used, learning rate was 0.01 learning rate decay 1E-06 and momentum 0.9. We can see how much the MSE varies with random seed and that glorot uniform weight initialization is better for this very small test.

	Random Seed 1	Random Seed 2	Random Seed 3
Glorot uniform weight initialization	0.1717	0.1706	0.1702

Uniform weight initialization	0.1708	0.1719	0.1714

Table 22 Mean absolute error (MAE) in hPa for different height classes for three models (marked with circles) from Table 20.

	30-15	30-20	40-30
All	61.9	61.1	60.9
Low	51.3	51.0	50.7
Medium	59.6	58.5	58.4
High	85.5	83.8	83.9