Norwegian Meteorological Institute

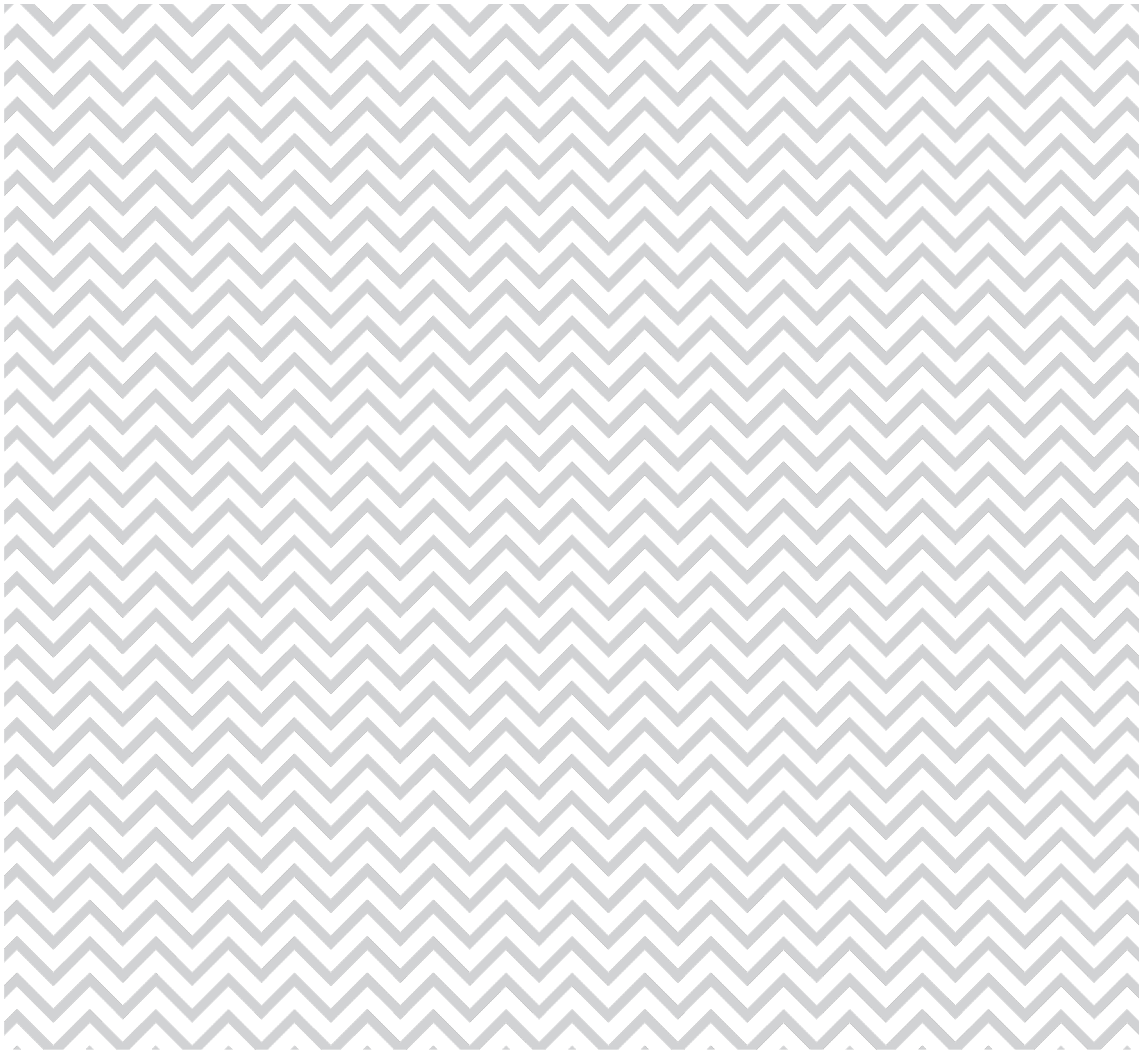
METreport

No. 12/2018
ISSN 2387-4201
Remote sensing

Improvement and Validation of the Optical Component Used for Snow Mapping in CryoClim

Sentinel4CryoClim Phase 2 Project Report

Mari Anne Killie, Steinar Eastwood, Atle M. Sørensen



Norwegian Meteorological Institute

METreport

Title Improvement and Validation of the Optical Component Used for Snow Mapping in CryoClim - Sentinel4CryoClim Phase 2 Project Report	**Date** February 12 2019
Section Remote Sensing	**Report no.** no. 12/2018
Author(s) M. A. Killie, S. Eastwood, A. M. Sørensen	**Classification** **X** Free O Restricted
Client(s)	**Client's reference**

Abstract

Sentinel4CryoClim Phase 2 is a follow-up to the CryoClim and Sentinel4CryoClim projects. The ultimate goal is to improve and advance the global snow cover time series developed in CryoClim and advanced in Sentinel4CryoClim Phase 1.

This report describes the work performed by the Norwegian Meteorological Institute (MET) in Sentinel4CryoClim Phase 2. The project is funded by ESA Prodex and is a cooperation with the Norwegian Computing Center (NR). NR delivers a separate report (see Solberg et al., 2018).

Keywords

Remote sensing, cryosphere

Disiplinary signature

Responsible signature

Meteorologisk institutt
Meteorological Institute
Org.no 971274042
post@met.no

Oslo
P.O. Box 43 Blindern
0313 Oslo, Norway
T. +47 22 96 30 00

Bergen
Allégaten 70
5007 Bergen, Norway
T. +47 55 23 66 00

Tromsø
P.O. Box 6314, Langnes
9293 Tromsø, Norway
T. +47 77 62 13 00

www.met.no

Table of contents

1 Introduction

In the CryoClim project, an operational service for long-term systematic climate monitoring of the cryosphere was developed. The service provides global sea ice and snow cover products, as well as Norwegian (mainland and Svalbard) glacier products. The focus of this follow-up project is to advance the current global snow mapping service in CryoClim. This includes improving the current algorithms, extending the time series, and adapting for improved input data series as well as for future sensors.

1.1 Status of the CryoClim snow monitoring service

The CryoClim project was running from 2008 to 2013, and included the partners the Norwegian Computing Center (NR), the Norwegian Meteorological Institute (MET), the Norwegian Water Resources and Energy Directorate (NVE) and the Norwegian Polar Institute (NPI). The aim of the project was to develop a web portal as well as cryosphere products. The portal would provide search, view and download functionality for cryospheric climate products, both indicator data as well as gridded data. The portal is available at `www.cryoclim.net`.

1.2 The structure of the project and of this report

The main objective of the Sentinel for global snow mapping in CryoClim (Sentinel4CryoClim) project is to advance the current global snow mapping service in CryoClim to be compatible with the Copernicus Climate Change Service objectives and position it as a candidate for snow monitoring in Copernicus.
The following sub-objectives have been set:
1. Mitigate weaknesses in the single-sensor components of the algorithm (optical and passive microwave radiometers) and multi-sensor multi-temporal data

fusion to further increase the accuracy and robustness of the product.
2. Extend the product with uncertainty estimates at the product and per-pixel levels.
3. Advance the algorithms and processing chains with the inclusion of Sentinel-3 OLCI and SLSTR data.
4. Perform more extensive validation of the product in space and time, including focus on inter-sensor issues in the time series.
5. Include the results in the CryoClim processing chain for snow and advance the operational level of the processing.
6. Position the CryoClim snow sub-service as a candidate for snow monitoring in Copernicus Climate Change Service.

In Phase 1 of the project, focus was on objectives 1, 2 and 6, as well as relevant parts of 4 and 5. For Phase 2, the focus is on objectives 3 and 6 plus relevant parts of 4 and 5. Phase 2 of the project is divided in the following work packages:

- WP 1. Management
- WP 2. Mitigation of retrieval algorithm weaknesses
- WP 3. Development of product uncertainty estimates
- WP 4. Integration of Sentinel-3 data
- WP 5. Validation
- WP 6. Processing environment
- WP 7. Exploitation

MET has been responsible for or involved in (some of) the tasks in work packages 2 – 6. METs work is described in this report.

2 Work package 2: Mitigation of retrieval algorithm weaknesses

The CryoClim daily multi-sensor multi-temporal snow cover product is based on an optical snow cover product and a passive microwave snow cover product, which are combined in a Hidden Markov Model (HMM). MET is responsible for the daily optical products, and NR is responsible for the fusion into the multi-sensor product. Development of the PMW component is a shared responsibility between MET and NR. All three data series are produced at MET. This chapter describes work related to improving the optical snow cover component.

2.1 A recap of the optical snow cover algorithm

The MET global optical snow cover chain processes all available swaths from the EUMETSAT Climate Monitoring Satellite Application Facility (CM SAF) AVHRR GAC data record. The calculations are based on a Bayesian approach using a set of *signatures* (instrument channel combinations) and statistical coefficients. For each pixel of the swath, the probabilities for the surface classes *snow*, *land* (snow-free ground) and *cloud* are estimated. The statistical coefficients are derived from pre-knowledge of the typical behaviour of the surface classes in the various parts of the spectrum. An example of an optical swath product can be seen in Figure 1 which shows a section of a NOAA-19 swath from March 15 2009, 08:50 UTC.

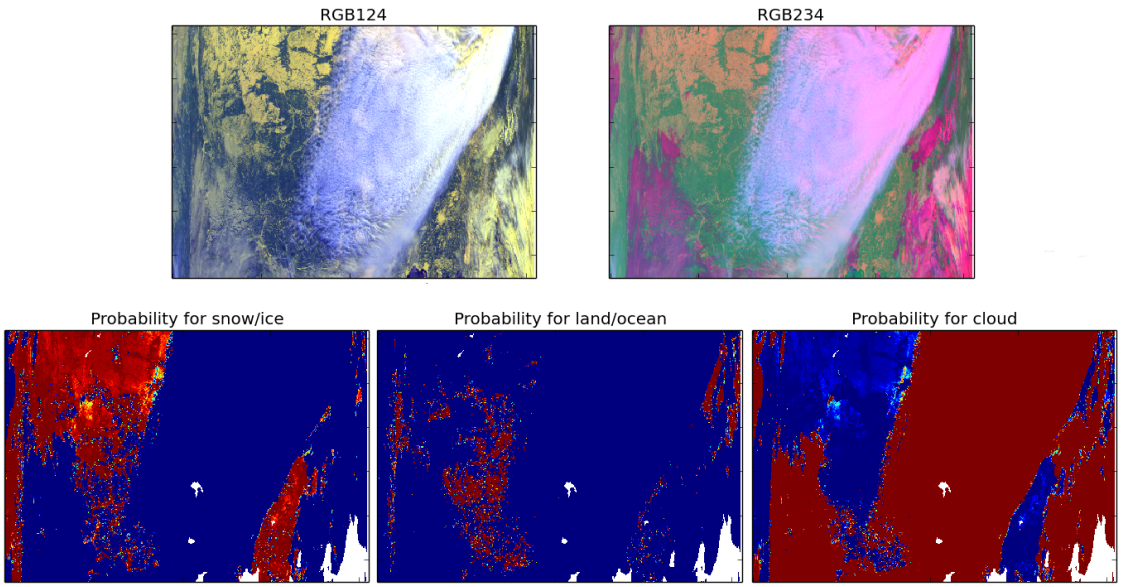


Figure 1: A section of a NOAA-19 swath from March 15 2009, 08:50 UTC. The top row shows RGB colour composites using AVHRR channels 1, 2 and 4 (left) and AVHRR channels 2, 3B and 4 (right). The bottom row shows the probabilities for the classes snow (left), land (middle) and cloud (right). Red colour indicates high probability, blue colour indicates low probability. White shows areas of no product (water).

The processed swaths are gridded and averaged. Swath product pixels with a probability for the class *cloud* larger than 40% are considered cloud-covered and are therefore not used in the averaged product. Averaging all available swath products from a 24-hour period gives daily, gridded, snow cover products containing probabilities for snow/no snow. A threshold is applied at 50% probability for snow, and a binary snow/no snow product is the result. The daily product has 5 km resolution. Figure 2 shows an example of a daily, gridded, optical product. See the snow sub-service report from CryoClim (Killie et al., 2013) for a more detailed description of the optical algorithm. As described further in Chapter 3, a first version uncertainty model was developed in Phase 1 of Sentinel 4 CryoClim (Solberg et al., 2017). In Phase 2 this is implemented for the global product.

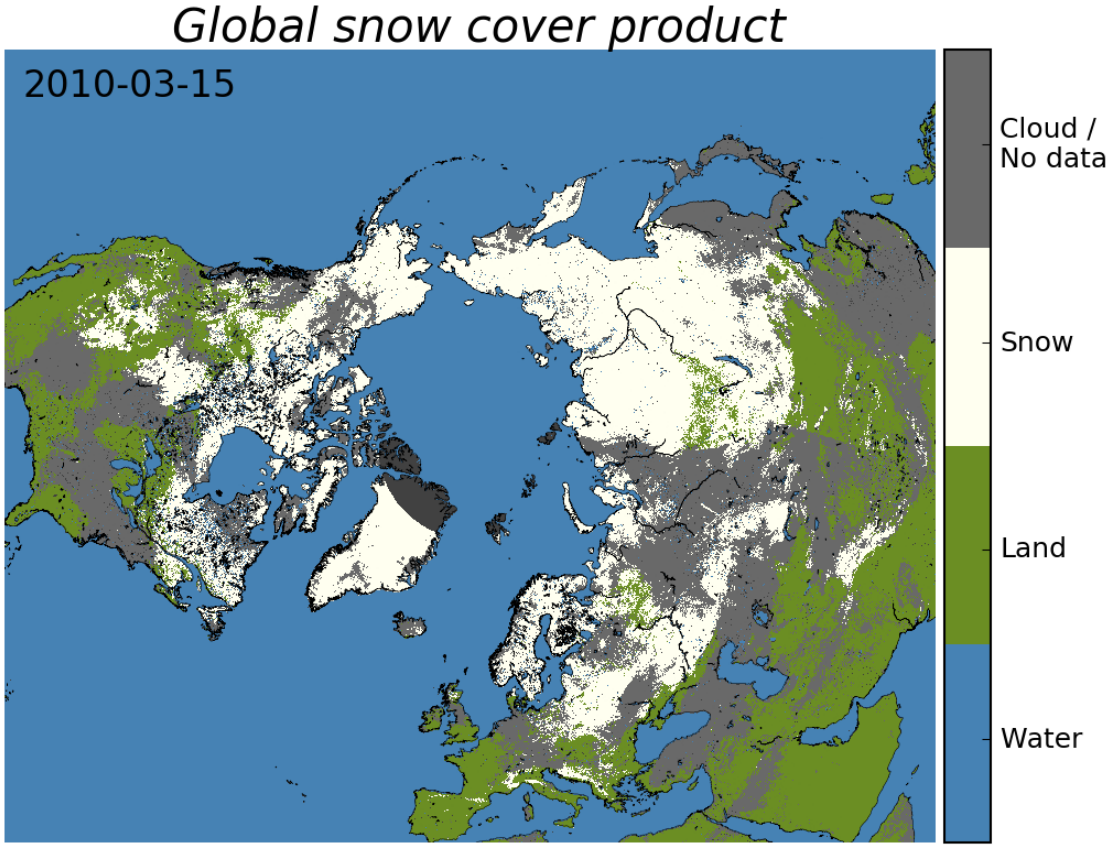


Figure 2: The figure shows the Northern hemisphere daily, optical snow cover product for March 15 2010.

An aggregation period of 24 hours is not sufficient to give a cloud-free product. Furthermore, the algorithm uses satellite measurements of reflected sunlight, and the products will therefore have relatively large areas at the poles that have no data during local winter-time. The optical product can therefore benefit greatly from being combined with a PMW product.

Overall, the optical algorithm performs well. Validation work carried out in the CryoClim project shows that when comparing daily, optical snow cover products with snow observations from ground stations spread on the Northern Hemisphere for the year 2005, there was overall agreement between the ground observation and the satellite-based optical snow cover product in 97% of the cases. In addition to the large-scale comparison against ground observations of snow depth, a local comparison between the optical snow cover product and 5 detailed snow maps for Jotunheimen was performed. The MET optical snow cover product correctly identified 94% of the pixels in these

snow maps. More details on this work can be found in the CryoClim snow sub-service documentation (Killie et al., 2013).

2.2 Adapting to a new version of the input data

The Advanced Very High Resolution Radiometer (AVHRR) has been operating since 1979 onboard operational meteorological satellites. The first version of the instrument measured in four spectral bands. From 1982 the 12.0 µm channel (Channel 5) was added (AVHRR/2), and in 1998 the 1.61 µm channel (Channel 3A) was added (AVHRR/3). Channel 3A is the most interesting channel for snow cover identification. The AVHRR/3 instrument can however only transmit at 5 channels at the time, and therefore switch between Channel 3A and the 3.74 µm channel (Channel 3B). Details on the different versions of the instrument can be found online[1].

The AVHRR instrument has ~1 km nadir resolution but only data at a reduced resolution (~4 km) is permanently archived and available with global coverage. This data is called Global Area Coverage (GAC) data. The CryoClim project used a fundamental climate data record (FCDR) for AVHRR GAC radiances and brightness temperatures covering 1982 - 2009, kindly provided by the EUMETSAT CM SAF. The CM SAF FCDR was produced as part of a project to derive climate data records for cloud, surface albedo and surface radiation budget products (CLARA-A1: CM SAF Cloud, Albedo And Surface Radiation dataset from AVHRR data, Karlson et al., 2013). The CM SAF has since made available a second edition (CLARA-A2) which covers the 34-year period from 1982 until 2015 (Karlson et al., 2017). Among the improvements in the updated version of the level 1c data record for radiances and brightness temperatures are improved geolocation, removal of corrupt data and no overlap between orbits. For each swath file (GAC orbit) calibrated reflectances and brightness temperatures, sun and satellite zenith and azimuth angles and scanline quality information are available. The output files are on HDF5 format and follow international Climate and Forecast Conventions. The Polar Platform System (PPS) cloud software [2] is used for the CLARA-A2 products. Detailed information can be found in the Algorithm Theoretical Basis Document (ATBD) and the Product User Manual, both accessible from `www.cmsaf.eu`. In addition to this data record the optical snow cover product in Sentinel4CryoClim uses historical model temperature data from ERA-Interim global reanalysis (Dee et al., 2011).

1 `https://www.wmo-sat.info/oscar/instruments/` is one suggestion.

2 The PPS cloud software is issued by the EUMETSAT Nowcasting Satellite Application Facility (NWC SAF).

The global optical data record is based on this AVHRR GAC data record from the period 1982 to 2015 , that is the entire AVHRR/2 era and AVHRR/3 up until 2015. The MET optical algorithm was developed using AVHRR/3 data only. AVHRR/2 has no Channel 3A, which is the preferred channel (over 3B) when available as it is particularly well suited for detection of snow. At 1.6 μm snow has very low reflectance while clouds still reflect well (Warren, 1982). This helps discrimination between clouds and snow. Figure 3 gives an overview of the AVHRR-carrying satellites contributing to CLARA-2.

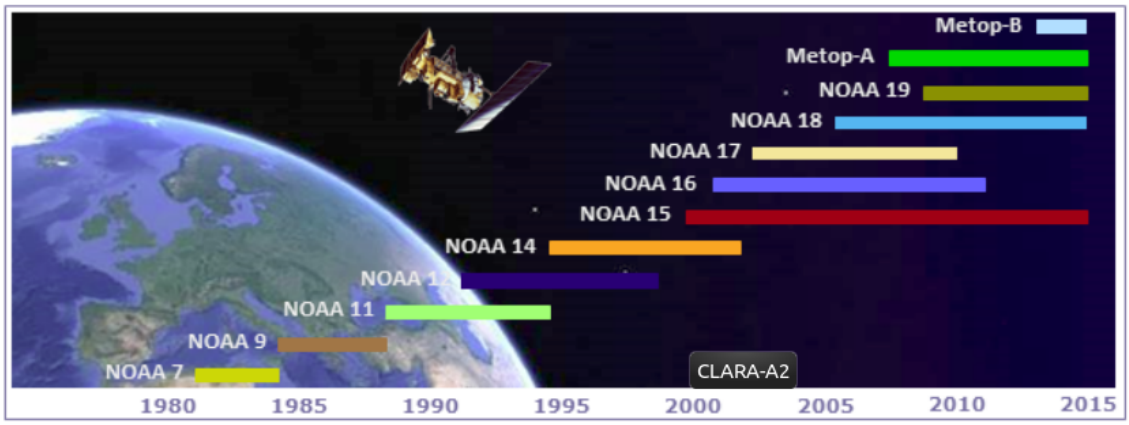


Figure 3: The AVHRR instruments used in CLARA-2. The figure is borrowed from www.cmsaf.eu. AVHRR/1 instruments are not included in CLARA-A2.

Figure 3 shows that for the first ~10 years of the dataseries there is only one satellite with AVHRR/2 in orbit at the time. The next ~10 years have two flying simultaneously (although there is a small but important gap in January 1996 with only one satellite in operation, see the discussion on false snow in Asia on page 24). From then on there are three or more instruments in orbit simultaneously. Multiple satellites available gives more swath data, meaning that the gridded, optical product is more likely to be cloud-free, and also that erroneous pixels from one swath product might be "corrected" by another swath product covering the same area.

2.3 Improving the optical snow cover component

In spite of an overall very good accuracy, the optical product has some known weaknesses. These include reduced performance at low solar elevations (due to shifts in the signatures) as well as in mountain areas during spring/summer (due to inaccuracies

in the model surface temperature for steep terrain). Earlier attempts have been made to replace the static coefficients with dynamic coefficients to achieve better performance at low solar elevations. This has so far not been a success due to insufficient amounts of training data available at the time. In the first version of the global snow cover product produced in Cryoclim optical satellite swath pixels with solar zenith angle up to 85 degrees were processed. This limit has now been reduced to 80 degrees in order to reduce errors at low solar elevation. In addition to the issue with inaccurate model surface temperature causing reduced classification in mountains, some time series issues were identified during Phase 1 of the project (conference Chapter 2.2 in Solberg et al., 2017). These will also have focus in the Phase 2 work to improve the optical snow cover component.

Inaccuracies in model surface temperature

Among the signatures/features used in the algorithm is “dT” - the difference between the model surface temperature and the 10.8 μm (AVHRR Channel 4) brightness temperature. The main role of this feature is to prevent pixels dominated by cold (icy) clouds from being wrongly classified as *snow*. The idea is that for cloud-free conditions the temperature difference between the brightness temperature observed at 10.8 μm and a model surface temperature should be small – *given that the model surface temperature is accurate and representative of the pixel area*. For cloudy pixel, the temperature difference is usually positive and often 10 degrees or larger. The statistical coefficients for this signature are chosen so that if the temperature difference is significant, the signature "dT" will influence the classification of the pixel towards the class *cloud*.

If the surface temperature from a numerical model deviates from the true surface temperature it can have consequences for our classification. The model surface temperature from ERA-Interim is on ~80 km resolution, far larger than the ~4 km resolution of the AVHRR GAC swath data. This makes it particularly difficult to achieve a representative model temperature where there are steep changes in the terrain. An artificially large difference between the surface model temperature and the 10.8 μm brightness temperature can force the class *cloud* to become the most probable class, and can thus make it difficult to identify snow-covered, cloud-free pixels. We have seen this effect earlier, in particular for springtime snow cover in mountain areas (the Alps, the Pyrenees etc.). A switch was therefore implemented in the code. The switch turns off the signature "dT" when the surface model temperatures is above 273 K. This does indeed allow more snow to be identified in mountains, but it also allows cold clouds to be classified as *snow*, as can be seen in Figure 4. The example shows a section of a NOAA-19 swath from March 15 2009, 08:50 UTC, with false colour composites,

probability plots and classified results. Central in the image is a large, cold cloud. The yellow circles have been added to lead focus to the large, cloud-covered area that has been wrongly classified as *snow*. Notice the square pattern in the transition from correctly classified *cloud* to the area of clouds wrongly classified as *snow*. Notice also that the warmer clouds (lower left corner and lower right side and corner) are indeed correctly classified as *cloud*. Warm clouds are separated from snow with the aid of the other participating signatures. The circled area shows pixels for which the signature "dT" has been switched off, ending in erroneous classification.

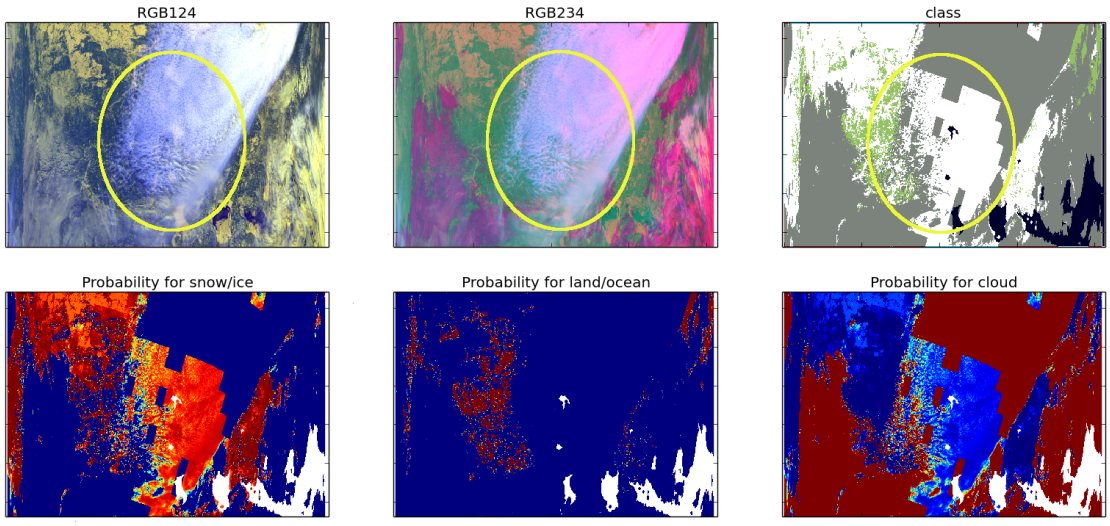


Figure 4: A section of a NOAA-19 swath from March 15 2009, 08:50 UTC. The top row shows RGB colour composites of AVHRR channels 1, 2 and 4 (left) and 2, 3B and 4 (middle), and the classified result (right). Here, grey refers to clouded pixels, snow-free land is shown in green, and white shows pixels classified as snow. The bottom row shows the probabilities for the classes snow (left), land (centre) and cloud (right). For the probability plots, red refers to high probability, while blue corresponds to low probability. In this image north is down and south is up. The white areas (water) in the lower, right corner show the Baltic Sea, Lake Ladoga and Lake Onega.

Figure 5 shows the same plots as Figure 4, except that now the signature "dT" is allowed to contribute regardless of model surface temperature (i.e., the switch has been switched off). The large clouded area encircled in Figure 4 is now correctly classified as cloud (grey area in upper, right panel of Figure 5).

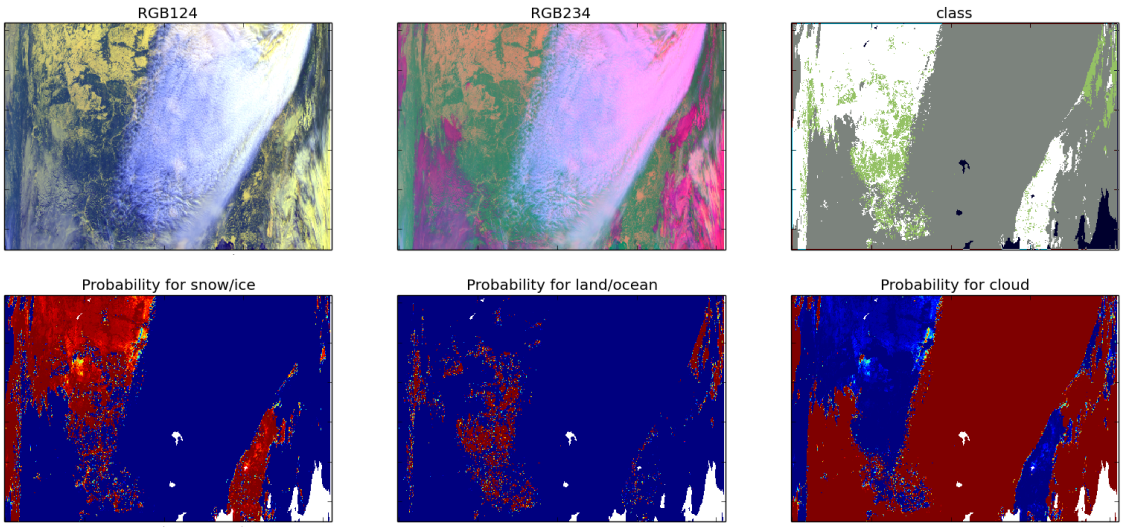


Figure 5: The figure shows a part of a NOAA-19 swath from March 15 2009, 08:50 UTC, and is organized identically as Figure 4. The difference between Figures 4 and 5 is that for Figure 5 the signature "dT" is fully included for all pixels.

Due to misclassification of pixels of the type seen in Figure 4 we decide to remove the switch. The signature "dT" is included for all temperatures. Instead we must seek other ways of solving the problem for which the switch was originally included. The following sub-chapters describe the efforts made to reduce the difference in temperature between the model surface temperature and the actual surface temperature. Ideally this will improve the identification of snow in steep terrain.

Interpolation of model surface temperature

ERA-Interim global reanalysis is the source of the model surface temperature. Global fields on ~80 km resolution is downloaded for the entire time series. In previous versions of the snow product (Cryoclim and Sentinel4CryoClim Phase 1) the model temperature is interpolated to each satellite swath pixel using fimex[3] with the default interpolation method. The default method is nearest neighbour, and this is the reason for the square pattern seen in Figure 4. The interpolation method can be replaced with a more advanced scheme. Figure 6 shows an example of interpolation of model surface temperature into a satellite swath. The left panel shows the model surface temperature

3 Fimex is a File Interpolation, Manipulation and EXtraction library for gridded geospatial data, developed at the Norwegian Meteorological Institute.

when nearest neighbour was used. The right panel shows the model surface temperature when using bilinear interpolation.

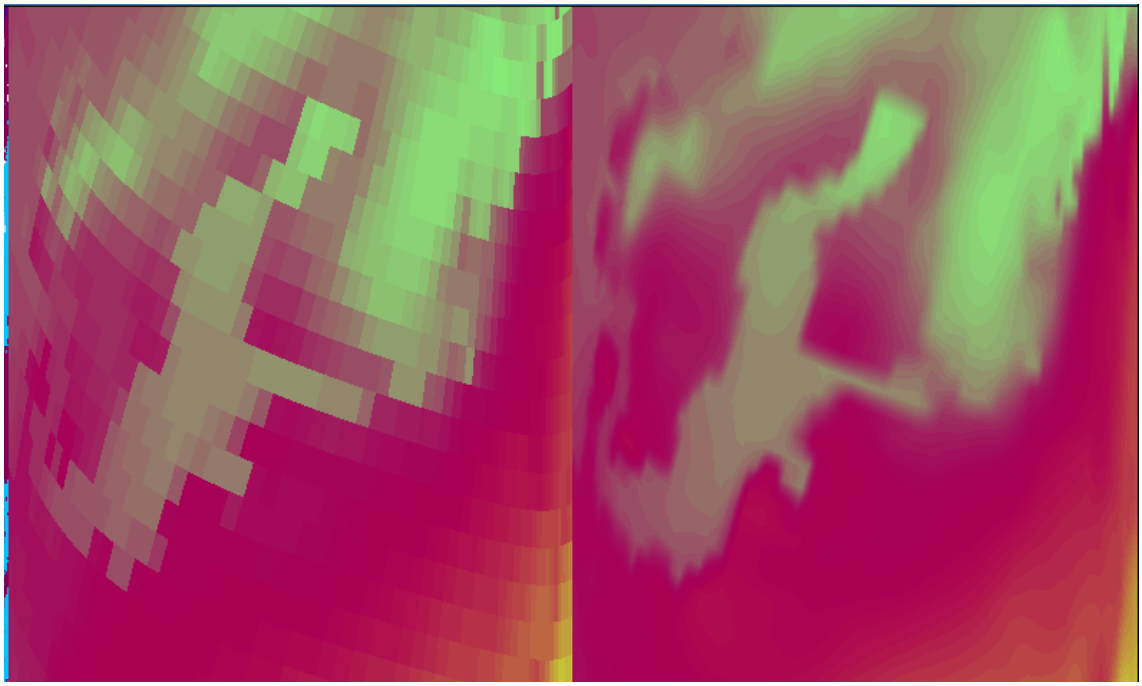


Figure 6: Examples of interpolated model surface temperature. In the left panel nearest neighbour has been used. The right panel shows the result using bilinear interpolation.

Should we decide to reintroduce the switch for the signature "dT", in combination with bilinear interpolation, the square pattern seen in Figure 4 would be replaced with a smoother transition. As long as information on pixel elevation is not included in the interpolation, the bilinear interpolation will not necessarily compensate adequately for temperature changes due to steep terrain.

Height correction of model surface temperature

In the new version CLARA-A2 data record swath files containing auxiliary data such as interpolated model surface temperature were collected. The notation "T_pps" will be used to specify when the model surface temperature is collected from this set of swath files. ERA-Interim model surface temperature is the original source for the T_pps files as well, and no height correction has been applied. The files were downloaded and tested. Figure 7 shows an example of result when T_pps is used as source for model

surface temperature. The upper row shows RGB colour composites from a part of a NOAA-9 swath from April 14 1988, 13:49 UTC. The image is upside down, and the lower left corner shows the Alps, while the Pyrenees are seen in the upper part of the image. The RGB composites show that the Alps are snow-covered, with a cloud cover over the western parts of the mountains. The Pyrenees are also snow-covered with clouds covering the western parts. In addition there are thin clouds covering central parts of the Pyrenees. The yellow circles have been added to highlight snow-covered mountain areas that are mostly cloud-free but not correctly identified. The misclassification is mostly due to a non-negligible difference between the model surface temperature and the true surface temperature.

Height correction of the model surface temperature was attempted to reduce differences between the model surface temperature and the true surface temperature in areas of steep terrain. For this purpose the swath pixel elevation is needed. This is available from the physiography swath files that are a part of the CLARA-A2 level 1c dataset. The elevation for which the surface model temperature is valid is also needed. The surface geopotential field is collected from ERA-Interim and interpolated to the satellite swath. Note that the interpolation method used for T_pps is unknown to us. We use bilinear interpolation for the geopotential field. Once a value for the difference in height is found, we can assume an adiabatic lapse rate and correct the model surface temperature. For dry adiabatic processes the temperature is reduced with 1 K per 100 meters increase in height. If the moisture in the air cools to saturation condensation occurs, which leads to a decreased cooling rate compared to the dry adiabatic case. An average value of 0.7 K reduction in temperature per 100 meters increase in elevation is used. The lower row of Figure 7 shows the result when height correction is applied. The left panel shows the probability for the class *snow*, and the right panel shows the classified product. For this example a larger fraction of the snow-covered pixels in the Alps is classified as *snow* when height correction of the model surface temperature is used. Parts of the Pyrenees are also identified as snow-covered.

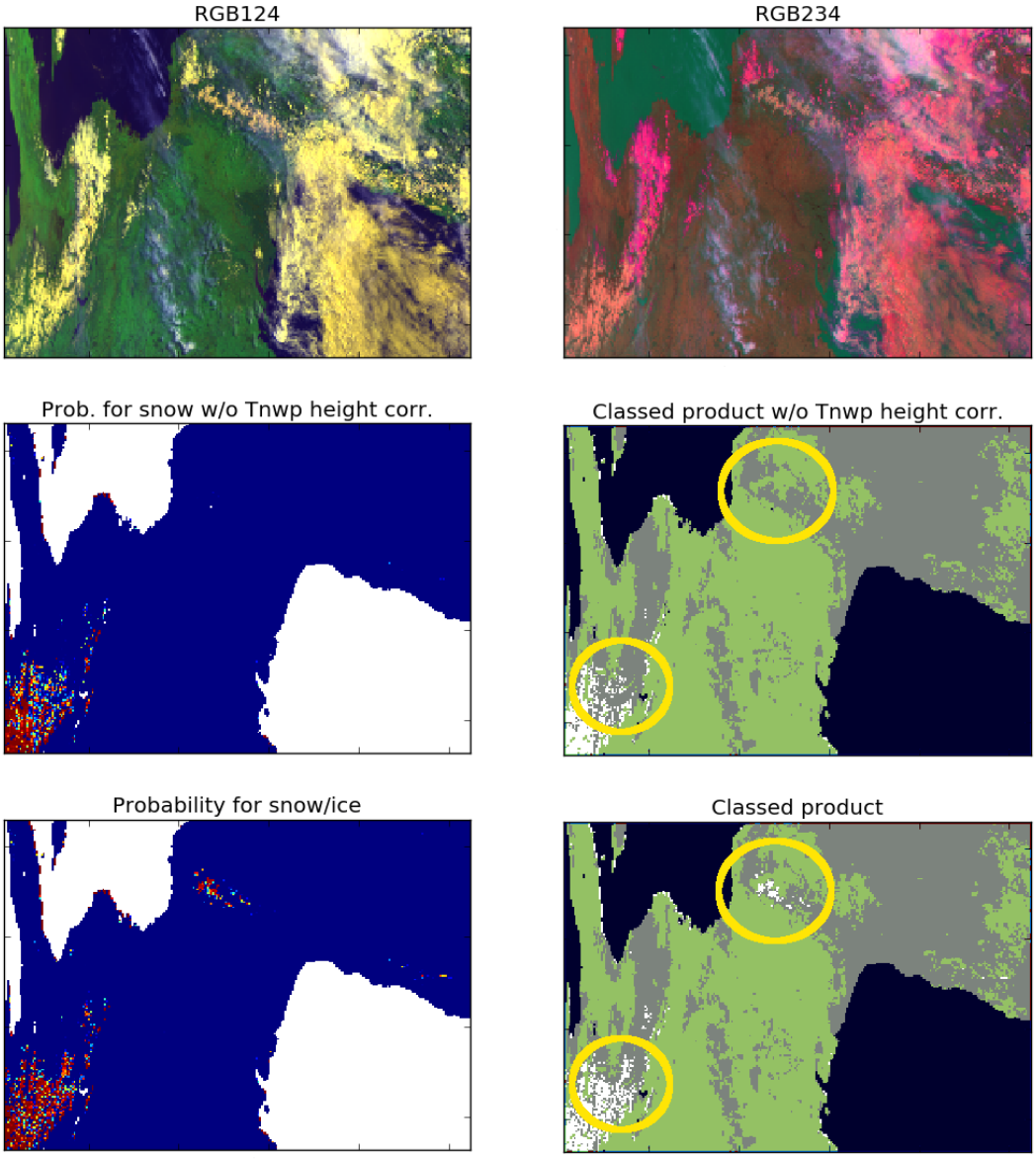


Figure 7: The figure shows a part of a NOAA-9 swath from April 14 1988, 13:49 UTC. The upper row shows RGB colour composites using AVHRR channels 1, 2 and 4 (left) and 2, 3B and 4 (right). The middle row shows the probability for snow (left panel, red colour indicates high probability and blue colour indicates low probability) and the classified product (right panel, snow is shown in white, snow-free land in green, and clouds in grey). T_pps is the source for model surface temperature. The lower row shows the same as the middle row, except that T_pps has been height corrected. The yellow circles have been added to highlight areas showing improvement.

Since the interpolation method used for T_pps is unknown to us, artefacts easily appear. An example can be seen in Figure 8 which shows a part of a NOAA-12 swath from January 16 1995, 01:04 UTC. The upper row shows RGB colour composites of the Himalayas. The solar zenith angle is very high. The lower row shows the original model surface temperature T_pps (left) and the height corrected T_pps (right). It is obvious from the lower, right panel that the geopotential height and the model surface temperature must be interpolated using the same method to avoid artefacts that can otherwise propagate into the classified product. Unfortunately there was no more time to work on this task as it was necessary to start the time consuming task of data processing. To avoid artefacts, height correction of T_pps had to be abandoned for now, even though examples showed increased identification of springtime snow cover in mountains.

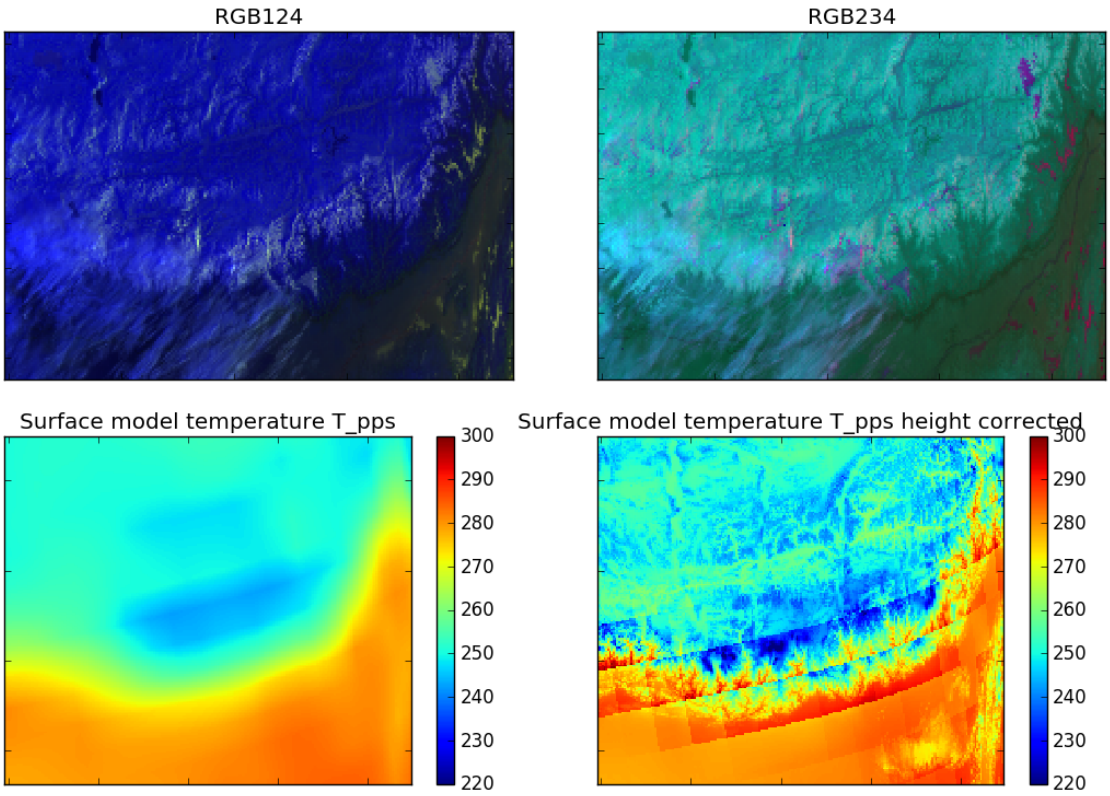


Figure 8: The figure shows a part of a NOAA-12 swath from January 16 1995, 01:04 UTC. The upper row shows colour composites using AVHRR channels 1, 2 and 4 (left) and 2, 3B and 4 (right). The bottom row shows the T_pps model surface temperature (left) and the height corrected T_pps model surface temperature (right).

Some swaths where processed using first T_pps as source for model surface temperature and then model surface temperature from ERA-Interim (bilinear interpolation). The swath products were compared. Figure 9, showing a part of a NOAA-11 swath from January 2 1990 at UTC 10:49, gives an example of result. The figure covers the Alps (upside down) which are cloud-free and snow-covered. The middle row shows the swath product when using model surface temperature from ERA-Interim (bilinear interpolation), and the bottom row shows the swath product when using model surface temperature from CLARA-A2. The yellow circles have been added to highlight an area of relatively large difference. More snowy pixels are correctly identified when model surface temperature from ERA-Interim (bilinear interpolation) is used, which is why T_pps was abandoned as well, and we returned to ERA-Interim with bilinear interpolation as the source for model surface temperature needed by the troublesome signature "dT".

It is clear that this area needs more work. For future work we will attempt to interpolate ERA-Interim data in 3 dimensions, or replace the signature "dT" altogether with something that does not rely on model temperature data.

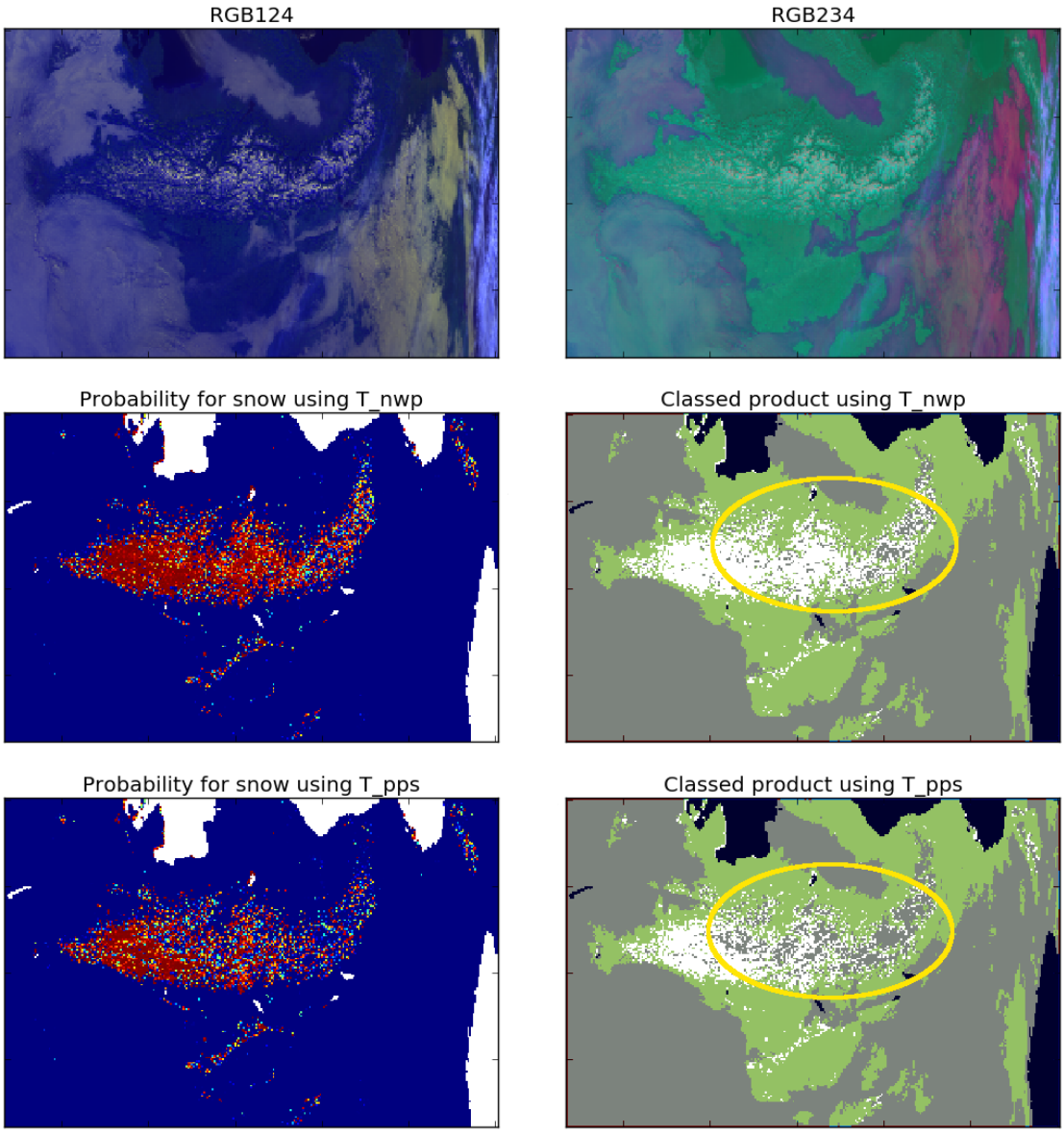


Figure 9: The figure shows a section of a NOAA-11 swath from January 2 1990, UTC 10:49. The top row shows RGB colour composites of AVHRR channels 1, 2 and 4 (left) and 2, 3B and 4 (right). The snow-covered area in the middle of the image is the Alps. The middle row shows the swath product result when using model surface temperature from ERA-Interim with bilinear interpolation (T_nwp). The lower row shows the swath product result when using the model surface temperature included in the CLARA-A2 level 1c dataset (T_pps).

Coastline inaccuracies

The Bayesian snow cover algorithm can be used over land and over water. When using the algorithm over water, the statistical coefficients for the classes *cloud*, *water* and *sea ice* are used, and the probabilities for these classes are estimated. Over land the probabilities for *cloud*, *land* and *snow* are estimated using the corresponding coefficients. It is essential to know whether the pixel in question is dominated by land or by water. In this project we are interested in climate series for terrestrial snow, and thus do not process ocean pixels. Pixels along the coast line can be a challenge. If an open water pixel is attempted processed with coefficients for the classes *land*, *snow* and *cloud*, it is likely to be classified as *snow/ice*. This happens because the true class (here *water*) is not available, and the algorithm therefore points towards the class that matches the most of the ones available (here *snow*). In Phase 1 of Sentinel4CryoClim the pixel land/water information was read from the CLARA-A2 level 1c cloud mask files. The information was in the form of binary values for land, coast, and ocean, and coast pixels were considered land. Figure 10 (left panel) shows a typical result. A narrow belt along the coast has been wrongfully classified as *snow*. During Phase 2 we downloaded auxiliary swath files containing fraction of land. When these physiography files are available the coast line can be more precisely located, and fewer pixels along the coast are misclassified (see right panel of Figure 10). The fraction of land is given as data type short with values from 0 to 255. Pixels with values above 100 are considered land and processed. Pixels with values equal or below are not processed.

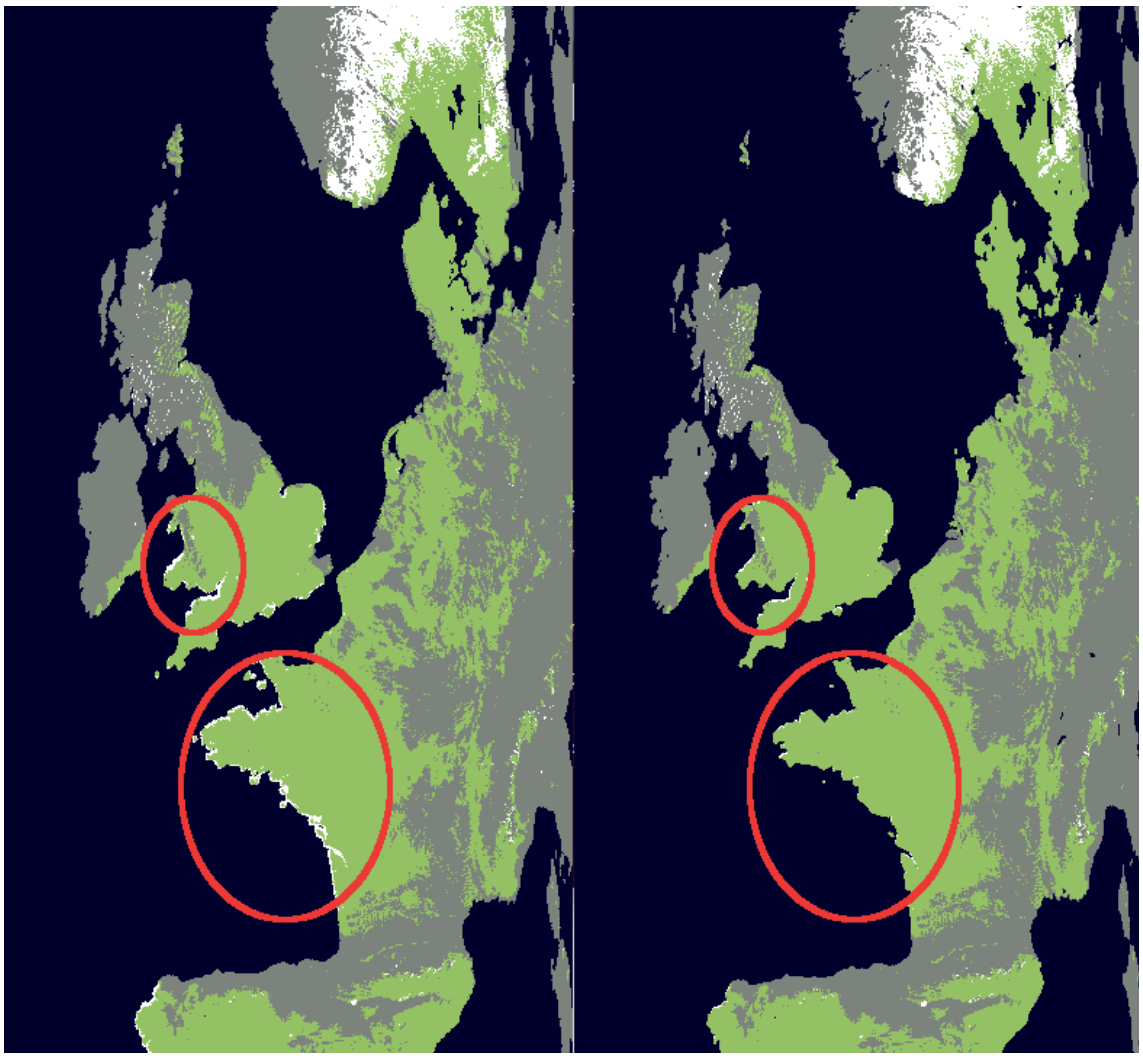


Figure 10: The figure shows a part of a swath product for NOAA-15 on March 15 2000, UTC 07:50. The left panel shows the classified product when land/ocean information is collected from cloud mask files (Phase 1 result). The right panel shows the classified product when fraction of land is used to decide whether the pixel is over land or ocean. The red circles have been added to draw attention to areas of typical improvement. Pixels assigned to the class snow are shown in white, land-covered pixels are shown in green, and clouded areas are shown in grey.

False snow in Asia

The first version of the optical snow cover products based on the CLARA-A2 dataset - produced during Phase 1 of Sentinel4CryoClim - revealed some issues, in particular for the earlier years of the time series (see Chapter 2.2.5 of the Phase 1 report: Solberg et al., 2017). Among the issues reported are cases of false snow cover in Sahara, particularly found during the first ~10 years of the data series, and there are also cases of false snow cover in South Asia.

The cases of false snow in South Asia are most frequent during the period for which NOAA-12 was the only satellite operating. NOAA-12 covers the period September 16 1991 to December 14 1998. From September 14 1994 to January 19 1995 NOAA-12 is the only satellite operating, which means that errors introduced by NOAA-12 swath products cannot be "corrected" by data from the AVHRR instrument on other platforms. NOAA-14 data enters from January 20 1995. Figure 11 shows a part of the optical daily product from January 20 1995. On this day NOAA-12 and NOAA-14 data exist. The left panel shows a part of the original, gridded product. The yellow circle is added to draw attention to snow-free areas misclassified as snow-covered in India and Bangladesh.

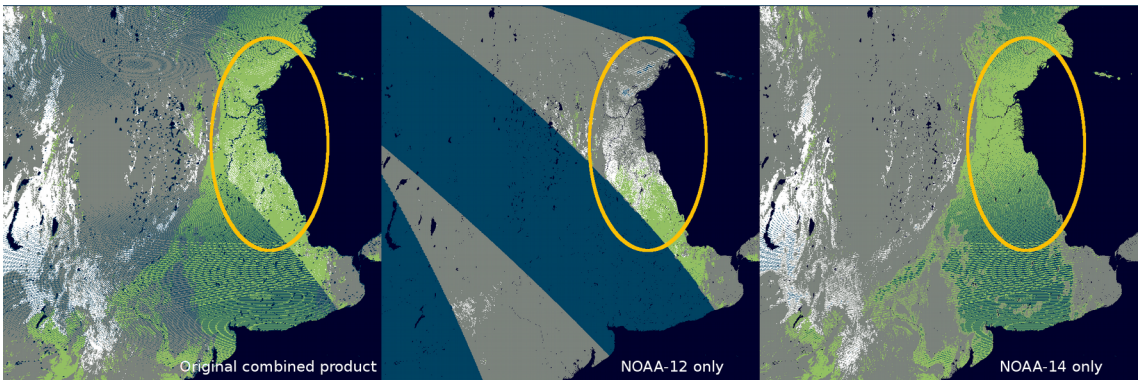


Figure 11: A part of the optical, daily product from January 20 1995. The left panel is the original gridded product, and is based on swath products from NOAA-12 and NOAA-14. The middle panel shows a gridded product based on NOAA-12 swath products only, and the right panels shows a gridded product based on NOAA-14 swath products only. For all panels snow is shown in white, snow-free land in green, and clouds in grey. The dark blue stripes in the middle panel show areas of no satellite cover.

The middle panel shows the result when gridding NOAA-12 only. The yellow circle highlights an area with cloud cover as well as (false) snow. When NOAA-14 only is gridded we get the result shown in the right panel. The (false) snow in India and

Bangladesh has disappeared. The middle and right panels show a large difference in satellite coverage. Figure 12, borrowed from the CLARA-A2 ATBD for cloud products (Karlson et al., 2016), shows the satellite equator passing times as function of time. During the period where NOAA-12 operates "alone" it has an equatorial passing time that is relatively early in the morning. This coincides with Northern hemisphere winter, meaning that the solar elevation is at it's lowest . For low solar elevations, the static coefficients can lead to reduced algorithm performance, and misclassifications of the type seen in the left and middle panel of Figure 11 typically appear.

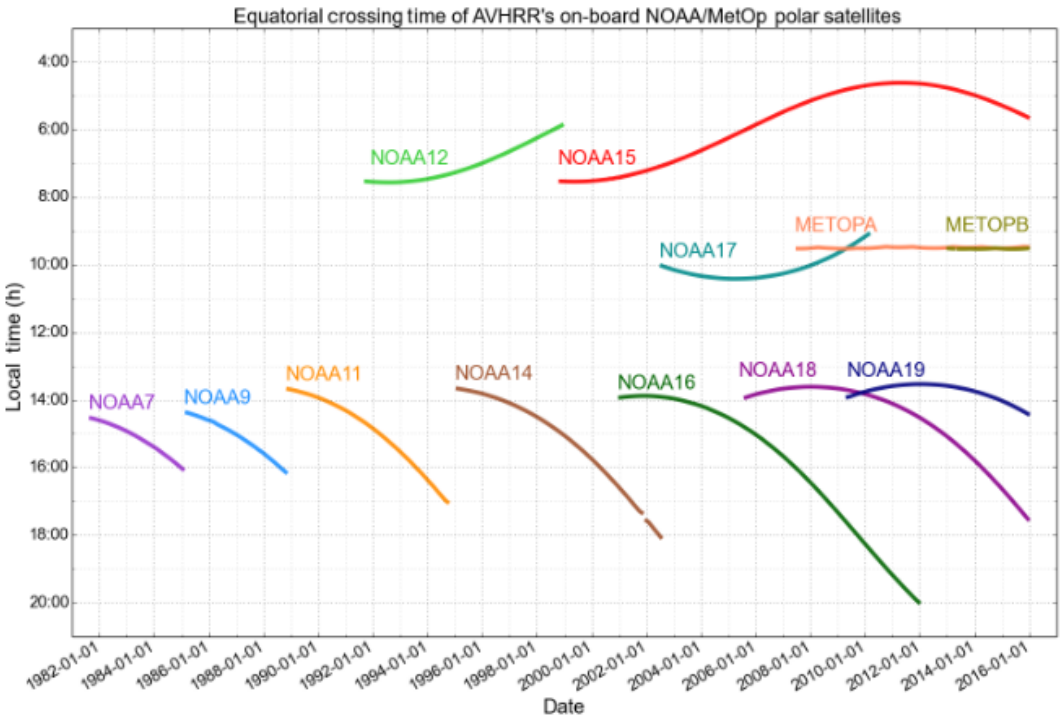


Figure 12: The figure is borrowed from Karlson et al., 2016 and shows equator passing times for the platforms contributing to the CLARA-A2 dataset.

To remove the false snow in Asia frequently occurring for NOAA-12 due to low solar elevation, we implement a test that masks all swath pixels that are classified as snow and have a solar zenith angle above 70 degrees. This will inevitably also mask some correctly classified snow pixels, but this is still considered the best solution for now since the ultimate goal is to improve the multi-sensor product. Figure 13 compares the previous (left panel) and new (right version) of the daily, optical product. Both gridded products are based on swath products from NOAA-12 and NOAA-14. The right panel

shows the result when the limitations on NOAA-12 have been implemented. All the false snow in south Asia (India, Bangladesh) has been removed.

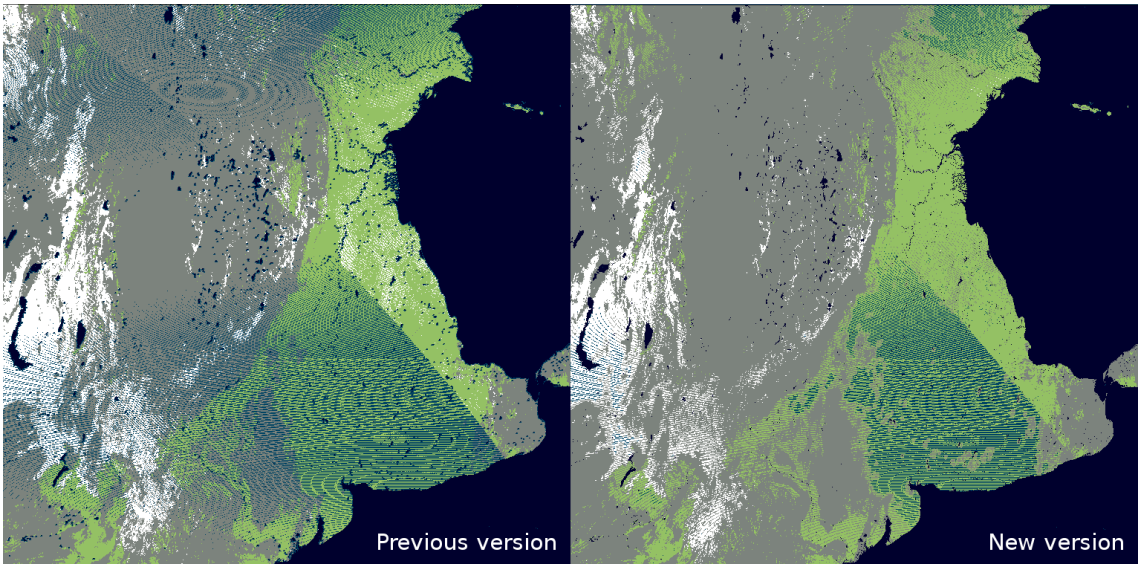


Figure 13: This figure shows a part of the optical, daily product from January 20 1995 before (left panel) and after (right panel) masking of snow pixels from NOAA-12 at low solar elevation has been implemented. Snow is shown in white, land in green, and clouds in grey.

False snow in Sahara

Results from Sentinel4CryoClim Phase 1 showed multiple cases of false snow cover in Sahara during the early part of the optical time series. The reason seems to be a combination of several factors: The instruments are not identical from platform to platform (the response functions differ). The AVHRR instrument onboard NOAA-7 to NOAA-14 is of type AVHRR/2 and has no Channel 3A. The set of statistical coefficients is based on NOAA-17 and NOAA-18 data mainly (AVHRR/3), but is used for all platforms. Figure 14 shows a typical example of false snow in Sahara (left panel).

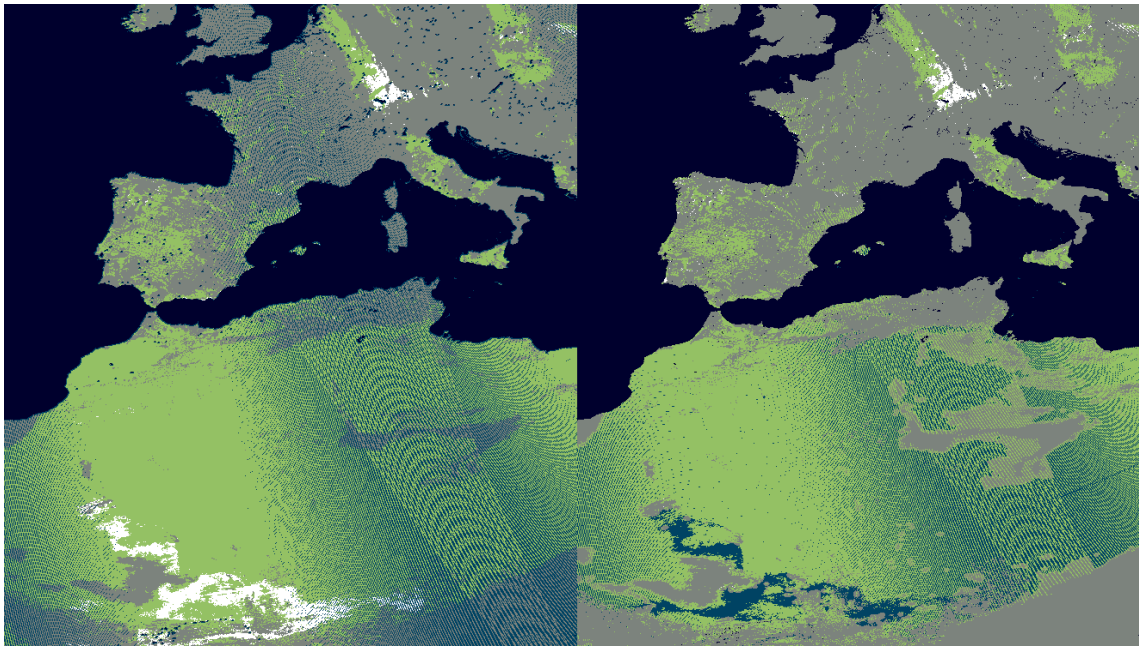


Figure 14: Example of result before (left panel) and after (right panel) a temperature filter was added to prevent false snow from appearing in Sahara. The example is taken from the daily optical product from March 22 1985. Snow cover is shown in white, snow-free land in green and clouds in grey. Areas of no data are shown in blue.

It is not straightforward to develop and implement a proper solution to this problem. We chose to implement a temperature filter that masks all pixels classified as *snow* occurring at surface model temperatures above 295K. The limit must be carefully selected. If the threshold is set too high, erroneous snow is not removed. Should it be set too low, true snow is in the risk of being masked. The right panel of Figure 14 shows an example of result when applying this filter.

3 Work package 3: Development of product uncertainty estimates

In Phase 1 of Sentinel4CryoClim a first version uncertainty model was developed for the optical component of the snow cover product. We followed an approach used in EUMETSAT OSI SAF for sea surface temperature (SST): an extensive match-up dataset with collocated satellite swath snow products and ground truth snow measurements was collected. The data was analysed, and verification measures[4] (hit rate, false alarm ratio etc.) were computed as function of various variables (time of year, solar elevation, class probability, snow depth and more). The verification measures showed variation with more or less all of the variables tested. Even though there is an obvious variation with time of day (solar elevation), the variation in hit rate with time of year is the most dominating. It was therefore reasonable to introduce a first version uncertainty model that was a function of time of year. Variation with solar elevation and other variables will be revisited in future work.

The first version uncertainty model is based on the monthly values for hit rate for each class (*snow* and *land*). For each pixel of the swath product, accuracy values for the most probable class are collected from a table with monthly values, and interpolating linearly between the two closest months. This is then added to the swath product files. More details can be found in the Phase 1 project report (Solberg et al., 2017). The work in Phase 1 was performed in the context of the regional snow cover chain at MET. In Phase 2 the uncertainty model has been implemented in the global snow cover processing chain. Figure 15 shows an example of the result. The left panel shows a section of the March 15 2005 daily optical product, and the right panel shows the corresponding uncertainty estimate. Snow-covered pixels get one value, while snow-free pixels are assigned another value. Pixels with no surface information (clouded, ocean, no data) have no uncertainty value.

4 Conference Chapter 5.2 for more on verification measures

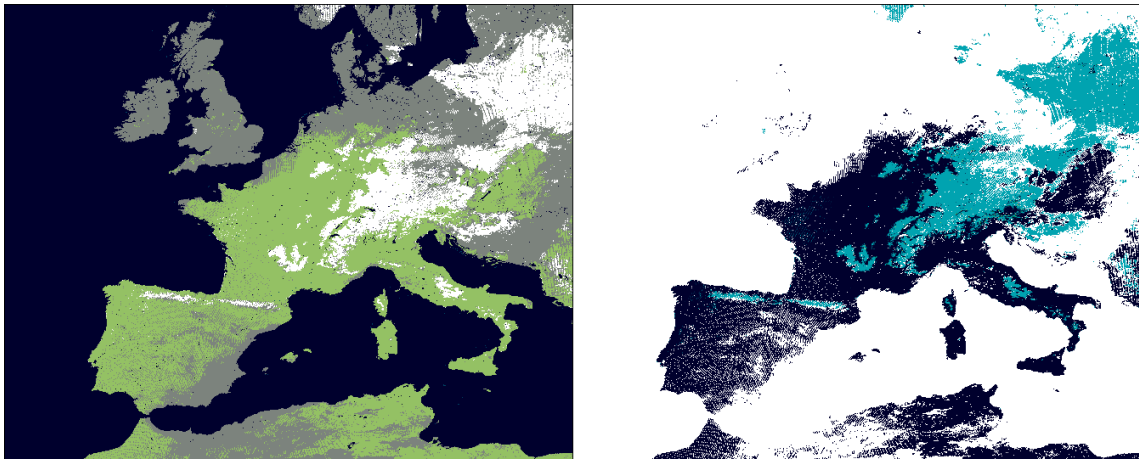


Figure 15: Example of uncertainty estimate result. The left panel shows the classified product for March 15 2005 (snow is shown in white, clouds are shown in grey, land pixels in green, and water is black), while the right panel illustrates the corresponding uncertainty estimate. Light blue illustrates the uncertainty value for the class snow (here 90.11%), and dark blue illustrates the uncertainty value for the class land (here 95.79%). White areas are areas with no uncertainty value (cloud or ocean).

Uncertainty estimates are now available on the daily optical snow cover products for the full time series 1982 - 2015, and the uncertainty estimates from all daily products for the year 2005 have been validated. The validation is performed as follows: snow depth observation data from the GHCN-D dataset is used as ground truth (see Chapter 5.1 for a description of GHCN-D). For each ground observation the corresponding product pixel cell is located. The product class (*snow*, *land*, *ocean*, *cloud*, *nodata*) as well as the probability for correct classification are collected. For 2005 there are ~870 000 Northern hemisphere observations of snow depth in the GHCN-D dataset. When observations that are located in satellite product grid cells with no value are removed (either cloudy, mostly water, immersed in wintertime darkness or for other reasons), ~430 000 samples remain. These are binned in 20 bins from 0 till 1.0 according to their uncertainty estimate value. The total hit rate[5] is calculated for each bin and plotted in Figure 16 (black line). For 97% of the data the probability for correct classification (hit rate) has a value of 0.8 or higher. The bars plotted in Figure 16 show the number of samples in each bin. The black line is dashed for the range where there is very little data (below 0.55 probability for correct classification) and therefore is questionable

5 The total hit rate is the ratio between the number of correct classifications and the total number of comparisons for that bin.

whether hit rate values should be calculated. The red line shows an ideal result. Pixels with low probability for correct classification should seldom be correctly classified (y-axis), while pixels with high probability for correct classification should often be correctly classified. The result seen in figure 16 does indeed show an increase in calculated probability for correct classification with increasing probability for correct classification as collected from the product. The figure also shows that for pixels with uncertainty in the range ~0.6 - 0.9 the uncertainty estimate overestimates the observed mismatch.

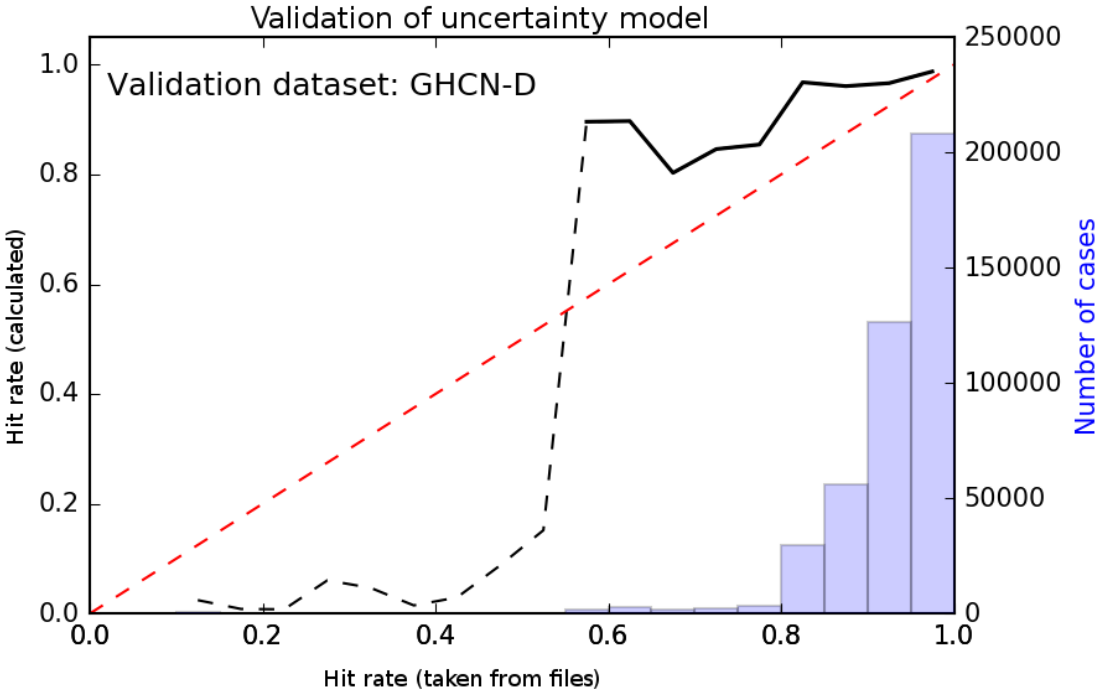


Figure 16: Comparison of product uncertainty estimate with actual (calculated) uncertainty. The black line shows the validation result. The red dashed line shows the theoretical result. The blue bars show the number of comparisons made for each bin. This is based on validation of all daily products for the year 2005.

4 Work package 4: Integration of Sentinel-3 data

The AVHRR instrument is a fundamental data source for the CryoClim snow cover time series. AVHRR has been active since the early 80-ies and is currently flying on NOAA-15, NOAA-18, NOAA-19, MetOp-A and MetOp-B. However, no more AVHRRs are planned for launch. To be able to continue the global snow cover monitoring and prolong the climate data set, an alternative source for optical data is needed.

The Visible Infrared Imaging Radiometer Suite (VIIRS) instruments has been flying since 2011 onboard Suomi NPP. A second VIIRS was recently launched on NOAA-20. NOAA-20 is the first of 4 planned polar-orbiting satellites of the Joint Polar Satellite System (JPSS-1, -2, -3, and -4), the US new generation polar-orbiting operational environmental satellite system. VIIRS has 22 spectral bands. 16 of these have a spatial resolution of 750 m at nadir, 5 have a spatial resolution of 375 m at nadir, and the final is the day/night band with a spatial resolution of 750 m. All AVHRR/3 bands are continued on VIIRS, and the regional optical snow cover chain at MET has earlier been adapted to VIIRS. Local AVHRR and VIIRS swaths are processed hourly.

The Sentinel-3 satellites are a part of the European Commision's Copernicus programme for Earth observation. Sentinel-3A was launched on February 16 2016 and Sentinel-3B on April 25 2018. The design life is 7.5 years, with additional fuel for another 5 years. Two more Sentinel-3s are planned, and the total programme lifetime is ~15 years. The satellites are polar orbiting at a height of 814.5 km with a local equatorial crossing time of 10:00 am. With two satellites in orbit, the revisit time is daily. Among the instruments on board Sentinel-3, is the optical Sea Land Surface Temperature Radiometer (SLSTR) instrument. SLSTR has two instrument views: nadir and oblique (rear). The nadir swath scanner footprint is 1400 km wide, while the oblique is 740 km. In the visible and shortwave infra-red range there are six channels recording at 500 m resolution. In addition to these solar reflectance bands there are five

channels with 1000 m resolution in the thermal range. The radiometric bands are listed in Table 1 below.

Band	Central Wavelength (μm)	Resolution (m)	Central Wavelength for similar channel on AVHRR/3 (μm)
S1	0.554	500	
S2	0.659	500	0.630 (Channel 1)
S3	0.868	500	0.862 (Channel 2)
S4	1.375	500	
S5	1.613	500	1.61 (Channel 3A)
S6	2.256	500	
S7	3.742	1000	3.74 (Channel 3B)
S8	10.854	1000	10.8 (Channel 4)
S9	12.023	1000	12.00 (Channel 5)
F1	3.742	1000	
F2	10.854	1000	

Table 1: the SLSTR radiometric bands and AVHRR/3 channels.

Table 1 contains a column for the AVHRR channels as well. All AVHRR/3 bands are continued on SLSTR, meaning that the Sentinels are relevant for CryoClim as a way to continue the global snow cover product beyond the lifetime of the remaining AVHRR platforms.

4.1 Porting the optical chain to Sentinel-3

The SLSTR data is organized fundamentally different from the AVHRR GAC data provided by CM SAF. Extensive rewrite of code was therefore necessary to adapt to the SLSTR input data. The AVHRR GAC data is organized according to swaths - with a handful of HDF5 files per swath, one containing the radiometric data, latitude and longitude and the rest for auxiliary data. The SLSTR data is organized as sectors, with

one directory for each sector, and the directory contains more than 100 files (see Table 2).

cartesian_an.nc	flags_in.nc	S1_quality_ao.nc	S5_quality_an.nc	S6_radiance_co.nc
cartesian_ao.nc	flags_io.nc	S1_radiance_an.nc	S5_quality_ao.nc	S7_BT_in.nc
cartesian_bn.nc	geodetic_an.nc	S1_radiance_ao.nc	S5_quality_bn.nc	S7_BT_io.nc
cartesian_bo.nc	geodetic_ao.nc	S2_quality_an.nc	S5_quality_bo.nc	S7_quality_in.nc
cartesian_cn.nc	geodetic_bn.nc	S2_quality_ao.nc	S5_quality_cn.nc	S7_quality_io.nc
cartesian_co.nc	geodetic_bo.nc	S2_radiance_an.nc	S5_quality_co.nc	S8_BT_in.nc
cartesian_in.nc	geodetic_cn.nc	S2_radiance_ao.nc	S5_radiance_an.nc	S8_BT_io.nc
cartesian_io.nc	geodetic_co.nc	S3_quality_an.nc	S5_radiance_ao.nc	S8_quality_in.nc
cartesian_tx.nc	geodetic_in.nc	S3_quality_ao.nc	S5_radiance_bn.nc	S8_quality_io.nc
F1_BT_in.nc	geodetic_io.nc	S3_radiance_an.nc	S5_radiance_bo.nc	S9_BT_in.nc
F1_BT_io.nc	geodetic_tx.nc	S3_radiance_ao.nc	S5_radiance_cn.nc	S9_BT_io.nc
F1_quality_in.nc	geometry_tn.nc	S4_quality_an.nc	S5_radiance_co.nc	S9_quality_in.nc
F1_quality_io.nc	geometry_to.nc	S4_quality_ao.nc	S6_quality_an.nc	S9_quality_io.nc
F2_BT_in.nc	indices_an.nc	S4_quality_bn.nc	S6_quality_ao.nc	time_an.nc
F2_BT_io.nc	indices_ao.nc	S4_quality_bo.nc	S6_quality_bn.nc	time_bn.nc
F2_quality_in.nc	indices_bn.nc	S4_quality_cn.nc	S6_quality_bo.nc	time_cn.nc
F2_quality_io.nc	indices_bo.nc	S4_quality_co.nc	S6_quality_cn.nc	time_in.nc
flags_an.nc	indices_cn.nc	S4_radiance_an.nc	S6_quality_co.nc	viscal.nc
flags_ao.nc	indices_co.nc	S4_radiance_ao.nc	S6_radiance_an.nc	xfdumanifest.xml
flags_bn.nc	indices_in.nc	S4_radiance_bn.nc	S6_radiance_ao.nc	
flags_bo.nc	indices_io.nc	S4_radiance_bo.nc	S6_radiance_bn.nc	
flags_cn.nc	met_tx.nc	S4_radiance_cn.nc	S6_radiance_bo.nc	
flags_co.nc	S1_quality_an.nc	S4_radiance_co.nc	S6_radiance_cn.nc	

Table 2: the list of files containing data for one SLSTR sector.

Data from the two instrument views are on separate files indicated by "_xn.nc" for nadir and "_xo.nc" for oblique. For now we focus on the nadir view. Furthermore, SLSTR has 500 m resolution for some of the channels, including S2, S3 and S5 which are needed in the snow cover algorithm. We do a reduction to 1000 m resolution for these channels. The snow cover algorithm uses reflectances from the AVHRR visible channels 1, 2 and 3A. SLSTR visible channel data is stored as radiances. We collect the top of atmosphere (TOA) solar irradiance from the corresponding "_quality_xn.nc" files and convert from radiance to reflectance values. AVHRR can only transmit one of Channel 3A and 3B at the time. SLSTR transmits both (i.e., S5 and S7). It could be useful to combine the information from both of these simultaneously. However, for now we implement only the use of S5. The number of files needed for each sector reduces to 12 (see Table 3).

Filename	content needed
S2_radiance_an.nc	S2 radiance
S2_quality_an.nc	TOA solar irradiance for S2
S3_radiance_an.nc	S3 radiance
S3_quality_an.nc	TOA solar irradiance for S3
S5_radiance_an.nc	S5 radiance
S5_quality_an.nc	TOA solar irradiance for S5
S7_BT_in.nc	S7 brightness temperature
S8_BT_in.nc	S8 brightness temperature
geodetic_an.nc	latitude, longitude, elevation
geometry_tn.nc	solar zenith angle
met_tx.nc	NWP data, skin temperature
flags_an.nc	land information

Table 3: the SLSTR sector files needed when using the optical snow cover algorithm on SLSTR data.

With all necessary input data in place, the statistical coefficients derived for AVHRR are read, and the snow cover algorithm estimates probabilities for *snow*, *land* and *cloud*. Figure 17 shows an example of snow cover classification from April 9 2017. The data archive was searched for daylight scenes containing snow and for which we had AVHRR or VIIRS swaths in near time for comparison. Not many scenes have been

processed at this point, but the results seen so far are very promising keeping in mind the simplifications that were made.

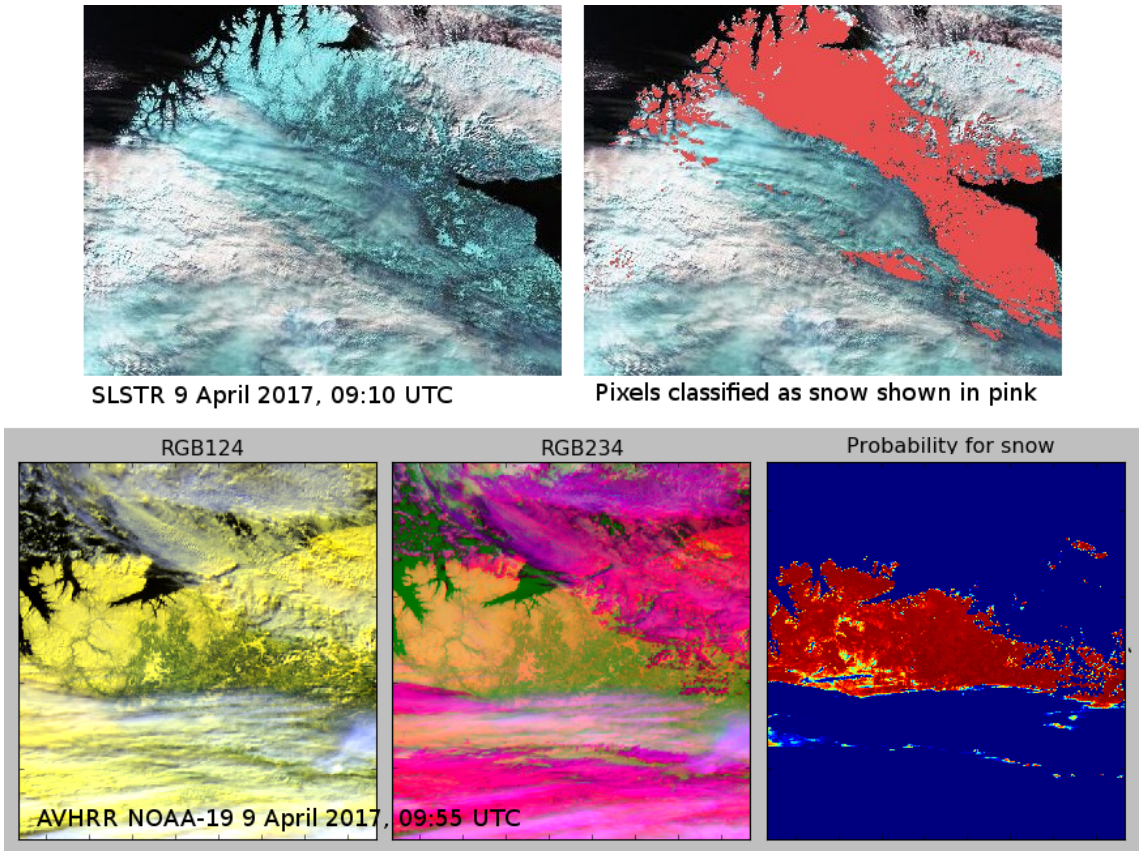


Figure 17: Example of result for SLSTR processing. The top row shows a part of an SLSTR sector crossing Northern Norway on April 9 2017, 09:10 UTC (left) and the corresponding classed product (snowy pixels in pink, right). The bottom panels show RGB colour composites of a nearby AVHRR swath (NOAA-19, April 9 2017, 09:55 UTC, left and middle), and the corresponding probability for snow (right panel, red indicates high probaility for snow, blue corresponds to low probaility for snow).

4.2 Implementing SLSTR in the processing chain

The optical and PMW snow service processing in CryoClim / Sentinel4CryoClim is heritage from EUMETSAT OSI SAF. It is a flexible system that allows data from several sensors to be processed individually, with one sub-module for each, and then combined into a multi-sensor product. Adding new sensors in this system is straight forward. The SLSTR version of the processing chain has been implemented as an independent modul in the optical snow cover processing chain. It has been tested and run in the same environment that is used for the processing of the AVHRR GAC, PMW and HMM components.

5 Work package 5: Validation of optical component

In the CryoClim project the global, optical daily snow cover product was validated against in situ snow depth observations from about 30 stations daily through 2005. Data from WMO stations from the Global Historical Climatology Network Daily (GHCN-D) database served as "ground truth", and was compared with the geographically and temporal corresponding satellite product pixel. The ground stations were carefully selected to only use stations that represent the surrounding terrain well. Stations with suspicious snow depth observations were removed, and all snow depths lower than 5 cm were avoided. All this was done to reduce the representation error. Figure 18 shows the variation in accuracy through the year found in CryoClim. A total of 8547 points were compared, and the total accuracy (hit rate) for the gridded optical product was 97%.

When developing the first version uncertainty estimate in Phase 1 of this project, an extensive validation of AVHRR Local Area Coverage (LAC) swath scenes covering Scandinavia was performed. The resolution of these regional swath products is 1 km, compared to 4 km for the global swath product. Furthermore, the terrain types encountered in the regional area are fewer, and closer to the terrain types for which the algorithm was originally developed. No selection of stations was done to avoid "difficult" terrain. All swaths from the 2015/2016 season were compared with ground snow observation. A total of 188 620 samples were collected. The overall hit rate was 94%. The total hit rates for classes *snow* and *land* were both also 94%. Seasonal variations very similar to that found in CryoClim were found here as well. Conference the report from Phase 1 (Solberg et al., 2017) for more details.

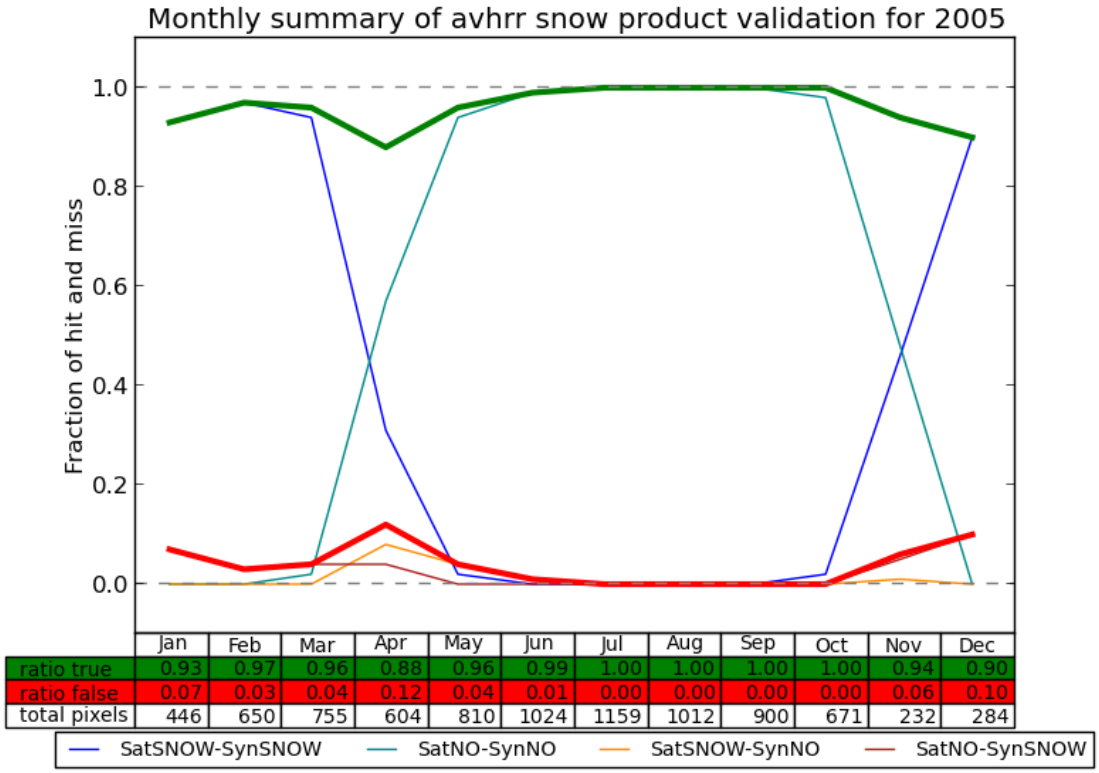


Figure 18: Validation results from CryoClim. The figure shows the monthly total hit rate for the CryoClim daily optical component for 2005.

In Phase 2 we increase the validation efforts substantially. Four different validation datasets are used, and we validate the entire 34 year long time series of daily optical products. This chapter provides a short description of each validation dataset, a section on the validation methods used, and then the validation results organized per dataset. Knowing that a particular ground snow observation can be listed in more than one of the validation datasets, we do not present results that are the sum for all four validation datasets.

In this phase there has been no screening of ground stations based on their location, which opens for representation error. It is not optimal to use a ground-based snow depth measurement as “truth” for the snow cover condition in an entire satellite product grid cell. There might be geographical or topographical factors that make the ground observation a poor representative for the entire satellite product pixel, such as large variation in altitude or land use within the pixel. Particular during melting and first

snowfall, the in situ snow observation might not represent the full satellite product pixel very well. This is something we must keep in mind when choosing to validate a satellite based product against ground point observations.

The validation chapter in this report focuses on the optical component only. NR is responsible for validation of the multi-sensor product (conf. Solberg et al., 2018). While the multi-sensor product has no gaps due to cloud cover or polar night, the optical daily product has gaps for both reasons. The number of compared samples will therefore be lower for the optical product than for the multi-sensor product.

5.1 Validation datasets

As part of the work package on validation in Phase 1 of Sentinel4CryoClim, NR looked into various available historical data records for ground snow observations. A total of four datasets were selected, including GHCN-D used in CryoClim. These are briefly described below. A thorough description can be found in Chapter 4 of the project report from Phase 1 of this project (Solberg et al., 2017).

The distribution of validation data per year is shown in Figure 19. Each source is indicated by a colour. GHCN-D is the dominating source throughout the satellite data period 1982 - 2015. The FSU data is limited in amount and not easy to spot, and is therefore encircled by a pink ellipse in Figure 19. The figure shows the number of actual comparisons, *after* the removal of cloudy, water-filled and night time pixels.

GHCN-D
Global Historical Climatology Network Daily is an integrated database of daily climate observations from land surface stations across the globe. The data records have undergone a common set of quality assurance reviews. GHCN-D contains records from over 90 000 stations in 180 countries and territories, and the variables include maximum and minimum temperature, total daily precipitation, snowfall and snow depth. About two thirds of the stations provide precipitation measurements only.

While data from GHCN-D version 3.01 was used in CryoClim, version 3.22 is used now. The overall quality is assumed to be higher with this version of the dataset. GHCN-D does not contain direct information on snow cover, so snow depth observations are used instead. A point measurement of snow depth does not necessarily correspond well to the full snow cover of the 5x5 km grid cell, which opens for

representation error. In CryoClim stations were hand-picked to reduce representation error, and suspicious data was removed. For the validation work in Sentinel4CryoClim there has been no screening of stations based on terrain, and no manual removal of suspicious data. We do, however, repeat the screening of data based on snow depth. Snow depths between 0 and 5 cm are not used.

The GHCN-D dataset contains data for the entire period 1982 - 2015. Snow depth observations from a total of 7829 unique stations are used for validation of the optical component. After removing samples for which the satellite product is clouded, dominated by water or for some other reason contains no data (most often due to winter-time darkness), a total of ~11 Million samples remain. GHCN-D is by far the dominating source of validation data, as shown in Figure 19.

SCCONE

Snow Cover Characteristics Over Northern Eurasia (SCCONE) is created from daily snow observations at some 600 meteorological stations, and maintained by the All-Russia Research Institute of Hydrometeorological Information - World Data Centre (RIHMI-WDC). The snow observations include snow depth measurements and visual determination of the amount of snow covering the area around a meteorological station.

The SCCONE snow extent observations are used for validation. Observed snow cover below 50% is considered snow-free while 50% and above is considered snow-covered. Data flagged as missing, doubtful or with undocumented flag value has been removed. SCCONE is the second largest of the four validation datasets. In total 528 unique stations contribute to the 1 875 039 samples used when validating the optical daily product for the period 1982 - 2015 (see Figure 19).

HSDSD

HSDSD is a dataset collected in the former Soviet Union and processed by the State Hydrometeorological Service (Obninsk, Russia). The dataset includes snow depth and snow cover measurements from some 280 stations. The data is quality controlled, and snow depth is recorded year-round, also when snow is absent. This makes it possible to accurately identify the onset and end of the periods with snow cover, in contrast to the regular WMO synoptic observations, which often omit snow depth observations when snow is absent.

The HSDSD dataset starts in 1881 and ends in 1995. We use data from 1982 to 1995. The report from Phase 1 of Sentinel4CryoClim contains further descriptions of pre-screening of the HSDSD data performed by NR, but it is worth mentioning that

geographical coordinates for the stations are taken from the GHCN-D dataset, meaning that stations that are not also described in GHCN-D are removed from the HSDSD validation dataset. 247 stations are used for validation of the optical component, and there are 336 862 samples in total.

FSU

The FSUHSS dataset is generated by the Institute of Geography, Russian Academy of Sciences. The dataset includes measurements of snow cover, snow density, snow depth and snow water equivalent. The observations are done along transects. The dataset contains data from a total of 1345 stations. There are no snow-free observations.

Available data for the period 1 August 1992 – 31 December 1996 is collected from FSU. Snow cover measurements below 50% snow cover are considered to be snow-free in order to compare with the binary satellite product. Data flagged as suspicious is removed. Like for the HSDSD data, the station coordinates are taken from the GHCN-D dataset, and stations not also available in GHCN-D are removed from the validation. For validation of the optical product a total of 231 unique stations contribute, and 7442 comparisons are made.

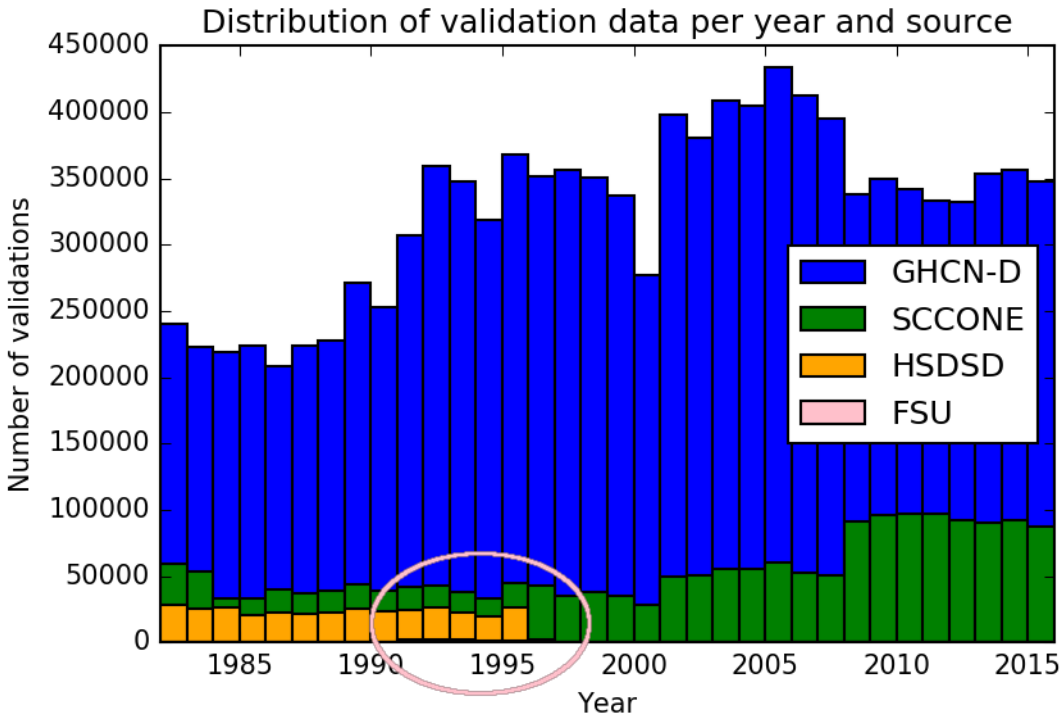


Figure 19: The amount of validation data from each datasets shown as function of year.

5.2 Validation methods

We follow the strategy used in CryoClim and Phase 1 of Sentinel4CryoClim. The ground observation is translated from snow depth if necessary - to a binary snow/no snow value. Samples with snow depths above 0 cm and below 5 cm are removed. The ground snow observation (*snow*, *no snow*) is compared with the class value (*snow*, *snow-free*) of the nearest pixel from the satellite product of the same day, and verification measures such as total accuracy, total hit rate for each class (*snow*, *land*), false alarm ratio etc. are computed. As discussed earlier such a comparison can introduce representation error.

Table 4 shows the 2x2 contingency table for matches and mismatches between the snow cover product derived from satellite and the ground observation of snow.

	Ground observation: **snow**	Ground observation: **no snow**
Satellite product: **snow**	A (hit)	B (false positive)
Satellite product: **no snow**	C (miss)	D (true negative)

Table 4: 2 x 2 contingency table (confusion matrix) for verification of satellite snow product.

A perfect match between the satellite product and the ground observations would produce only hits and true negatives, and no cases of alternatives B and C. Statistical scores (verification measures) can be computed from the contingency table values and used to describe the products' performance. Among these scores are hit rates, false alarm ratio, probability of false detection and bias:

- Total hit rate (accuracy): (A+D)/(A+B+C+D)
- The hit rate (accuracy) for snow: A/(A+C)
- The hit rate (accuracy) for snow-free: D/(B+D)
- The false alarm ratio: B/(A+B)
- The probability of false detection: B/(B+D)
- Bias: (A+B)/(A+C)

Below we present confusion matrices, tables and various plots of the validation results. The results are presented for each validation dataset source.

5.3 Validation results: GHCN-D 1982 - 2015

Table 5 shows a confusion matrix of the comparison between daily optical products and snow observations from the GHCN-D dataset. Some 11 million data points are compared, and the overall accuracy is 96%. Hit rate for each surface class (*snow* and *snow-free*) is 94% and 97% respectively. 7829 unique stations contribute.

	GHCN-D: snow (SD > 5 cm)	GHCN-D: no snow (SD = 0 cm)	Total
Satellite product: snow	1 970 411	289 236	2 259 647
Satellite product: no snow	136 763	8 651 587	8 788 350
Total	2 107 174	8 940 823	11 047 997
Hit rates	Hit rate snow: 94%	Hit rate land: 97%	Total hit rate: 96%

Table 5: Summary of validation results against GHCN-D.

Table 6 shows a summary of the verification measures for each month of the year. The majority of data samples belong to the summer season. The northern areas are dark during winter, and many ground observation stations have insufficient daylight for a satellite product to be generated. We see that the overall accuracy is 92% or higher for each month, which means that the algorithm generally does a good job at correctly classifying the pixels. Looking, however, at accuracy (hit rate) for the classes *snow* and *land* independently, we see that the hit rate for snow is very poor during summer, indicating that snow observed at the ground stations during summer time is often not found by the algorithm. There are several factors that can contribute to this. The extent of the summertime snow-covered areas might be too small to dominate the 5x5 km grid cell (representation error), the surface appearance of the snow might be too different from the winter/spring snow[6], and therefore not recognized by the current set of static statistical coefficient, or there might be other shortcomings in the algorithm that cause this. Chapter 5.8 contains a further discussion on poor hit rate for the class *snow* during summer. We return also to the topic later in this sub-chapter.

The hit rate for *land* is 87% at the lowest (below 90% for the winter months December through February), and above 95% for most of the remaining year (May through October). We suspect that onset of snow cover and snow melt has an influence on the reduced performance during the winter season, and representation error can be a factor.

6 The majority of the training data on which the statistical coefficients are based are collected from swaths during spring (February to May).

Month	Accuracy	Hit rate snow	Hit rate land	False alarm ratio	Prob. of false detection	Bias	Number of samples
January	0.93	0.96	0.87	0.052	0.13	1.0	559 280
February	0.93	0.96	0.87	0.046	0.13	1.0	730 475
March	0.93	0.94	0.91	0.056	0.085	1.0	886 898
April	0.93	0.89	0.94	0.17	0.059	1.1	919 075
May	0.96	0.84	0.96	0.49	0.036	1.7	1 142 222
June	0.98	0.69	0.98	0.81	0.017	3.5	1 294 466
July	0.99	0.14	0.99	0.99	0.012	10	1 420 403
August	0.99	0.13	0.99	0.99	0.01	14	1 368 531
September	0.99	0.49	0.99	0.91	0.014	5.7	1 095 938
October	0.96	0.84	0.97	0.52	0.031	1.7	791 757
November	0.93	0.91	0.93	0.20	0.066	1.1	440 682
December	0.92	0.95	0.88	0.089	0.12	1.0	398 270
Total	**0.96**	**0.94**	**0.97**	**0.13**	**0.032**	**1.1**	**11 047 997**

Table 6: Monthly verification measures for validation of daily optical snow cover product against GHCN-D for 1982 - 2015.

The false alarm ratio describes how many of the satellite snow classifications that are snow-free according to the ground observation. The table shows that the false alarm ratio is very high during summertime. The false alarm ratio is considered a measure of reliability. We select a random summer date: July 1 2005. For this date there are 28 cases of false positives (ref. Table 4 for definition of "false positive"). 9 of these are located in Norway, and listed in Table 7. Each is looked up and found to be located near a road in a valley or along the coast, in proximity to mountain areas (see also Figure 20 where the locations of the 9 stations are indicated by red circles). Based on this we conclude that representation error is very likely the main reason for the high false alarm ratio during summer time seen in Table 6. Due to the definition of bias (see Chapter 5.2) we are for the same reason not alarmed by the high bias values seen for the same months.

Date	Station ID	latitude	longitude	sample type
20050701	NO000110851	61.33	6.93	SNOW_NOSNOW
20050701	NO000133554	60.90	6.72	SNOW_NOSNOW
20050701	NO000110977	62.34	8.05	SNOW_NOSNOW
20050701	NO000110833	61.71	6.62	SNOW_NOSNOW
20050701	NO000110770	61.69	6.81	SNOW_NOSNOW
20050701	NO000133578	62.33	7.51	SNOW_NOSNOW
20050701	NO000110788	61.42	6.38	SNOW_NOSNOW
20050701	NO000110743	60.90	7.20	SNOW_NOSNOW
20050701	NO000134694	62.67	9.20	SNOW_NOSNOW

Table 7: Cases of false positives for July 1 2005 in Norway when validating against GHCN-D. The station locations are also plotted in Figure 20 which shows a section of the daily optical snow cover product from this date.

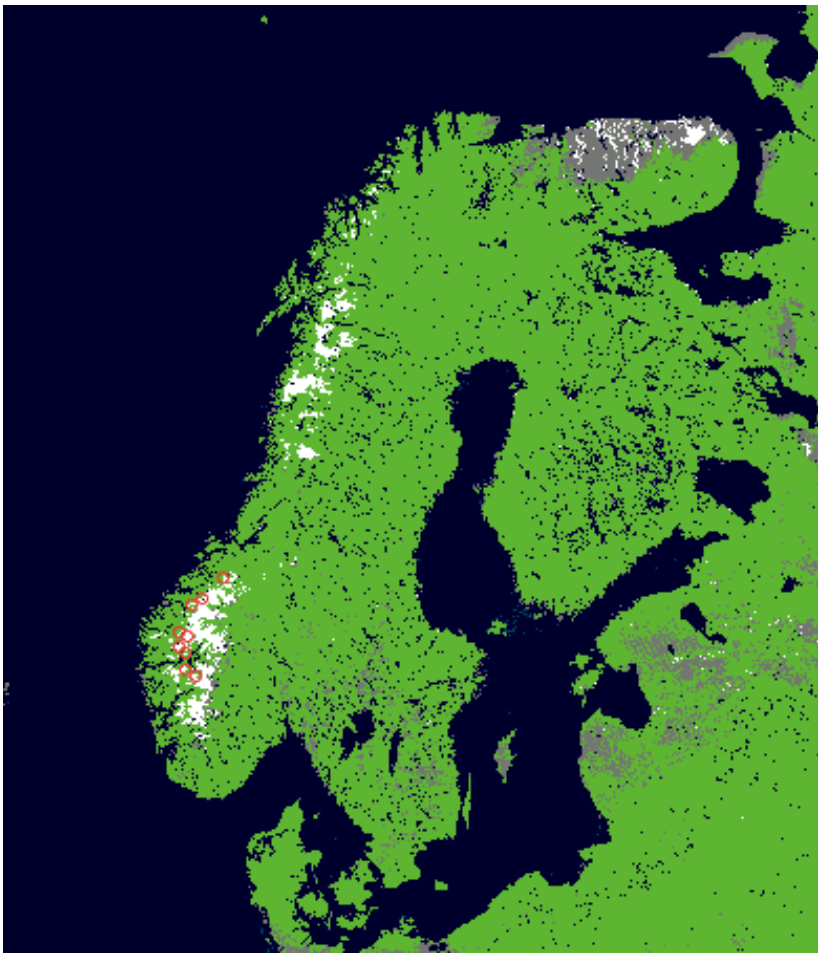


Figure 20: The Figure shows a section of the daily, optical product from July 1 2005. Snow cover is seen in white, land is shown in green, and areas that are clouded for all available satellite swaths of this day are shown in grey. The red circles show the location of the 9 cases of false positives in Norway on this date (ref. Table 7).

The probability of false detection gives an indication of the fraction of snow-free observations that are classified as *snow* in the satellite product. This ratio is above 10% during winter (DJF), but otherwise very low. Here, again, representation error is expected to have at least some influence.

So far we have discussed the monthly verification measures. Figure 21 shows the total accuracy as function of time of year with one line for each year of the period 1982-2015. All data within one month contributes to the monthly value.

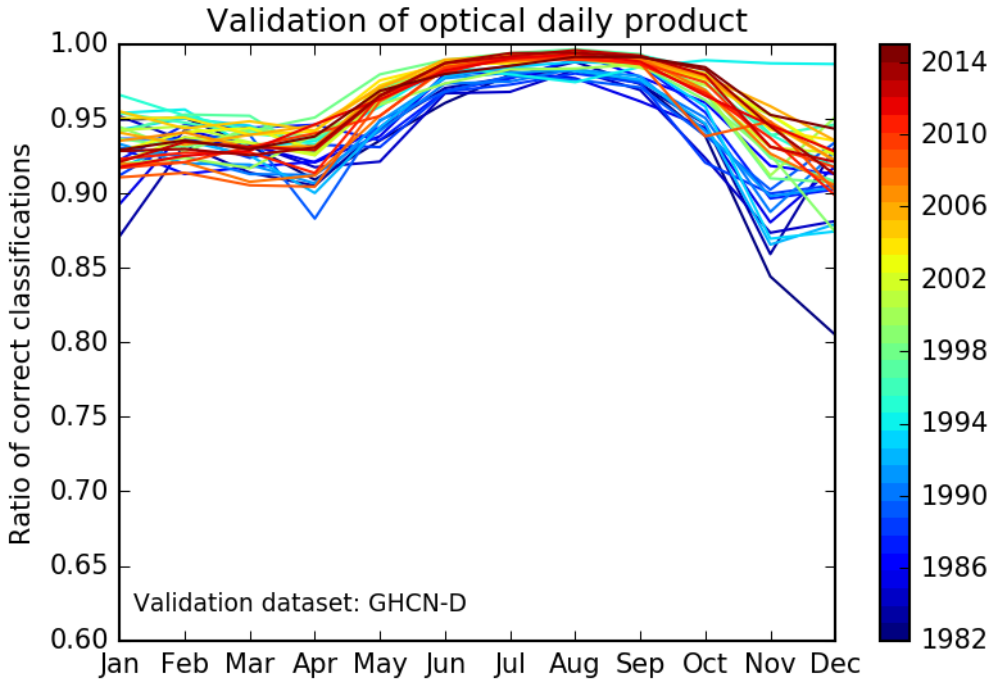


Figure 21: Monthly total hit rate when validating against GHCN-D. Each year is plotted as a separate line.

The figure shows that during summer time the total accuracy has improved with time. This is also the case for the fall season, while the results are not so clear for the first 4 months of the year. For two reasons we expect the validation to improve from the start towards the end of the dataset period: the algorithm is trained using data from AVHRR/3 (first carried on NOAA-15 launched in May 1998), and there is much more data available for the latter half of the dataset period (see Figure 3).

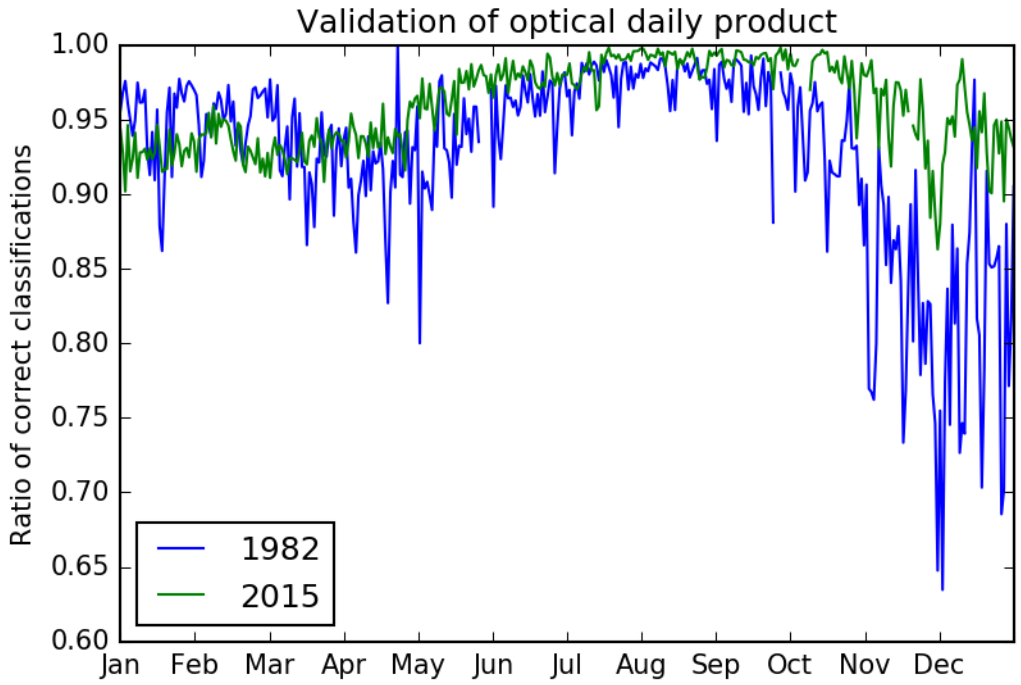


Figure 22: Daily total hit rate when validating against GHCN-D.

Figure 21 is a plot of monthly averaged values for total hit rate. We show a similar plot in Figure 22, but now the daily hit rate values are shown. This is to illustrate the level of variations from day to day. To avoid a very cluttered figure we show only the years 1982 and 2015. There is less variation from day to day in 2015 than in 1982.

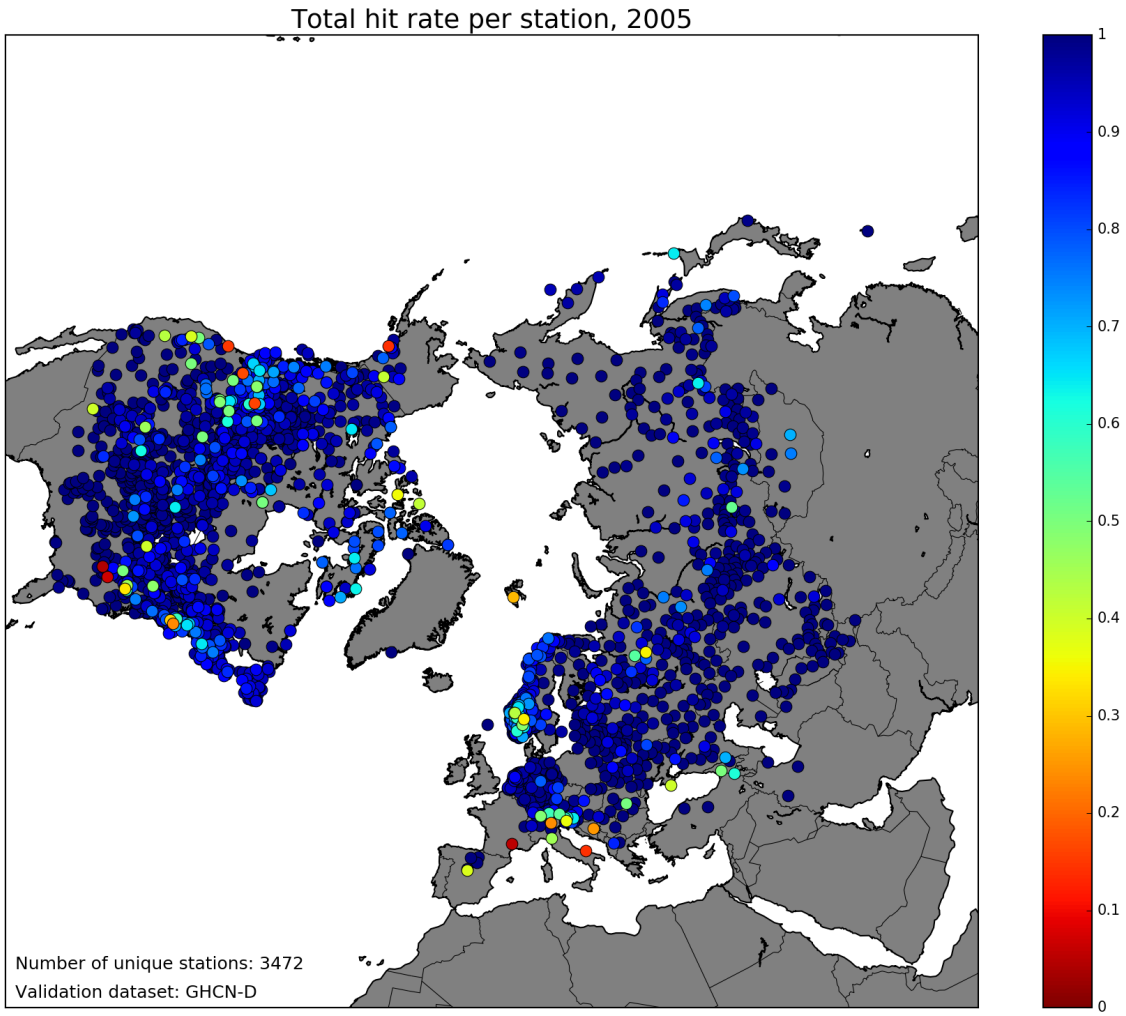


Figure 23: Total hit rate per ground station when validating against GHCN-D for the year 2005. Each dot represents a ground station. The colour of the dot gives the hit rate for the ground station.

Figure 23 shows a map plot of the total accuracy (hit rate) for the year 2005 when validating against GHCN-D, and Figure 24 shows the corresponding plot for hit rate for the class *snow* (similar plots can be made for other statistical scores as well). Figure 25 shows the number of samples from each station. Each dot in Figures 23 and 24 represents a ground station, and the colour of the dot shows the hit rate value for that particular ground station. Blue colour corresponds to high accuracy, while red shows low accuracy. Note that these plots are designed as an instrument to investigate cases of poor statistical scores. The stations with the poorest scores are plotted on top so that they do not "drown" underneath all the blue dots from all the well performing stations.

This must be kept in mind so that the plots are not misleading the reader. The "carpet" of blue dots seen in Figures 23 and 24 illustrates again the overall high accuracy achieved. We must also keep in mind that hit rates have been calculated for all stations, even for those with just a few samples in total, meaning that the results from the stations with few ground observations might not be representative.

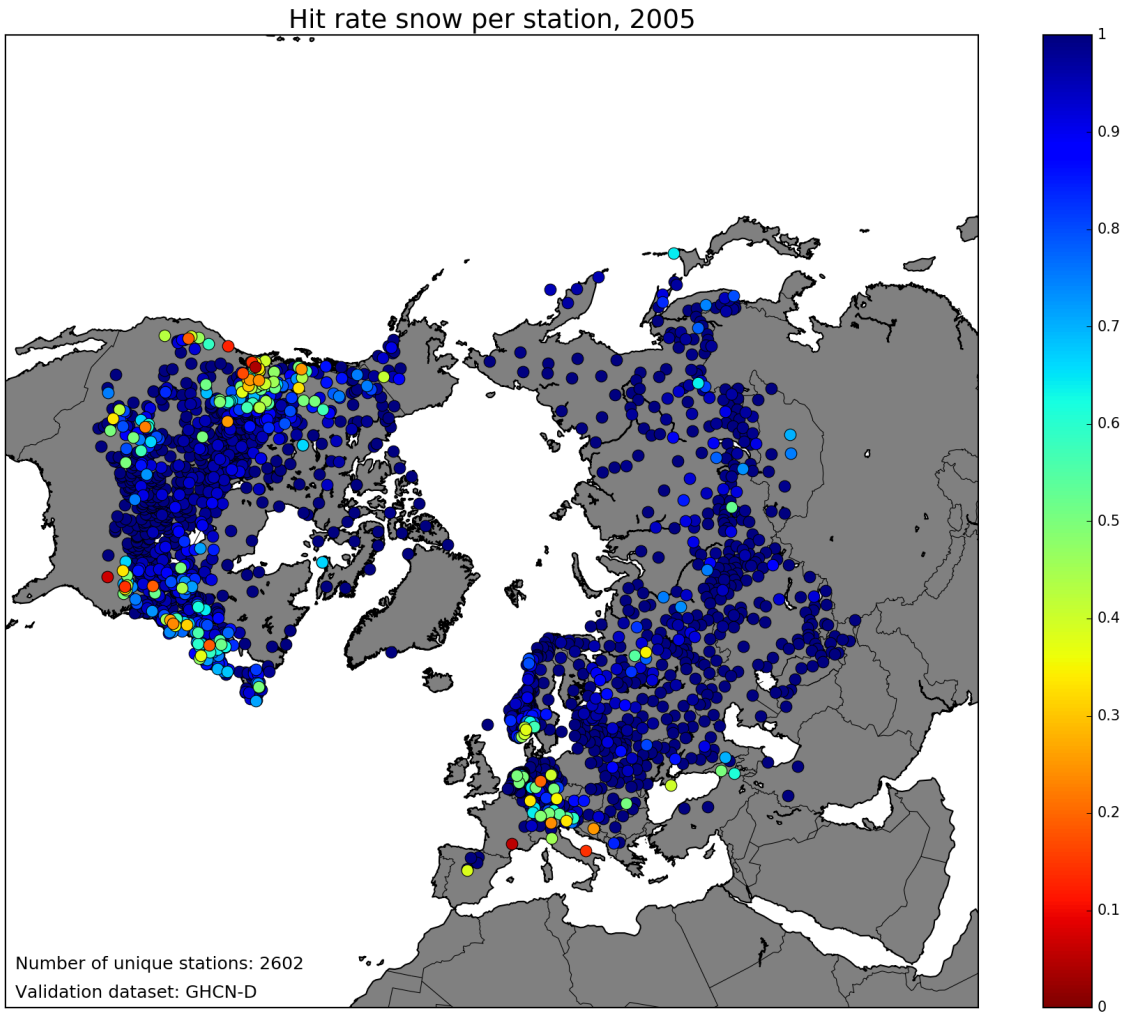


Figure 24: Hit rate for the class snow when validating against GHCN-D for the year 2005. Each dot represents a ground station. The colour of the dot gives the hit rate. Note that stations with no observation of snow during 2005 is not plotted. This figure therefore shows 2602 unique stations, while Figure 23 shows 3472 unique stations.

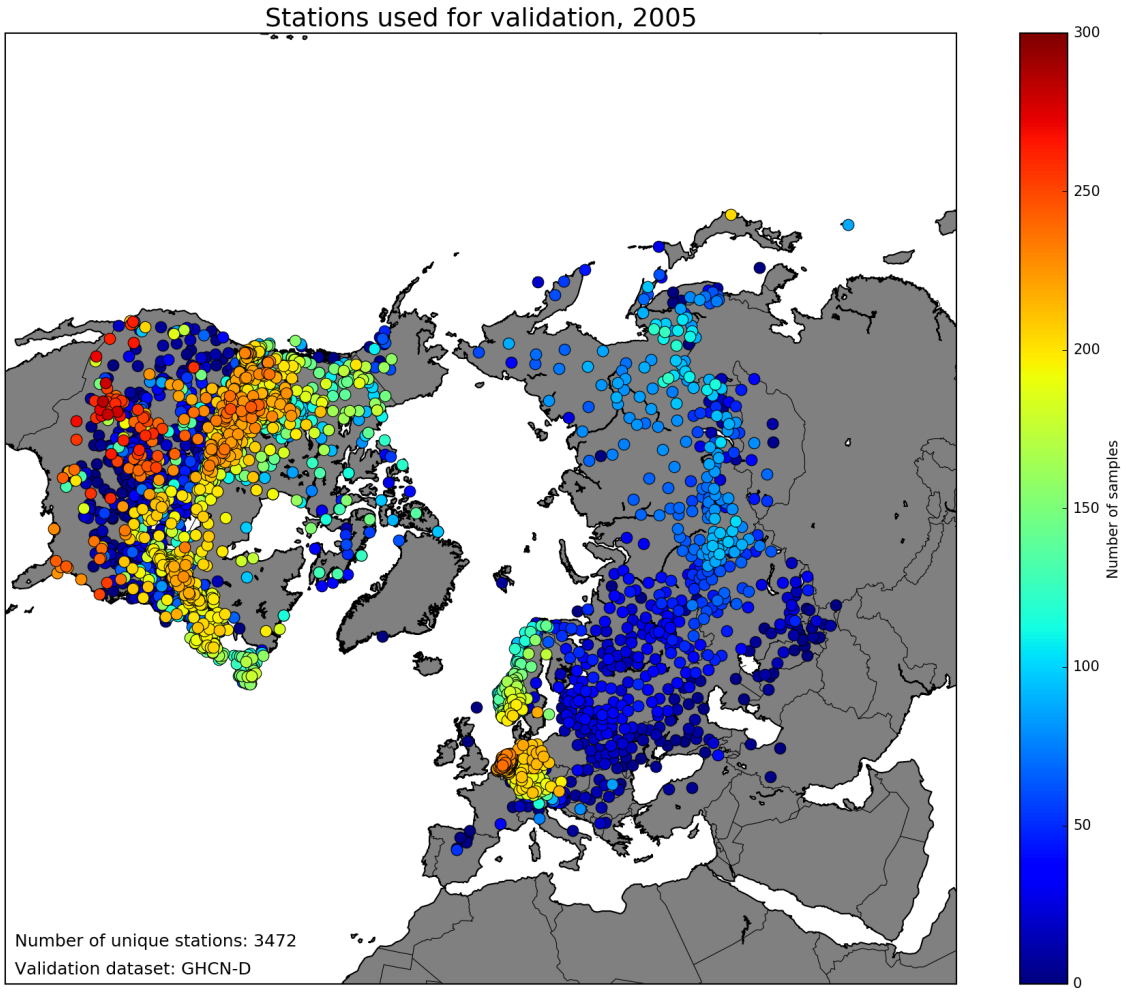


Figure 25: Number of samples from each ground station for the year 2005. This plot shows all stations, also those with only snow-free ground observations.

We notice that the red, orange and yellow dots in Figure 24 - indicators of low hit rate for *snow* - to some extent are gathered in certain areas: the western part of North America, the eastern part of North America and the central part of Europe. This might be due to geographical (terrain) reasons, or it might be due to differences in data quality from one region (country) to another. 7 stations are selected for a closer look. They are selected either based on their colour (i.e., hit rate for *snow*) in Figure 24 or from scanning through the validation output files looking for stations showing frequent cases of missed snow observations. To fully understand the mechanisms causing reduced performance we would have to repeat the swath processing using a version of the code that stores extra data (such as each signature value for each swath pixel) for these stations. Due to time constraints this is not doable. We will, however, look at the

validation results for the year 2005 as function of time of year, compare ground snow depth (if easily available) with the validation results, and assess the location of the station in terms of surrounding terrain and topography. Hopefully this will improve the understanding of the algorithm weaknesses and be of help in future work.

1) CA001078209: Topley Landing, British Columbia, Canada
This station is found near Babine Lake at an elevation of 722 m, on a plateau between the Canadian Coast mountains and the Rocky Mountains. The terrain surrounding the station looks relatively uniform, and consists of forest. 199 comparisons are made between observed snow depth and the nearest satellite product grid cell during 2005. The overall accuracy is 92%. From April through October the majority of the samples are of type "true negative" (ref: Table 4). There are 151 true negatives in total. On 4 occasions during the late spring and summer, the satellite grid cell is classified as *snow* while no snow is recorded at the station. These are believed to origin from erroneous classifications in the satellite product. 32 of the 44 in situ observations of snow coincide with *snow* in the satellite product. This gives a hit rate for *snow* of 73%. The majority of the 12 missed snow observations occurs during the month of most intense melting at the ground station. This suggests that both algorithm shortcomings (see Chapter 5.3) and representation error are likely candidates causing the missed snow observations.

2) CA001105658: Grouse Mountain, British Columbia, Canada
The station is located just north of Vancouver near a mountain resort at Grouse mountain at an elevation of 1128 m. There are in situ observations of snow from the start of year until late-April, and then from mid-November through December. Only 2 out of 47 ground observations of snow are seen in the satellite product, giving a very poor hit rate for snow (4%). The total accuracy for this station is 68%. The station is located approximately 2 km from North Vancouver and no more than 2 product grid cells away from the sea front. This makes it very difficult to get "good" validation results for this station. Although the mountain terrain continues north of Grouse Mountain the station location is not typical for the surrounding terrain.

3) US000014755: Mount Washington, New Hampshire, USA
This station is located at Mount Washington in New Hampshire, at an elevation of 1910 m. Mount Washington is the highest peak in the north-eastern United States (1916 m) and according to Wikipedia also the most topographically prominent mountain east of the Mississippi River. Based on this description we expect the validation results to be heavily influenced by representation error, and the ground station is probably not well suited for validation of a satellite product of 5 km resolution. The overall accuracy is 65%, and the accuracy for the class *snow* is 31% for 2005. The station is snow-free from early May to late October. Of the 96 in situ observations of snow, the nearest pixel of the daily satellite product show snow on 30 occasions. When the dates for these

occurrences are compared with the observed snow depth at the station it seems to coincide with recent snow fall. Representation error is very likely to cause the poor accuracy for snow for this station.

4) RO000015280: Varfu Omul, Romania

This station is located at Omu Mountain, the highest mountain top of the Romanian mountain range Bugeci. The station elevation is 2504 m. The validation data record contains observed snow only, and there are 85 samples for 2005. 43 of these have the class *snow* in the satellite product pixel containing the station coordinates, giving a total hit rate, as well as hit rate for *snow*, of 51%. The in situ data shows a snow cover that lasts until late June and reappears in late November. Comparisons of the observed snow depths and the satellite product grid cell class show similarities to that seen for the station at Mount Washington. Fresh snow falls can make the snow extent sufficiently large to be detected by the satellite product, but only a more limited area have the conditions needed to maintain the snow cover.

5) SP000008215: Navacerrada, Spain

The station is located near the municipality Navacerrada, ~50 km north-west of Madrid, at an elevation of 1894 m. The area is dominated by the Sierra de Guadarrama (the Guadarrama Mountains). For the year 2005 52 comparisons are made between the ground observation of snow depth from this station and the daily optical product. There are no snow-free in situ observations. The satellite product pixel is assigned to the class *snow* for 20 of the samples and to the class *land* for 32. This gives a hit rate for *snow* of 38%. The nearby terrain is not very uniform. Comparisons of the measured snow depth value with the validation results indicate that the success rate is larger for larger snow depth values. Representation error and perhaps also algorithm shortcomings seem relevant here.

6) FR000007560: Mount-Aigoual, France

This station near the French Riviera is very noticeable in Figures 23 and 24 due to its dark red colour (i. e., very low hit rate value for *snow*). The station is the highest manned weather station in France and is located at 1567 m on Mount-Aigoual, 70 km from the Mediterranean. The in situ snow depth measurements contain no snow-free observations. Of the ~40 samples from 2005, there are only 2 cases for which the satellite product grid cell is classified as snow-covered. This gives a hit rate (accuracy) for *snow* of 5%. Again representation error seems to be the main reason for the poor result from this station. The mountain top location of the ground station is not representative for the 5x5 km satellite product pixel in which the station is located.

7) AM000037682: Amasia, Armenia

This station is located in Northwest Armenia near the border to Turkey at an elevation of 1866 m. The country Armenia is a part of the Armenian Highland, the highest plateau in the Northern Middle East. The terrain is mostly mountainous and flat, and the highland continental climate allows for hot summers and cold winters. The station reports snow for the first ~4 months of 2005. No snow-free observations are recorded. In total 41 comparison are made between the in situ snow depth value and the satellite product grid cell. For 25 of these the product grid cell is classified as snow-covered. The remaining 16 is snow-free in the satellite product. The hit rate for *snow* is 61%. The observed snow cover lasts until mid-April, and reappears in December. From ~mid-March an onward, however, the satellite product fails to identify this grid cell as snow. Scattered snow or algorithm shortcomings may be among the reasons why the melting season is not recorded. The terrain is considered "undramatic" and the station validation data is considered good.

The locations for the seven ground stations are shown in Figure 26. Among these seven we have seen examples of stations that due to their location probably should not have been used for validation of a 5 km resolution satellite product. We have also seen example of stations that in terms of topography seem well suited for validation purposes, but for which the algorithm still fails to record the melting season, leading to reduced validation results during spring/early summer. This might be due to scattered snow, or an indication that the static set of statistical coefficients used in the Bayesian method does not describe the fading snow cover well. The algorithm shows signs of underestimating snow cover at the end of the melting season.

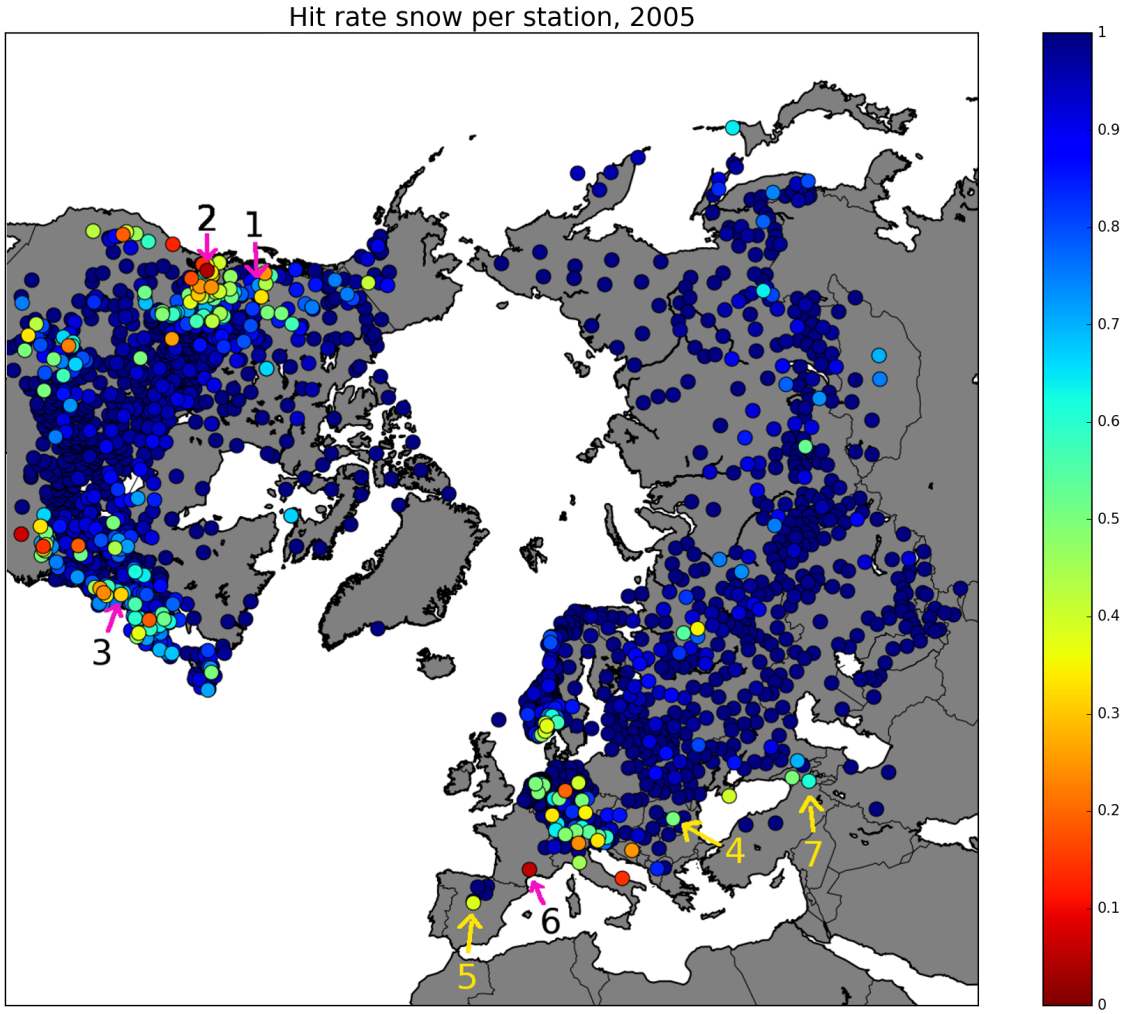


Figure 26: The figure is a copy of figure 24 on which the 7 selected ground stations are identified. Validation results from these stations are discussed at the end of Chapter 5.3.

The plots shown in Figures 23-25 are repeated for the years 1985, 1995 and 2015, and included in Appendix A. The results for 1985, 1995 and 2015 generally support what is seen for 2005.

5.4 Validation results: SCCONE 1982 - 2015

An overall number of 1 875 039 validations against snow extent data from SCCONE are done for the period 1982 - 2015. The validation results are distributed as shown in the confusion matrix below (Table 8). The total accuracy is 97%, and 528 unique stations contribute. The overall hit rate for the class *snow* is 95%, and for the class *land* 98%.

	SCCONE: snow	SCCONE: no snow	Total
Satellite product: snow	787 364	23 740	811 104
Satellite product: no snow	39 015	1 024 920	1 063 935
Total	826 379	1 048 660	1 875 039
Hit rates	Hit rate snow: 95%	Hit rate land: 98%	Total hit rate: 97%

Table 8: Summary of validation results against snow extent data from the SCCONE dataset.

Month	Accuracy	Hit rate snow	Hit rate land	False alarm ratio	Prob. of false detection	Bias	Number of samples
January	0.97	0.98	0.75	0.0076	0.25	0.99	113 942
February	0.97	0.98	0.70	0.0081	0.30	0.99	188 621
March	0.96	0.97	0.84	0.012	0.16	0.98	232 031
April	0.94	0.94	0.96	0.027	0.044	0.96	188 007
May	0 96	0.90	0.98	0.096	0.023	1.0	185 504
June	0.99	0.74	0.99	0.61	0.012	1.9	190 161
July	0.99	0.21	0.99	1.0	0.011	9.0	194 155
August	0.99	0.25	0.99	0.99	0.011	2.0	180 771
September	0.98	0.55	0.99	0.52	0.013	1.2	143 944
October	0.92	0.79	0.97	0.077	0.029	0.86	113 176
November	0.92	0.91	0.94	0.022	0.065	0.93	75 888
December	0.96	0.97	0.87	0.011	0.13	0.98	68 839
Total	**0.97**	**0.95**	**0.98**	**0.029**	**0.023**	**0.98**	**1 875 039**

Table 9: Monthly verification measures for validation against SSCONE.

Table 9 shows the monthly verification measures. The overall monthly accuracy is 92% or higher for each month. During summer time the hit rate for the class *snow* drops to 21% (July), which resembles the behaviour found for GHCN-D. The other verification measures also seem to match the seasonal variations of the GHCN-D results well (conf. Table 6). Low hit rate for *snow* during summer is discussed in Chapter 5.8.

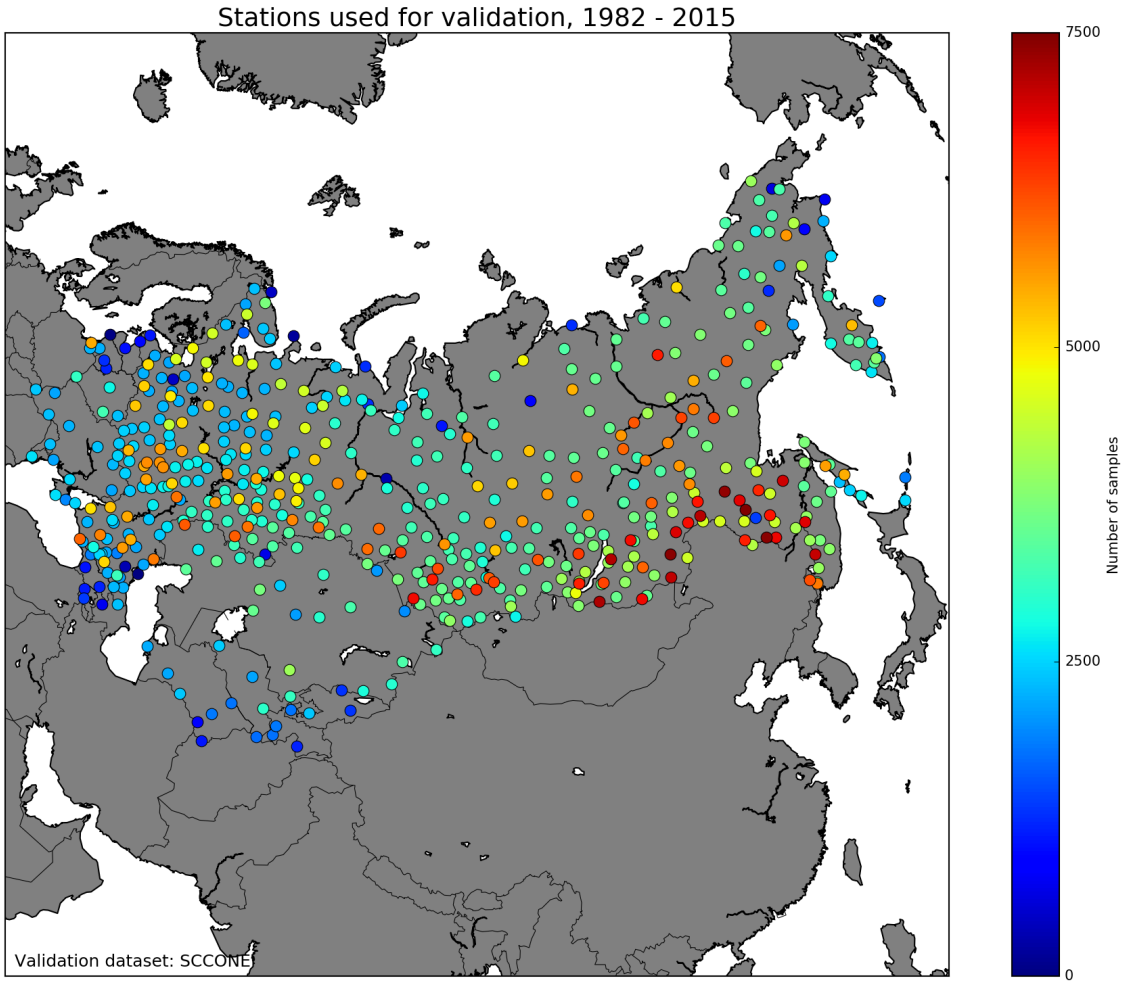


Figure 27: Ground stations from SCCONE used for validation. Each dot represents a station. The colour indicates the number of samples from each station over the period 1982-2015.

The SCCONE ground observation stations used to validate the optical daily snow cover component are shown in Figure 27. The colour of the station symbol indicates the total number of observations used from this station. Figure 28 shows the total accuracy for each station, and Figure 29 shows the hit rate for the class *snow*. Both figures show a

very large amount of dark blue dots, i.e. stations with high accuracy, both total accuracy and accuracy for the class *snow*.

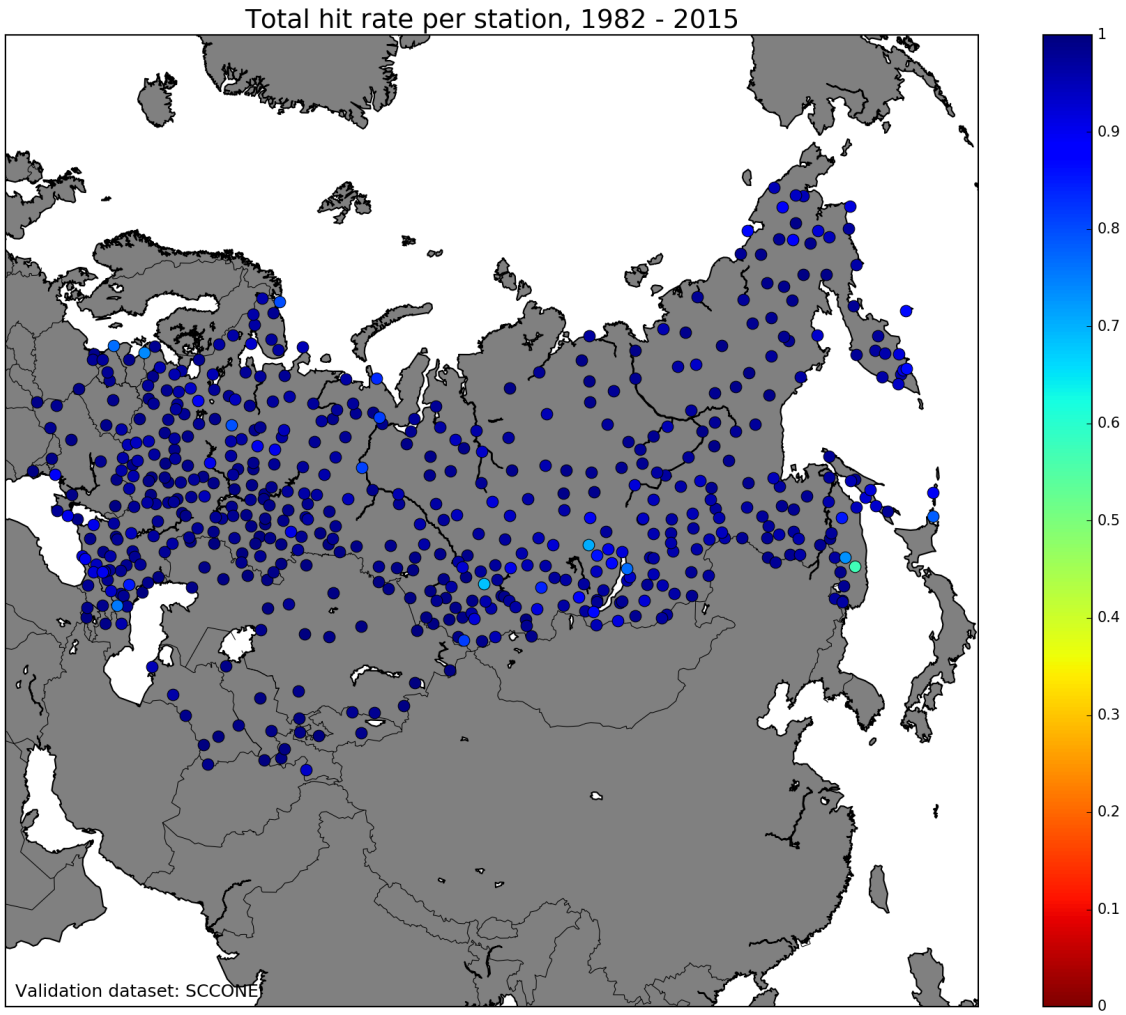


Figure 28: Total hit rate per station from the SCCONE dataset used for validation.

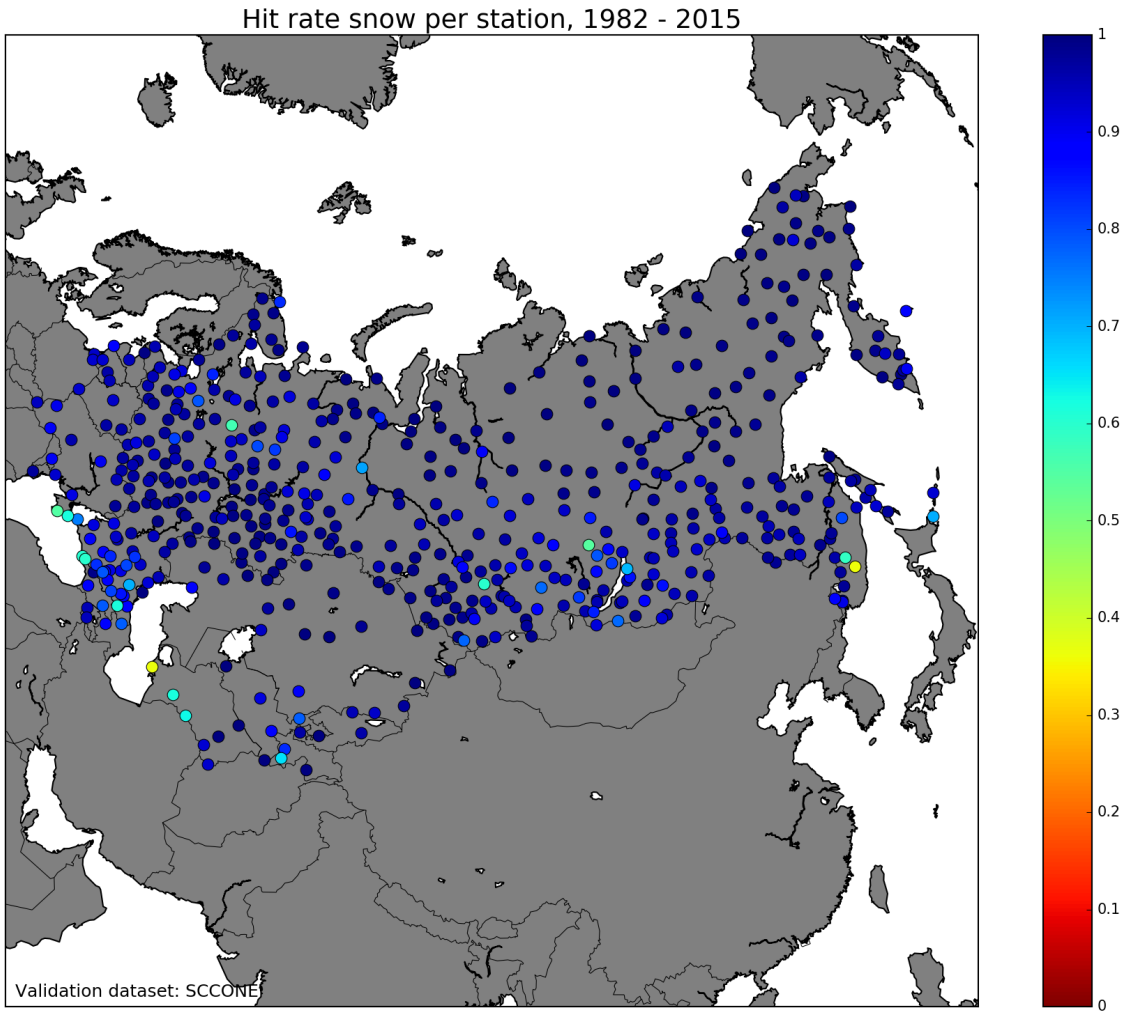


Figure 29: Hit rate (accuracy) for the class snow when validating against snow extent data from the SCCONE dataset.

Comparing the figures for SCCONE (28, 29) with the corresponding figures for GHCN-D (23 and 24, respectively) it is easily seen that there are many more stations from the GHCN-D dataset showing low hit rate (total or for snow) values than what is the case for SCCONE. SCCONE contains snow extent, while from GHCN-D snow depth values was used.

5.5 Validation results: HSDSD 1982 - 1995

The HSDSD dataset contributes validation data for the period 1982 - 1995. An overall number of 336 862 cloud free validations are available for the period. The results are distributed as shown in Table 10 below. The total accuracy is 97%, which is also the accuracy for classes *snow* and *land*. 247 unique stations contribute.

	HSDSD: snow	HSDSD: no snow	Total
Satellite product: snow	83 679	6 740	90 419
Satellite product: no snow	2 699	243 744	246 443
Total	86 378	250 484	336 862
Hit rates	Hit rate snow: 97%	Hit rate land: 97%	Total hit rate: 97%

Table 10: Summary of validation results against HSDSD data.

Month	Accuracy	Hit rate snow	Hit rate land	False alarm ratio	Prob. of false detection	Bias	Number of samples
January	0.97	0.99	0.86	0.022	0.14	1.0	12 933
February	0.98	0.99	0.92	0.0077	0.077	1.0	24 510
March	0.98	0.98	0.94	0.012	0.055	0.99	30 567
April	0.96	0.96	0.97	0.046	0.033	1.0	27 219
May	0.96	0.89	0.96	0.36	0.037	1.4	36 303
June	0.98	0.91	0.98	0.92	0.02	11	44 601
July	0 .98	nan	0.98	1.0	0.016	inf	44 617
August	0.98	nan	0.98	1.0	0.016	inf	42 630
September	0.98	0.65	0.98	0.83	0.017	3.9	32 759
October	0.95	0.83	0.96	0.21	0.036	1.1	23 273
November	0.93	0.94	0.92	0.081	0.081	1.0	10 305
December	0.95	0.98	0.86	0.051	0.14	1.0	7 145
Total	**0.97**	**0.97**	**0.97**	**0.075**	**0.027**	**1.0**	**336 862**

Table 11: Monthly summaries of validation results against HSDSD data.

The monthly values for the statistical scores are listed in Table 11. The results support what is seen for GHCN-D and for SCCONE. Figure 30 shows the geographical position of the ground stations that are used from the HSDSD data. The colour of the station symbol indicates the total number of samples from this station. Figures 31 and 32 show the total accuracy for each station and the hit rate for the class *snow*, respectively.

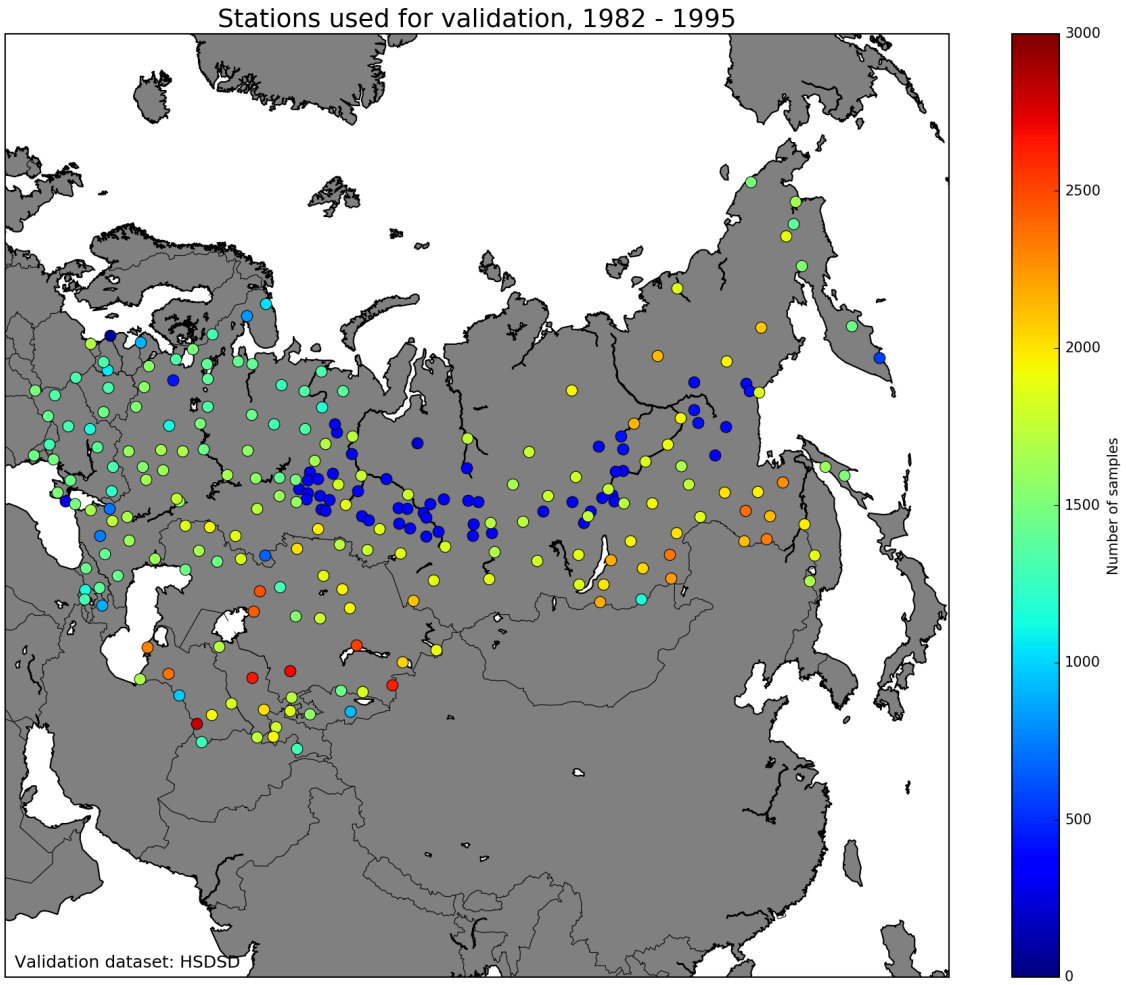


Figure 30: Ground stations from the HSDSD dataset used for validation. The colour of the station symbol indicates the number of samples this station contributes with for the period 1982-1995.

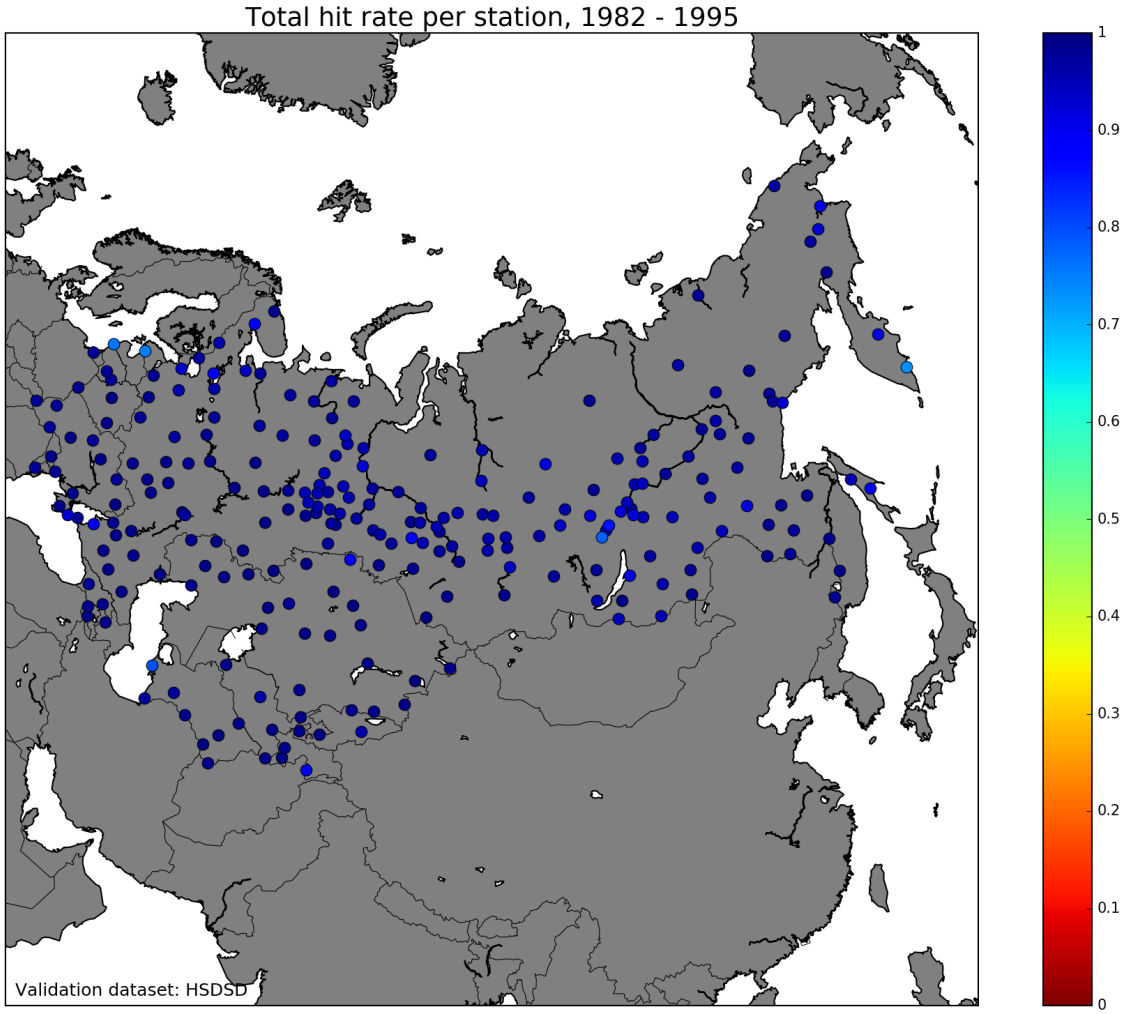


Figure 31: Total hit rate for each ground station used from the HSDSD dataset.

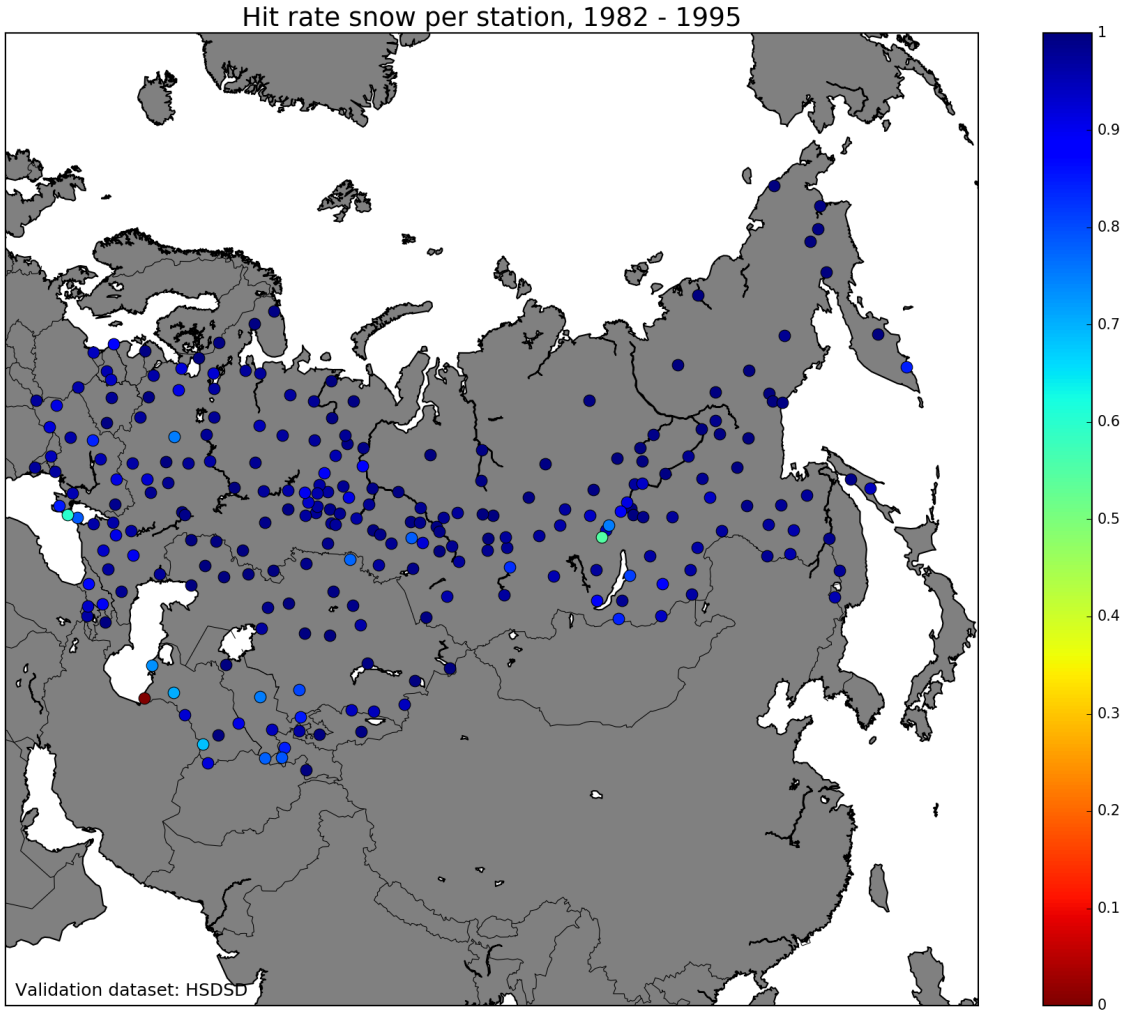


Figure 32: Hit rate for snow for each ground station used from the HSDSD validation dataset.

5.6 Validation results: FSU 1991 - 1996

The FSU dataset contains snow observations only. There are no snow-free observations. An overall number of 7442 validations are performed for the period 1991 to 1996. The results are distributed as shown in the confusion matrix below (Table 12). The total accuracy is 94% which is necessary the same as the hit rate for the class *snow* given that there are no snow-free ground observations. 231 unique ground stations contribute.

	FSU: snow	FSU: no snow	Total
Satellite product: snow	6 967	0	6 967
Satellite product: no snow	475	0	475
Total	7 442	0	7 442
Hit rates	Hit rate snow: 94%	Hit rate land: nan	Total hit rate: 94%

Table 12: Summary of validation results for FSU.

The following table shows the monthly values for the verification measures. When reading the table one should keep in mind that there is limited amount of data in this dataset. It follows from the expressions for hit rate *land* and for probability of false detection that these values are all undefined. Given that the validation dataset does not include observation of snow-free samples, all hit rates for the class *land* is also undefined. The false alarm ratio is bound to be 0. Furthermore, the expression for bias reduces to become equal to the expression for hit rate *snow*, which the monthly values in the table confirm.

Month	Accuracy	Hit rate snow	Hit rate land	False alarm ratio	Prob. of false detection	Bias	Number of samples
January	0.99	0.99	-	0	-	0.99	448
February	0.99	0.99	-	0	-	0.99	1 633
March	0.98	0.98	-	0	-	0.98	2 611
April	0.87	0.87	-	0	-	0.87	1 887
May	0.74	0.74	-	0	-	0.74	472
June	-	-	-	-	-	-	2
July	-	-	-	-	-	-	0
August	-	-	-	-	-	-	0
September	0.92	0.92	-	0	-	0.92	13
October	0.82	0.82	-	0	-	0.82	153
November	0.99	0.99	-	0	-	0.99	147
December	1.0	1.0	-	0	-	1.0	76
Total	0.94	0.94	-	0	-	0.94	7 442

Table 13: Monthly summaries of verification measures when validating against the FSU dataset.

Figure 33 shows a map plot of the ground stations used from the FSU dataset. Figure 34 shows the total hit rate for each station. In the absence of snow-free ground observations, the hit rate for the class *snow* equals the total hit rate for this dataset. A separate figure showing the hit rate for *snow* is therefore not included.

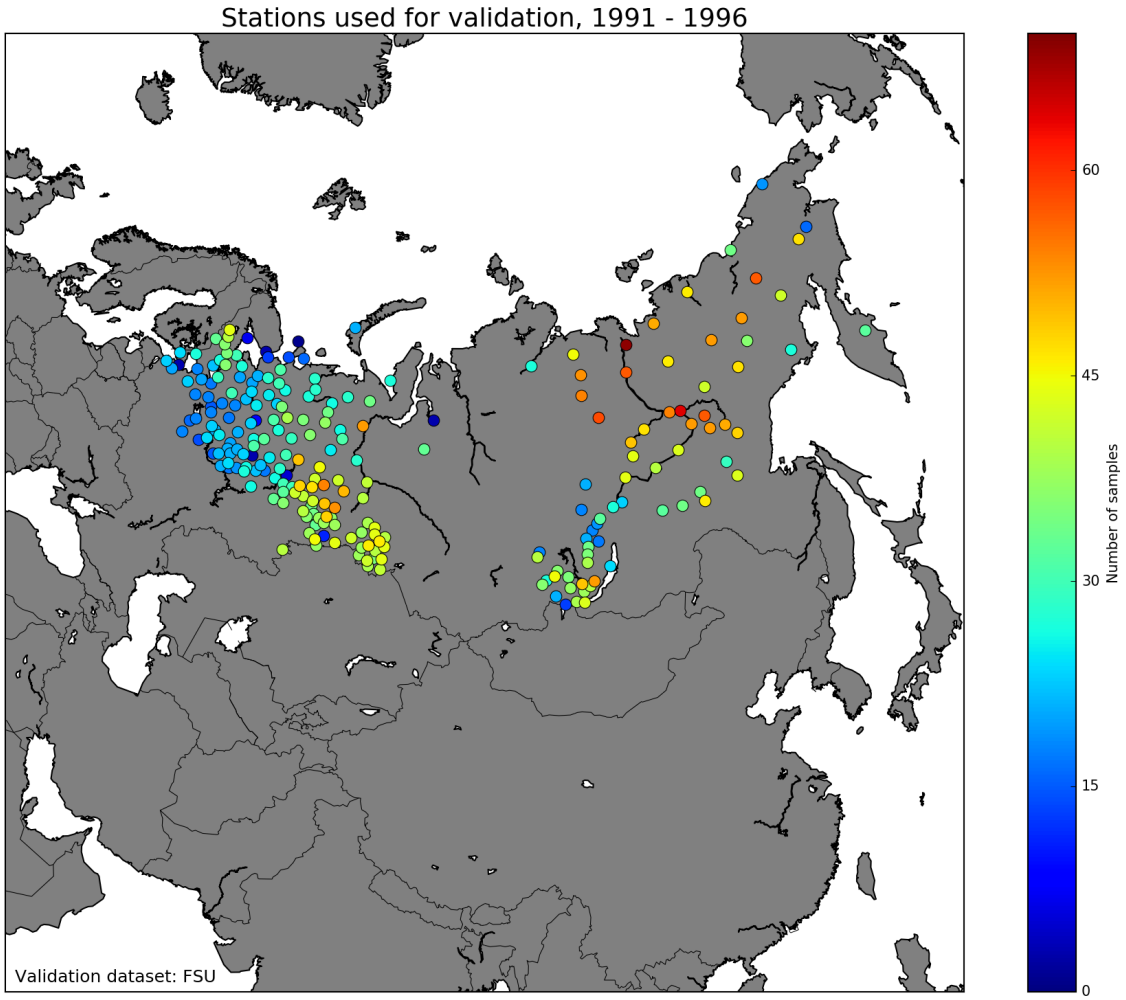


Figure 33: Ground stations used for validation against the FSU dataset. The colour of the station marker indicates the number of samples this station contributes with.

Figure 34 shows that the satellite product overall shows a very good match with the FSU snow extent data.

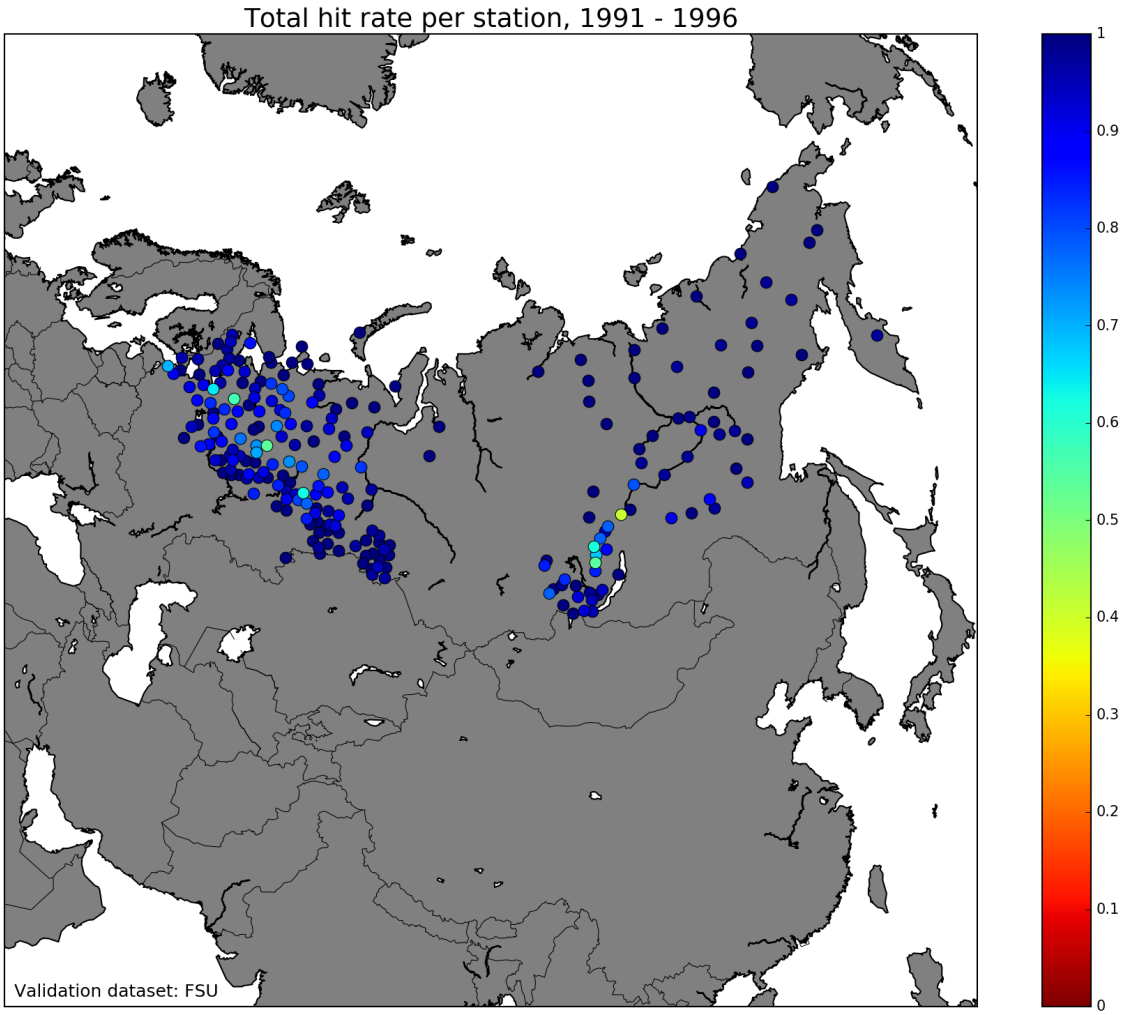


Figure 34: The total hit rate computed for each ground station separately. This would be the same as the hit rate for snow for this dataset.

5.7 Comparison of validation dataset results

Figure 35 shows the distribution of validation data per year and category (each element of the contingency table: hit, miss, false negative, and true positive, conf. Table 4). There is one panel for each of the four validation datasets. GHCN-D (upper left panel) is by far the largest of the validation datasets. It contributes to 11 million samples and is one of two which covers the entire satellite product period 1982 - 2015. The other is SCCONE (upper right panel) which provides 1.9 million samples. SCCONE has a larger relative amount of snow ground observations compared to GHCN-D (green + red bars in Figure 35, upper right). Data from GHCN-D is used for the entire Northern hemisphere, while SCCONE has data for the former USSR only (ref.: Figures 25 and 27).

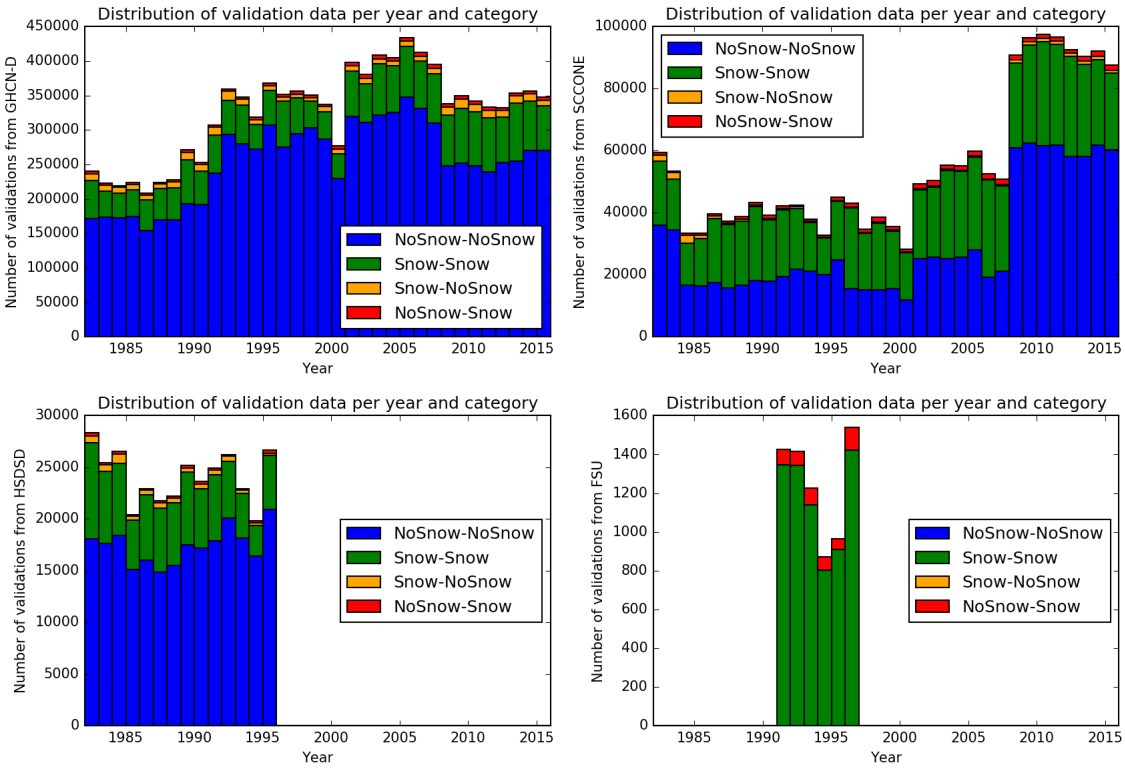


Figure 35: the figure shows the distribution of validation data per year and category for each of the four validation datasets used. GHCN-D is shown in the upper left panel, SCCONE in the upper right, HSDSD in the lower left, and FSU in the lower right.

Figure 36 shows the monthly values for total hit rate for GHCN-D (left) and SCCONE (right). For both cases there is one line for each year of the period 1982 - 2015. The figure shows that validation against SCCONE gives better results during wintertime (December - March) than validation against GHCN-D. From April through September the results are more or less the same.

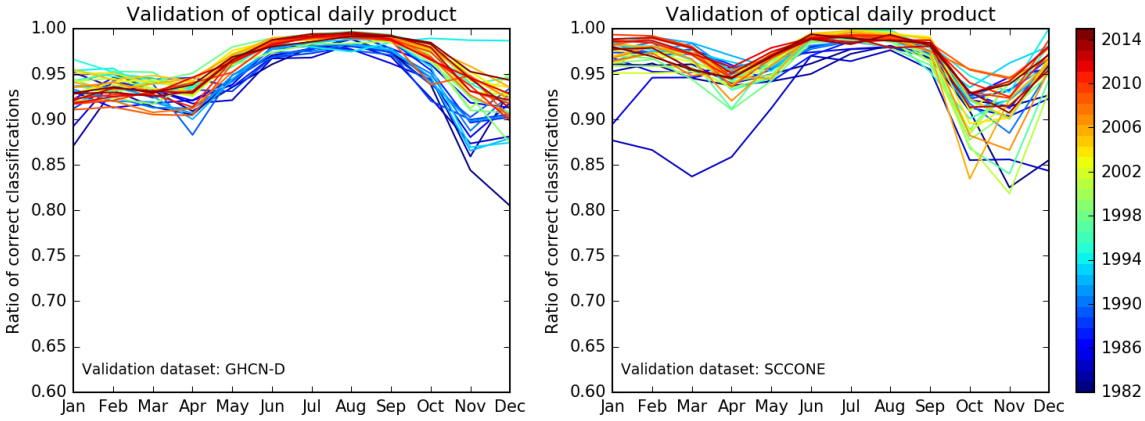


Figure 36: A comparison of monthly total hit rate when validating against GHCN-D (left) and SCCONE (right). There is one line for each year of the dataset period 1982 - 2015.

5.8 Low hit rate for snow during summer

This chapter focuses on the extraordinary low success rate of identifying snow cover during summer seen for GHCN-D and SCCONE (ref.: Tables 6 and 9). We select the year 2005, the month July and the validation source GHCN-D, and investigate the details of the comparisons. Table 14 shows the confusion matrix for this data. 54 000 data points are compared, and the overall accuracy is 99%. 38 of the 40 samples of observed snow depth above 5 cm are *not* seen in the satellite product. Output from the validation code shows that 3 individual ground stations are the source for all 38 cases of missed snow observations in July. These stations are listed in Table 15. Appendix B contains a table with the full list of missed snow observations during July 2005.

	GHCN-D: snow 7 (SD > 5 cm)	GHCN-D: no snow (SD = 0 cm)	Total
Satellite product: snow	2	447	449
Satellite product: no snow	38	53 548	53 586
Total	40	53 995	54 035
Hit rates	Hit rate snow: 5%	Hit rate land: 99%	Total hit rate: 99%

Table 14: Confusion matrix for July 2005 when validating the daily satellite product against snow depth values from GHCN-D.

Date	Station ID	latitude	longitude	elevation	name
200507X	AU000015410	47.05	12.95	3106 m	Sonnblick
200507X	GM000004155	47.42	10.99	2964 m	Zugspitze
200507X	SZ000002220	47.25	9.35	2502 m	Sàntis

Table 15: These 3 ground stations are the source of all cases of missed snow observations during July 2005.

All 3 stations are located well above 2000 m, in or nearby the Alps (Austria, Germany and Switzerland). Figure 37 shows a part of a NOAA-17 satellite swath from July 3 2005, 09:18 UTC. The panels in the upper row show RGB colour composites. The three stations are indicated by yellow circles. Snow cover can be seen in a strong pink color in the upper, right panel. The western parts of the Alps are cloud-free in this scene, and the pixels range from fully to partially snow-covered to snow-free. The Swiss station (Sàntis, the westernmost of the three) might suffer from algorithm shortcomings in identifying late spring snow cover. For all three stations, representation error is most probably the cause. The extent of the snow cover at the station location is not sufficiently large to dominate the 5 km satellite product grid cell.

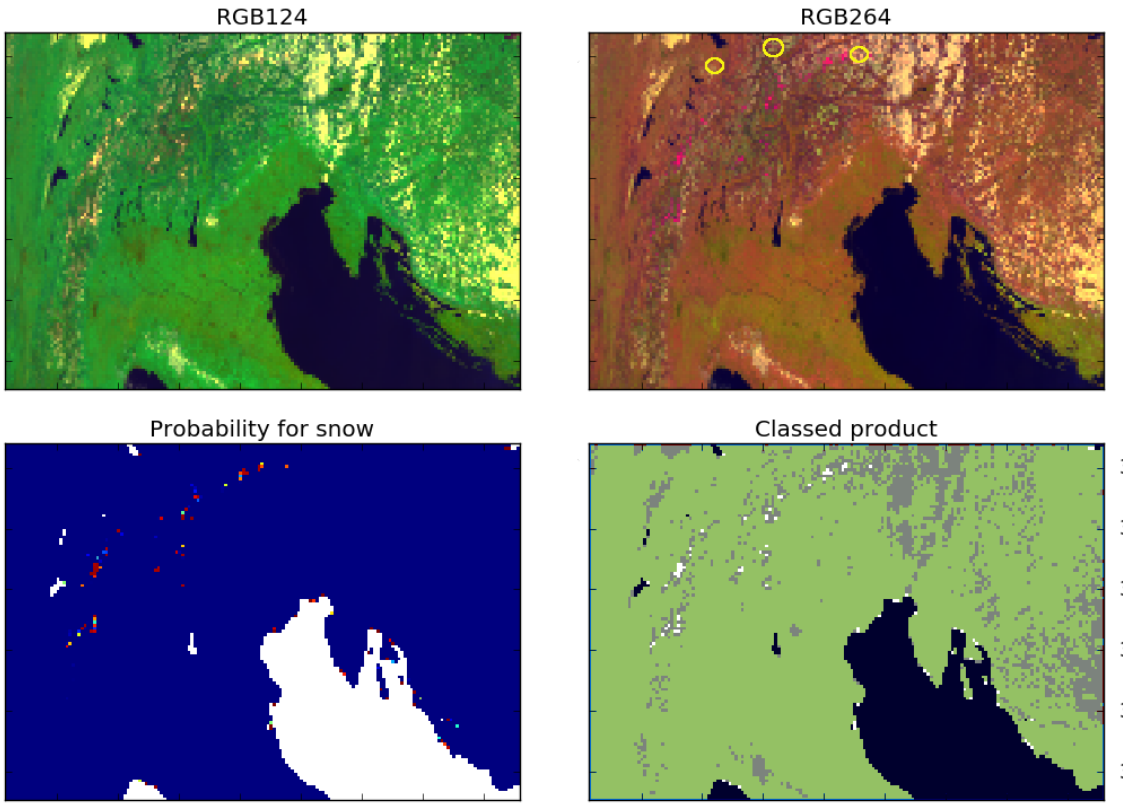


Figure 37: The figure shows a part of a NOAA-17 satellite swath from July 3 2005, 09:18 UTC. The upper row shows RGB colour composites of AVHRR channels 1, 2 and 4 (left) and 2, 3A (6) and 4 (right). The ground stations providing 100% of the missed snow observations for July 2005 are indicated by yellow circles. The lower left panel shows the probability for snow (dark blue refers to low probability, red indicates high probability). The lower right panel shows the classified swath product. Here, white shows snow, green shows snow-free land, black shows ocean/water, and grey indicates clouds.

Summary of the validation results

In this chapter we have described and discussed the validation work for the optical component. The full time series 1982 - 2015 of Northern hemisphere daily products has been validated. Four different validation datasets have been used. Two of these cover the entire time series, one of which contains full hemispheric snow depth data. The three other datasets contain snow cover observations from the former USSR. The validation results have been assessed in time and space. The overall results are very good. Total

accuracy range from 94% (FSU) to 97% (SCCONE, HSDSD). For hit rate *snow*, the results are in the range 94% (GHCN-D, FSU) to 97% (SCCONE, HSDSD). Total hit rate for *land* is 97% (GHCN-D, HSDSD) or 98% (SCCONE).

Looking at variation in total accuracy as function of time since the start of the time series one can see some improvements (conf. Figures 21 and 36). These improvements are likely due to the algorithm being trained using data from AVHRR/3 (first carried on NOAA-15 launched in May 1998), and the fact that there is much more satellite data available for the latter half of the dataset period (see Figure 3). Variation with season is prominent for some of the verification measures, in particular for hit rate for *snow* and for false alarm ratio. Some of these variations are likely due to difficulties at onset of the snow season and melting season, both algorithm challenges in terms of identifying thin snow or old, melting snow and difficulties because of scattered snow. The high false alarm ratio during summer, seen in particular for GHCN-D and SCCONE, is likely a direct result of representation error. This conclusion is based on results seen in Figure 20 and Table 7 which point to multiple examples of ground stations that are placed in a valley surrounded by mountains which can keep their snow cover well into summer. A parallel argument can be used to explain - at least partially - the very low hit rate for *snow* seen during summer for datasets GHCN-D and SCCONE, where ground stations at high mountains are not representative of the surrounding lower altitude areas.

The validation method is not optimal. We have not done any screening of stations based on their location. During the validation work we have seen repeated examples of disagreement between the in situ snow observation and the satellite product which very likely is a result of the location of the ground station.

6 Work package 6: Processing environment

Work package 6 includes updating processing chains, updating algorithms, implementation of new or improved algorithms, processing, and time series production. All algorithms and processing chains are now installed at MET. The work has been a joint effort between NR and MET.

6.1 Optical component

In Phase 2 we have downloaded the new input dataset and made the necessary adaptions needed in the swath processing code to handle the new format of the input files. As described in Chapter 2 various changes to the optical algorithm were implemented and tested. Furthermore we have replaced the former C code for aggregation of swath products into daily products with a new python code. The new code uses the python package `pyresample` and Gaussian resampling. Figure 11 shows a comparison of an optical daily product using the previous gridding routine and the new gridding routine. The main differences are 1) that the many undefined pixels near water (yellow dots seen in the left panel) found when using the C gridding disappears, and 2) that the edges of clouds (grey colour in right panel) are more smoothed.

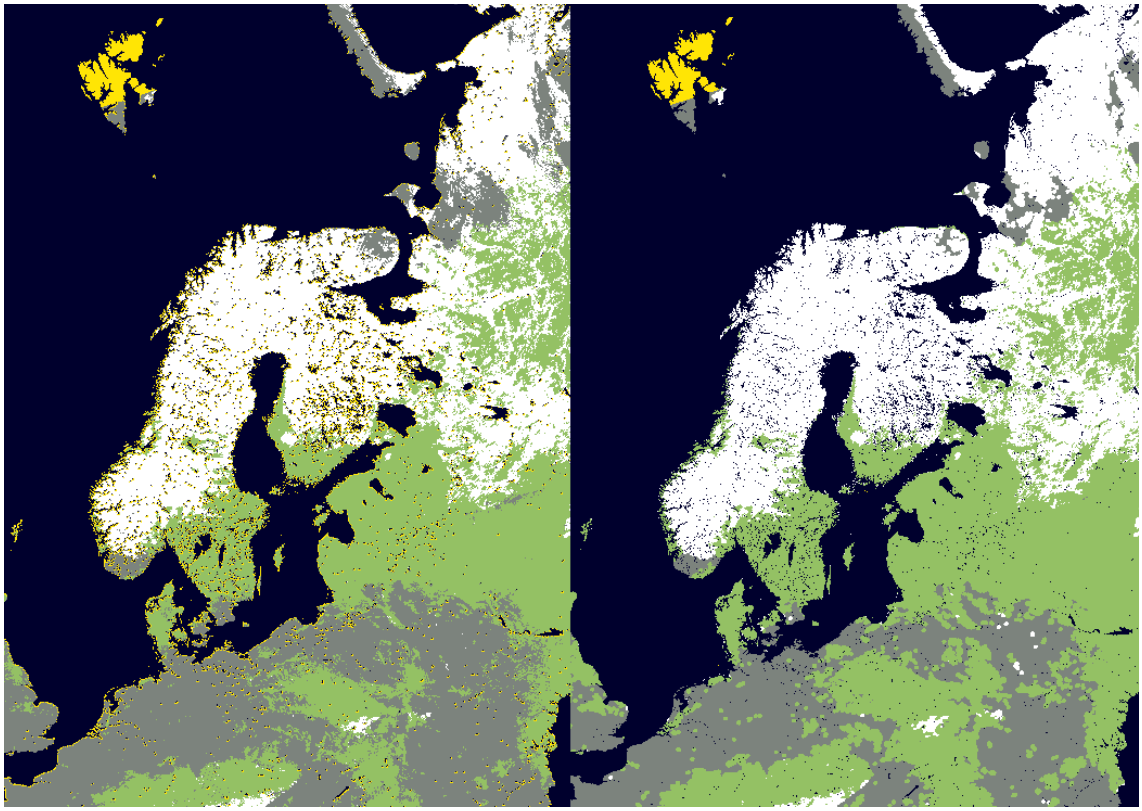


Figure 38: The figure shows a part of the optical, daily, gridded product from March 15 2005. The left panel shows the product when gridding with the C code used in CryoClim and in Phase 1 of Sentinel4CryoClim. The right panel shows the result using the new python code for gridding. Snow-free land is shown in green, snow-covered areas in white, clouds are grey, ocean/water is black, and areas of no data (due to insufficient sunlight or other reason) are shown in yellow.

6.2 PMW component

In CryoClim and Sentinel4CryoClim phase 1, the source for PMW data was a SMMR and SSM/I data record delivered by Remote Sensing System (RSS). The data covers 1978 – 2009. Since then, EUMETSAT CM SAF has released a FCDR for PMW data from SMMR, SSM/I and SSMIS covering 1978 to 2015 (Fenning et al., 2017) In cooperation with NR, the PMW code was rewritten to handle the new input data files collected from EUMETSAT CM SAF, and implemented at MET.

The new PMW data was first processed using the statistical coefficients derived in CryoClim for the PMW data from RSS. The results showed a reduction in performance compared to the CryoClim results, making it necessary to collect new training data and derive new coefficients. Using the new set of coefficients, the PMW swath data was processed again. Due to various issues with several of the SSM/I instruments, a range of considerations had to be made for the time series processing. Details can be found in the project report delivered by NR (Solberg et al., 2018).

6.3 HMM component

NR is responsible for the HMM algorithm development, and has implemented the HMM chain at MET. Following updates/improvements of the algorithm, NR implemented the new version at MET. The algorithm takes daily PMW envi files and daily AVHRR netCDF files as input and can be run when a full season of input data (August - July) is ready. MET has produced the full multi-sensor time series.

6.4 Southern Hemisphere

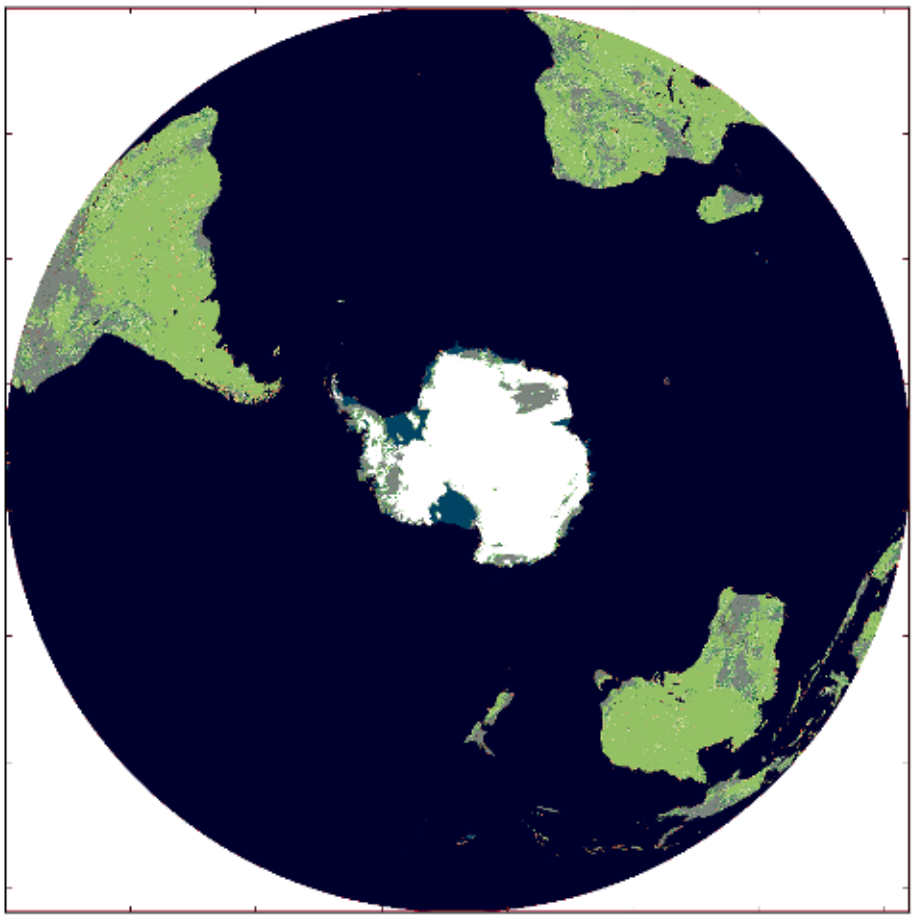


Figure 39: Example of Southern Hemisphere optical daily snow cover product. Snow is shown in white, snow-free land in green and clouds in grey. Areas of no data is seen in dark blue (here indicating sea ice shelves outside of Antarctica).

At the very end of the project the first version of daily, gridded products for the Southern Hemisphere was generated. Little time has been available to assess the results, and no proper validation has been performed. Figure 39 shows a Southern Hemisphere daily optical snow cover product.

7 Discussion and summary

In this project MET has focused on improvements of the optical component (algorithm improvements and adding an uncertainty estimate), extensive validation of the optical product, and time series processing. In addition the optical chain has been adapted to and tested for SLSTR data. The results seen so far for SLSTR are promising even though some simplifications were made on the way.

The work on optical algorithm improvements is described in Chapter 2 . The main issue we attempted to fix was the discrepancy between model surface temperature and true surface temperature which appears due to the coarse resolution of the global model surface temperature data. Should there be a gap between model and true temperature for a snow-covered pixel the algorithm tends to believe that the pixel is cloud-covered even when it is in fact cloud-free. Several approaches were tested to reduce the gap, but in the end we achieved only limited improvements. This is not a closed topic, but definitely in need of future work. Height correction should be revisited using proper interpolation, but even so it is not obvious that this would give the necessary accuracy in the model surface temperature for each swath pixel. The satellite swath pixel size is large, and steep terrain is very difficult.

The validation work performed in Work Package 5 has been a very valuable exercise. The full time series has been validated, and the validation work has been repeated using four different validation dataset. Although the overall validation results are very good (above 95% for total accuracy), we have seen that during summertime snow-covered ground stations are often not identified as snow-covered in the satellite product (low hit rate for *snow*). We have also seen a high false alarm ratio for the summer months. A common factor is steep terrain, and it is likely that (a part of) the reason is representation error.

The geographical plotting of verification measures has been particularly helpful, providing easy focus on areas of reduced performance. There has been no prior screening of the ground stations due to station location and surrounding terrain, even though in many cases the local topography will not be representative for a 5 km satellite product grid cell. Disagreements between the satellite product and the in situ observation are bound to happen. Validation results also confirm our previous belief that the spring time snow cover is somewhat underestimated.

At the very end of the project products for the Southern hemisphere was produced. The Southern hemisphere daily gridded optical products have not been validated. A few products have been assessed.

Appendix A

This appendix contains map plots of validation results (see Chapter 5.3) for the years 1985, 1995 and 2015. The validation source is GHCN-D (see Chapter 5.1).

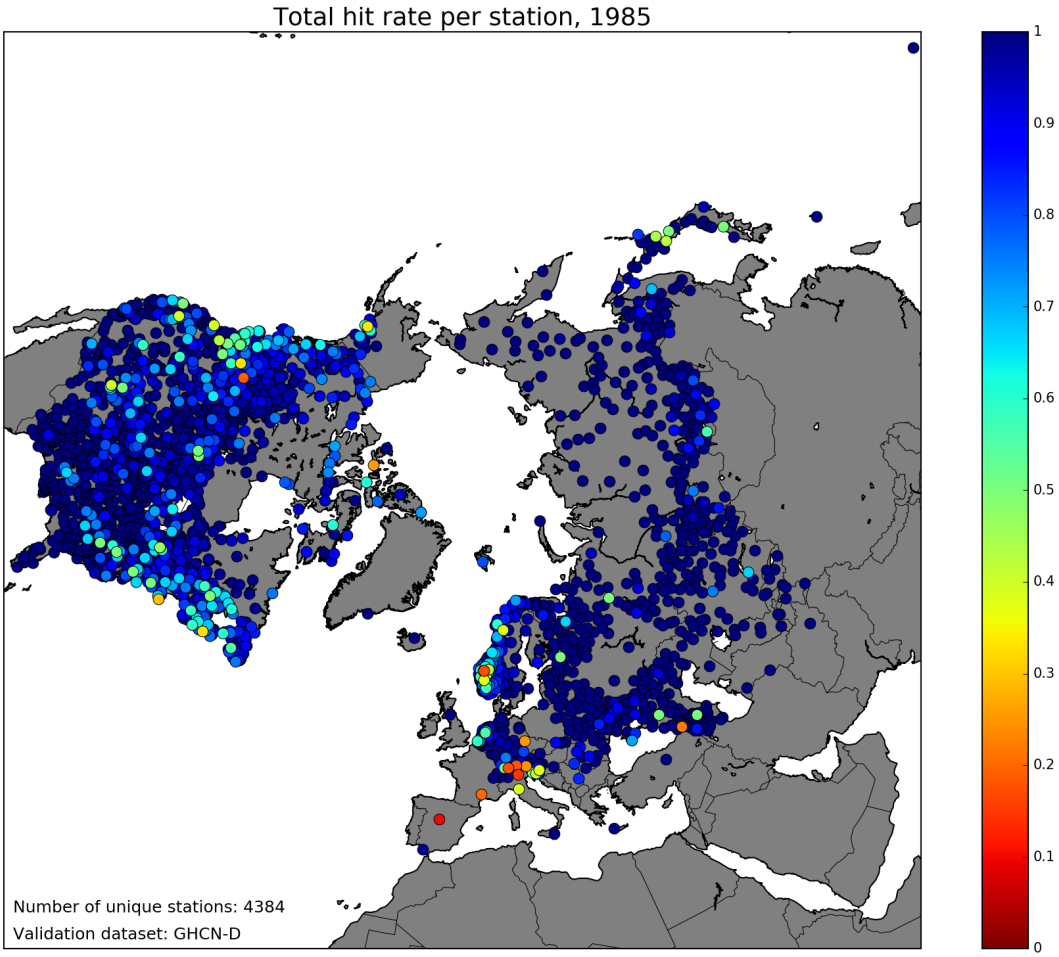


Figure 40: Total hit rate per ground station when validating against GHCN-D for the year 1985. Each dot represents a ground station. The colour of the dot gives the hit rate for the ground station.

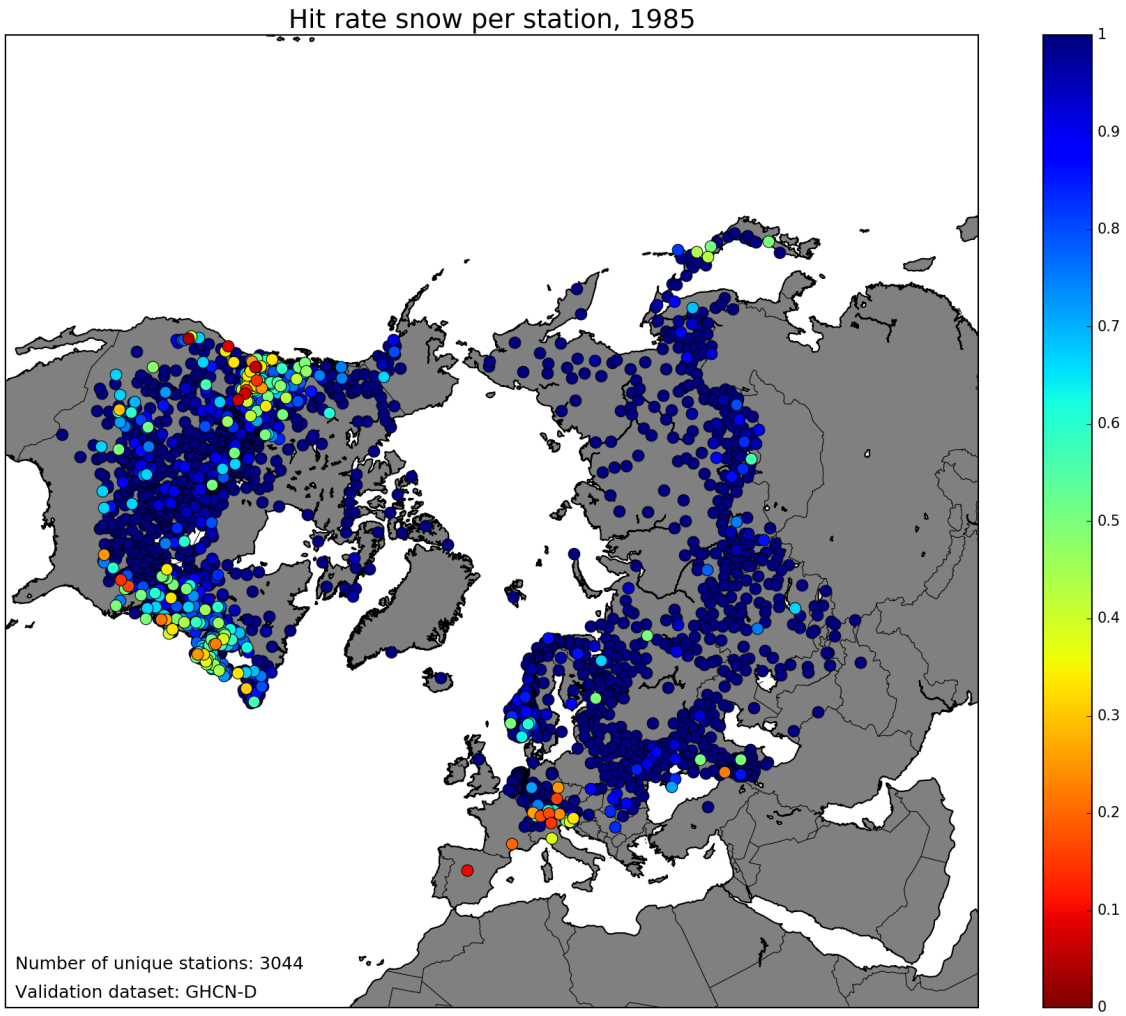


Figure 41: Hit rate for the class snow when validating against GHCN-D for the year 1985. Each dot represents a ground station. The colour of the dot gives the hit rate. Note that stations with no observation of snow during 1985 is not plotted. This figure therefore shows 3044 unique stations, while Figure 40 shows 4384 unique stations.

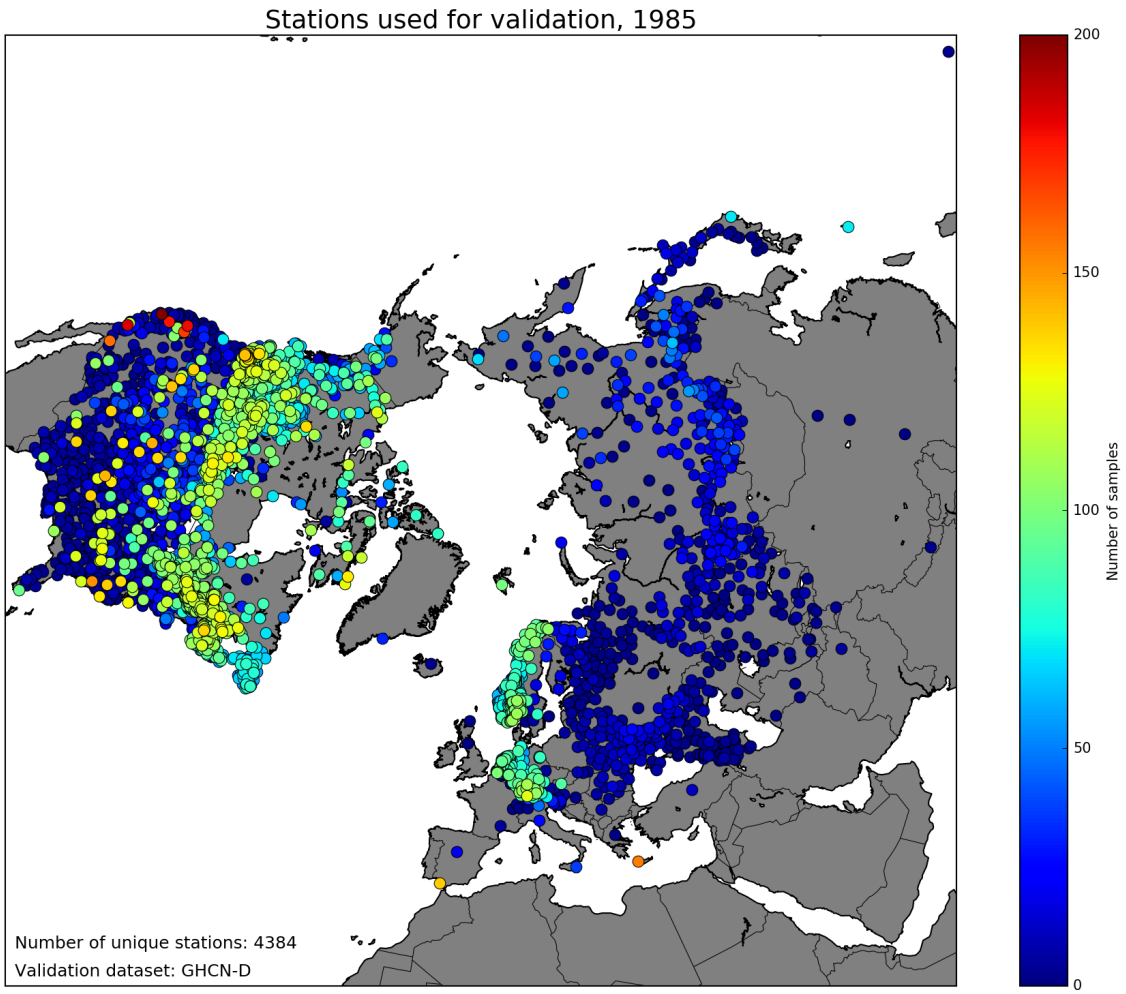


Figure 42: Number of samples from each ground station for the year 1985.

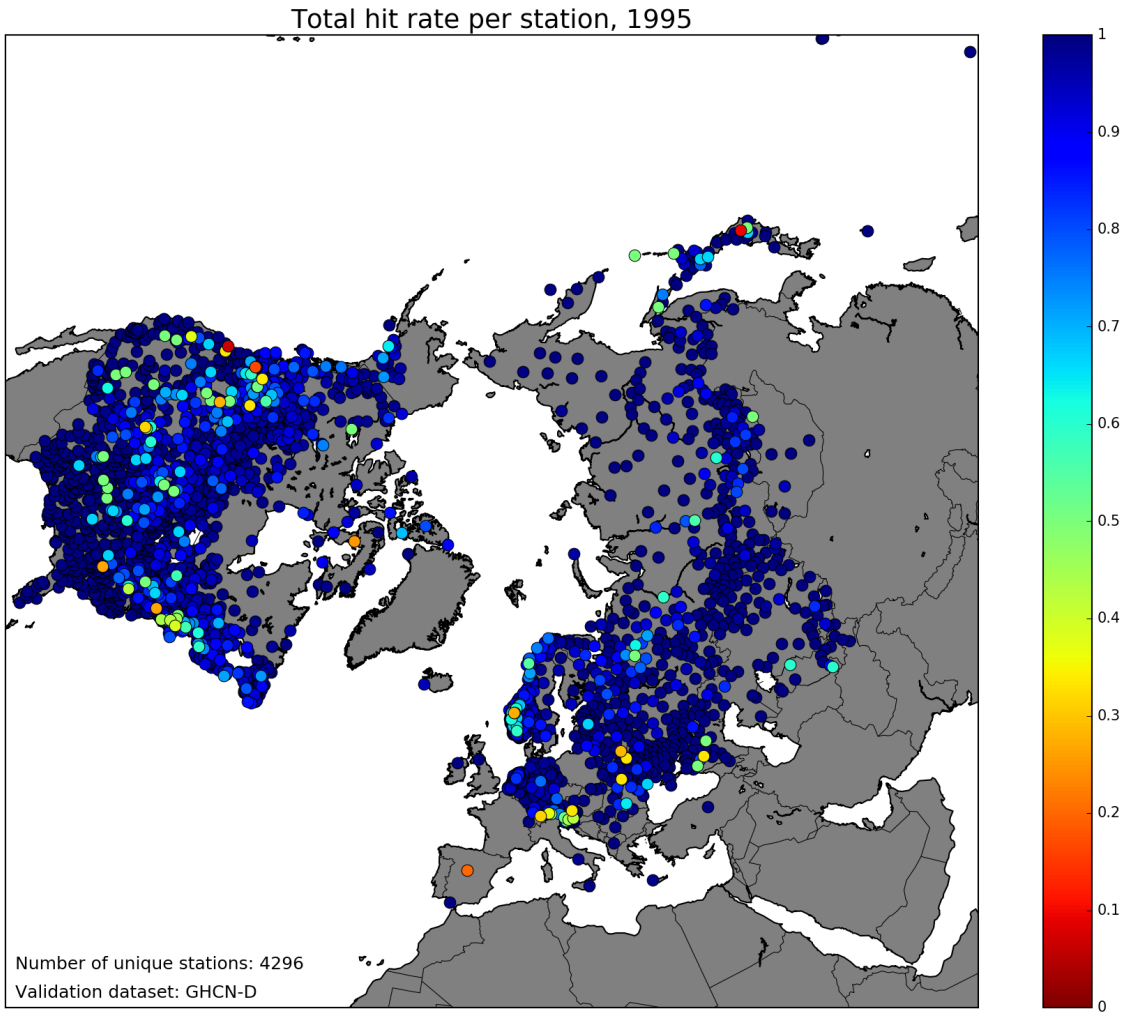


Figure 43: Total hit rate per ground station when validating against GHCN-D for the year 1995. Each dot represents a ground station. The colour of the dot gives the hit rate for the ground station.

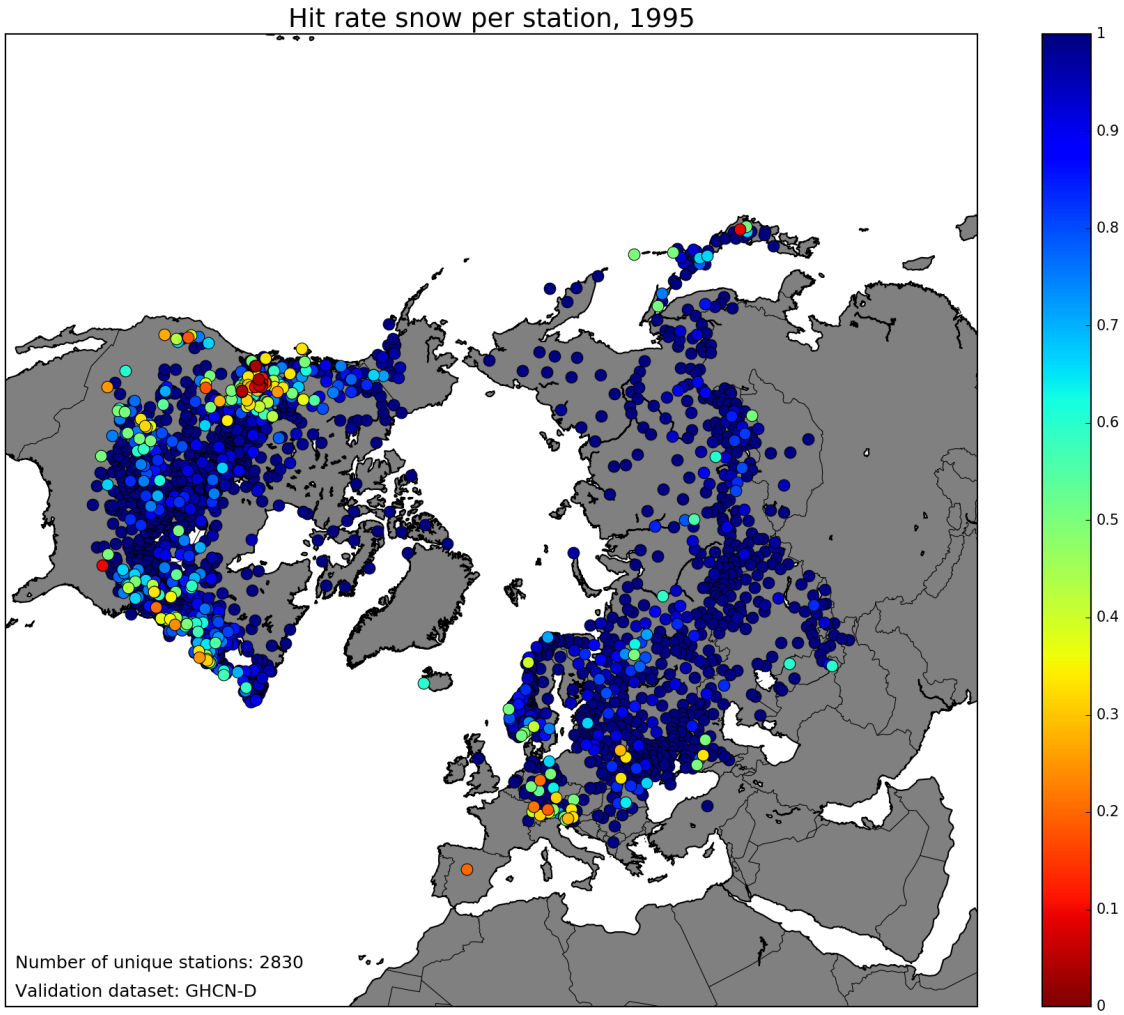


Figure 44: Hit rate for the class snow when validating against GHCN-D for the year 1995. Each dot represents a ground station. The colour of the dot gives the hit rate. Note that stations with no observation of snow during 1985 is not plotted. This figure therefore shows 2830 unique stations, while Figure 36 shows 4384 unique stations.

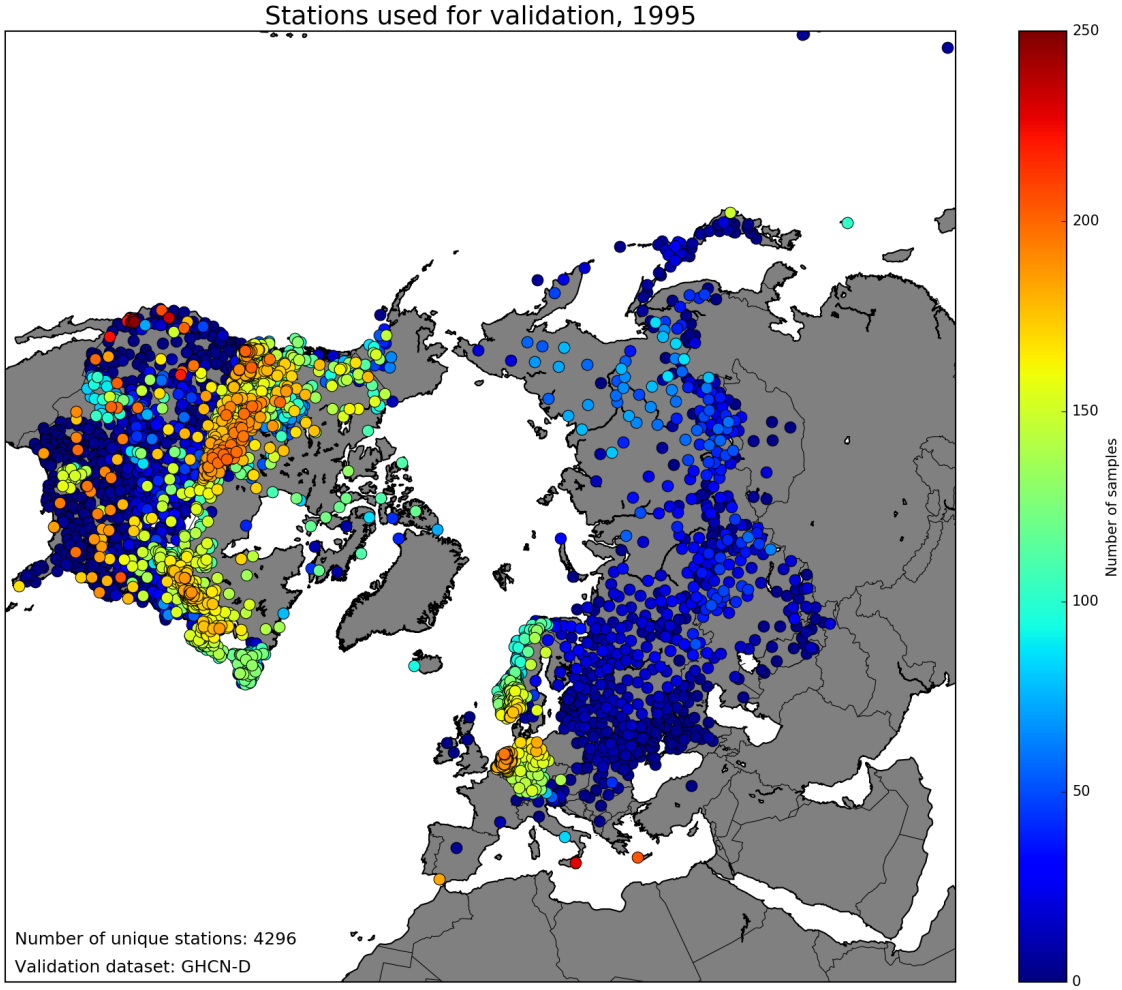


Figure 45: Number of samples from each ground station for the year 1995.

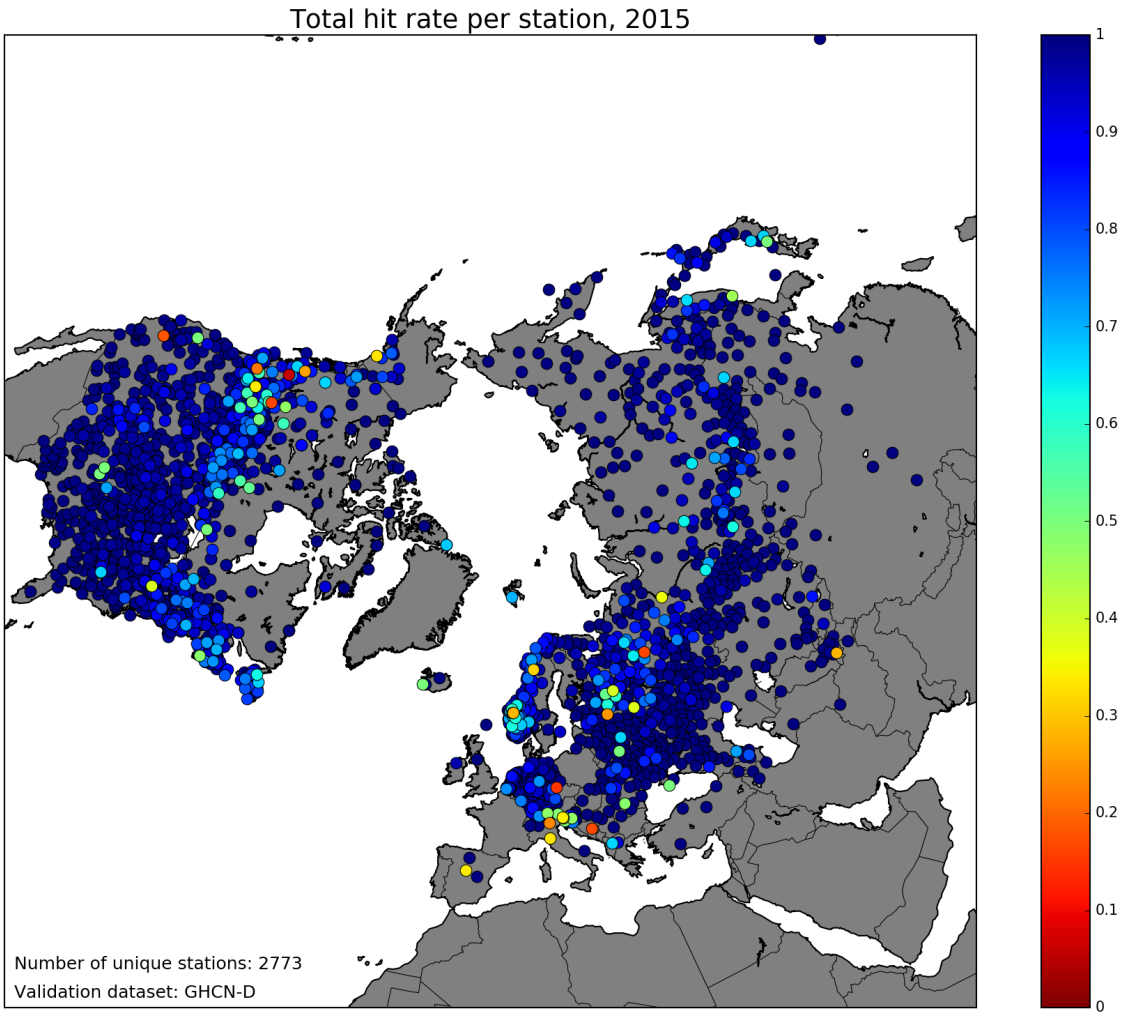


Figure 46: Total hit rate per ground station when validating against GHCN-D for the year 2015. Each dot represents a ground station. The colour of the dot gives the hit rate for the ground station.

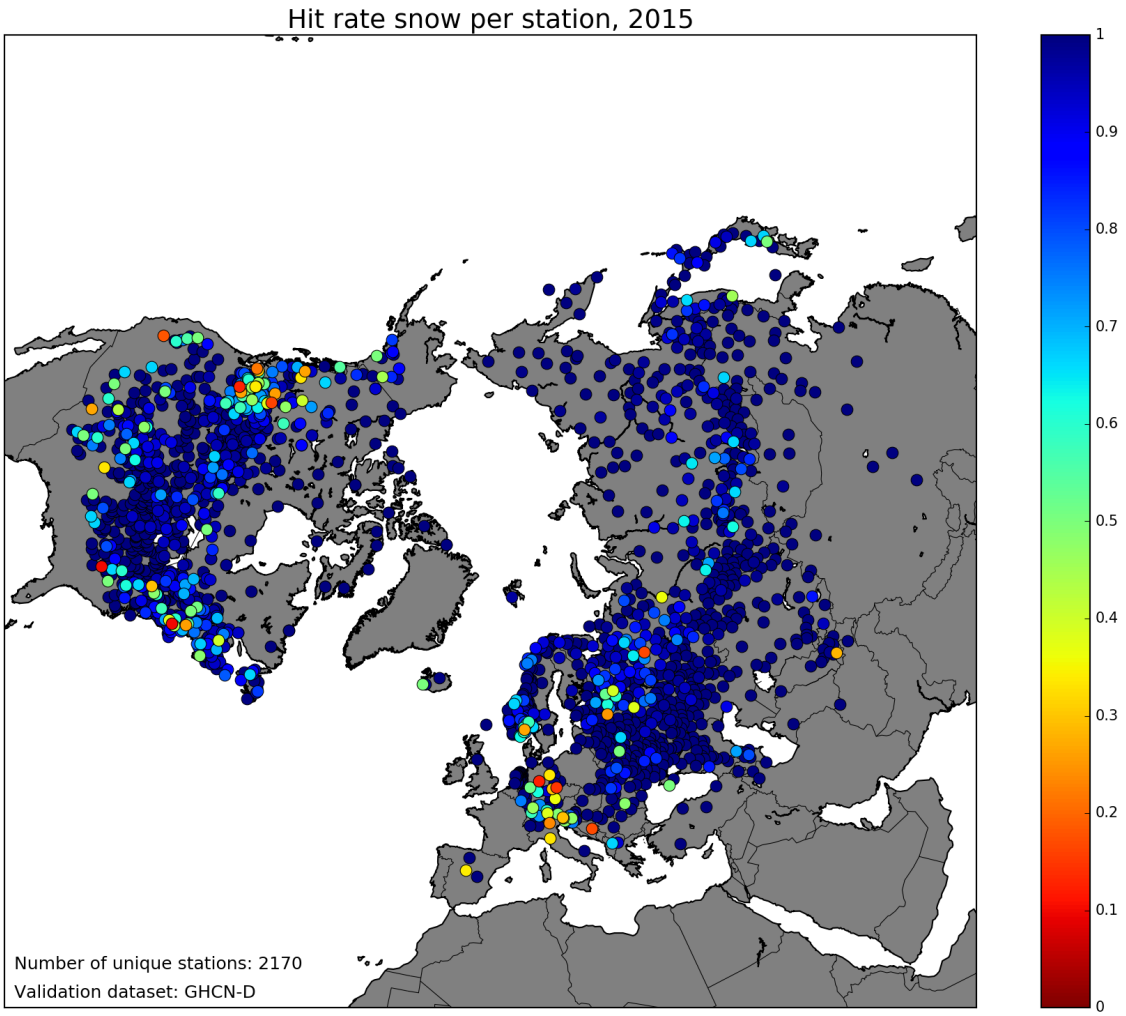


Figure 47: Hit rate for the class snow when validating against GHCN-D for the year 1985. Each dot represents a ground station. The colour of the dot gives the hit rate. Note that stations with no observation of snow during 1985 is not plotted. This figure therefore shows 3044 unique stations, while Figure 36 shows 4384 unique stations.

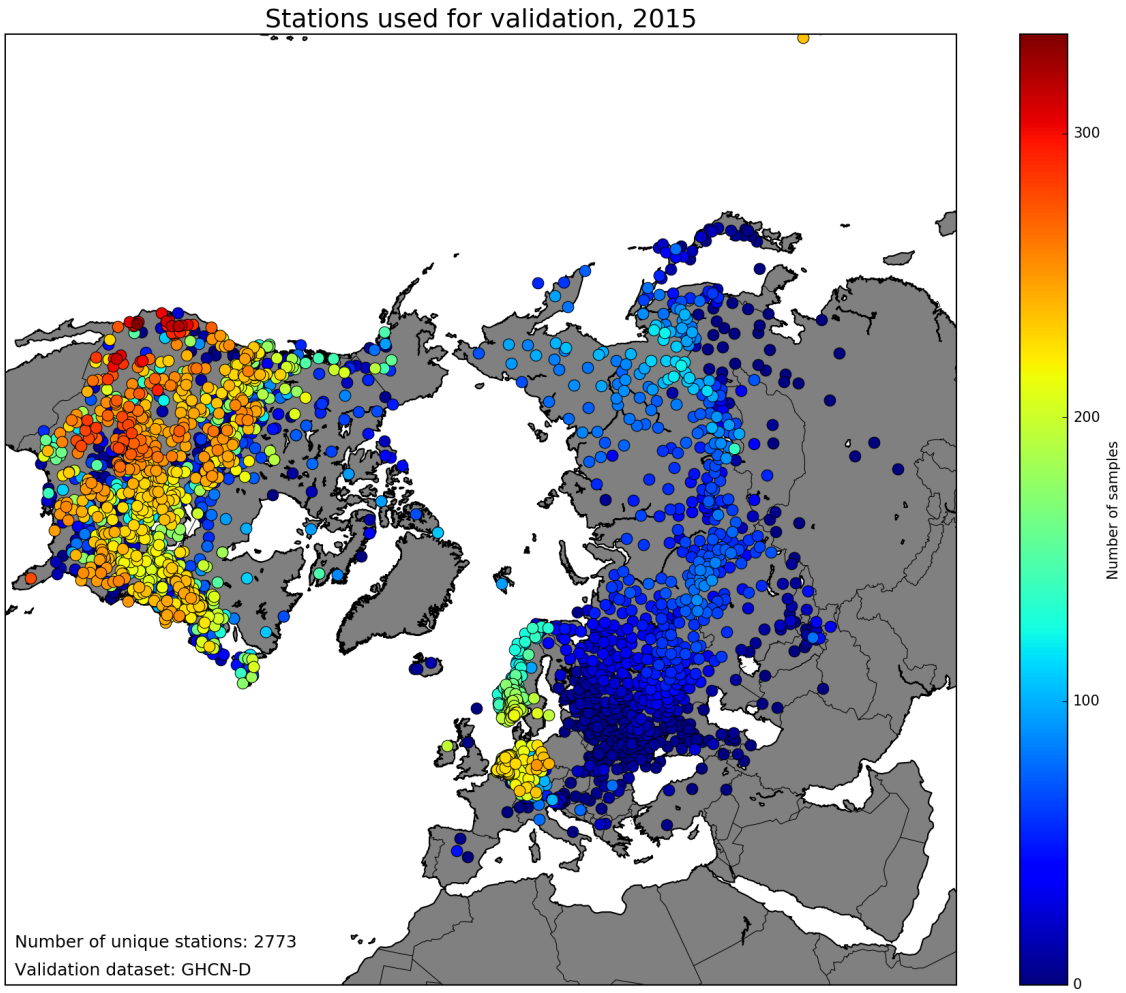


Figure 48: Number of samples from each ground station for the year 2015.

Appendix B

The following table lists the validation samples for July 2005 and validation source GHCN-D. As described in Chapter 5, the hit rate for snow during summer is very low, in particular when validating against GHCN-D and SCCONE. A closer look at July 2005 revealed that all cases of missed snow observations could be attributed to one of three ground stations. These are described in 5.8. The full list of samples for July 2005 is included in the table below.

Date	Station ID	latitude	longitude	sample type
20050703	AU000015410	47.05	12.95	NOSNOW_SNOW
20050704	AU000015410	47.05	12.95	NOSNOW_SNOW
20050713	AU000015410	47.05	12.95	NOSNOW_SNOW
20050714	AU000015410	47.05	12.95	NOSNOW_SNOW
20050715	AU000015410	47.05	12.95	NOSNOW_SNOW
20050717	AU000015410	47.05	12.95	NOSNOW_SNOW
20050718	AU000015410	47.05	12.95	NOSNOW_SNOW
20050720	AU000015410	47.05	12.95	NOSNOW_SNOW
20050721	AU000015410	47.05	12.95	NOSNOW_SNOW
20050727	AU000015410	47.05	12.95	NOSNOW_SNOW
20050728	AU000015410	47.05	12.95	NOSNOW_SNOW
20050729	AU000015410	47.05	12.95	NOSNOW_SNOW
20050730	AU000015410	47.05	12.95	NOSNOW_SNOW

20050703	GM000004155	47.42	10.99	NOSNOW_SNOW
20050704	GM000004155	47.42	10.99	NOSNOW_SNOW
20050706	GM000004155	47.42	10.99	NOSNOW_SNOW
20050713	GM000004155	47.42	10.99	NOSNOW_SNOW
20050714	GM000004155	47.42	10.99	NOSNOW_SNOW
20050715	GM000004155	47.42	10.99	NOSNOW_SNOW
20050717	GM000004155	47.42	10.99	NOSNOW_SNOW
20050719	GM000004155	47.42	10.99	NOSNOW_SNOW
20050720	GM000004155	47.42	10.99	NOSNOW_SNOW
20050727	GM000004155	47.42	10.99	NOSNOW_SNOW
20050728	GM000004155	47.42	10.99	NOSNOW_SNOW
20050729	GM000004155	47.42	10.99	NOSNOW_SNOW
20050703	SZ000002220	47.25	9.35	NOSNOW_SNOW
20050704	SZ000002220	47.25	9.35	NOSNOW_SNOW
20050706	SZ000002220	47.25	9.35	NOSNOW_SNOW
20050708	SZ000002220	47.25	9.35	NOSNOW_SNOW
20050711	SZ000002220	47.25	9.35	NOSNOW_SNOW
20050712	SZ000002220	47.25	9.35	NOSNOW_SNOW
20050713	SZ000002220	47.25	9.35	NOSNOW_SNOW
20050714	SZ000002220	47.25	9.35	NOSNOW_SNOW
20050715	SZ000002220	47.25	9.35	NOSNOW_SNOW
20050717	SZ000002220	47.25	9.35	NOSNOW_SNOW
20050718	SZ000002220	47.25	9.35	NOSNOW_SNOW
20050720	SZ000002220	47.25	9.35	NOSNOW_SNOW
20050724	SZ000002220	47.25	9.35	NOSNOW_SNOW

Table 16: The table shows all samples of category "nosnow_snow" for July 2005 when validating against GHCN-D data.

Acknowledgements

The CLARA-A2 level 1c dataset (Karlsson et al., 2017) is kindly provided by the EUMETSAT CM SAF. The dataset was collected and made available to this project with the help of the NORMAP project.

We thank NR for the fruitful cooperation on this project, and in particular for sharing code and validation datasets as part of Work Package 5.

References

Dee, D.P., Uppala, S.M., Simmons, A.J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M.A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A.C.M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A.J., Haimberger, L., Healy, S.B., Hersbach, H., Hólm, E.V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A.P., Monge-Sanz, B.M., Morcrette, J.J., Park, B.K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.N. and Vitart, F. (2011) The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Q J Roy Meteor Soc*, 137, 553–597. https://doi.org/10.1002/qj.828.

Fennig, Karsten; Schröder, Marc; Hollmann, Rainer (2017): Fundamental Climate Data Record of Microwave Imager Radiances, Edition 3, Satellite Application Facility on Climate Monitoring, DOI:10.5676/EUM_SAF_CM/FCDR_MWI/V003, https://doi.org/10.5676/EUM_SAF_CM/FCDR_MWI/V003.

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

Karlsson, K.G., Timo Hanschmann, Martin Stengel, Jan Fokke Meirink 2016 EUMETSAT Satellite Application Facility on Climate Monitoring Algorithm Theoretical Basis Document: CM SAF Cloud, Albedo, Radiation data record, AVHRR-based, Edition 2 (CLARA-A2) Cloud Products (level-1 to level-3)

Karlsson, Karl-Göran; Anttila, Kati; Trentmann, Jörg; Stengel, Martin; Meirink, Jan Fokke; Devasthale, Abhay; Hanschmann, Timo; Kothe, Steffen; Jääskeläinen, Emmihenna; Sedlar, Joseph; Benas, Nikos; van Zadelhoff, Gerd-Jan; Schlundt,

Cornelia; Stein, Diana; Finkensieper, Stephan; Håkansson, Nina; Hollmann, Rainer; Fuchs, Petra; Werscheck, Martin (2017): CLARA-A2: CM SAF cLoud, Albedo and surface RAdiation dataset from AVHRR data - Edition 2, Satellite Application Facility on Climate Monitoring, DOI:10.5676/EUM_SAF_CM/CLARA_AVHRR/V002, https://doi.org/10.5676/EUM_SAF_CM/CLARA_AVHRR/V002.

Killie, M. A., Eastwood, S., Solberg, R., Rudjord, Ø. and Salberg, A. B., 2013 Snow sub-service development: Report on CryoClim WP 3 progress and results. Project deliverables no. D3.1-D3.10, CryoClim Phase 4, ESA/NSC PRODEX. Norwegian Computing Center and Norwegian Meteorological Insittute.

Solberg, R., Rudjord, Ø., Salberg, A. B., Killie, M. A., Eastwood, S. and Breivik, L. A., 2017 Advancement of global snow mapping in CryoClim. Sentinel4CryoClim Phase 1, Deliverables 1-6.

Solberg R., Rudjord, Ø., Salberg, A. B., and Reksten, J. H., 2018 Further advancement of global snow mapping in CryoClim. Sentinel4CryoClim Phase 2, Deliverables 1, 3-7.

Warren, S. G., 1982 Optical property of snow. *Journal of Geophysical Research: Space Physics*, 20(1), 67–89.

Glossary

ATBD	Algorithm Theoretical Basis Document
AVHRR	Advanced Very High Resolution Radiometer
CM SAF	Climate Monitoring Satellite Application Facility
EUMETSAT	The European Organisation for the Exploitation of Meteorological Satellites
FCDR	Fundamental Climate Data Record
GAC	Global Area Coverage
GHCN-D	Global Historical Climatology Network - Daily
HMM	Hidden Markov Model
LAC	Local Area Coverage
NOAA	National Oceanic and Atmospheric Administration
OLCI	Ocean and Land Colour Imager
PMW	Passive MicroWave
SCCONE	Snow Cover Characteristics Over Northern Eurasia
SLSTR	Sea Land Surface Temperature Radiometer
TOA	Top Of Atmosphere
WMO	World Meteorological Organization