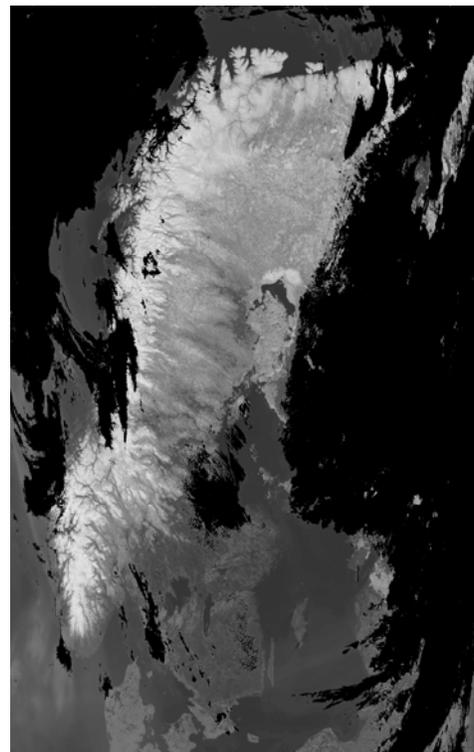
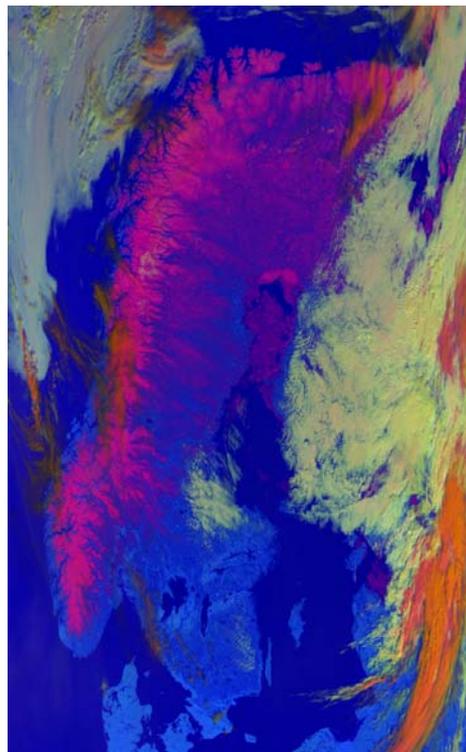


Cloud detection in MODIS images



Note no

SAMBA/28/09

Authors

Hans Koren

Date

August 2009

Norsk Regnesentral

Norsk Regnesentral (Norwegian Computing Center, NR) is a private, independent, non-profit foundation established in 1952. NR carries out contract research and development projects in the areas of information and communication technology and applied statistical modelling. The clients are a broad range of industrial, commercial and public service organizations in the national as well as the international market. Our scientific and technical capabilities are further developed in co-operation with The Research Council of Norway and key customers. The results of our projects may take the form of reports, software, prototypes, and short courses. A proof of the confidence and appreciation our clients have for us is given by the fact that most of our new contracts are signed with previous customers.

Title	Cloud detection in MODIS images
Authors	Hans Koren
Date	August
Year	2009
Publication number	SAMBA/28/09

Abstract

One significant problem when studying satellite images produced from optical sensors is the presence of clouds. The surface of the earth can be seen only in areas which are not obscured by clouds. Therefore, it is important to be able to detect the clouds in the images. Norwegian Computing Center has for several years produced maps of snow cover based on satellite images from optical sensors, NOAA AVHRR and Terra MODIS. It can be difficult to discriminate clouds from snow in the visual part of the spectrum. With infrared bands included, we have succeeded in creating an effective automatic algorithm for cloud detection. In this report there is a description of the theory behind the algorithm and also a description of the training procedure for MODIS images.

The results of cloud detection are evaluated and compared with the MODIS MOD35_L2 cloud mask produced by NASA. In general our algorithm gives better results than MOD35, and especially over snow.

Keywords	Cloud detection, snow, satellite images, MODIS
Target group	Snow hydrology, climatology, meteorology
Availability	Open
Project number	54
Research field	Earth observation
Number of pages	34
© Copyright	Norsk Regnesentral

Contents

1	Introduction	7
2	Methods	7
2.1	Theory.....	7
2.2	Creating the labelled codebook.....	9
2.3	Vector Quantization Codebook	9
2.4	Training procedure and data sets.....	10
3	Results	12
3.1	Comparing different codebooks	12
3.1.1	Discussion.....	13
3.2	Comparing NR products with MODIS MOD35_L2.....	21
3.2.1	Discussion.....	22
4	Conclusions.....	31
5	References.....	32
	Appendix	33

1 Introduction

At NR we have been studying satellite images for many years. Although we have had projects using radar satellites, most of the work has been done on images from optical sensors. A significant problem using optical sensors is the presence of clouds. Radar signals penetrate the clouds, but optical signals are stopped. The surface of the earth can be studied from satellites by optical sensors only in areas which are not obscured by clouds. Also, shadows of clouds will reduce the incoming and reflected light. Therefore, it is important to have available methods to find the clouds in the images. In satellite images of high spatial resolution (< 100 m) it is usually easy to detect the clouds visually. For lower resolution it can be more difficult, especially with snow present. Normally one will try to detect the clouds automatically. In areas without snow the clouds will separate well from the darker ground in the visual wavelength region. In areas with snow, both the clouds and the snow will have high reflection in the visual region and it may be difficult to distinguish the two. In other parts of the spectrum, the reflectance of clouds and snow will be different. To be able to separate clouds from bare land, water, snow and ice one has to use algorithms where both visual and thermal infrared wavelength bands are included.

For many years NR has produced maps showing the daily snow extent on land in the Nordic countries, and Norway especially. We started using NOAA AVHRR images of 1 km resolution and developed a method for cloud detection in these images. In 2000 we started using MODIS images because of the 250 m resolution and free data (Solberg et al. 2006). In addition to the 36 bands MODIS images, a cloud product, MOD35_L2 also was delivered on the web (http://modis-atmos.gsfc.nasa.gov/MOD35_L2/index.html). We tried to use this product, but found quite soon that it was far from perfect, especially over snow and near the snow borders. Therefore, we developed a cloud detection algorithm which has been gradually improved during the following years until the present version from 2007. MOD35_L2 has also been improved during the years, and the present version is far better than the first one. This algorithm should find clouds over all parts of the earth at all times of the year, with and without snow. One can not expect it to be perfect at all conditions. We have tried to make an algorithm specialised for Norway/Scandinavia in the snowmelt season (March – July), and we expect it to be better than MOD35_L2 under these conditions. Comparisons of the two algorithms can be found in later chapters.

2 Methods

2.1 Theory

It is a great challenge to detect clouds automatically in satellite images, especially over snow. NR has experimented with several approaches, and the current best cloud detection algorithm is based on k Nearest Neighbour (k-NN) classification of MODIS data. In a k-NN classifier a pixel, represented by a vector of band values, is assigned the

label, which represents the most prevalent among the k nearest labelled vectors from a reference set. A k -NN classifier is an asymptotically optimum (Maximum Likelihood) classifier as the size of the reference set increases (Duda et al. 2001).

The classifier has been trained based on a set of partially cloudy images acquired through snowmelt seasons. For each image the bands 1, 4, 6, 19, 20, 26 and 31 are being used. Band 1 is available with a spatial resolution of 250 m, band 4 and 6 with 500 m and the other bands with a resolution of 1 km. Band 1, 4 and 6 are used in versions aggregated to 1 km resolution, and this is the resolution of the resulting cloud detection mask. In addition, information about the sun angle is collected from the MODIS metadata file. The wavelengths of the used bands can be found in Table 1.

Band	Bandwidth (μm)
1	0.620 – 0.670
4	0.545 – 0.565
6	1.628 – 1.652
19	0.915 – 0.965
20	3.660 – 3.840
26	1.360 – 1.390
31	10.780 – 11.280

Table 1 MODIS bands used in cloud classification

The seven bands applied, cover wavelengths from 0.545 μm to 11.280 μm . The bands are chosen to get useful information for distinguishing clouds from snow, land and water. Band 1 and 4 are the red and green bands, which separates snow and clouds (together) from land and water. Band 6 contains information discriminating snow and clouds. For this band clouds have high, and snow low reflectance. Band 19 is sensitive to atmospheric water vapour. Band 20 and especially 31 contain information about the temperature of the land surface and the clouds. Band 26 is sensitive to water vapour and cirrus clouds. NASA uses 19 bands to make the MODIS cloud mask product (MOD35_L2), but using the seven bands mentioned seems to be more suitable for this regional application. Note that the MODIS products are developed for global applications and had to be optimised for this purpose.

In order to obtain good performance of the k -NN classifier, and since we have no prior information for weighting (as to which channels are the most important), we have chosen to use band variance normalization. A statistical analysis is performed over the training set, yielding means and variances of the image bands. The statistics is made for pixels where the sun elevation is higher than a certain specified angle. 10 degrees has been used as elevation limit.

The k -NN based (cloud) classification system consists of the following main elements:

- A set of tools used to create a labelled spectral reference set or codebook, where each spectral vector in the codebook has a unique spectral shape and represents an instance of the corresponding class label.

- A k-NN based classifier. Classification of a satellite image (cloud detection) is done for each pixel by determining the most prevalent label among the k nearest neighbours in the codebook.

A somewhat heterogeneous processing chain has been developed for creating the labelled codebook. The processing chain consists of functionality implemented in both IDL/ENVI, Matlab and also some compiled C routines. The resulting labelled codebook is written in a machine independent “raw” format and may be used by a k-NN classifier on any platform. Currently the k-NN classifier is implemented both in Matlab, IDL, and as C/ .exe.

2.2 Creating the labelled codebook

The creation of a codebook consists of the following stages (for MODIS):

- IDL/ENVI: Perform conversion to ENVI file format. In addition to the 36 original channels, two channels are added (37 and 38) with sun positions (azimuth and elevation) for each pixel. The sun position is used for exclusion of pixels having sun elevation below the given limit, and for scaling of the visual bands.
- Matlab: A vector quantization codebook (not to be confused with the final labelled codebook) is created for each individual image. Using a k-means similar clustering approach a set of representative spectral vectors (cluster centres) is found for each individual training set image. The image may then be approximated by an index image addressing the codebook vectors. This is similar to traditional lookup-tables for colour imaging.
- Matlab: A labelling tool is used to visually identify the class and then assign labels to most or all of the quantization codebook vectors. The visual identification is facilitated both by a reduced dimension spectral representation (a selection of a combination of the original bands is used to create a three-channel RGB image) and the context information naturally observed by the operator in the representation of the image.
- Matlab: A classification codebook is finally obtained by combining the labelled quantization codebooks from a set of training images into a single (classification) codebook. This combination stage currently only consists of a re-clustering of the spectral representation vectors for each class over all training images. The combination is performed in order to reduce the final k-NN classification computational load.

2.3 Vector Quantization Codebook

A vector quantization codebook has been made for each training image, performing the following steps:

1. Correct for error pixels (found by statistical analysis)
2. Correct visual bands by scaling for sun angle
3. Compute standard deviation for each band
4. Normalize each band (scale to equal variance)

5. Cluster image vectors into specified number of clusters using a particular type of k-means named Myers Clustering (due to the initialization stage where extreme outliers are sought). This initialization method was first presented by Wayne L. Myers (Myers and Patil, 2006)
6. Rescale the cluster centres (removing the previous variance normalization)

We have used 1500 clusters. Each cluster is represented by a 7-dimensional vector. In the current version we have used 65 MODIS images as input to statistical analysis. The images are specified in the Appendix.

Codebooks have been created with and without correction for sun angle. In the current version sun correction is not included.

In a configuration file the number of nearest neighbours can be specified. In the tests made in this study the number has been set to 4.

2.4 Training procedure and data sets

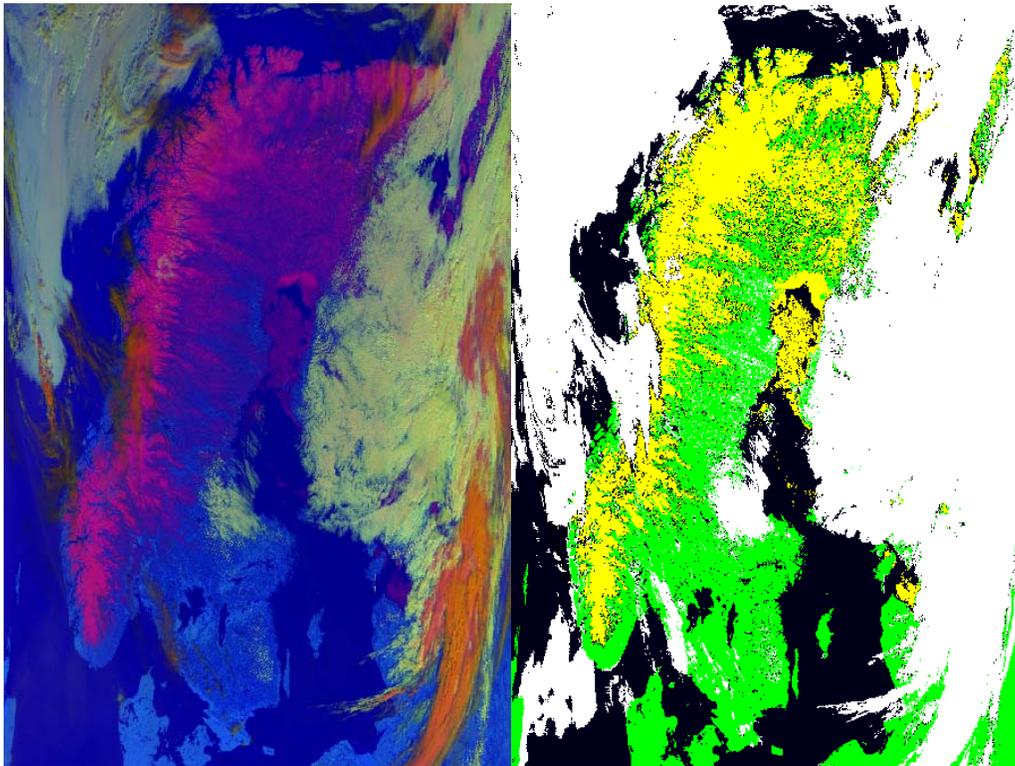
A number of training images have been labelled manually. Each pixel should, if possible, be given a class label. In our initial reference set we use the following classes: cloud, land, ocean and snow. Reference vectors are extracted using a manually controlled spectral-distance based region-growing procedure. The procedure enables an accurate positioning of the spectral transition between different classes by utilizing the operator's ability to interpret both the pixel context and the pixel colour.

The pixel "colour" in this case is the RGB image obtained through a transform of each pixel vector. The colours RGB are assigned to the bands 1, 6 and 31, respectively (see Figure 3.4). This choice of colours shows ocean and land in blue, snow and ice in red/violet, and clouds in nuances of green/brown/beige. The clouds are in most cases easily separated visually from the other classes. A tool is developed to ease the manual labelling of images, a procedure which typically takes a couple of hours per image.

The labelled pixels change colour during the process, cloud – white, snow/ice – yellow, ocean – black, land – green, and multi labelled – red (see Figure 1). The operator selects a pixel with the cursor and selects a class. All pixels within the same cluster will then get the colour of this class. The operator can then select a region-growing threshold value. Then all pixels with cluster vectors spectrally close to the initial cluster vector will get the same colour. The operator can see if all these pixels seem to belong to the same class and label them, or one can extend or reduce the threshold value to extend or reduce the coloured regions before classification. Then the operator moves the cursor to another unclassified pixel and repeats the operations. Experience shows that one should not classify too large regions at the same time. Even if all pixels within the region close to your selected pixel obviously belong to the same class, there may be regions at other places in the image which are on the border to other classes (mostly snow/ice). If you label these as cloud, you may later during classification of snow label the same pixels as snow. The pixels will then be labelled to more than one class. These pixels will be marked red, and the operator can later make corrections and relabel them. In the best case, all pixels should be labelled, but there will always be doubts, and the operator can

leave pixels unlabelled. Unlabelled and multi-labelled pixels (cluster vectors) will not be included in the training set.

For each image a codebook is created, showing the relations between cluster vectors and classes. With a number of training images and 1500 cluster vectors per image the reference set is too large to carry out a cloud classification of a MODIS image in reasonable time. The final reference set size is reduced to a manageable size using standard vector-quantization (k-Means). A total of 500 representation vectors were used for each class.



Figur 1 Intermediate steps in the cloud codebook development. Left: MODIS image before labelling. Right: Labelled image. The image is from 17 April 2003.

In the current version a number of 44 images have been labelled. These are from dates between 4 February and 23 July from the years 2000 – 2006. We are trying to estimate the snow cover in the mountainous areas outside forests. The conditions of snow and temperatures vary throughout the season. In February and most of March there is much snow and low temperatures, so most of the snow is dry. From the end of March until the end of May there is still plenty of snow, but the snow is getting wet. In June and July there is less snow, and it is wet.

One could suspect that the classification of clouds near the snow could be improved by making a special codebook for each part of the season, using training images only from a corresponding time interval. We have made a simple experiment by dividing the number of training images into three and made specified codebooks for the early,

medium and late part of the season. For the early part we have used 10 images of dates between 4 February and 13 April. For the medium part we have used 23 images taken between 16 April and 12 June, and for the late part we have used 11 images taken between 17 June and 23 July. The images used are specified in the Appendix. In addition we have made a codebook covering the complete season where all the 44 labelled images have been included. In our case we have no overlap between the three seasonal codebooks. A more refined version of seasonal codebooks could use more images in each book, and also allow for using some of the same images in two adjacent codebooks.

3 Results

We have tested the cloud classification on a number of images. We have used all four codebooks to see if the results differ. The classification has been done on images which are included in the codebooks and also other images from dates between February and September. We have found that the latest version gives better results than former versions of our classifier. The latest version has more test images, and the images have a better time coverage of the melting season than the former versions. We have also compared the results with the latest available version of the MODIS cloud product, MOD35_L2, version 5. To find the 'correct' cloud mask, we have inspected the MODIS images visually. In many cases it is quite easy to see the clouds. In other cases, especially with small clouds close to snow it is difficult to discriminate snow and clouds. In several cases we have used corresponding Landsat images to separate the clouds from the snow.

3.1 Comparing different codebooks

The following figures show the cloud classification for a series of MODIS images taken from February till August. Each image has been classified with the four codebooks to see which codebook gives the best result in each case, and to see which codebook should be used in the different parts of the season. The results are shown in images showing the southern part of Norway, corrected to UTM zone 33 projection. The classified snow cover is also included in the images to show how the cloud classification works over and near the snow. The clouds are shown in dark gray, bare ground in medium gray, and snow in white and nuances of lighter gray. The sea is black. A water mask has been added, which excludes the clouds over the sea. Inland lakes are not included in the water mask. In addition band 1 of the original MODIS image is shown in the same projection. The codebooks are named '*early*', '*medium*', '*late*', and '*total*'.

2002.02.25 (Figure 2)

The MODIS image shows almost full snow cover in the mountains and practically no clouds over the western part. The snow is probably cold and dry. As expected, *early* reflects this situation well, and so does *total*. *Medium* shows some extra clouds over the snow, and *late* covers nearly the complete snow covered area with clouds

2003.04.18 (Figure 3)

The MODIS image shows very few clouds. The snow cover is nearly full in the higher areas, but in the lower areas there is only fractional snow cover. The snow is wet in the lower regions and moist high up in the mountains. *Early* shows some extra clouds at the border of the snow, and also over the snow in the north. *Late* shows few clouds over the highest areas, but many of the areas with fractional snow cover are classified as clouds. *Medium* and *total* gives the best results.

2000.05.04 (Figure 4)

The MODIS image shows some cloudy areas in the north and a group of clouds outside the snow near the southern end of Norway. *Early* and *late* have some extra clouds along the borders of the snow, *early* also over the snow. *Medium* and *total* gives very good results over the snow. All codebooks detect the clouds in the south. The cloudy areas in the north east are classified differently with the four codebooks. It is not easy to tell which is best.

2003.06.02 (Figure 5)

Early shows too many clouds over the snow. Only the highest areas are without clouds. All codebooks detect the cloudy areas well. There are only minor differences in the classifications.

2005.07.11 (Figure 6)

Early shows too many clouds over the snow. *Late* is best regarding clouds outside the snow. But there are no large differences between *medium*, *late* and *total*.

2003.08.09 (Figure 7)

Early shows no clouds over the snow. This seems to be correct. *Medium* has nearly the same result. *Total* shows some clouds close to the borders of the snow, while *late* has many clouds over the snow. For clouds outside the snow, *late* is clearly best, followed by *total*, *medium* and *early*. The result for clouds outside the snow could be expected. For the snow areas, it is quite unexpected that *early* should give so much better result than *late*, as the image is taken late in the season, in fact after the period 4 February – 23 July of labelled test images.

The summer 2003 was very special. Almost all snow from the last winter had melted. Some old snow was still present on the upper part of glaciers. However, the reflectance was low in all regions. The temperature the 9 August was high, and the snow was wet. These conditions should favour the *late* codebook. Still, the best cloud classification is done by the *early* codebook.

3.1.1 Discussion

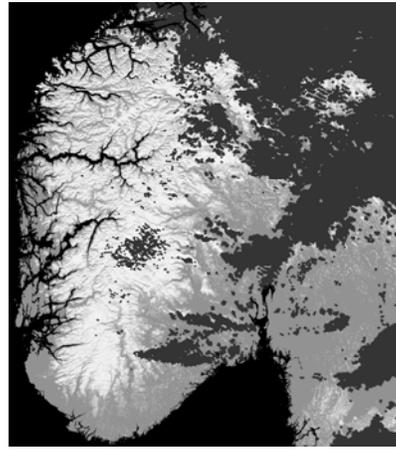
From the shown examples and some more tests, one can conclude that broadly speaking, the *early* codebook is best for the winter season (February, March), *medium* is best for the spring (April, May), and *late* and *medium* are about equal in the summer (June, July), with *late* somewhat better on clouds outside the snow. The example from August is special. One should usually not expect the early codebook to be best in the late season. During the entire period from February till August, the *total* codebook has about equal or even better results than the best of the seasonal codebooks. It can be difficult to determine the end of one season and the start of the next. This will differ

between the years. As the *total* codebook is close to best during all seasons, we have chosen to use this codebook for the whole time period in our production line for snow cover estimation.

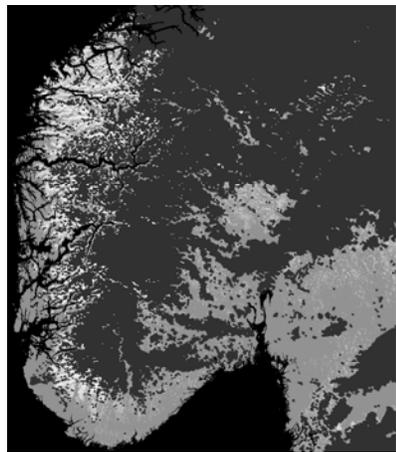
From the shown examples and other tests, the *total* codebook gives very good results for detection of clouds over and near snow. There are very few extra clouds classified, and the existing clouds seem to be found. Small clouds outside the snow are sometimes not found. The result is that these are being classified as snow. The training set could need more images with this type of clouds. Another way to avoid such misclassifications is to introduce contextual information. In the lower parts of Scandinavia there is usually no snow from the beginning of May, and the objects are probably clouds.



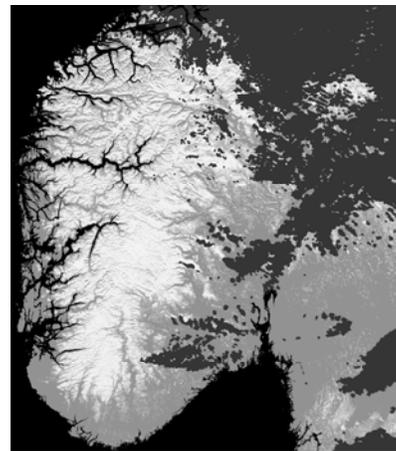
Early



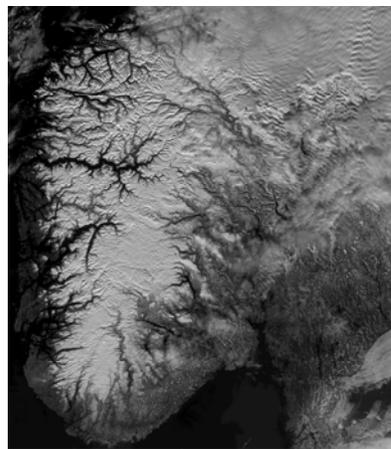
Medium



Late



Total

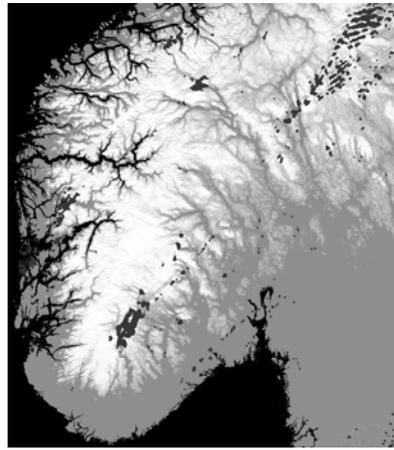


MODIS band 1

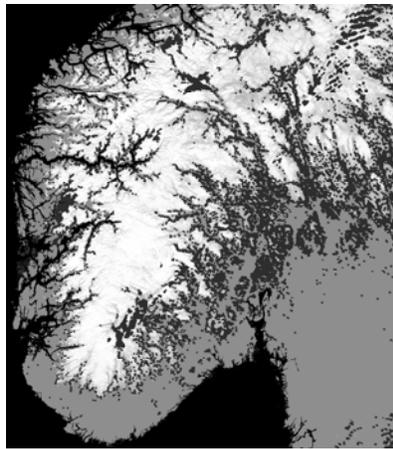
Figure 2 Snow cover and cloud classification of MODIS image from 2005.02.25 11:05



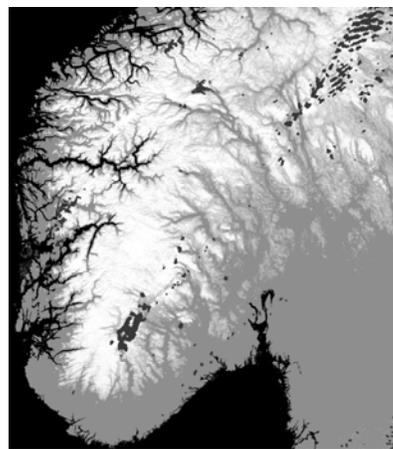
Early



Medium



Late

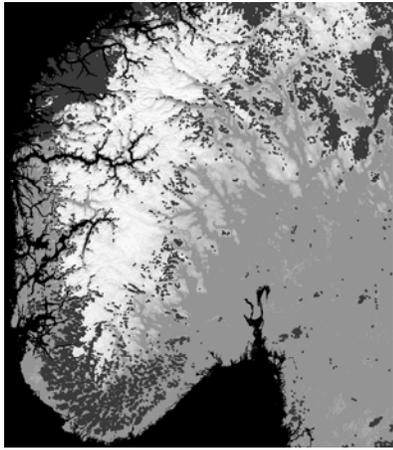


Total

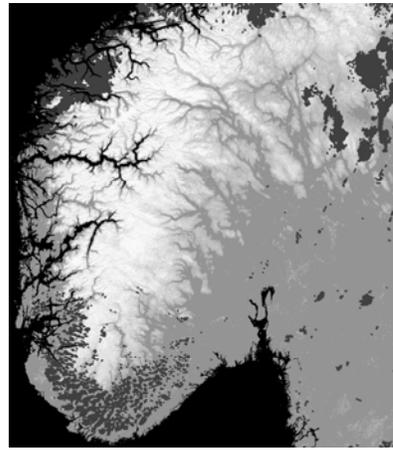


MODIS band 1

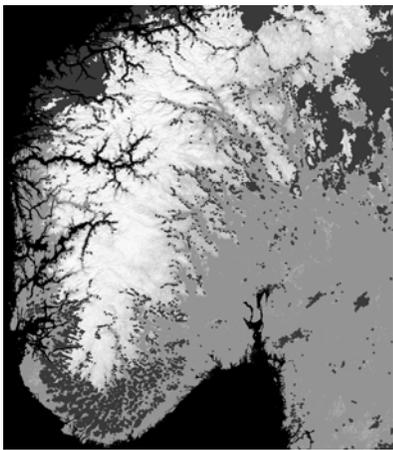
Figure 3 Snow cover and cloud classification of MODIS image from 2003.04.18 11:00



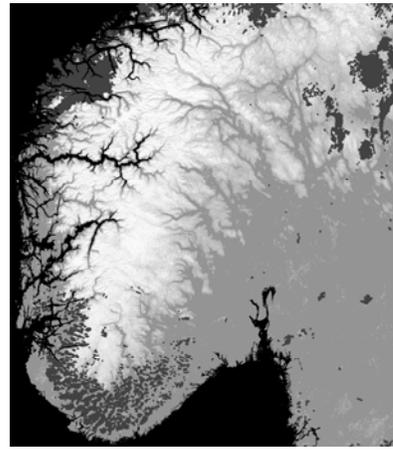
Early



Medium



Late

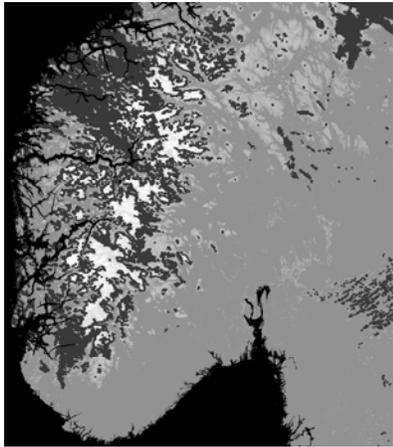


Total



MODIS band 1

Figure 4 Snow cover and cloud classification of MODIS image from 2000.05.04 11:10



Early



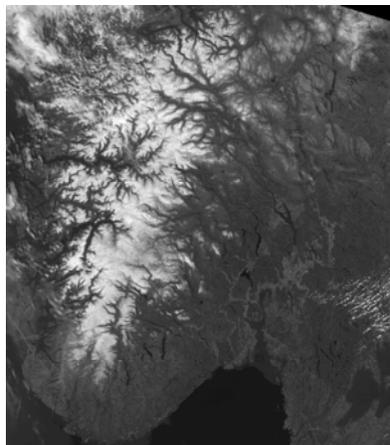
Medium



Late



Total

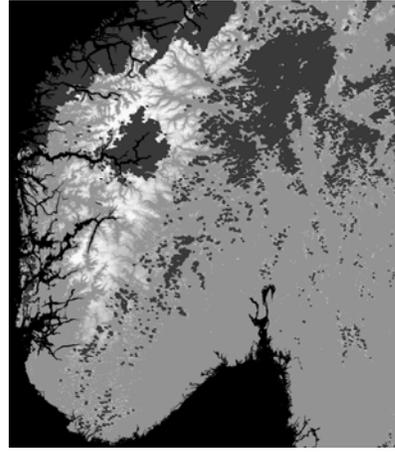


MODIS band 1

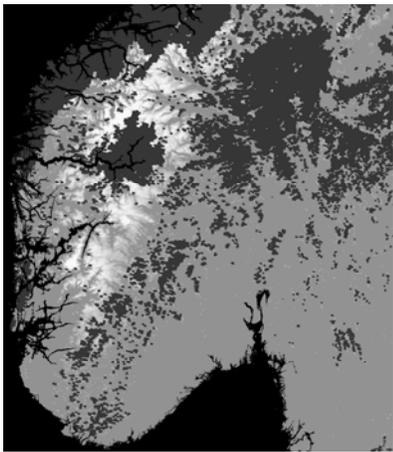
Figure 5 Snow cover and cloud classification of MODIS image from 2003.06.02 10:30



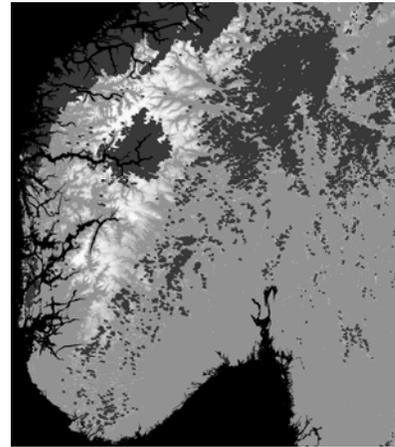
Early



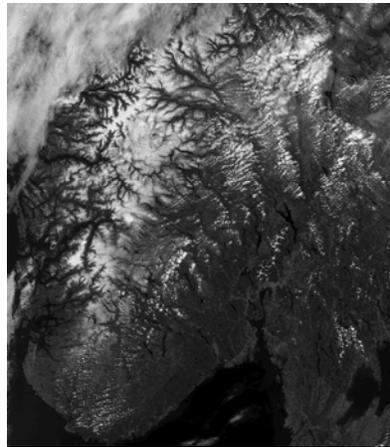
Medium



Late

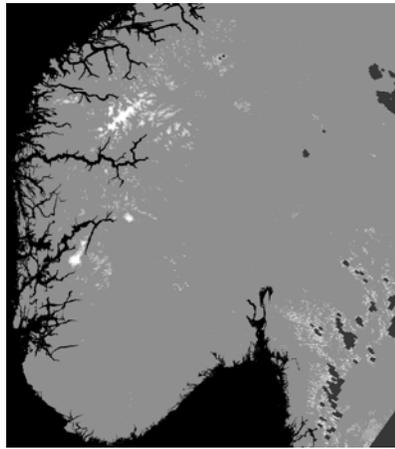


Total

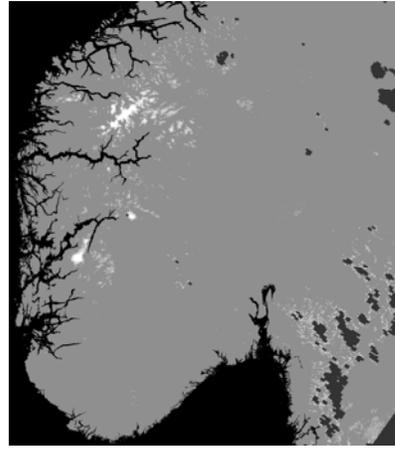


MODIS band 1

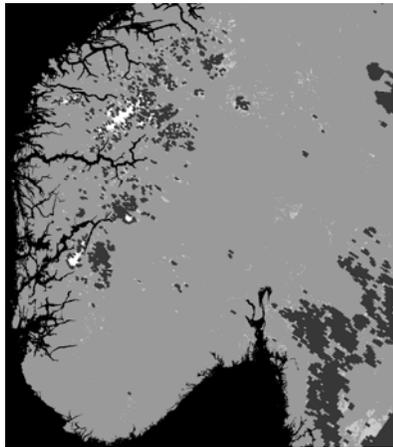
Figure 6 Snow cover and cloud classification of MODIS image from 2005.07.11 10:15



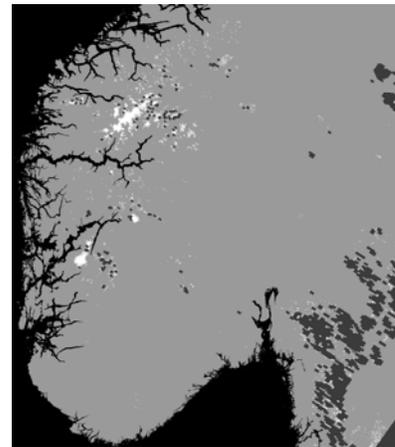
Early



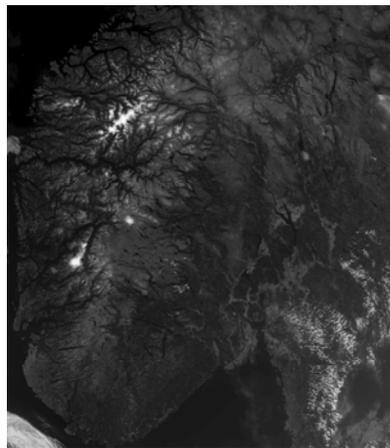
Medium



Late



Total



MODIS band 1

Figure 7 Snow cover and cloud classification of MODIS image from 2003.08.09 11:40

3.2 Comparing NR products with MODIS MOD35_L2

The cloud classification results have also been compared to the cloud mask from the MOD35_L2 MODIS product provided by NASA. We have used the classifications by the *total* codebook, even if this did not provide the best results in all cases. The cloud masks are derived from the swath images and are not geometric corrected. This is because the special way of resampling, used by our method gives a small extension of the cloud cover in the corrected masks. Because cloud classification is quite time consuming, the classification in the NR production chain is not performed on the full MODIS images, but only inside a mask covering Norway and parts of Sweden and Finland.

In Figure 8 to Figure 15 the NR cloud masks are compared with the corresponding MOD35 product for a subsection covering South Norway. The MOD35 product shows clouds in black. The NR clouds are shown in dark gray, and the areas outside the classification mask are shown in black. Band 1 and 6 of the MODIS image are presented as references. Band 6 gives high signals for clouds and low for snow, thus clouds are clearly visible over snow-covered areas. In band 6 one can detect even very thin layers of clouds which are not visible in band 1 and hardly disturbs the snow classification.

2005.02.25 (Figure 8)

The most significant difference between the NR and MOD35 clouds is that MOD35 shows many small clouds over the snow, especially along the borders of the snow, while NR's does not. At the border a pixel may contain information both from snow and bare ground, 'mixed pixels'. Such pixels seem to be classified as clouds by the MOD35 classification.

Looking at the MODIS band 6 image, it is evident that the size of the clouds in the upper part of the image is underestimated by both methods. MOD35 gives better estimates than NR's.

2003.04.18 (Figure 9)

MOD35 finds non-existing clouds at the snow borders as for 2005.02.25. NR's does not. The clouds visible in band 6 are better detected by MOD35. Some of these clouds are very thin and the terrain is clearly visible through them. These are not detected by NR's. The fractional snow cover (FSC) found beneath the clouds have reasonable values, so these clouds probably have little influence on the snow cover classification.

2000.05.04 (Figure 10)

MOD35 shows extra clouds along the borders of the snow area. It also shows clouds over many lower areas with fractional snow cover where there are no clouds. The NR-algorithm detects no non-existing clouds. Existing clouds are somewhere best detected by MOD35 and somewhere best by NR. Many of the thin cloud layers are missed by both methods.

2003.06.02 (Figure 11)

Large cloud-free-areas with snow have been detected as clouds in MOD35. The clouds in the eastern part of the image are mainly better detected by NR's algorithm. Some

other clouds have been detected by NR's and not MOD35 and vice versa. There are areas of thin clouds which have not been detected by either method.

2005.07.11 (Figure 12-14)

In Figure 12 MOD35 has classified clouds instead of snow in the western part of Norway. The real clouds have been quite similarly detected by the two methods, except for two areas at the top of the image, where NR's has detected clouds and MOD35 not. The clouds over the sea in the southern part are better detected by MOD35. Both methods do not detect the cloud stripes which are visible in band 6 and not in band 1.

In Figure 13, showing Nordland (Norway) and Norrland (Sweden), MOD35 has detected clouds instead of snow. The clouds along the border of the image in west, north and east have been detected quite similarly by the two methods. NR's gives a far better detection of smaller cloud areas in the southern and eastern part of the image. None of the methods have detected the thin cloud layer in the east.

In Figure 14, showing the western part of Finland without snow, NR's generally detects more of the existing clouds than MOD35.

2003.08.09 (Figure 15)

MOD35 has a better detection of the clouds over water. Some of them are outside the NR mask. MOD 35 has detected some clouds in the snow area. NR's has detected a few. In the shown example, the classification codebook covering the whole melting season has been used. There are no clouds in this area, and using another codebook, these clouds will disappear (see Figure 7).

The stripes of clouds in the eastern part, hardly visible in band 1, have not been detected by any method.

3.2.1 Discussion

Large systems of clouds and smaller cloud systems over bare land are detected quite similarly by both methods. There are differences, which one will see when studying the examples, but the main difference lies in detection of clouds over and at the borders of snow-covered areas.

This is the largest difference between the two methods: MOD35 makes incorrect detections of clouds over snow areas, with full and also fractional snow cover where NR's does not. The same may happen along the borders of the snow covered areas.

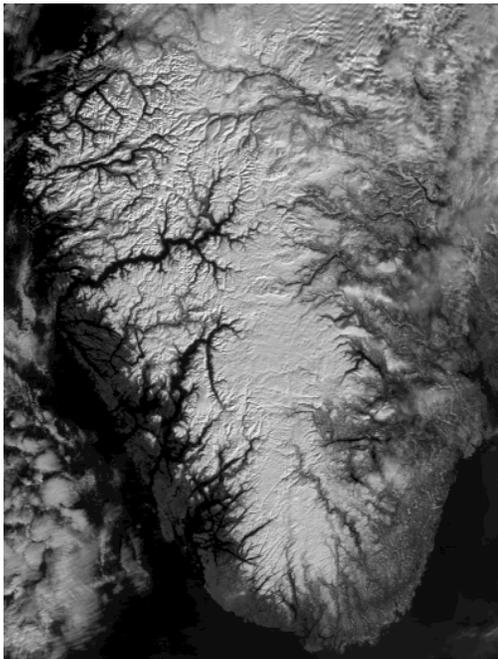
MOD35 has a better detection of thin, transparent clouds.



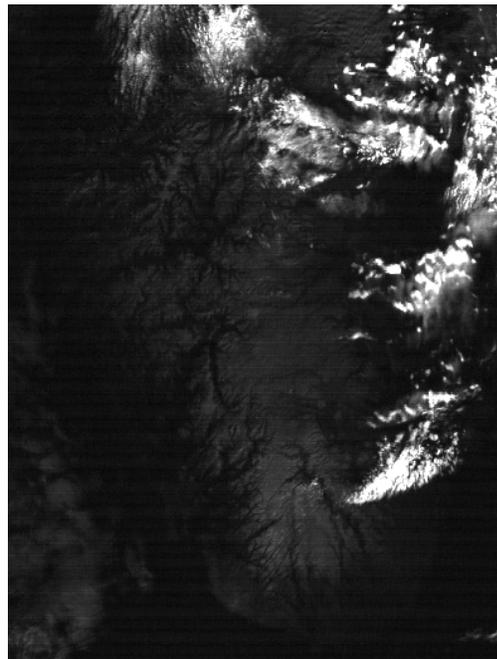
Clouds NR



Clouds MOD35

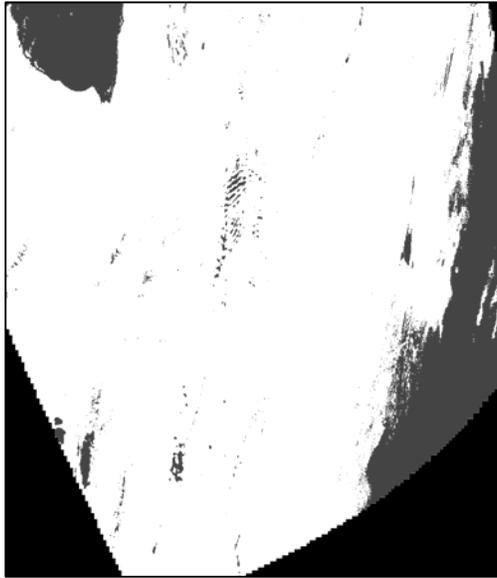


MODIS band 1



MODIS band 6

Figure 8 Classified clouds by NR and MOD35_L2, MODIS image from 2005.02.25 11:25



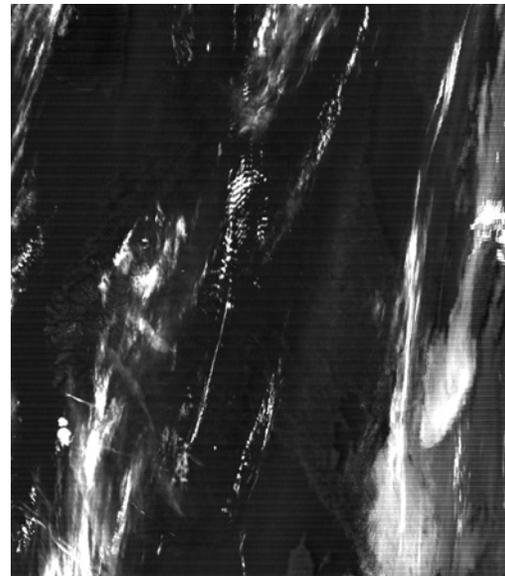
Clouds NR



Clouds MOD35

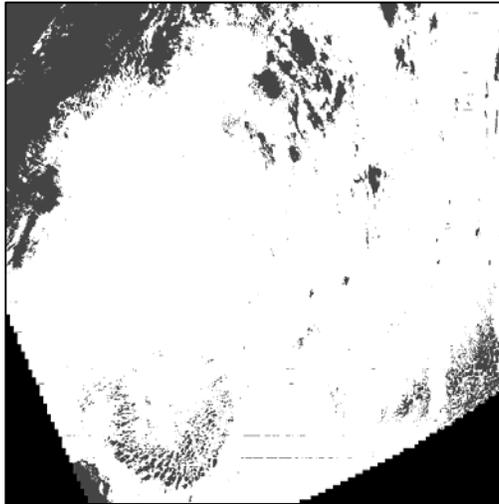


MODIS band 1

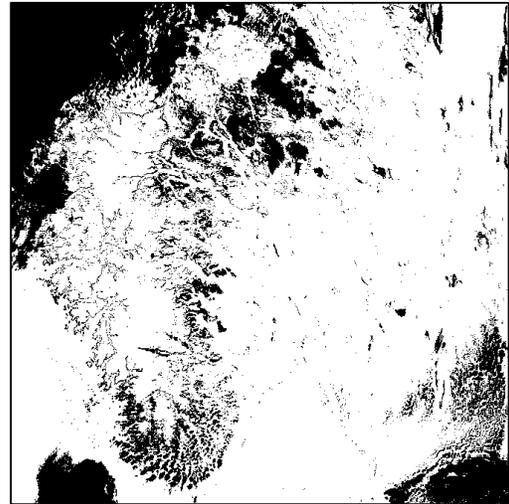


MODIS band 6

Figure 9 Classified clouds by NR and MOD35_L2, MODIS image from 2003.04.18 11:00



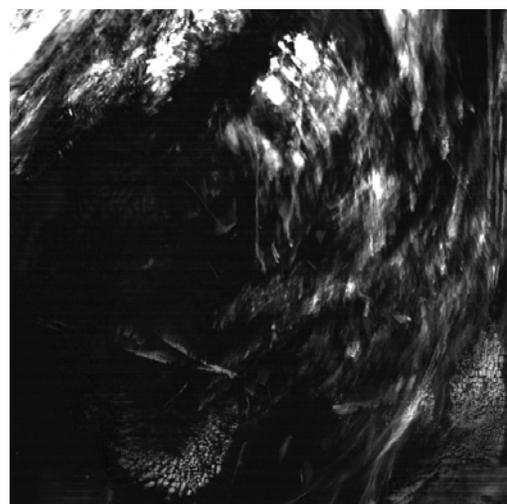
Clouds NR



Clouds MOD35

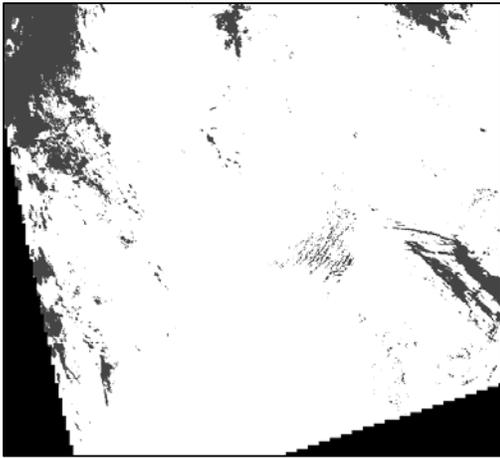


MODIS band 1

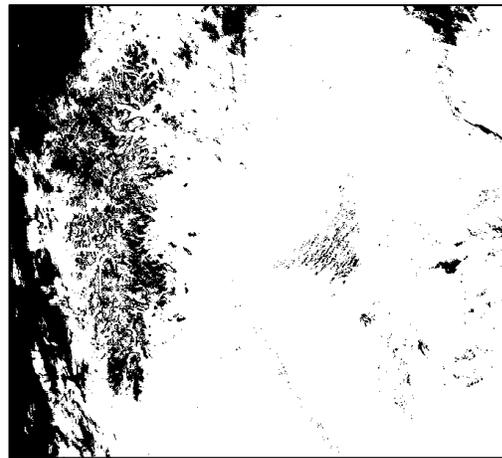


MODIS band 6

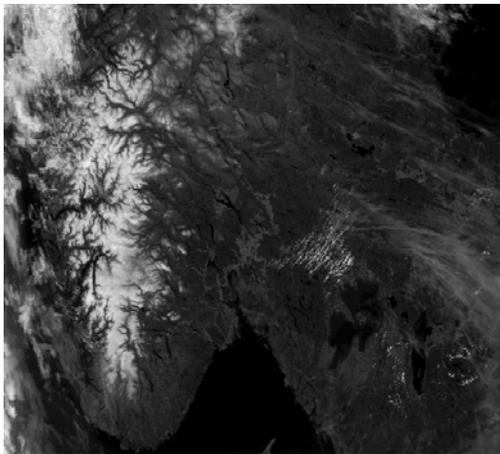
Figure 10 Classified clouds by NR and MOD35_L2, MODIS image from 2000.05.04 11:10



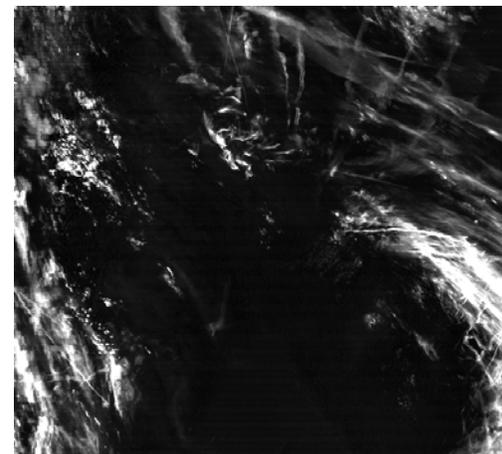
Clouds NR



Clouds MOD35



MODIS band 1



MODIS band 6

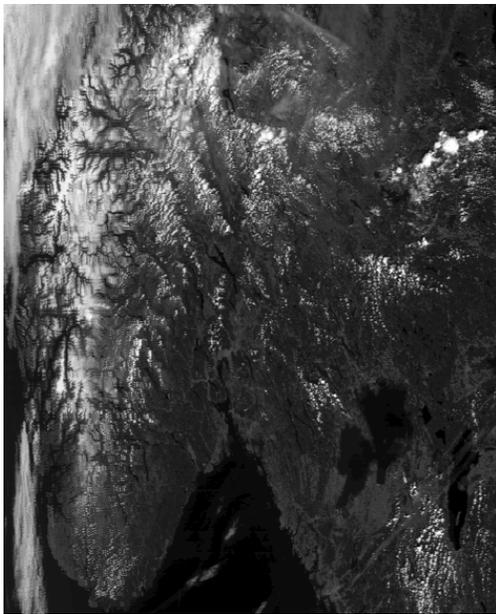
Figure 11 Classified clouds by NR and MOD35_L2, MODIS image from 2003.06.02 10:30



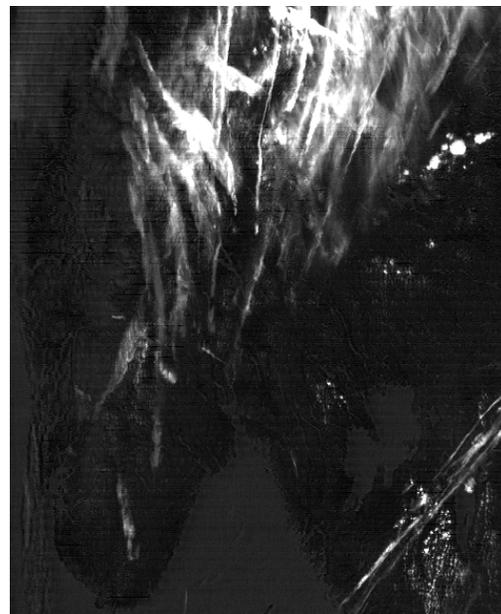
Clouds NR



Clouds MOD35



MODIS band 1

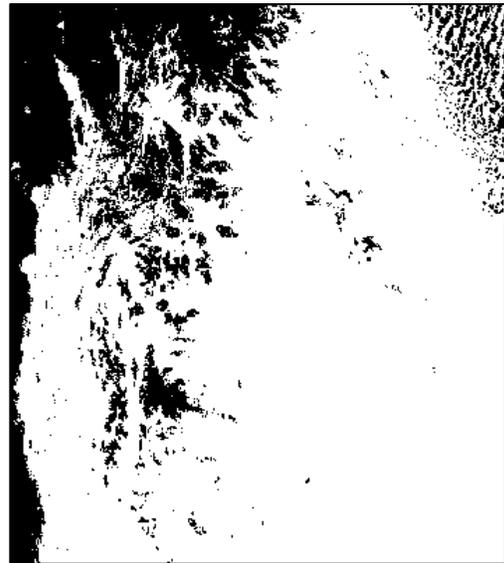


MODIS band 6

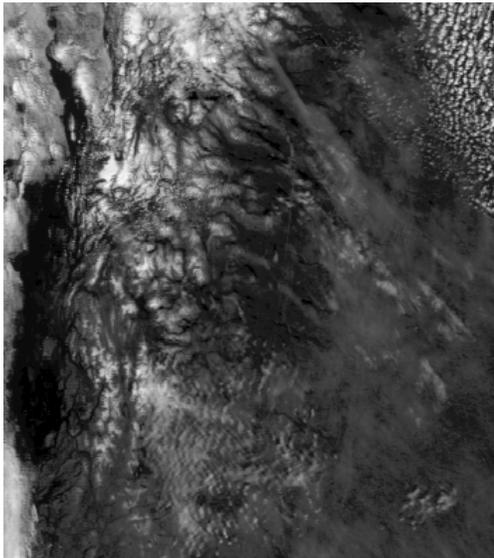
Figure 12 Classified clouds by NR and MOD35_L2, MODIS image from 2005.07.11_1015, South Norway and Sweden.



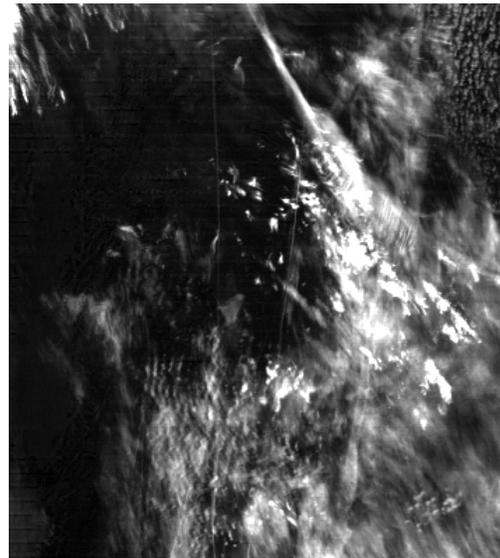
Clouds NR



Clouds MOD35

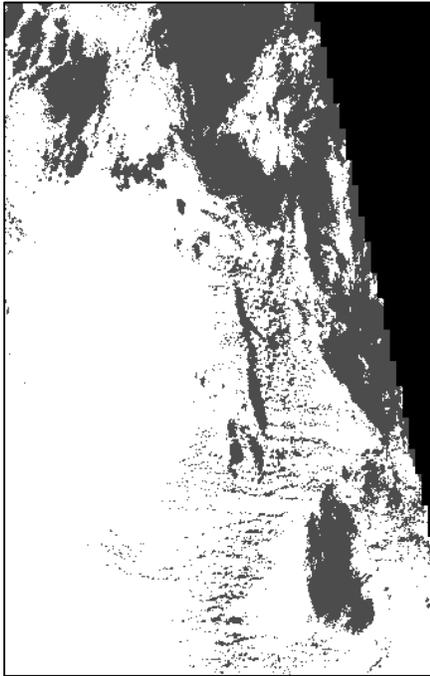


MODIS band 1



MODIS band 6

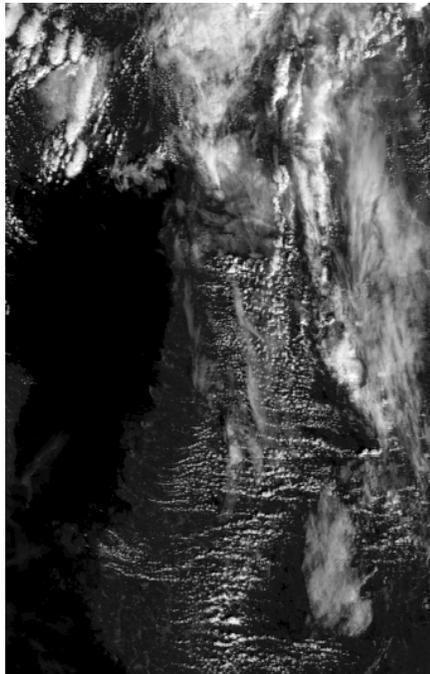
Figure 13 Classified clouds by NR and MOD35_L2, MODIS image from 2005.07.11 10:15, Nordland (Norway) and Norrland (Sweden)



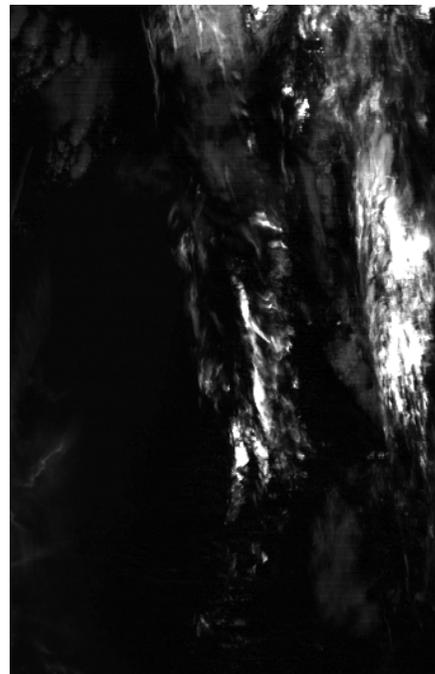
Clouds NR



Clouds MOD35

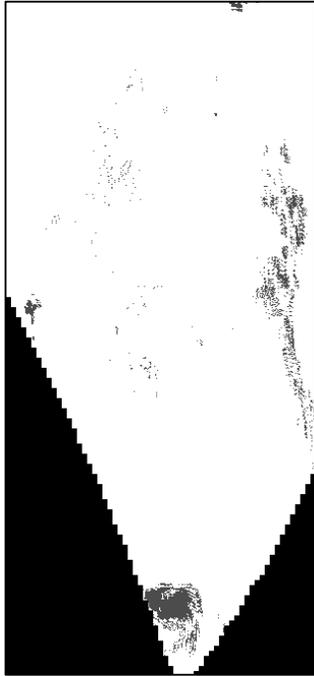


MODIS band 1

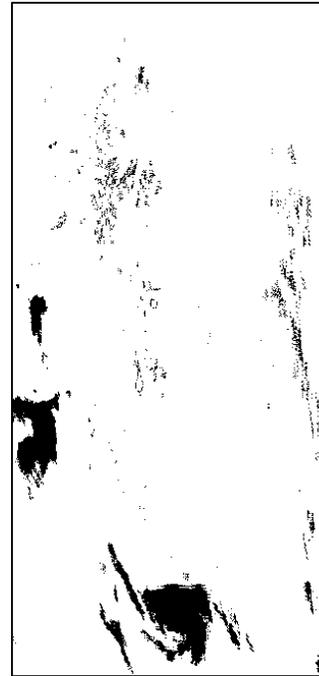


MODIS band 6

Figure 14 Classified clouds by NR and MOD35_L2, MODIS image from 2005.07.011 10:15, Finland



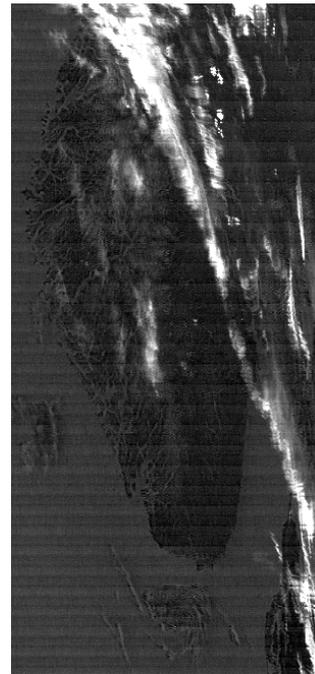
Clouds NR



Clouds MOD35



MODIS band 1



MODIS band 6

Figure 15 Classified clouds by NR and MOD35_L2, MODIS image from 2003.08.09 11:40

4 Conclusions

The main conclusion is that the k-NN method, based on a set of training images spanning the snowmelt season in Scandinavia, produces a fairly correct cloud mask over Scandinavia in this season. The existing clouds are detected and there are few erroneous classifications of non existing clouds over the snow covered areas.

Selected snow maps have been compared with the MOD35_L2 MODIS cloud mask product provided by NASA. As a rule, the k-NN-classifier based method produces a better cloud mask than the MODIS cloud mask product. There are small differences in general, but the main difference lies in detection of clouds over and at the borders of snow-covered areas. MOD35 frequently shows clouds along most of the edges of the snow covered area. In some cases the k-NN classifier also produces a few extra clouds near the snow edges, but it is considered as a minor problem. MOD35 will in many cases produce clouds over areas with wet snow and fragmented snow cover.

The MOD35 algorithm has been developed for global applications for all types of clouds all over the Earth in all seasons. NR's algorithm has on the other side been tailored for the snowmelt season in Norway. Therefore, it could be expected to produce generally better results under these conditions. Using different codebooks for the different seasons (winter, spring and summer) may further improve the cloud detection, as shown in the examples above. We have chosen to use a combination of three codebooks, covering the whole melting season in our production chain, because this codebook gives the best result in most cases. Then we don't have to check the date, temperatures and snow conditions before running a snow classification.

Close studies of the results from generation of time series of SCA products by the NLR algorithm show that in some cases clouds over snow-free land have not been detected. In the snow production programme the cloud pixels are classified as snow and give a wrong value of the snow coverage. This often happens in late spring and summer with small cumulus clouds. The problem can be solved by using contextual information. In the summer and late spring, there will be no snow in the lowest regions in Scandinavia.

The examples show that thin, more or less transparent clouds usually are not detected. This is mainly due to the creation of the training set. There are some difficulties with cloud labelling in the training process of the algorithm. There are two main types of clouds – transparent and opaque clouds. If the thin transparent clouds are labelled as clouds, the classification algorithm will probably detect them as clouds and they will be included in the cloud mask. When aiming at mapping the snow cover, one might try to neglect these clouds because the ground below is visible. But the clouds will reduce the reflected light from the snow, and one might end up with incorrect values for the snow cover percentage. When labelling, it is often difficult to see if there are thin clouds or not. One could probably use more than one cloud class and separate the opaque from the transparent clouds. If the classification procedure manages to discriminate the two types of clouds, one might be able to detect the snow also in areas with thin clouds. But one should then reduce the confidence of the results in such areas.

5 References

Duda, R.O., Hart, P.E. and Stork, D.G. *Pattern Classification*, John Wiley & Sons, Inc., 2001.

Myers, W.L. and Patil, G.P. *Pattern-Based Compression of Multi-Band Image Data for Landscape Analysis*, Springer, 2006.

Solberg, R., Koren, H. and Amlien, J. *A review of optical snow cover algorithms*, NR note SAMBA/40/06, 2006

http://modis-atmos.gsfc.nasa.gov/MOD35_L2/index.html

Appendix

MODIS images used for creation of vector quantization codebook.

65 images, specified by date and time, sorted by time of year.

20030204_1105
20030206_1055
20030214_1005
20030301_1100
20030312_1040
20030313_1125
20030328_1040
20030406_1035
20030407_1115
20030411_1050
20030413_1040
20030415_1030
20030416_1110
20030417_1015
20030418_1100
20030420_1045
20030422_1035
20030423_1115
20030424_1020
20030429_1040
20040501_1040
20030502_1110
20020505_1035
20030505_1005
20000506_1055
20020506_1120
20010507_1100
20020507_1025
20010508_1005
20010509_1050
20040510_1035
20000511_1115
20010514_1110
20020517_1100
20060517_1115
20020519_1045
20040523_1005

20040530_1010
20030531_1040
20030601_1125
20020602_1100
20030602_1030
20010604_1125
20030607_1045
20040609_1045
20040612_1115
20040613_1020
20030616_1040
20040617_1135
20030623_1045
20030626_1115
20030627_1020
20030628_1105
20040628_1115
20040630_1105
20040703_1135
20040704_1040
20030705_1110
20030706_1015
20030713_1020
20030714_1105
20030715_1010
20030716_1050
20050723_1040
20030809_1140

MODIS images used in early knncb

10 images sorted by time of year

20030204_1105
20030206_1055
20030214_1005
20030301_1100
20030312_1040
20030313_1125
20030328_1040
20030406_1035
20030407_1115
20030413_1040

MODIS images used in medium knncb

23 images sorted by time of year

20030416_1110
20030417_1015
20030418_1100
20030420_1045
20030422_1035
20030424_1020
20030429_1040
20030502_1110
20030505_1005
20000506_1055
20010507_1100
20010508_1005
20010509_1050
20000511_1115
20040523_1005
20040530_1010
20030531_1040
20030601_1125
20030602_1030
20020602_1100
20010604_1125
20030607_1045
20040612_1115

MODIS images used in late knncb

11 images sorted by time of year.

20040617_1135
20030623_1045
20030626_1115
20030628_1105
20040628_1115
20040630_1105
20040704_1040
20030705_1110
20030715_1010
20030716_1050
20050723_1040