LIFE Project Number

ENV/FIN/000133

Progress report on filtered time-series (2001-2008)

Reporting Date

31/05/2010

Action

Action 7: Methodological development and implementation

LIFE+ PROJECT NAME or Acronym

SNOWCARBO

Author

Name Beneficiary	Finnish Environment Institute (SYKE)
Contact person	Ms Kristin Böttcher
Postal address	P.O. Box 503, FI-00101 Helsinki, Finland
Telephone	+358-401-876447
Fax:	+358-9-40300690
E-mail	Kristin.Bottcher@ymparisto.fi
Project Website	http://snowcarbo.fmi.fi

Table of contents

List of abbreviations

NDVI	Normalized Difference Vegetation Index
SCA	Snow Covered Area
MVC	Maximum Value Compositing
MODIS	Moderate Resolution Imaging Spectrometer
SYKE	Finnish Environment Institute
METLA	Finnish Forest Institute
FMI	Finnish Meteorological Institute
CLC	CORINE Land Cover
AG	Asymmetric Gaussian function
DL	Double logistic function
SG	Savitzky-Golay filter

Summary

This report describes methods and first results for the filtering and interpolation of time-series of Normalized Difference Vegetation Index (NDVI) and Snow Covered Area (SCA) derived from MODIS satellite data for the years 2001 until 2008.
Gaps in cloud-filtered time-series of NDVI and SCA were filled using linear interpolation and curve fitting methods. Smoothing and gap-filling was applied to NDVI and SCA time-series from homogenous areas around flux measurement stations and phenological stations in Finland, representing different land cover types, such as coniferous forest, broadleaved forest, open bogs, agricultural land.
Interpolated time-series from these measurement sites will serve as calibration and validation basis for the extraction of carbon-balance-related features.

Introduction and objectives

Action 7 provides time-series of Snow Covered Area (SCA) during snow melting period and vegetation status as described by a vegetation index. Both, SCA and vegetation index (Normalized Difference Vegetation Index, NDVI) are derived from MODIS satellite data.
The retrieval of information on snow cover and vegetation status using satellite data in the optical domain is impeded by cloud cover leading to gaps in the time-series. In addition, sub-pixel cloud contamination, varying atmospheric conditions and bidirectional effects cause noise in the temporal profile. These perturbing factors make the extraction of carbon-balance related features and the use of these data in models difficult.

Therefore, the objective of this report was to derive filtered and interpolated daily time-series of Normalised Difference Vegetation Index (NDVI) and Snow Covered Area (SCA).
This first deliverable concentrates on extraction of time-series from *in situ* measurement sites for phenology and $C0_2$ fluxes. Filtered and interpolated time-series from these sites will be used as calibration and validation basis for the determination of carbon-balance related features (beginning of growing season, maximum of growing season and end of growing season) in the next deliverable of action 7.

The report will give a short overview about the literature in the first section. Data sets will be described in the second section. Methods for filtering and interpolation of NDVI and SCA time-series and preliminary results will be presented in section 4.

Literature review

Noise reduction in NDVI time-series

The Maximum Value Compositing technique (MVC), introduced by Holben (1986), might be the most commonly use method for noise-reduction in time-series of NDVI data. The main assumption of the maximum value compositing technique is that non-optimal atmospheric, soil, view, and illumination angle conditions depress the NDVI and that the maximum NDVI in the composite period best represents vegetation status.
Other methods for filtering of cloud cover and noise have been proposed such as the Best Index Slope extraction (BISE) method (Viovy *et al.*, 1992), different function fitting methods (Zhang *et al.*, 2003; Jönsson and Eklundh, 2004; Beck *et al.*, 2006) and Fourier analysis.
Both double logistic function (DL) and asymmetric Gaussian functions (AG) mimic well vegetation development in the boreal zone whereas Fourier analysis was less appropriate for this biome (Jönsson and Eklundh, 2002; Beck *et al.*, 2006) and will therefore not be considered here. Gao et al. (2008) examined both AG and DL approaches and found that they produced similar results, with the exception that the AG approach is less sensitive to the incomplete time-series data with many gaps.
Recently, Hird and McDermid (2009) conducted a model-based empirical comparison of six selected NDVI time-series noise-reduction techniques including the following methods: asymmetrical Gaussian Function fitting, Double Logistic function-fitting, Savitzky-Golay filter (SG), "4253H, twice filter", Mean-value iteration filter and ARMD3-ARMA5 filter. MOD13Q1 (version 004) 16-day 250 m NDVI and NDVI Quality insurance products from Terra MODIS were used in the analysis. Performance was assessed using both analysis and metric analysis. The metric analysis involved the comparison of phenology-based metrics such as start, end and length of growing season from modelled time-series and the same

metric derived from experimental noise-reduced time-series. The two function-fitting techniques performed better overall. When stratifying by biogeographical region, results for the boreal biome indicate best overall performance of AG, followed by "4253H, Twice filter" and DL function fitting. Performance stratified by metric showed large amount of variation, revealing the notable sensitivity of noise reduction performance to individual metrics. The extraction of time-based metrics was consistently improved with the application of noise reduction, whereas the extraction of non time-based metrics (max NDVI, max green-up, integrated NDVI) was not consistently improved with noise reduction.

Filtering and interpolation of SCA time-series

The utility of SCA products is limited by cloud-cover causing gaps in the daily time-series. Recently, simple mapping methods and the combination of MODIS data from Terra and Aqua were used to fill gaps due to cloud cover in MODIS snow cover products (binary products: snow-covered and snow-free).

Tong et al. (2009) adopted a spatial filter (SF) to reduce cloud cover in MODIS 8-day snow products (MOD10A2) for the years 2000 to 2007 in Quesnel River Basin in Bristish Columbia, Canada. SF is used to reclassify the 8-day snow product. A cloud-covered pixel is replaced by the eight closest neighbouring pixels. The SF reduced cloud coverage for the study area from 15 % for MOD10A2 products to 9 % for SF products.

Parajka and Blöschl (2008) used temporal and spatial filters, that reduce cloud coverage by using information from neighbouring non-cloud covered pixels in time or space, and combined MODIS data from Terra and Aqua satellites. The approach was evaluated based on snow depth measurements at 754 climate stations in Austria. Cloud coverage was reduced considerably with this approach and the snow maps were still in good agreement with ground measurements.

A similar method was described by Gafurov and Bárdossy (2009). They developed a six-step methodology to estimate the snow cover of cloud covered areas. The methodology included: i) combination of MODIS Aqua and MODIS Terra images, ii) gap-filling based on data from up to 2 days forward and backward, iii) determination of the snowline (snow is continuous above a specific elevation), iv) spatial combination of neighbouring pixels (4 pixels), v) spatial combination of 8 neighbouring pixels using information on elevation and vi) determination of 'complete snowmelt' and 'snow accumulation start' for each pixel based on the time-series over an entire year. Each step removed progressively more cloud cover. The method was validated using least cloud covered original MODIS snow cover products, which were artificially filled with clouds. Average accuracy between 90 – 96 % was obtained when applying step 2 to 5. The last step led to a decrease in accuracy (78 %) but removed all remaining cloud cover.

Hall et al. (2010) proposed a simple temporal cloud-gap filling algorithm, which was applied to MODIS 0.5˚ resolution grid climate modelling daily snow cover product (MOD10C1). The algorithm tracks cloud persistence, to account for the uncertainty created by the age of snow observation. The effectiveness of the method was evaluated with assimilation experiments using Noah land surface model.

Dozier et al.(2008) described a method for filtering and interpolation of daily maps of fractional snow cover derived from MODIS data. This is the only study known to the author on filtering and interpolation of fractional snow cover data. They used an algorithm consisting of i) filtering of noise due to stripes and ii) interpolation and smoothing. Noise, occurring as stripe of pixels, was identified in each daily image by a two-dimensional adaptive Wiener filter combined with a gradient of estimated snow cover. Due to the expansion of the pixel at off-nadir view angles, the derived snow cover incorporates values from neighbouring pixels. The resulting error is not systematic as neighbours may have more or less snow cover. MODIS sensor zenith angles oscillate between near-nadir and more than 65˚ in a regular 16-day pattern, whereas there is only smooth variability in viewing angles across an image on a particular day. Therefore, smoothing spline was applied along the time axis. Smoothing spline interpolates well when data is well distributed along abscissa, but leads to unlikely values in case of large gaps. Therefore, after smoothing daily values for fractional snow cover, interpolation between missing days was done using a piecewise interpolant (monotone cubic interpolation). After interpolation and smoothing by date, the whole data cube was smoothed with a Gaussian filter.

Data sets

MODIS time-series

The MODIS NDVI and SCA time-series used here are described in detail in the report "1st Document on existing datasets" (SYKE composite products and time-series data, p. 27). Clouds were masked using operational cloud algorithms and visually verified cloud masks from GSE PolarView project. Multiple NDVI and SCA products per day were combined to one daily estimate after the masking of clouds.
In order to further reduce noise in the NDVI time series, weekly NDVI composites were produced using the maximum composite technique (Holben, 1986).

Land cover information

In order to facilitate the generation of time-series for homogenous areas (according to land cover) around *in situ* measurement sites, the fraction of each land cover class per MODIS pixel was calculated from CORINE Land Cover 2000 (CLC) for Finland. As an example the fractions of broadleaved forest, agricultural area and coniferous forest are shown in Figure 1.

Measurement sites

Time-series of NDVI and SCA were extracted at phenological measurement stations and at flux-measurement stations in Finland. Only stations with homogenous area (minimum 9 NDVI pixels[1]) around the station were selected. Selected sites are presented in Figure 1 and details for sites are given in Table 1.
Most of the *in situ* sites are located in coniferous forest (7 stations). In addition, two sites are located in open bogs. Agricultural land and broadleaved forest are represented each with one site. Number of MODIS pixels and thresholds used for the determination of the area from CLC 2000 are listed in Table 1. Note that number of pixels is different for NDVI and SCA time-series due to the different pixel size of the data.

[1] except for Lompolojänkkä

Table 1. Measurement sites used for time-series analysis.

Site	**Observation/ Organisation**	**Latitude/ Longitude [°]**	**Vegetation type**	**Number of pixels NDVI/ SCA***	**CLC 2000 class (class number**)**	**Threshold for LC fraction [%]**
Kevo	Phenology/ METLA	69°45' N 27°01' E	broadleaved forest (birch)	24/9	broadleaved forest (17,18)	90
Parkano	Phenology/METLA	62°01' N 23°03' E	coniferous forest (Scots pine)	11/8	coniferous forest (19,20,21)	90
Paljakka	Phenology/METLA	64°40' N 28°03' E	coniferous forest (Scots pine)	65/17	coniferous forest (19,20,21)	100
Äkäslompolo	Phenology/METLA	67°35' N 24°12' E	coniferous forest (Scots pine)	472 / 70	coniferous forest (19,20,21)	100
Saariselkä	Phenology/METLA	68°24' N 27°23' N	coniferous forest (Scots pine)	69/21	coniferous forest (19,20,21)	90
Hyytiälä	CO_2 flux/ University of Helsinki (CARBOEUROPE)	61° 51' N 24°17' E	coniferous forest (Scots pine)	175/23	coniferous forest (19,20,21)	100
Sodankylä	CO_2 flux /FMI	67°21.712'N 26°38.270'E	coniferous forest (Scots pine)	27/9	coniferous forest (19,20,21)	100
Kenttärova	CO_2 flux /FMI	67°59.234'N 24°14.583'E	coniferous forest (Spruce)	19/8	coniferous (19, 20,21)	95
Lompolojänkkä	CO_2 flux /FMI	67°59.832'N 24°12.551'E	Aapa mire	8/-	open peat land (38)	70
Kaamanen	CO_2 flux /FMI	69°08.441'N 27°16.230'E	Aapa mire	30/7	open peat land (38)	70
Jokionen	CO_2 flux /FMI	60°53.956'N 23°30.933'E	Agricultural peat field, managed	11/10	arable land (13)	90

* Pixel size for NDVI 0.0025 degrees and for SCA 0.005 degrees
** Class number according to CORINE Land Cover 2000 Finland

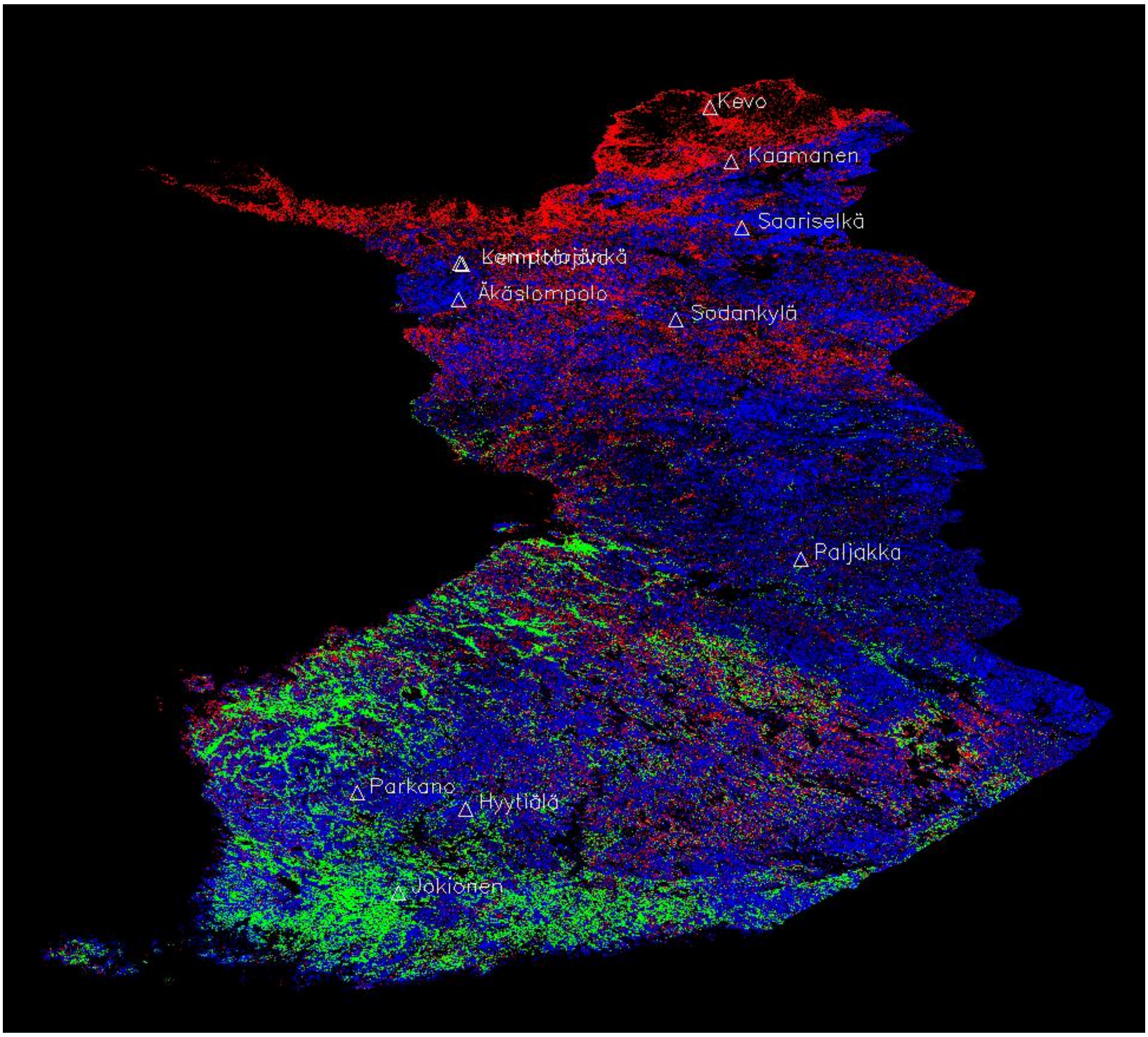


Figure 1. Fraction of broadleaved forest (Red), agricultural land (Green) and coniferous forest (Blue) per MODIS pixel for Finland. Sites used for time-series analysis are shown with a white triangle.

Filtering and interpolation of NDVI time-series

Description of methodology

The software TIMESAT version 2.3 (Jönsson and Eklundh, 2004), developed for seasonality extraction and noise removal by function fitting, was used for interpolation of NDVI time-series from homogenous sites near *in situ* measurement stations.
Asymmetric Gaussian function fitting, logistic function fitting as well as an adaptive Savitzky-Golay filter are implemented in the software based on least-squares fits to the upper envelope of the vegetation index data. Detailed description on the methodology can be found in the 'Users Guide for TIMESAT 2.3'(Jönsson and Eklundh, 2006).
We applied both, asymmetric Gaussian function fitting and logistic function fitting to cloud-filtered NDVI time-series, since both approaches were found suitable for boreal ecosystems (see Literature review). Another advantage of using double logistic function or Gaussian function for describing NDVI temporal profiles is that estimated model parameters, i.e. for inflection points) can be interpreted in terms of vegetation phenology (Soudani *et al.*, 2008).

Prior to function fitting with TIMESAT, cloud-filtered NDVI time-series were interpolated to daily time steps using linear interpolation. Function fitting was done for each year for the period from February until end of October. Interpolated values were given low weight in the function fitting procedure. Parameter settings for function fitting with TIMESAT are given in Table 2.

Table 2. Parameter settings for function fitting in TIMESAT.

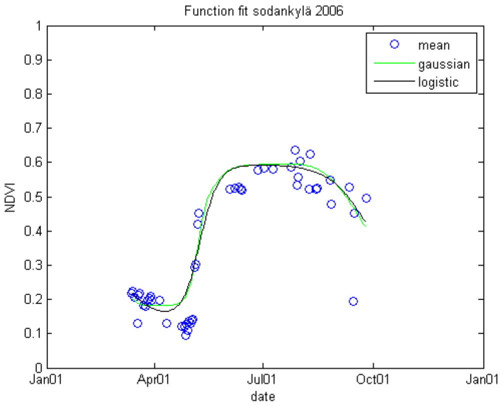
Parameter	Description	Value
Number of fitting steps	If more than 1 step the fitted functions are forced to the upper envelope	3
Strength of envelope adaption	Value between 1 -10, where 10 strongly forces the fit to follow the upper envelope of the data	2

Results for *in situ* measurement sites

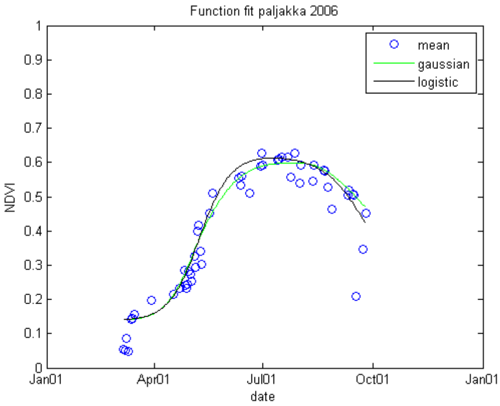
Examples for interpolated time-series for the year 2006 from 6 measurement sites are shown in Figure 2. The overall shape of NDVI-time profiles is well described by logistic and Gaussian functions and differences between both functions are small. However, interpolated NDVI values are overestimated for the period before the summer peak, especially for broadleaved forest (e) and for the open bog sites Kaamanen (c) and Lompolojänkkä (d).

This has implications for the use of these interpolated NDVI temporal profiles for the determination of the time of maximum NDVI. Hence, for this purpose the original NDVI values should be used.

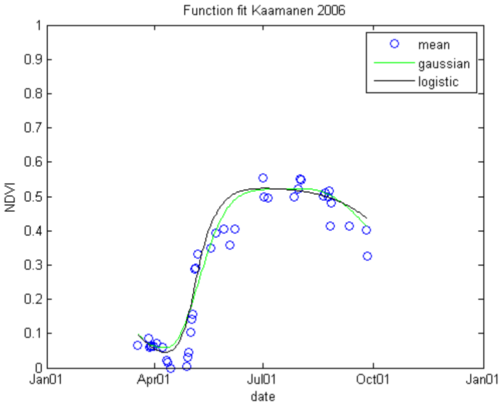
Due to low sun angles no useful NDVI observations were obtained after the end of growing season (sun angles > 75 degrees were excluded in pre-processing), which makes the curve fitting for this period difficult. Furthermore, a drop of NDVI values during the snow melting in spring is observed [for example Kaamanen (c) and Lompolojänkkä (d)]. This behaviour was also reported by Delbart et al. (2005) for the boreal region, who used therefore the Normalized Difference Water Index for the extraction of the beginning of the growing season. Beck et al. (2006) suggested another approach to mitigate this problem and to stabilise curve fitting during winter. They derived a pixel-specific winter NDVI (NDVI in winter without snow cover), which is constant over years. All pixels in the NDVI temporal profile lower than the winter NDVI and missing data during polar night were replaced by the winter NDVI. This allowed improved fitting results for the beginning of growing season and curve fitting after the end of growing season.



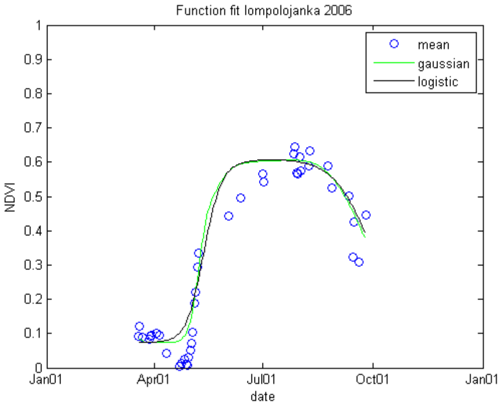
a)



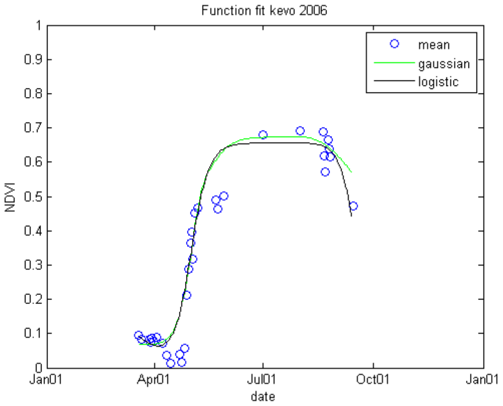
b)



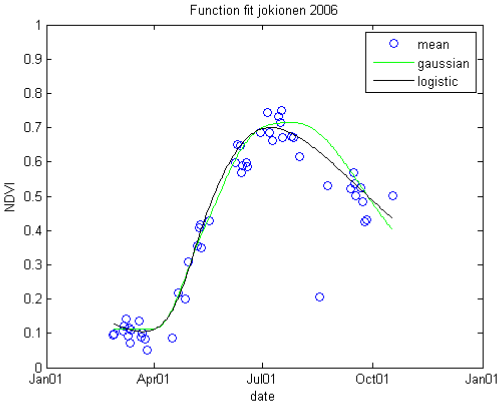
c)



d)



e)



f)

Figure 2. Mean NDVI time-series after cloud-filtering and interpolated time-series with logistic and Gaussian functions from in situ measurement sites for the year 2006: a) Sodankylä; b) Paljakka; c) Kaamanen; d) Lompolojänkkä; e) Kevo and f) Jokionen.

Hence, further work will investigate about a winter NDVI for different regions and land cover classes, which could be used to improve curve fitting algorithm.

Filtering and interpolation of SCA time-series

Description of methodology

Cloud-filtered time-series of SCA from measurement sites were interpolated to daily time steps using linear interpolation. Furthermore, SCA time-series from northern and middle boreal regions were interpolated using a sigmoid function.
The basic function is given below

$$g(t;x_1;x_2) = \frac{1}{1+\exp\left(-\frac{x_1-t}{x_2}\right)} \qquad (1)$$

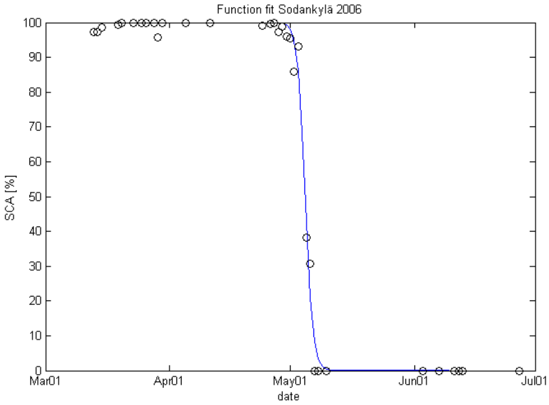
where t is the time variable, x_1 determines the position of the inflection point and x_2 gives the rate of change.
This approach follows the assumption that seasonal snow cover in these regions is persistent (100% snow cover) in winter time and decreases gradually after beginning of snowmelt in spring until complete snow clearance (0 % snow cover).
Gaps due to clouds and noise due to varying viewing and illumination conditions, atmospheric influence or stripes can be smoothed using this approach. However, a possible interruption of snow melting during colder periods or an increase of SCA due to new snow fall events can not be captured with this simple function.

Curve fitting was applied to SCA temporal profiles for each year for the period from February until end of July. All stations given in Table 1 were used, besides Hyytiälä, Parkano and Jokionen, which belong to southern boreal region. Fitting was done by minimizing the sum of squares of differences between Eq. (1) and measured SCA.

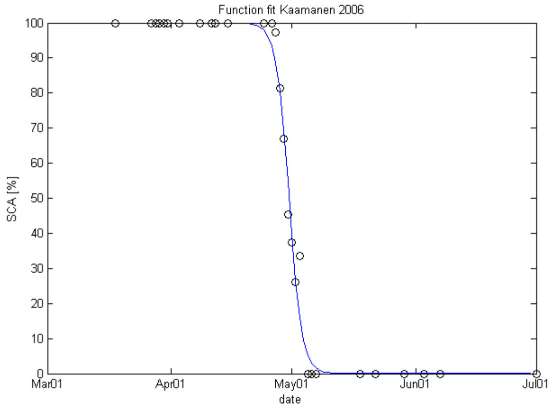
The software MATLAB version 7.8 was used for interpolation and curve fitting of SCA time-series.

Results for *in situ* measurement sites

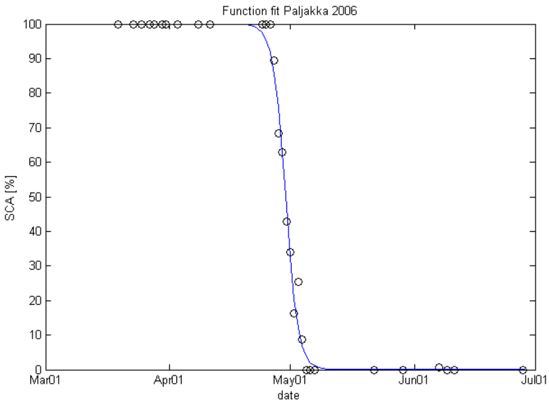
Results of SCA curve fitting for four stations for the snow melting period in 2006 are presented in Figure 3. Logistic functions mimic well SCA temporal profiles for the northern and middle boreal region.



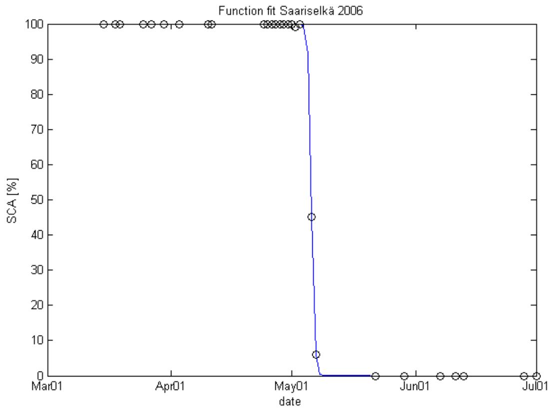
a)



b)



c)



d)

Figure 3. Mean SCA (black circle) and fitted logistic function (blue line) for year 2006 at *in situ* sites: a) Sodankylä; b) Kaamanen; c) Paljakka and Saariselkä.

Large gaps in the time-series due to cloud cover and/ or missing data lead to high uncertainty for interpolated values. In particular, for the years 2001 and 2002 long periods without SCA observation during the snow melting were found. This is illustrated with an example for Sodankylä in Figure 4, where the time period without SCA observation was more than 50 days.

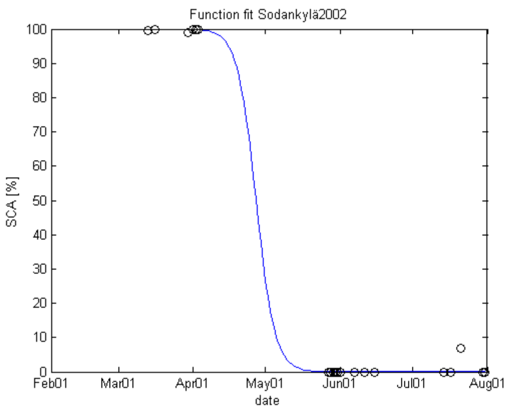


Figure 4. SCA temporal profile for Sodankylä site in 2002. Mean values are shown with black circles and interpolated values as a blue line.

Further work will therefore include a confidence measure for daily values, which tracks cloud persistence, similar to the approach by Hall et al. (2010) and allows the exclusion of time-series with large gaps.

The simple logistic model for the melting period is not applicable for the southern region in Finland (and in general for the southern part of SnowCarbo study area) as periods with full snow cover and snow-free ground may alternate. An example of snow depth measurements from the meteorological station in Jokionen in 2008 is shown in Figure 5. Hence, other methods for data interpolation and smoothing have to be developed and approaches described by Dozier et al. (2008) may be applicable for our study area (see Literature review, p.7).

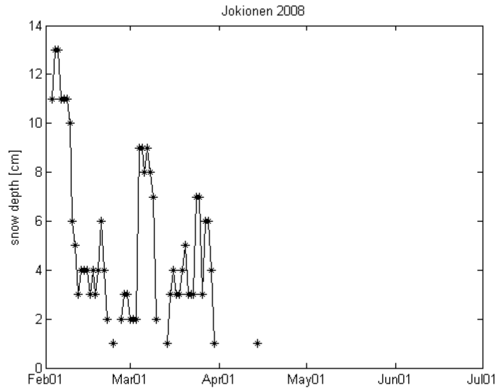


Figure 5. Snow depth measurement [cm] at meteorological station in Jokionen for the year 2008.

Data storage and access

Filtered and interpolated time-series of SCA and NDVI are stored in MATLAB or ASCII format on shared network folders for SYKE's internal use. The data is available to project partners upon request.

Conclusions and Outlook

Time-series of SCA and NDVI need to be processed to remove data gaps and low quality data caused by cloud contamination before they can be used for the extraction of carbon-balance related features, such as beginning and end of growing season, and before they can be used directly in models.

NDVI time-series from *in situ* measurement sites were smoothed and gap-filled using logistic and Gaussian functions, which were found suitable to describe vegetation development in boreal ecosystems. Further work is needed to stabilise winter NDVI values for the improvement of curve fitting.

SCA time-series for the northern boreal and middle boreal region were smoothed and gap-filled using a simple logistic function. In the next step confidence information will be added to the daily values and accuracy of interpolated SCA temporal profiles will be evaluated against *in situ* measurements at meteorological stations. Further methodological development for the smoothing and gap-filling of SCA temporal profiles is needed for the southern part of the SnowCarbo study area.

Large gaps in the time-series during snow melting period and during critical plant growing stages are a strong limitation for the use of both SCA and NDVI time-series. In this work, neighbouring pixels with the same land cover class were used to increase the number of non-cloudy observations. The search area for the inclusion of neighbouring pixels may be increased to the modelling grid in the next step. Another possibility would be to use historical data in the gap-filling algorithm based on the assumption that temporal profiles are similar in different years.

References

Beck, P.S.A., Atzberger, C., Høgda, K.A., Johansen, B., Skidmore, A.K., 2006. Improved monitoring of vegetation dynamics at very high latitudes: A new method using MODIS NDVI. Remote Sensing of Environment 100, 321-334.

Delbart, N., Kergoat, L., Le Toan, T., Lhermitte, J., Picard, G., 2005. Determination of phenological dates in boreal regions using normalized difference water index. Remote Sensing of Environment 97, 26-38.

Dozier, J., Painter, T.H., Rittger, K., Frew, J.E., 2008. Time-space continuity of daily maps of fractional snow cover and albedo from MODIS. Advances in Water Resources 31, 1515-1526.

Gafurov, A., Bárdossy, A., 2009. Cloud removal methodology from MODIS snow cover product. Hydrol. Earth Syst. Sci. 13, 1361-1373.

Hall, D.K., Riggs, G.A., Foster, J.L., Kumar, S.V., 2010. Development and evaluation of a cloud-gap-filled MODIS daily snow-cover product. Remote Sensing of Environment 114, 496-503.

Hird, J.N., McDermid, G.J., 2009. Noise reduction of NDVI time series: An empirical comparison of selected techniques. Remote Sensing of Environment 113, 248-258.

Holben, B., 1986. Characteristics of maximum-value composite images from temporal AVHRR data. International Journal of Remote Sensing 7, 1417-1434.

Jönsson, P., Eklundh, L., 2002. Seasonality extraction and noise removal by function fitting to time-series of satellite sensor data. IEEE Transactions of Geosciences and Remote Sensing 40, 1824-1832.

Jönsson, P., Eklundh, L., 2004. TIMESAT--a program for analyzing time-series of satellite sensor data. Computers & Geosciences 30, 833-845.

Jönsson, P., Eklundh, L., 2006. TIMESAT - a program for analyzing time-series of satellite sensor data;

Users guide for TIMESAT 2.3. Malmö and Lund.

Parajka, J., Blöschl, G., 2008. Spatio-temporal combination of MODIS images; potential for snow cover mapping. Water Resour. Res. 44, W03406.

Soudani, K., le Maire, G., Dufrêne, E., François, C., Delpierre, N., Ulrich, E., Cecchini, S., 2008. Evaluation of the onset of green-up in temperate deciduous broadleaf forests derived from Moderate Resolution Imaging Spectroradiometer (MODIS) data. Remote Sensing of Environment 112, 2643-2655.

Tong, J., Déry, S.J., Jackson, P.L., 2009. Interrelationships between MODIS/Terra remotely sensed snow cover and the hydrometeorology of the Quesnel River Basin, British Columbia, Canada. Hydrol. Earth Syst. Sci. 13, 1439-1452.

Viovy, N., Arino, O., Belward, A.S., 1992. The Best Index Slope Extraction (BISE): A method for reducing noise in NDVI time-series. International Journal of Remote Sensing 13, 1585 - 1590.

Zhang, X., Friedl, M.A., Schaaf, C.B., Strahler, A.H., Hodges, J.C.F., Gao, F., Reed, B.C., Huete, A., 2003. Monitoring vegetation phenology using MODIS. Remote Sensing of Environment 84, 471-475.