The EUMETSAT Satellite Application Facility on Land Surface Analysis (LSA SAF)

Validation Report (VR)

NORMALIZED DIFFERENCE VEGETATION INDEX

PRODUCTS: LSA-410 (ENDVI10)

Reference Number: SAF/LAND/VITO/VR_endvi/1.0
Issue/Revision Index: Issue 1.0
Last Change: 05/03/2013

DOCUMENT SIGNATURE TABLE

	Name	Date	Signature
Prepared by :	**VITO (E. Swinnen, H.Eerens)**	**05/03/2013**	
Approved by :	**Land SAF Project Manager**	**DD/MM/2013**	

DOCUMENTATION CHANGE RECORD

Issue / Revision	Date	Description:
Version 0.0	**2009**	**Theoretical comparison between METOP-AVHRR and SPOT-VGT.**
Version 0.1	**2011**	**General relationship established between NDVI from METOP-AVHRR and SPOT-VGT.**
Version 1.0	**05/03/2013**	**Version prepared for LSA SAF ORR.**

DISTRIBUTION LIST

Internal Consortium Distribution		
Organisation	**Name**	**No. Copies**
IPMA	**Luís Pessanha**	
IPMA	**Isabel Trigo**	
IPMA	**Isabel Monteiro**	
IPMA	**Sandra Coelho**	
IPMA	**Carla Barroso**	
IPMA	**Pedro Diegues**	
IPMA	**Teresa Calado**	
IPMA	**Benvinda Barbosa**	
IPMA	**Ana Veloso**	
VITO	**Herman Eerens**	
VITO	**Else Swinnen**	
VITO	**Tim Jacobs**	
VITO	**Sara Verbeiren**	
VITO	**Bart Deronde**	
VITO	**Lieven Bydekerke**	

External Distribution		
Organisation	**Name**	**No. Copies**
EUMETSAT	**Frédéric Gasiglia**	
EUMETSAT	**Dominique Faucher**	
EUMETSAT	**Lorenzo Sarlo**	
EUMETSAT	**Lothar Schueller**	
EDISOFT	**Teresa Cardoso**	
EDISOFT	**Carlos Vicente**	
EDISOFT	**Joaquim Araújo**	
GMV	**Mauro Lima**	
GMV	**José Freitas**	

13

Steering Group Distribution		
Nominated by:	**Name**	**No. Copies**
IPMA	**Pedro Viterbo**	
EUMETSAT	**Lorenzo Sarlo**	
EUMETSAT	**Harald Rothfuss**	
EUMETSAT	**Lothar Schueller**	
EUMETSAT	**Kenneth Holmlund**	
IPMA	**Pedro Viterbo**	
MF	**Jean-François Mahfouf**	
RMI	**Rafiq Hamdi**	
VITO	**Eric Gontier**	

14
15
16

Table of Contents

1 Introduction

This document provides the validation of the METOP-AVHRR NDVI data. The validation is done relatively against SPOT-VGT and a justification for this approach is provide in the next section. In the next section the VGT data are described. Section 3 provides a theoretical assessment of the different factors that can influence the similarity of the two NDVI data sets. Section 4 details the methodology used to perform the comparison. The results are shown and discussed in section 5, where the link to section 3 is made. Section 6 concludes the analysis.

1.1 Justification of the validation approach

The Normalized Difference Vegetation Index (NDVI) is a measure of vegetation photosynthetic activity which is based on the RED and NIR reflectances only (see Eq. 1).

$$NDVI = \frac{NIR-RED}{NIR+RED} \qquad \textit{Eq. 1}$$

Besides the surface conditions, these reflectances are also largely dependent on the specifications of the spectral bands of a sensor, defined by the spectral response functions (SRFs). This means that there is not a true NDVI for a specific surface, but only an NDVI, according to sensor-specific SRFs (see e.g. Steven, Malthus, Baret, Xu, & Chopping, 2003; A. Trishchenko, 2009; A. P. Trishchenko, Cihlar, & Li, 2002). For this reason, it is difficult to validate the NDVI using in situ measurements. In situ measurements should be sufficiently large (> 1km²), and observed with a sensor with the same SRFs or measured by a hyperspectral sensor from which the signal can be filtered using the SRFs afterwards.

For the validation of the 10-daily composited of METOP-AVHRR NDVI data, we did not have these in situ measurements available. Therefore, the METOP-AVHRR NDVI is validated through an inter-comparison with the NDVI derived from SPOT-VEGETATION (VGT).
This is a valid approach for a number of reasons:

- The interband and absolute calibration accuracy of VGT is high due the combination of an onboard calibration device and vicarious calibration. This calibration accuracy makes sure that the quality of the 10-daily composites is high and stable over time.
- The platform is well-stabilized and is not subject to orbital drift for the analysis period used.
- VGT has a similar spatial resolution compared to METOP-AVHRR, which facilitates the comparison.
- The pre-processing approach of VGT is similar to the one used to derive the 10-daily composites of METOP-AVHRR NDVI.
- Because the NDVI depends on the sensor characteristics, it is often used for a relative analysis of its behaviour over time, contrary to quantitative assessments.

1.2 Related documents

ADD SAF_LAND_VITO_ADD_ENDVI_1.0.pdf

ATBD	SAF_LAND_VITO_ATBD_ENDVI_1.0.pdf
OP	SAF_LAND_VITO_OP_ENDVI_1.0.pdf
PUM	SAF_LAND_VITO_PUM_ENDVI_1.0.pdf

2 Data description

The two SPOT-VEGETATION sensors (VGT1 and VGT2) provide daily global observations since April 1998. The imagery are distributed in the form of 10-daily composites (S10), 1-daily composites (S1) and segments (P-products) and are widely used in research activities and operational monitoring (e.g. Geoland2, MARS in Baruth et al. 2008) for crop monitoring (e.g. Weiss, Baret, Eerens and Swinnen, 2010), burnt area detection (e.g. Zhang, 2003, Tansey et al., 2008), vegetation dynamics monitoring (e.g. Carreiras and Pereira, 2005), etc.
From April 1998 until the end of January 2003, the VGT data distributed were acquired by the VGT1 sensor. Since then, the data from VGT2 is distributed to the users.

The pre-processing steps of the VGT imagery is explained in more detail on http://www.vgt.vito.be. A short summary of the pre-processing of VGT is discussed in the next section and compared to that of METO-AVHRR.

3 Factors affecting the agreement between NDVI of different sensors

This section describes the differences between the two satellites/sensors and their processing and briefly discusses how these differences can affect the resulting data. It builds further on material presented in Swinnen and Veroustraete (2008) and Eerens *et al*. (2009).
In section 5, an attempt is made to attribute the observed differences between the two NDVI data sets to these factors (where possible).

3.1 Orbital characteristics: Overpass time

Both METOP and SPOT were launched in a circular, near-polar, sun-synchronous orbit at an altitude of between 817 and 832 km with an orbital inclination angle of around 99°. The overpass time is similar, and 1 hour earlier for METOP compared to SPOT. The overpass time determines surface illumination conditions and hence the observed radiance by the satellite. The one hour difference in overpass time will affect the imagery very little.
For both sensors, the orbital behaviour is controlled and neither of the satellite experience extensive orbital drift. Orbital drift is the process in which the satellite's overpass time is gradually shifted to a later time, and is one of the major problems for the data from NOAA-AVHRR. For this sensor, the imagery is therefore acquired progressively later in the day with increasing satellite age, causing a systematic change in illumination conditions.

	SPOT-VEGETATION	METOP-AVHRR
ORBITAL CHARACTERISTICS		
Altitude (km)	832	817
Inclination angle (degrees)	98.7	98.7
Equator crossing time (LST) – at descending node – at ascending node	 10:30	 9:30 21:30
Stability of the platform	No orbital drift	No orbital drift
SPECTRAL CHARACTERISTICS		
Spectral channels		
Blue	Ch1: 0.43 – 0.47	
Red	Ch2: 0.61 – 0.68	Ch1: 0.58 – 0.68
NIR	Ch3: 0.78 – 0.89	Ch2: 0.725 – 1.0
SWIR	Ch4: 1.58 – 1.75	Ch3a: 1.58 – 1.64 (day time)
MIR		Ch3b: 3.55 – 3.93 (night time)
TIR1		Ch4: 10.3 – 11.3
TIR2		Ch5: 11.5 – 12.5
Calibration of shortwave channels	Onboard calibration	Vicarious calibration *a posteriori* (Rao & Chen, 1999)
Atmospheric correction	SMAC	SMAC
Inputs for atmospheric correction	– 6-hourly water vapour from MeteoServices – Ozone climatology – AOD derived from blue band – Pressure from DEM	– 6-hourly water vapour from MeteoServices – Ozone climatology – AOD derived from climatology – Pressure from DEM
SPATIAL CHARACTERISTICS		
Swath width (km)	2250	2400
Total Earth scan angle (degrees)	101	110.8
Nominal resolution (km)	1.15	1.09
Maximum off-nadir resolution – Along track direction (km) – Across track direction (km)	 Not specified 1.7	 2.4 6.9
Attitude of satellite	Known	Known
PSF	Broad	Narrow
Scanning system	Pushbroom	Wiskbroom
Spatial resolution after resampling	1°/112 at equator	1°/112 at equator
Resampling method	Cubic convolution	Nearest neighbour

Sources:
METOP-AVHRR: http://www.ncdc.noaa.gov/oa/pod-guide/ncdc/docs/klm/html/d/app-d5.htm
http://www.eumetsat.int/HOME/Main/What_We_Do/EUMETCast/index.htm
http://oiswww.eumetsat.org/WEBOPS/eps-pg/AVHRR/AVHRR-PG-0TOC.htm
SPOT-VGT: http://www.vgt.vito.be/

3.2 Scanning system

The nominal (sub-nadir) resolution of both the VGT and AVHRR sensors is approximately 1x1 km. With a swath width of more than 2000 km, there is global coverage with a near daily frequency at the equator, to a daily frequency at higher latitudes. The difference in scanning system and optics between the AVHRR and VGT affects the off-nadir spatial resolution of the imagery. The AVHRR sensor is an across-track scanner, which scans the Earth in a series of lines, perpendicular oriented to the direction of the orbit. Each line is scanned from one side to the other, using a rotating mirror placed in front of a sensor. The mirror sweeps with a constant angular velocity, resulting in the same angular resolution for every measurement. The sensors instantaneous field-of-view (IFOV) remains the same, and when sweeping away from the nadir position, the distance to the Earth increases and so does the ground surface resolved by the satellite. Because of the large swath width, the Earth's curvature adds an additional panoramic distortion to the off-nadir pixels. This leads to large off-nadir spatial deformations and the bow tie effect (Meyer, 1996). The VGT scanning system operates with an array of 1728 detectors positioned in one row perpendicular to the satellite orbital track simultaneously scanning one full scan line (aka pushbroom scanner). There is a fixed geometrical relationship between the detector elements in each scan line and the ground resolution measured by the sensor, accounting for a large part of off-nadir pixel deformation. Each individual detector measures the energy for a single ground resolution cell and thus the size and IFOV of the detectors determines the spatial resolution of the system. The VGT data are acquired essentially distortion free up to a 50° off-nadir angle if the Earths curvature is not taken into account (SPOT VGT User Guide).

3.3 BRDF

Surface reflectance varies with illumination and viewing geometry for anisotropic surfaces, like most of the natural surfaces. A Bi-directional Reflectance Distribution Function (BRDF) describes this dependency for each surface type corresponding with a pixel. Correction for BRDF effects was considered by Cihlar *et al.* (1998, 2004) as one of the most important requirements for long-term and multi sensor time series analysis with satellite imagery.
One can assume that the BRDF-effect can be of a similar magnitude for VGT as well as for AVHRR (up to a viewing zenith angle of 45°), but not identical due to the difference in scanning system (especially for off-nadir conditions). The difference in overpass time could contribute to a difference in BRDF, but seen the small time difference this will be very limited.
The 10-daily composited of METOP-AVHRR and VGT are not normalized for BRDF effects.

3.4 Point spread function

An important factor influencing the spatial resolution of satellite imagery is the Point Spread Function (PSF) of the optical system. The PSF defines the propagation of radiation from a point source. It is the result of the physical, optical and electronic properties of the sensor and of the atmosphere at the time of image acquisition (Ruiz and Lopez, 2002). The PSF of VGT is narrower than those of the AVHRR sensors. For an identical nominal spatial resolution, a

broad PSF implies a larger area being sensed than the one implied by the IFOV. The result of a broader PSF is an increased auto-correlation or smearing between pixels of an image segment. Since the PSF of VGT is narrower than that of the AVHRR's, the smearing in a VGT image segment is smaller than that of the AVHRR.

3.5 Geometric accuracy

Geometric accuracy is another important issue to be considered when comparing time series of remote sensing data. Mis-registration, together with the previously discussed spatial characteristics, induces image blur in an image time series (Meyer, 1996). For VGT, a high absolute, multi-temporal and multi-band registration accuracy is obtained using an elaborate database of ground control points (GCP's). The absolute location accuracy is 330 m as root mean square error (RMSE) (Sylvander *et al.*, 2000). For VGT2, the stars tracker onboard SPOT5 allows for an accurate geometric modelling without any need for GCPs.
EUMETSAT provides very accurate lat/lon-planes for the data from the METOP-AVHRR sensor. This accuracy is achievable because the attitude of the satellite is known.
For METOP-AVHRR, the geolocational accuracy is of high quality and probably of the same magnitude as for SPOT-VGT.
The evaluation of the geometric accuracy is not part of this validation report, but is treated in the ATBD.

3.6 Spectral response functions

Figure 1 shows the Red and NIR spectral bands of VGT and AVHRR. These channels are of particular interest, because they are used to derive the NDVI, a broadly used vegetation index. The spectral response function (SRF) of the sensor describes what part of the electromagnetic spectrum is measured. Though similar, the SRF's show different shapes, central wavelength locations, bandwidths and degrees of overlap between respective channels, especially with respect to the transition from the chlorophyll absorption band to the foliage reflection band (0.68-0.72 μm) (Trishchenko *et al.*, 2002). It is clear that this leads to differences in NDVI among the sensors for the same surface. In general, smaller bandwidths lead to higher NDVI values. The VGT Red channel extends beyond the 0.7 μm limit, which has a significant impact on the Red reflectance as well as on the NDVI. Consequently, the effect of varying SRFs induces radiometric errors imposed on the natural variability in land surface attributes. The difference in SRF-definition of the Red and NIR channels between VGT and AVHRR is probably the largest source of difference between data of both sensors.

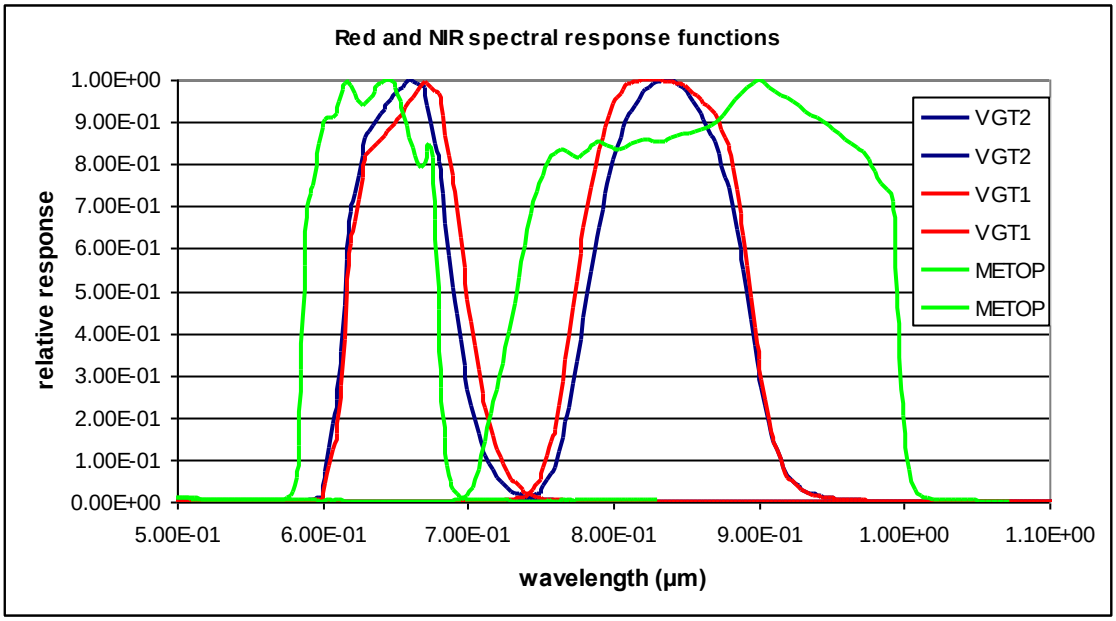


Figure 1 spectral response functions of Red and NIR for VGT1, VGT2 and AVHRR sensors

Trishchenko (2009) analyses this difference and provides adjustment functions to reduce them. However, it is better to accommodate for the difference by using biophysical parameters (e.g. fAPAR), extracted from the reflectance data. Most methods (e.g. CYCLOPES (Baret et al., 2007)) take sensor characteristics like the SRF into account for fAPAR extraction.

The AVHRR sensor has two thermal infrared bands. Thermal information is not acquired by the VGT sensor. The availability of thermal data facilitates cloud detection and leads to potentially different applications.

3.7 Calibration Accuracy

The radiometric performance of a sensor in the visible and near-infrared region usually degrades in orbit (e.g. Gutman, 1999). The VGT sensor, unlike the AVHRR, has an onboard calibration device for these channels, allowing accurate radiometric calibration (Henry and Meygret, 2000). For AVHRR, the estimation of calibration coefficients is performed *a posteriori*. Sensor response is then vicariously determined with stable terrestrial targets of which radiance can be measured or inferred. For METOP-AVHRR, the same vicarious calibration method is used as for NOAA-AVHRR to guarantee similarity between data from the same sensor family.

3.8 Atmospheric correction

Due to the different spectral band definitions, the influence of the atmospheric perturbation will be different for data from AVHRR and VGT. The broad NIR spectral band of AVHRR will suffer more from water vapour absorption, whereas that from VGT was specifically designed to avoid the 0.935µm water vapour absorption band. Van Leeuwen *et al.* (2006) investigated the influence of Rayleigh scattering, water absorption, ozone absorption, and aerosol optical thickness on multi-sensor NDVI (AVHRR, MODIS and VIIRS), and concluded that NDVI data continuity across sensors would be largely enhanced if adequate

atmospheric corrections are applied, because different sensors are differently affected by the atmosphere. Accurate atmospheric correction largely depends on the knowledge about the atmospheric conditions at the time of image acquisition.
The VGT and AVHRR imagery are processed with the same method (SMAC (Rahman and Dedieu, 1994)) using identical atmospheric inputs, except for aerosol optical depth. This approach limits differences due to processing strategies.

4 Intercomparison method

4.1 Data preparation

Global, ten-daily NDVI composites of VGT and AVHR from the entire years 2008 to 2011 were used for the intercomparison. These images were first reduced in size through a systematic sample and secondly using specific constraints on accompanying data layers, such as viewing angles.

4.1.1 Spatial filtering

The NDVI composites from both METOP-AVHRR and VGT have identical spatial resolution (1km²) and follow the same framing. A systematic subsample of the global images was taken using a moving window of 21x21 pixels from which the centre pixel was selected. In this way, all global conditions are present in the evaluation data set, but the processing time is considerably reduced. The size of the thinned images is 1920 columns and 698 rows.
This was done not only for the NDVI, but also for a number of accompanying layers, necessary in the sampling approach explained in the next section (angles, registration date, land cover).

4.1.2 Sampling design

The thinned images were further subjected to different sampling designs, in order to test different influences on NDVI differences mentioned in section 3.
The following constraints were used on each image pair:

(1) Same day of registration within the 10-day period.
(2) Identical value before and after smoothing of the NDVI time series per pixel. Smoothing is done to remove non-detected clouds and also accounts for spurious high observations due to anisotropy effects. It also interpolates missing values identified in the status map. The smoothing method used is the modified Swets method (Swets, Reed, Rowland, & Marko, 1999) and (Klisch, Royer, Lazar, Baruth, & Genovese, 2006). To assess the similarity of the NDVI values from both sensors, it is important to select original values, and not interpolated ones. Therefore, the original and smoothed values per pixel were compared and only identical values are selected.
(3) Limit the observation angle close to nadir. This is done by selecting pixels with a viewing zenith angle (VZA) less than 30°. This threshold is still broad, but otherwise an insufficient sample size was obtained.
(4) Always looking in the same direction, i.e. east or west of nadir. This was done using the viewing azimuth angle (VAA > 180° or < 180° for both sensors).

Per NDVI image pair, several masks were created that identified the pixels that met the different sets of conditions according to .

Table 1 Sampling constraints

constraint	all	noVZA	noVAA	noVA
(1) Same day	■	■	■	■
(2) Identical value before and after smoothing	■	■	■	■
(3) VZA<30°	■		■	
(4) VAA<180° or VAA>180°	■	■		

4.1.3 Intercomparison metrics

The verification method was defined to be the R between the NDVI from METOP-AVHRR and VGT. In this report, the results from different intercomparison metrics are reported, because each provide specific information on the similarity between the two data sets.

The metrics used are:

- R
- Geometric mean regression
- Agreement coefficient (AC), systematic agreement (ACs) and unsystematic agreement (ACu)
- Root Mean Squared Error (RMSE)
- Mean Bias Error (MBE)
- Scatterplots

Pearson product-moment correlation coefficient (R)

In statistics, the Pearson product-moment correlation is a measure of the correlation (linear dependence) between two variables X and Y, and its result ranges between 1 and -1. It expresses the strength of a linear dependence between two variables. A value of 1 implies a perfect linear relationship, with all points on a straight line for which Y increases when X increases. A value of -1 represent also a perfect linear relationship, but for which Y decreases when X increases.

R was considered a good accuracy measure, because it measures the linear relationship even if this relationship is not equal to the line of perfect fit (Y=X), which is to be expected seen the differences in sensor and processing characteristics. Nevertheless, it is important to assess the consistency of the relationship over time.

Geometric mean regression

Model I regression models (e.g. Ordinary Least Squares) are appropriate for predicting one data set out another and one data set is assumed error-free. This is not the case when comparing two similar data sets of remote sensing images, because both are subjected to noise. In this case, model II regression models are more suited. Different regression models II exist, such as the geometric mean (GM), orthogonal and OLS bisector regression models. The difference between the models is in the way the errors are minimized

- OLS minimizes the sum of the squared vertical distances (errors on Y) from the data points to the regression line
- GM minimizes the sum of the products of the vertical and horizontal distances (errors on Y and X)
- Orthogonal regression minimizes the sum of the squared perpendicular distance from the data point to the line (errors on Y and X)
- OLS bisector regression bisects the angle between the Y on X OLS regression line and the X on Y OLS regression line.

Each of the model II regression analysis methods has its merits and deficiencies. A comparison of MODIS NDVI with AVHRR NDVI in which the four different methods were used, showed that the model II approaches results were very similar, and that the difference was the largest compared to the simple OLS regression method (Ji, Gallo, Eidenshink, & Dwyer, 2008). Therefore, the choice is somewhat arbitrary. Here, the **GM regression** model was used because of its simplicity.

The GM model is of the form

$$Y = a + b \cdot X$$

With

$$b = sign(R)\frac{\sigma_Y}{\sigma_X} \qquad \textbf{(GMRslo)}$$

$$a = Y - b \cdot X \qquad \textbf{(GMRint)}$$

The **agreement coefficient (AC)** is based on the evaluation of the distances between the actual observations, the line of perfect fit and the regression line between both data sets. Thus, there are three distances among the points and the two lines:

- The distance between the point (X_i,Y_i) and the line of perfect fit.
- The distance between the point (X_i,Y_i) and the regression line. This is the unsystematic difference between X_i and Y_i.
- The distance between the 1-1 line and the linear regression line. This is the systematic difference between X_i and Y_i. The systematic difference is attributed to the fixed difference between the two data sets.

To make the coefficient unit independent and bounded, the distance is standardized by a quantity referred to as the potential difference, which is measured by the range of X and Y.

$$AC = 1 - \frac{\sum_{i=1}^{n}(X_i - Y_i)^2}{\sum_{i=1}^{n}(|\bar{X} - \bar{Y}| + |X_i - \bar{X}|)(|\bar{X} - \bar{Y}| + |Y_i - \bar{Y}|)}$$

Or

$$AC = 1 - \frac{SSD}{SPOD}$$

With SSD the sum of the squared differences and SPOD the sum of the potential differences.

The agreement coefficient evaluates the total agreement, which includes the systematic and unsystematic agreement between the two data sets. Using the geometric mean regression line,

the SSD can be further partitioned into the **systematic sum of product-difference (SPD_s)** and the **unsystematic sum of product-difference (SPD_u)**.

$$SPD_u = \sum_{i=1}^{n} \left(|X_i - \hat{X}_i|\right)\left(|Y_i - \hat{Y}_i|\right)$$

With $\hat{X}_i$ and $\hat{Y}_i$ according to the GM regression line.

Then, $SPD_s = SSD - SPD_u$.

The systematic and unsystematic agreement coefficients are then defined as follows:

$$AC_s = 1 - \frac{SPD_s}{SPOD}$$

$$AC_u = 1 - \frac{SPD_u}{SPOD}$$

The agreement coefficient could provide more information regarding the strength of the linear relationship, but also on the evolution of the form of the regression line over time.

Whereas the previous measures express similarity, it is equally important to investigate the magnitude of the differences between the data sets. Two measures were selected, the root mean squared error and the mean bias error.
The **Root Mean Squared Error (RMSE)** measures how far the difference between the two data sets is from 0 and is defined as

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i - Y_i)^2}$$

The **Mean Bias Error (MBE)** measures the average actual difference between two data sets and positive and negative differences between observations. It is defined as

$$MBE = \frac{1}{n}\sum_{i=1}^{n}(X_i - Y_i) = \bar{X} - \bar{Y}$$

To visually inspect the agreement/similarity of the two data sets, **scatterplots** were created, on which the geometric mean regression line is displayed.

For all analysis, the number of pixels of the sample is always taken as a quality criterion. Results based on less than 1000 pixels are omitted from the analysis.

4.1.4 Intercomparison approach on the NDVI time series

The metrics were derived for different analysis schemes. There are in essence two different ways to compare the two time series:

(1) Based on a sample derived from one of more paired scenes
(2) Based on paired observations of the time series per pixel.

The first approach provides the measures that are valid for a spatial aggregation of paired observations, and the second provides the metrics per pixel based on a sufficiently long time series. The first can be represented in graphs, the second as images.

For the first approach, different alternatives were used, i.e.

(a) Overall: the entire time series 2008-2011, all paired observations that are selected in the different sampling designs (see 4.1.2)
(b) Per year: same as overall, but then for each calendar year separately
(c) Per month: the three composites per month for the four years are taken as a sample
(d) Per scene
(e) For all land covers: this is done for (a), (b) and (c)
(f) Land cover specific: cropland, grassland, shrubland, forest. This is done only for (a) and (b)

For the different analysis schemes and sampling designs, the sample size was controlled to be equal through a random sample of all selected paired observations. There was no control of the sampling on latitude or other parameters.

The land cover is defined by a simplified version of the GLC2000 land cover map. The definition of the simplified classes is as follows:

- Cropland (GLC2000 classes 16-18)
- Grassland (GLC2000 classes 13-14)
- Shrubland (GLC2000 classes 11-12)
- Forests (GLC2000 classes 1- 6)

5 Results and discussion

The results for the sampling design using all conditions (all) are discussed first. Subsequently, these results are compared with those obtained with other sampling designs and the differences are related to a number of the factors that affect the agreement described in section 3.
The focus of this validation report is on the consistency over time, rather than the magnitude of the agreement. Nevertheless, a high agreement is important to be able to assess the consistency over time with sufficient reliability.

5.1 Sample with all conditions (all)

Table 2 provides a summary of the metrics used for the evaluation of the disagreement/similarity between the NDVI of METOP-AVHRR and VGT. The Pearson correlation coefficient R is always higher than 0.95 except when only forests are considered. The slope of the geometric mean regression line (GRM slope) is close to one, and stable over

the four years. The agreement coefficient AC shows somewhat more variation between the different years and land cover types, but also shows a good correspondence between the two data sets. The unsystematic agreement ACu (i.e. the deviation of the single points wrt to the GM regression line) is quite high. The systematic agreement ACs (i.e. the deviation of the regression line from the line of perfect fit) is also high, but slightly lower than the systematic agreement (ACs). The root mean squared error (RMSE) is around 0.09 NDVI.

Table 2 Summary of the agreement/disagreement metrics between the NDVI from METOP-AVHRR and VGT for the sample design with all conditions for different periods and land covers.

Period	Land cover	R	GMR intercept	GRM slope	AC	ACs	ACu	RMSE
4 years	All	0.975	0.064	1.008	0.915	0.947	0.968	0.087
2008	All	0.978	0.064	1.012	0.917	0.946	0.971	0.088
2009	All	0.973	0.071	1.002	0.908	0.942	0.966	0.091
2010	All	0.978	0.061	1.013	0.922	0.950	0.971	0.085
2011	All	0.977	0.062	1.005	0.920	0.950	0.970	0.083
4 years	Cropland	0.961	0.083	0.965	0.891	0.939	0.952	0.087
4 years	Grassland	0.976	0.068	1.012	0.905	0.933	0.972	0.087
4 years	Shrubland	0.961	0.082	0.981	0.868	0.912	0.956	0.088
4 years	Forest	0.932	0.116	0.936	0.812	0.883	0.929	0.095

When looking at these metrics per land cover, then a lower agreement and higher error is observed for pixels belonging to the forest class. The GM regression slope is more variable than considering all land cover classes. This is normal, since each of the classes has a dominant NDVI range, whereas the sample of the overall analysis contains the largest possible range of NDVI-values, which provides a more stable relationship.

Figure 2 shows the scatterplot and the geometric mean regression line between the NDVI from both data sets for all land covers and all years. It can be observed that the relationship between the two NDVI data sets is not entirely linear and that the variability of the points is higher for higher NDVI values.

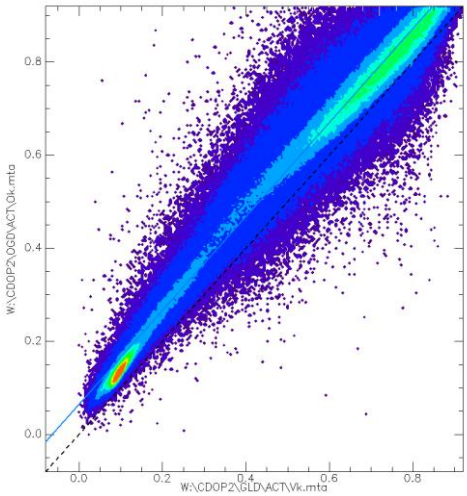


Figure 2 Scatterplot between the NDVI of VGT (X) and of METOP-AVHRR (Y) for all land covers and all years. The dotted black line represents the line of perfect fit, the blue line is the GM regression line.

The scatterplots for all years, but per land cover type are shown in Figure 3. Grasslands and shrublands show the strongest linear relationship. For forests, the majority of the points are associated with high NDVI values and the variability is high, which results in a higher RMSE and a lower R and AC.

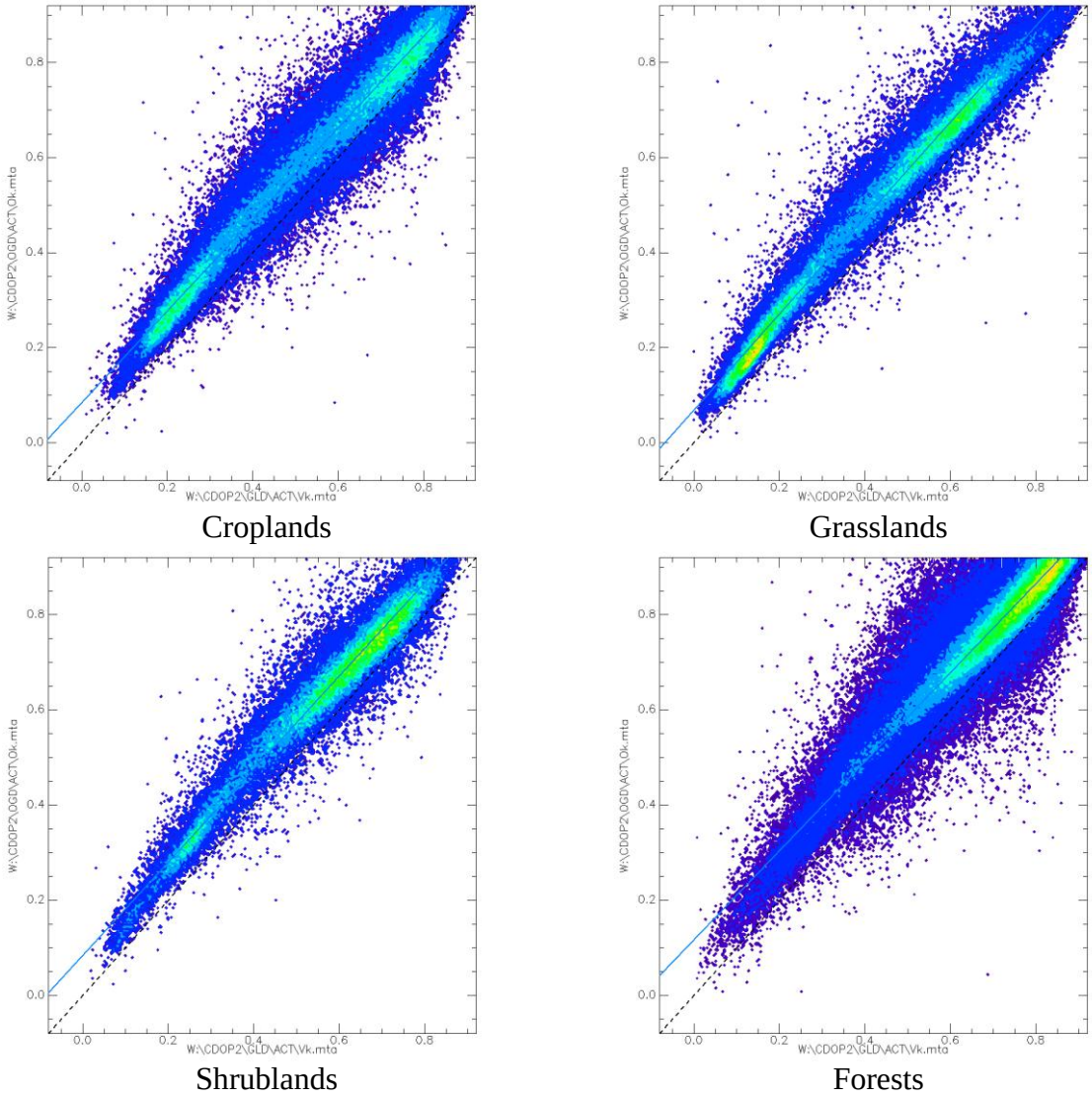


Figure 3 Scatterplot between the NDVI of VGT (X) and of METOP-AVHRR (Y) for all years, but for the differen land covers. The dotted black line represents the line of perfect fit, the blue line is the GM regression line.

The results for the different years and land cover types is also shown in Figure 4. Also here, the different metrics are stable over the four years and the forest class is associated with the largest differences. Also here, the agreement coefficients show a larger variability, suggesting that these metrics are more sensitive to variations in agreement than other metrics (such as R). Except for forests, the R is always higher than 0.95. Also here, the unsystematic agreement is very high, indicating a large correspondence.

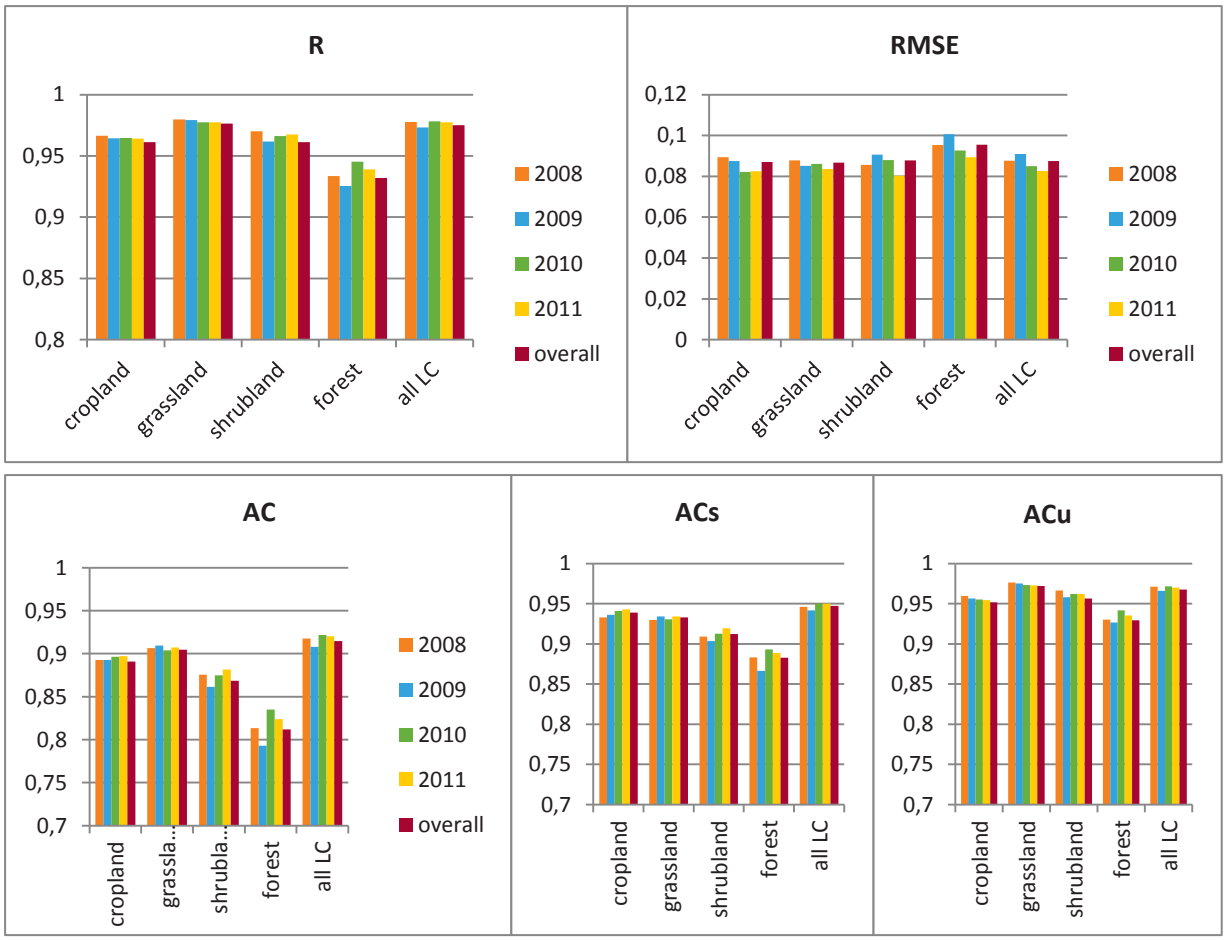


Figure 4 R, RMSE, AC, ACs and ACu between the NDVI from METOP-AVHRR and VGT for the sample design with all conditions for different periods and land covers.

When looking at the metrics over the different months (Figure 5), it is observed that the R and AC measures show some seasonality and are lowest around October. Also here, the AC-measures shows a larger variability compared to R.

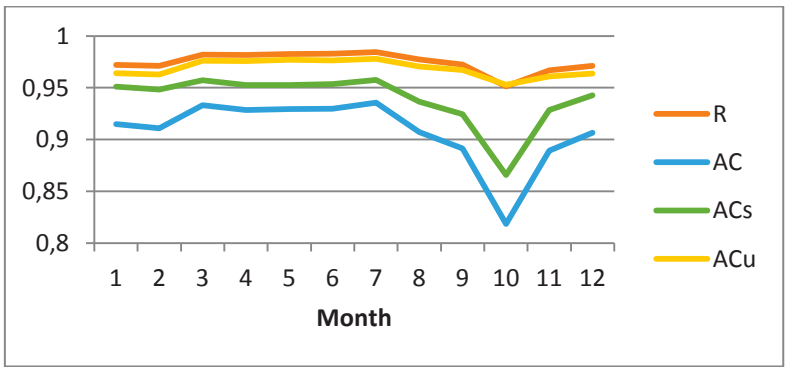


Figure 5 R, AC, ACu and ACs between the NDVI of METOP-AVHRR and VGT for all four years, all classes, but per month.

The slope of the GM regression deviates more from in October, Novermber and December (Figure 6). For the remainder of the year, this is close to 1.

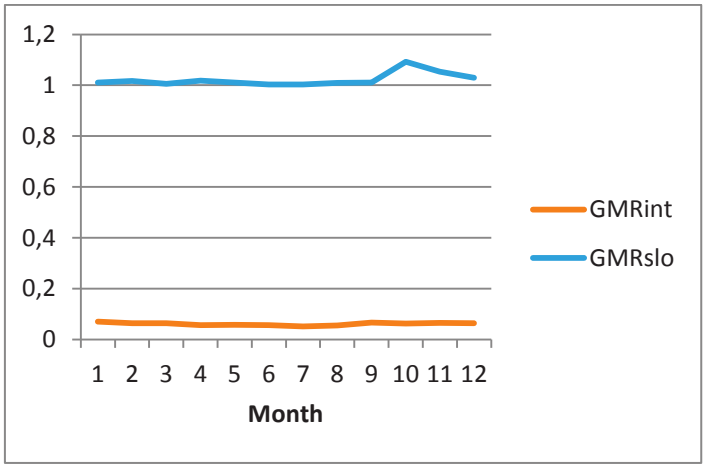


Figure 6 Intercept (GMRint) and slope (GMRslo) of the GM regression between the NDVI of METOP-AVHRR and VGT for all years, all land cover types, but per month.

The same tendency is observed when looking at the results of the subsequent image pairs (see Figure 7, Figure 8, Figure 9).

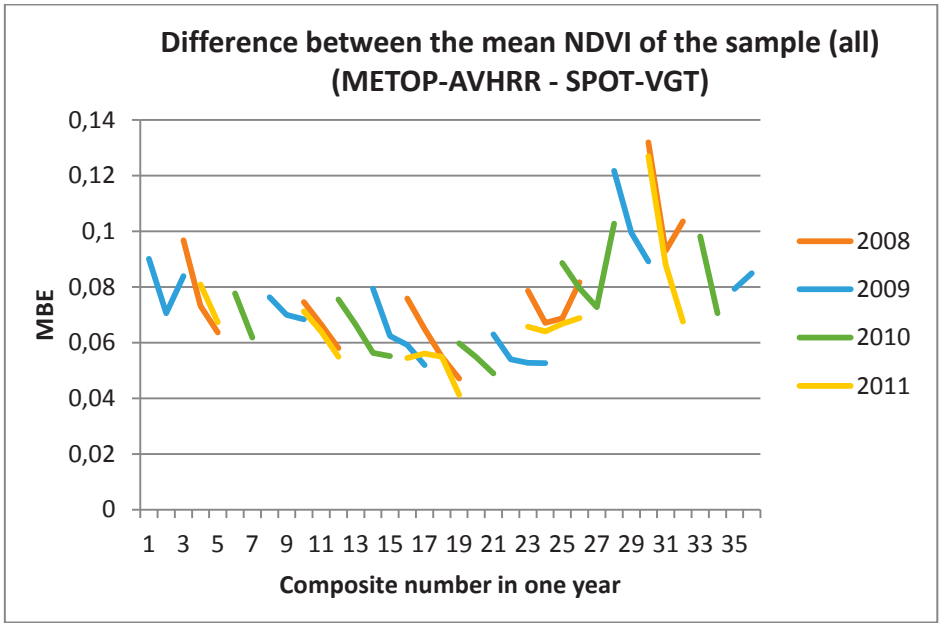


Figure 7 MBE between the NDVI of METOP-AVHRR and VGT for all years, all land cover types, but per image pair

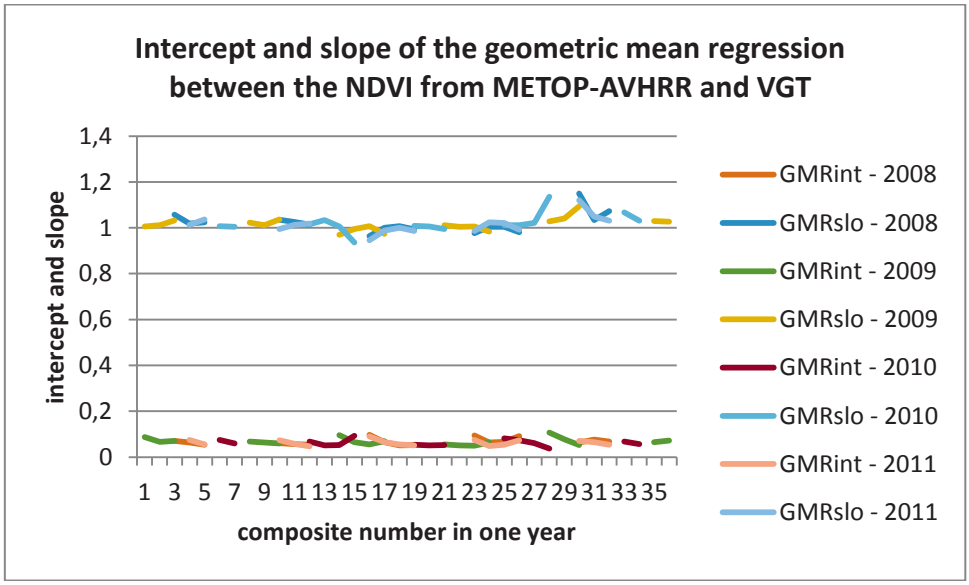


Figure 8 Intercept (GMRint) and slope (GMRslo) of the GM regression between the NDVI of METOP-AVHRR and VGT for all years, all land cover types, but per image pair.

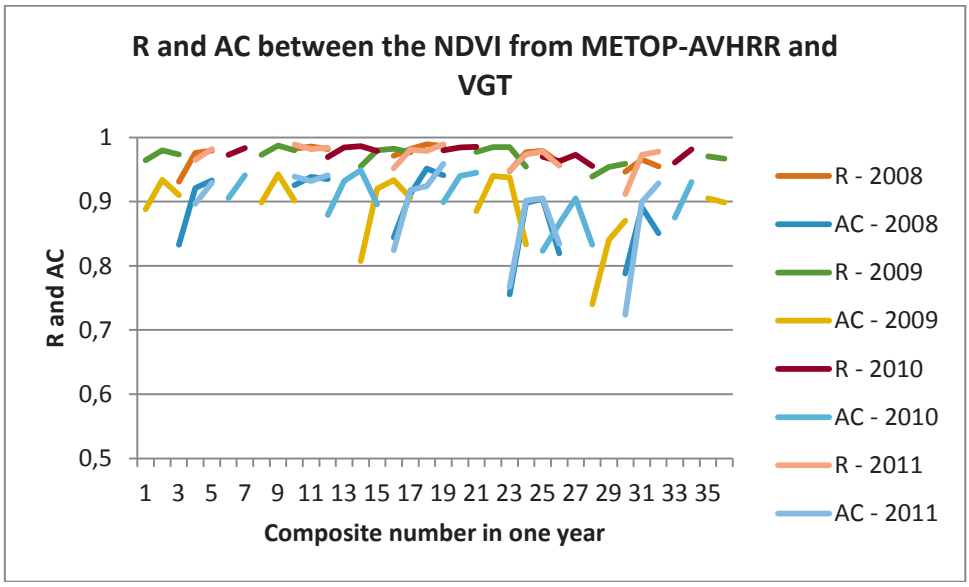


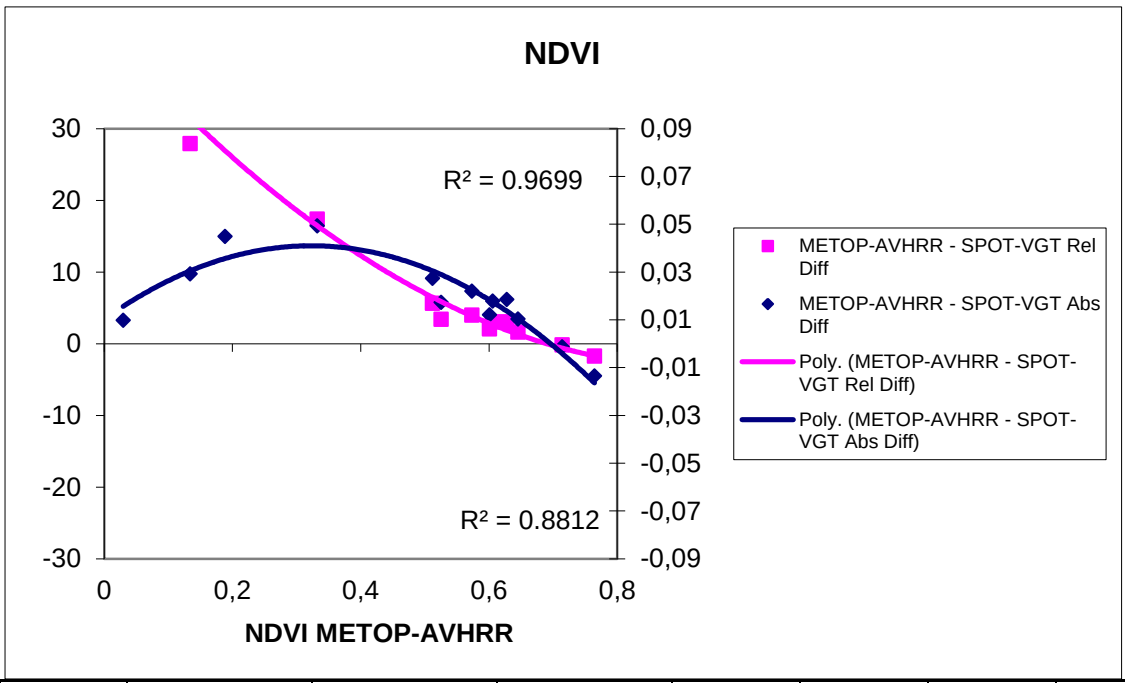
Figure 9 R and AC between the NDVI of METOP-AVHRR and VGT for all years, all land cover types, but per image pair

From the analysis of the NDVI-pairs selected using the most restrictive sample design, we observe that the similarity of the NDVI between METOP-AVHRR and VGT is generally high, but more importantly, consistent over the years, and throughout the majority of the year (except Oct, Nov and Dec). This is further investigated in the next section, where the disagreement is interpreted in terms of the influencing factors on the similarity discussed in section 3.

5.2 Effect of the spectral response functions (SRFs)

The scatterplots shown in Figure 2 and Figure 3 show a near-linear relationship between the NDVI of METOP-AVHRR and VGT. This non-linearity is largely due to the difference in SRFs. To assess this, two different approaches were followed based on the work published by Trishchenko (2009). He published empirical correction functions based on the SRFs of different sensors and a spectral library of global biomes. These correction functions are-second-order polynomials that are a function of the NDVI. One drawback is that the correction functions are all estimated relative to sensor responses of NOAA-9 AVHRR, which means that the shape of the correction function between METOP-AVHRR and VGT directly cannot be assessed, and that both datasets have to be corrected to test this hypothesis.

First, empirical correction functions based on a spectral library derived from different vegetation types in southern Africa was used to calculate new empirical correction functions between NDVI from METOP-AVHRR and VGT directly. This was done on surface reflectances, which is not entirely correct as an approach, but this is sufficient to evaluate the shape of the differences between the two NDVI data sets.



	from	to	a	x	x²	r²	σ
absolute	METOP-AVHRR	SPOT-VGT	0.037532	-0.03286	-0.00165	0.111917	0.011558
	SPOT-VGT	METOP-AVHRR	-0.04063	0.056844	-0.02825	0.079219	0.011135
relative	METOP-AVHRR	SPOT-VGT	49.95806	-180.731	172.9135	0.935961	5.490071
	SPOT-VGT	METOP-AVHR	R -34.542648	125.149	-123.119	0.910955	3.445695

Figure 10 Empirical correction fucntions between the NDVI of METOP-AVHRR and VGT according to the approach of Trishchenko (2002). Left axis : relative difference in % (upper R²), Right axis : absolute difference in NDVI units (lower R²).

Figure 10 shows a positive absolute difference between the NDVI data sets (METOP-AVHRR higher) and the opposite for low and high NDVI values. This is in line with the observations from the actual data set comparison and therefore, the empirical correction functions from Trishchenko (2009, 2002) were applied on both data sets and the results evaluated.

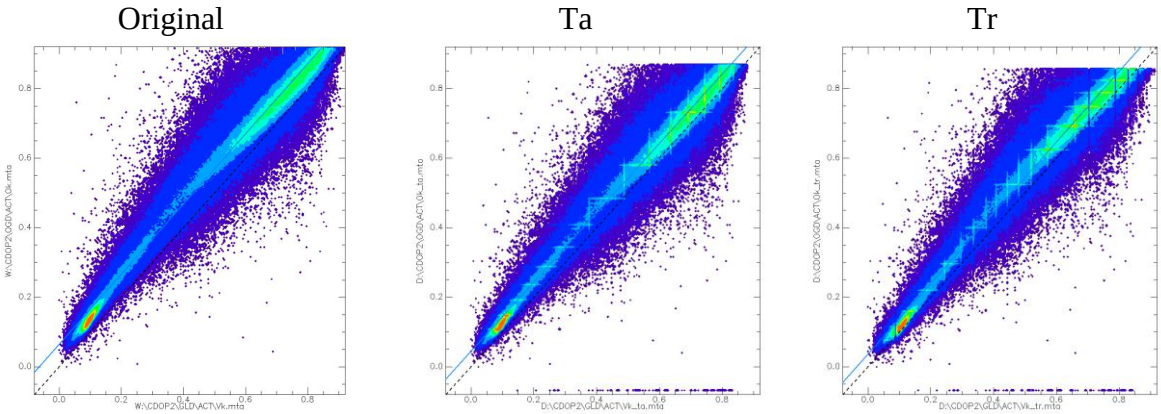


Figure 11 Scatterplots between the NDVI of METOP-AVHRR (Y) and VGT (X) for all years, all land covers and all constraints in the sampling design. Left : original, middle : absolute correction functions applied on both data sets, right : relative correction functions applied on both data sets.

Figure 11 shows that after application of the two sets of empirical correction functions, the agreement is slightly more linear, especially for the absolute correction functions (Ta).
For the absolute correction functions, the GM regression slope is closer to 1, and has a lower difference (RMSE in Figure 15 and MBE in Figure 14). The AC metric is significantly higher, especially the systematic difference (which was to be expected), but the R remains unchanged. This is not surprisingly because is only expresses the relationship between the two NDVI data sets, irrespective of the position of the regression line.

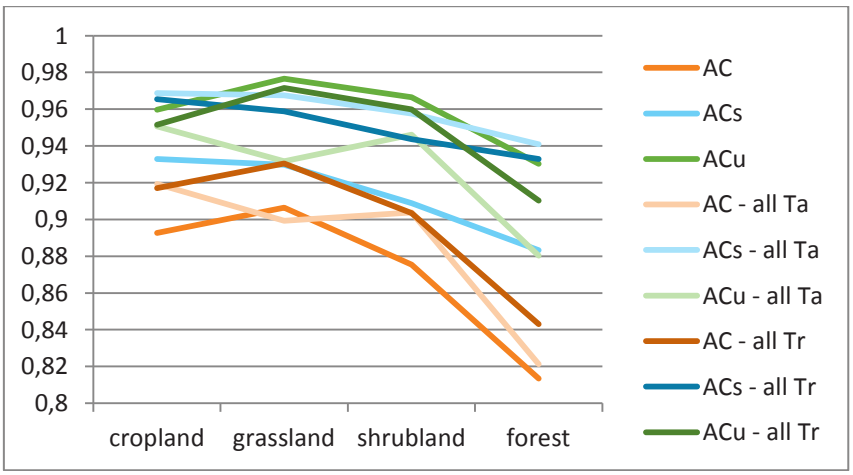


Figure 12 Agreement coefficients between the NDVI of METOP-AVHRR and VGT for all years, per land cover type, for the original data sets and for the NDVI data sets corrected for differences in SRF according to the absolute (Ta) and relative (Tr) correction functions of Trishchenko (2002, 2009)

The relative correction results in a slope that is not closer to 1, but the AC metrics are also higher, especially the systematic one. Like the absolute correction, the results have also lower differences (RMSE in Figure 15 and MBE in Figure 14). The different land cover types are differently affected by the correction functions (Figure 12).
For both correction functions, there is still a seasonal pattern in the differences that is not due to the SRFs (Figure 13).

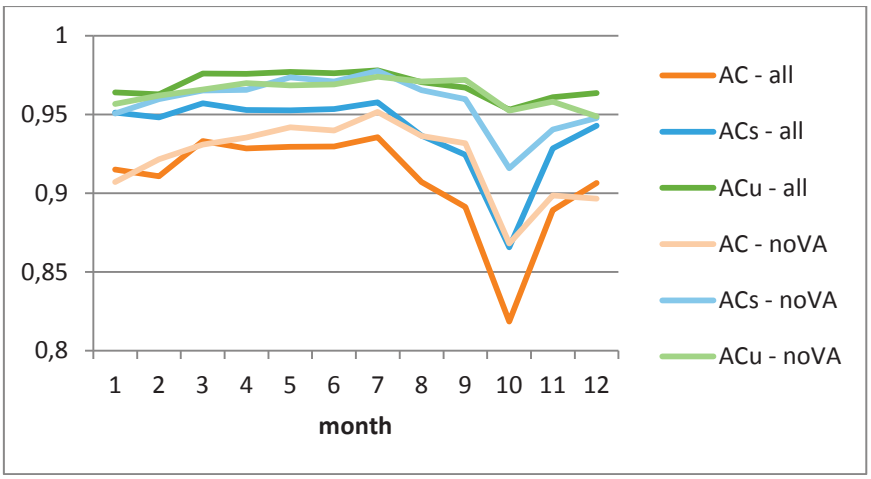
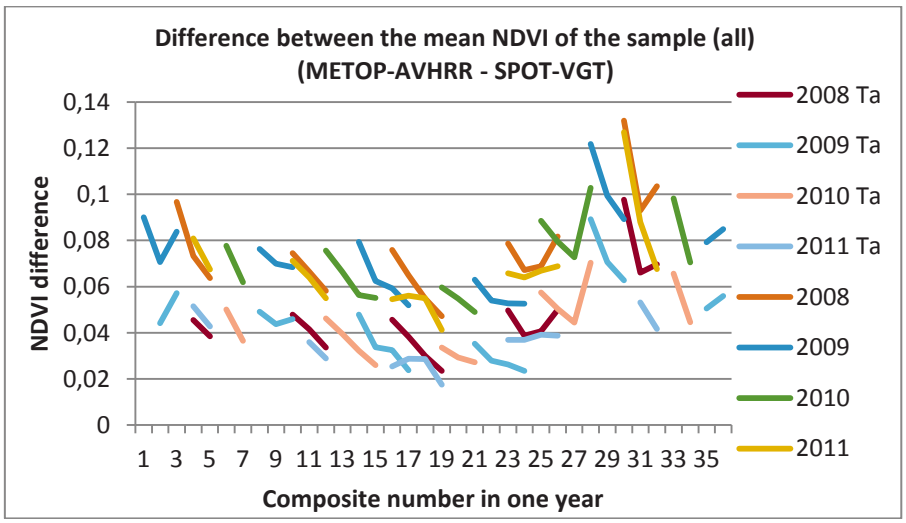


Figure 13 Agreement coefficients between the NDVI of METOP-AVHRR and VGT for all years, all land cover types, but per month, for the original data sets and for the NDVI data sets corrected for differences in SRF according to the absolute (Ta) and relative (Tr) correction functions of Trishchenko (2002, 2009)



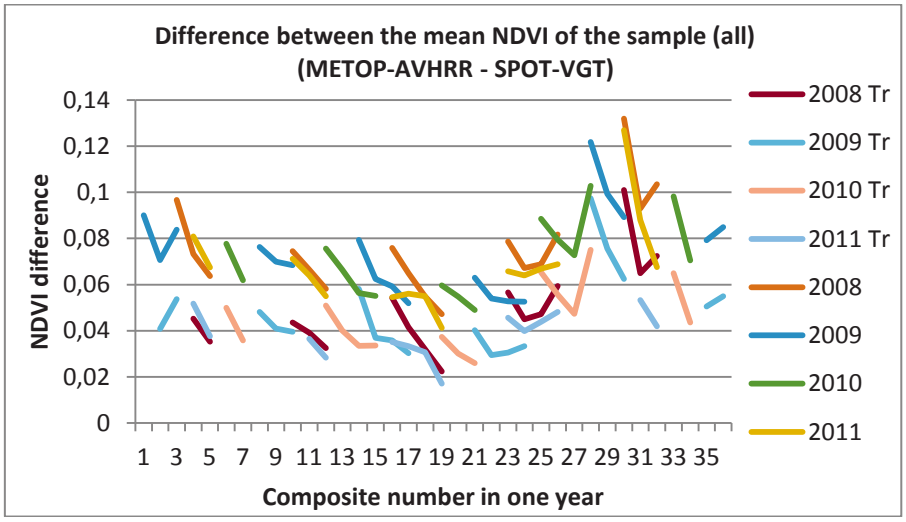


Figure 14 MBE between the NDVI of METOP-AVHRR and VGT for all years, all land cover types, but per month, for the original data sets and for the NDVI data sets corrected for differences in SRF according to the absolute (Ta) and relative (Tr) correction functions of Trishchenko (2002, 2009)

We can conclude that part of the difference between the NDVI of both sensors is due to differences of the their SRFs.

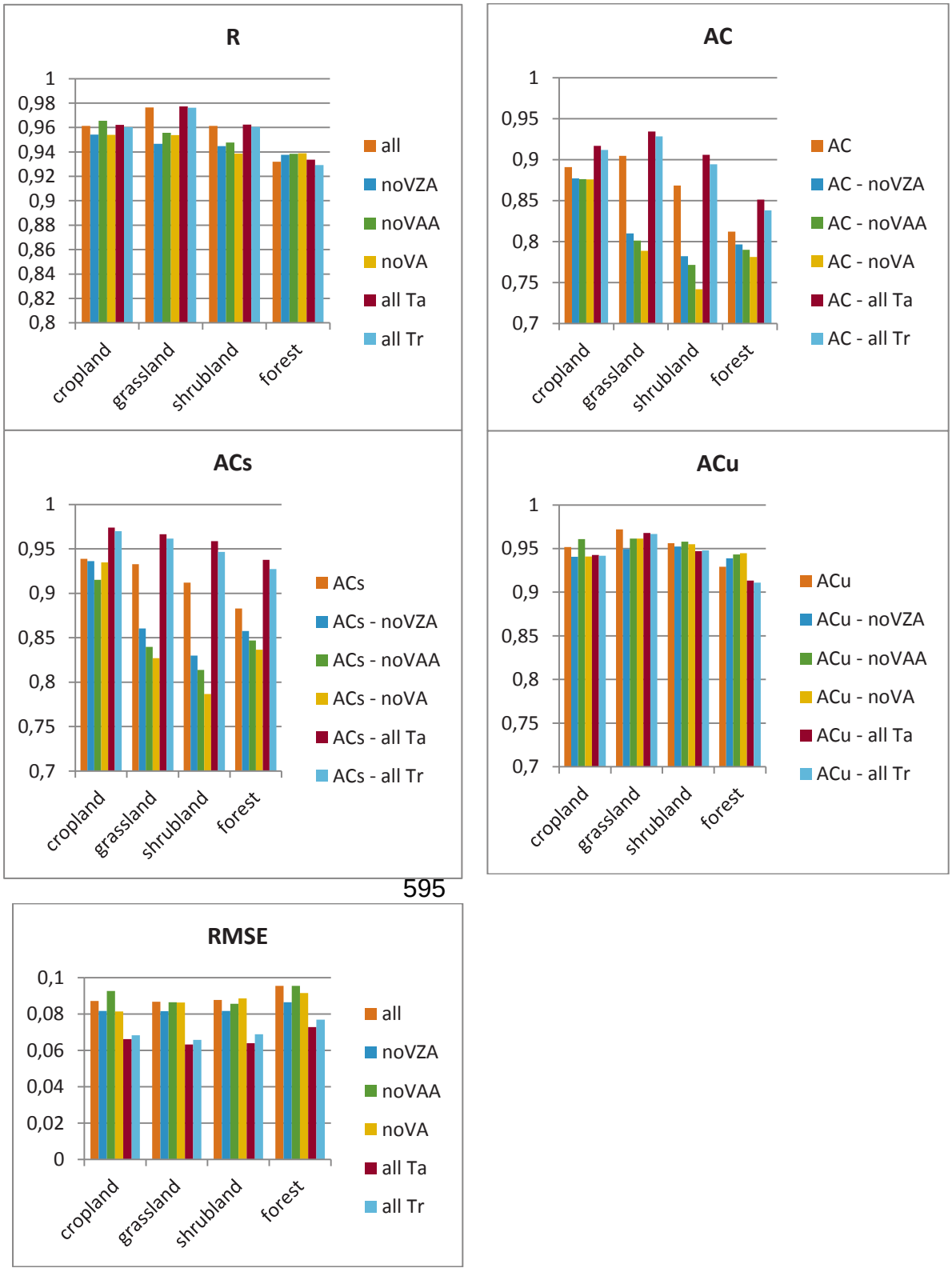
5.3 BRDF-effects

The effect of viewing angles was mainly investigated. Sun position is predominantly determined by the overpass time, the day in the year, the latitude and partly by the position of a pixel in the image. By limiting the viewing zenith angle (VZA) in the sample design, the sun zenith angle (SZA) is also limited to a smaller range. Likewise, but limiting the viewing azimuth angle (VAA), the SZA is also limited. Therefore, the solar angles were not taken into account explicitly.

The sample designs noVZA does not limit the VZA, noVAA does not limit the viewing direction and noVA does not limit the VZA or VAA. The same sample size was taken for all designs.

The results for all four years combined are summarized in Figure 15. Similar results were obtained for the years separately (not shown) and the results were stable over the different years. The R is adversely affected for the sample designs that have less constraints on the viewing geometry, and more when the constraint on VZA is not used (noVZA, noVA). The differences between the different sample designs is larger when considering the AC measures. Here, is becomes apparent that the viewing geometry mainly influences the systematic agreement, and not the unsystematic agreement. The RMSE shows slightly larger errors compared to the sample using all constraints.

We can conclude that part of the difference between the NDVI of both data sets is related to viewing geometry, and possibly also sun position.



595

596

Figure 15 R, AC, ACs, ACu, and RMSE between the NDVI from METOP AVHRR and VGT for all years, per land cover and for different samling designs (all, noVZA, noVAA, noVA) and for the NDVI data sets corrected for differences in SRF according to the absolute (Ta) and relative (Tr) correction functions of Trishchenko (2002, 2009).

5.4 Atmospheric correction

Atmospheric correction can have a large impact on the similarity between different data sets (van Leeuwen et al., 2006). For the processing of the METOP-AVHRR 10-daily composites, the same atmospheric correction method was used as for VGT, i.e. SMAC (Rahman & Dedieu, 1994), (Berthelot & Dedieu, 1997). In addition, the same inputs for ozone, and water vapour is used for both data sets. The only data set that differs is the aerosol optical depth (AOD). For METOP-AVHRR, a simple bell-shaped latitudinal function with the peak above the equator is used, whereas for VGT this is estimated from the images directly.

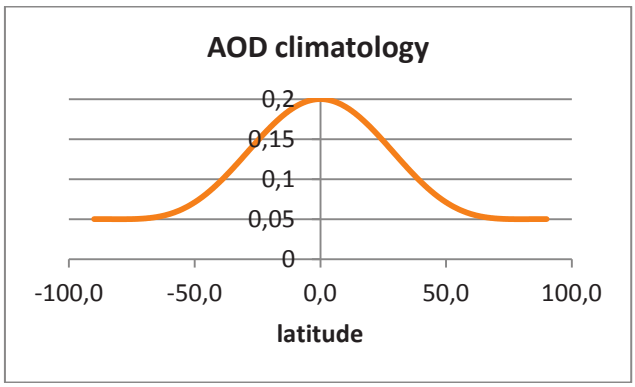


Figure 16 AOD climatology used for atmospheric correction of METOP-AVHRR RED and NIR bands.

The data from VGT are not available to compare to. Therefore we compare this function with the one assessed by (Hsu et al., 2012), who estimated the AOD from SeaWifs data and analysed seasonal effects and trends over the years. Figure 17 is copied from their publication and shows that the general AOD peak is not above the equator, but above the northern hemisphere, there is clear seasonality of the AOD and there is variability over the years. The used AOD climatology is thus not representative.

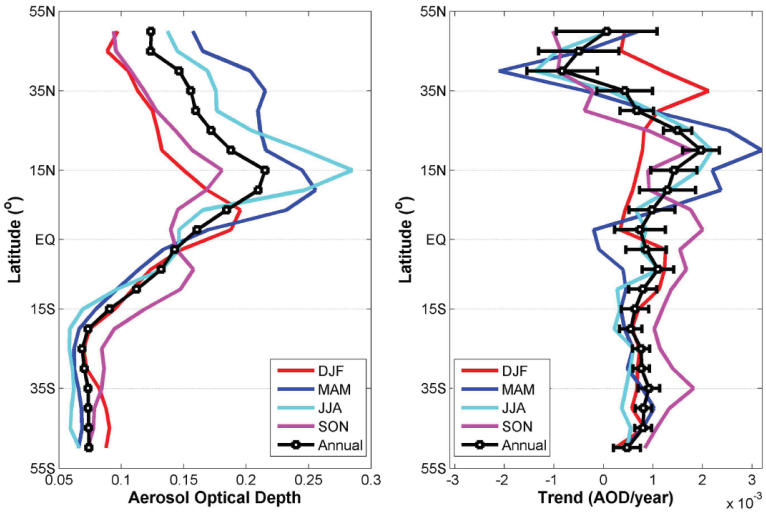


Figure 17 AOD measurements from SeaWifs data overall and per season (left) , and the variability / trend per year (right) (from Hsu et al., 2012).

Since AOD can have a substantial impact on the NDVI (RED and NIR are differently affected), this might be one of the causes of the seasonality found in the differences between the NDVI data sets.

5.5 Investigation of the seasonal differences between the NDVI data sets

If we want to conclude that the data sets have a consistent relationship over time, it is important to investigate the seasonality in the differences between the two NDVI data sets in more detail. Several hypothesis can be formulated for the cause of this seasonality.

(1) AOD effects (treated in the previous section)
(2) SZA effect
(3) Erronuous calculation of d^2 (d=Sun-Earth distance) in the calibration of VGT data.

To investigate this seasonality, the RMSE per pixel was calculated between paired observations per month over the four years. This analysis shows the location of the largest differences. Figure 18 shows the RMSE for Europe and Africa for all months (January upper left till December lower right. The highest errors can be found in October in the boreal region. These errors would have been found in November, December and January, but due to the low SZA there are no observations. Although the difference in SZA is the lowest in the boreal region and the highest above the equator, the impact of the difference is the highest in at high latitudes because of the large SZA. Likewise, around June, higher RMSE are also found in southern Africa, but the sample size is dominated by pixels frome the boreal region and therefore it is impacts the overall analysis much more.

On the other hand, SMAC is only valid for SZA values up to 60°, but in the processing of VGT and METOP-AVHRR, this is not taken into acocunt. During the northern hemispherical winter, the errors resulting from the atmospheric correction will be higher, and the same for the southern hemisphere.

The erronuous calculated of the d^2 in the calibration of the VGT images cannot be assessed. This error will be rectified in the next re-processing of the archive.

We conclude that the seasonality in the differences between the two NDVI data sets can be due to a combination of AOD, SZA and validity of the atmospheric correction.

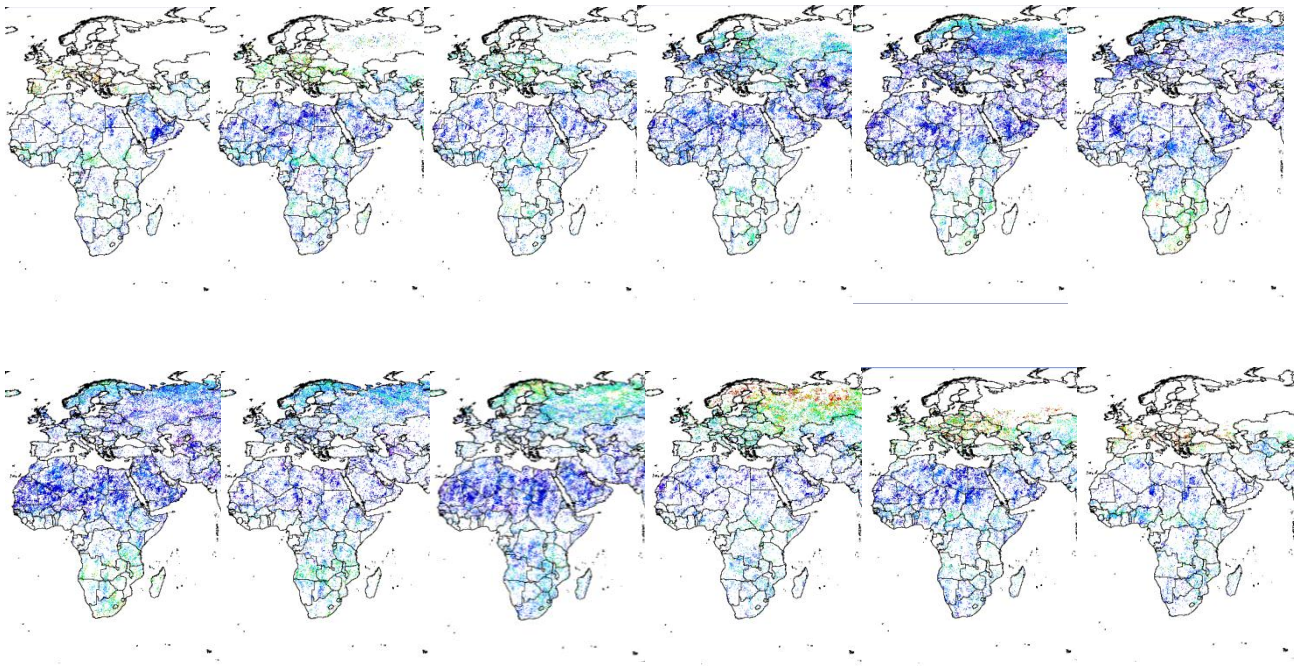


Figure 18 RMSE between the NDVI of METOP-AVHRR and VGT for all years, but per month. Blue : low RMSE, Red : high RMSE (values in between cyan-green-yellow)

5.6 Other influences

Calibration accuracy should be apparent through a systematic trend in the agreement between the NDVI data sets, which is not the case. Therefore, we conclude that there is no difference related to calibration accuracy and that both sensors are properly calibrated. .

The impact of the **point spread function** on the image agreement is difficult to assess. Although the higher spatial auto-correlation in the AVHRR images, the images look 'sharper'. This is because the nearest neighbour (NN) resampling is used for projection, instead of cubic convolution (CC) for VGT. The use of NN does not introduce additional spatial-autocorrelation, limiting the effect of the difference in PSF between the sensors.

Geometric accuracy can have a substantial influence on the agreement between two spatial data sets. This is however treated in the ATBD.

6 Conclusions

A high, near linear agreement was found between the NDVI of VGT and METOP-AVHRR, and the results were stable over time. The relationship expressed through the geometric mean regression has a slope very close to 1.

A seasonality of the differences exists but is likely caused by a combination of SZA and validity of the atmospheric correction.

The slight non-linearity could be attributed to the differences between the spectral response functions. Other influencing factors, such as viewing geometry were also demonstrated.

Concerning the best metric to evaluate the similarity and the consistency of the agreement over time, it is concluded that the agreement coefficient (AC) is superior to R.

7 References

Berthelot, B., & Dedieu, G. (1997). Correction of atmospheric effects for VEGETATION data. In G. Phulpin (Ed.), *Physical Measurements and Signatures in remote Sensing* (pp. 19–25). Amsterdam: Balkema.

Cihlar, J., J.M. Chen, Z. Li, F. Huang, R. Latifovic, and R. Dixon, 1998, Can interannual land surface signal be discerned in composite AVHRR data. *J. Geophys. Res.*, vol. 103, no. D18, p. 23163-23172

Cihlar, J., R. Latifovic, J. Chen, A. Trishchenko, Y. Du, G. Fedojesevs, and B. Guidon, 2004, Systematic corrections of AVHRR image composites for temporal studies. *Remote Sens. Environ.*, vol. 89, no. 2, p. 217-233

Eerens, H., Baruth, B., Bydekerke, L., Deronde, B., Dries, J. Goor, E., Heyns, W., Jacobs, T., Ooms, B. Piccard, I., Royer, A., Swinnen, E. Timmermans, A., Van Roey, T., Vereecken, J. and Y. Verheijen, 2009, Ten-daily Global Composites of METOP-AVHRR, *Proceedings of the ISDE Conference, Beijing, in press.*

Gutman, G.G., 1999, On the use of long-term global data of land reflectances and vegetation indices derived from the advanced very high radiometer. *J. Geophys. Res.*, vol. 104, no. D6, p. 6241-6255

Henry, P. and A. Meygret, 2000, Calibration of VEGETATION cameras on-board SPOT4. *Proceedings of the VEGETATION-2000 conference, Belgirate, 3-6 April,* p. 23-32

Hsu, N. C., Gautam, R., Sayer, A. M., Bettenhausen, C., Li, C., Jeong, M. J., Tsay, S.-C., et al. (2012). Global and regional trends of aerosol optical depth over land and ocean using SeaWiFS measurements from 1997 to 2010. *Atmospheric Chemistry and Physics*, *12*(17), 8037–8053. doi:10.5194/acp-12-8037-2012

Ji, L., Gallo, K., Eidenshink, J. C., & Dwyer, J. (2008). Agreement evaluation of AVHRR and MODIS 16-day composite NDVI data sets. *International Journal of Remote Sensing*, *29*(16), 4839–4861. doi:10.1080/01431160801927194

Klisch, A., Royer, A., Lazar, C., Baruth, B., & Genovese, G. (2006). IMPACT OF DIFFERENT TEMPORAL VEGETATION INDEX SMOOTHING METHODS ON PHENOLOGICAL PARAMETER EXTRACTION. *ISPRS Archives Vol. XXXVI, Part 8, "Remote Sensing Applications for a Sustainable Future."* Haifa, Israel: ISPRS.

Meyer, D.J., 1996, Estimating the effective spatial resolution of an AVHRR time series. *Int. J. Remote Sensing*, vol. 17, no. 15, p. 2971-2980

Rahman, R., & Dedieu, G. (1994). SMAC: a simplified method for the atmospheric correction of satellite measurements in the solar spectrum. *International Journal of Remote Sensing*, *15*(1), 123–143.

Rao, C.R.N. and J. Chen, 1999, Revised post-launch calibration of the visible and near-infrared channels of the Advanced Very High Resolution Radiometer (AVHRR) on the NOAA-14 spacecraft. *Int. J. Remote Sensing*, vol. 20, no. 18, p. 3485-3491

Rosborough, G.W., D.G. Baldwin, and W.J. Emery, 1994, Precise AVHRR image navigation. *IEEE Trans. Geosci. and Remote Sens.* Vol. 32, no. 3, p. 644-657

Ruiz, C.P. and F.J.A. Lopez, 2002, Restoring SPOT images using PSF-derived deconvolution filters. *Int. J. Remote Sensing*, vol. 23, no. 12, p. 2379-2391

Steven, M. D., Malthus, T. J., Baret, F., Xu, H., & Chopping, M. J. (2003). Intercalibration of vegetation indices from different sensor systems. *Remote Sensing of Environment*, *88*(4), 412–422.

Swets, D. D. L., Reed, B. C. B., Rowland, J. J. D., & Marko, S. E. (1999). A weighted least-squares approach to temporal NDVI smoothing. *Proceedings of the 1999 ASPRS Annual Conference: From Image to Information, Portland, Oregon*. Retrieved from http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:A+Weighted+Least-Squares+Approach+to+Temporal+NDVI+Smoothing#0

Swinnen, E., F. Veroustraete , 2008, Extending the SPOT-VEGETATION NDVI time series (1998-2006) back in time with NOAA-AVHRR data (1985-1998) for Southern Africa, *IEEE Transactions on Geoscience and Remote Sensing* - Vol. 46 (2): 558-572

Sylvander, S., P. Henry, C. Bastien-Thiry, F. Meunier, and D. Fuster, 2000, VEGETATION geometrical image quality. In: *Proceedings of the VEGETATION-2000 conference, Lago Maggiore, Italy*, p. 33-44

Trishchenko, A. (2009). Effects of spectral response function on surface reflectance and NDVI measured with moderate resolution satellite sensors: Extension to AVHRR NOAA-17, 18 and METOP-A. *Remote Sensing of Environment*, *113*(2), 335–341. doi:10.1016/j.rse.2008.10.002

Trishchenko, A.P., J. Cihlar, and Z. Li, 2002, Effects of spectral response function on surface reflectance and NDVI measured with moderate resolution satellite sensors. *Remote Sens. Environ.*, vol. 81, no. 1, p. 1-18

van Leeuwen, W.J.D., B.J. Orr, S.E. March and S.M. Herrmann, 2006, Multi-sensor NDVI data continuity: uncertainties and implications for vegetation monitoring applications. *Remote Sens. Environ.*, vol. 100, p. 67-81