Climate Data Record (CDR) Program

Climate Algorithm Theoretical Basis Document (C-ATBD)

Leaf Area Index (LAI) and
Fraction of Absorbed Photosynthetically Active Radiation (FAPAR)
## REVISION HISTORY

<table>
<thead>
<tr>
<th>Rev.</th>
<th>Author</th>
<th>DSR No.</th>
<th>Description</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Martin Claverie, NASA</td>
<td>DSR-726</td>
<td>Initial Submission to CDR Program</td>
<td>11/12/2014</td>
</tr>
</tbody>
</table>
# TABLE of CONTENTS

1. **INTRODUCTION** ................................................................................................................................. 7  
   1.1 Purpose .................................................................................................................................................. 7  
   1.2 Definitions ........................................................................................................................................... 7  
   1.3 Referencing this Document .................................................................................................................. 7  
   1.4 Document Maintenance ......................................................................................................................... 8  

2. **OBSERVING SYSTEMS OVERVIEW** ........................................................................................................ 9  
   2.1 Products Generated ............................................................................................................................... 9  
   2.2 Instrument Characteristics .................................................................................................................... 9  

3. **ALGORITHM DESCRIPTION** .................................................................................................................. 10  
   3.1 Algorithm Overview ............................................................................................................................. 10  
   3.2 Processing Outline ............................................................................................................................... 10  
   3.3 Algorithm Input ................................................................................................................................... 12  
      3.3.1 Primary Sensor Data ....................................................................................................................... 12  
      3.3.2 Ancillary Data ............................................................................................................................... 12  
      3.3.3 Derived Data ............................................................................................................................... 14  
      3.3.4 Forward Models ............................................................................................................................ 14  
   3.4 Theoretical Description ....................................................................................................................... 14  
      3.4.1 Physical and Mathematical Description ........................................................................................ 14  
      3.4.2 Data Merging Strategy ............................................................................................................... 21  
      3.4.3 Numerical Strategy ...................................................................................................................... 21  
      3.4.4 Calculations .................................................................................................................................. 21  
      3.4.5 Look-Up Table Description ........................................................................................................ 22  
      3.4.6 Parameterization ......................................................................................................................... 22  
      3.4.7 Algorithm Output ........................................................................................................................ 22  

4. **TEST DATASETS AND OUTPUTS** ....................................................................................................... 24  
   4.1 Test Input Datasets ............................................................................................................................... 24  
   4.2 Test Output Analysis ............................................................................................................................ 24  
      4.2.1 Reproducibility ............................................................................................................................. 24  
      4.2.2 Precision and Accuracy ................................................................................................................ 27  
      4.2.3 Error Budget ............................................................................................................................... 28  

5. **PRACTICAL CONSIDERATIONS** ......................................................................................................... 30  
   5.1 Numerical Computation Considerations ............................................................................................. 30  
   5.2 Programming and Procedural Considerations ..................................................................................... 30  
   5.3 Quality Assessment and Diagnostics ................................................................................................... 30  
   5.4 Exception Handling ............................................................................................................................. 30  
   5.5 Algorithm Validation ........................................................................................................................... 31  
   5.6 Processing Environment and Resources ............................................................................................ 31  

6. **ASSUMPTIONS AND LIMITATIONS** ............................................................................................... 32  

A controlled copy of this document is maintained in the CDR Program Library.  
Approved for public release. Distribution is unlimited.
7. FUTURE ENHANCEMENTS........................................................................................................33
  7.1 Land Cover map update ...........................................................................................................33
  7.2 Improvement of the algorithm for broadleaf forest .................................................................33

8. REFERENCES............................................................................................................................34

APPENDIX A. ACRONYMS AND ABBREVIATIONS..................................................................36
LIST of FIGURES

Figure 1: Timeline of the NOAA platform numbers. ................................................................. 9

Figure 2: Algorithm flowchart ............................................................................................... 11

Figure 3: Landcover for year 2004. Grey: water; blue: shrubland; yellow: Grasslands & Croplands & Non vegetated; light green: broadleaf forest; green: needleleaf forest; dark green: evergreen broadleaf forest .................................................................................................................................. 13

Figure 4: Global map of the V (a) and R (b) parameters derived by Vermote et al. (2009) applied to the time series of Terra MODIS band 2 (2000–2004). V and R are shown for highest NDVI values of each pixel. ................................................................................................................................. 16

Figure 5: BELMANIP-2 and DIRECT network sites location .............................................. 18

Figure 6: Domain definition for the 5 classes (red polygons). Grey images represent the density function for each 0.01 surface reflectance bin (white = no value; black = high density). ......................................................... 19

Figure 7: Conceptual representation of the Artificial Neuron Network, including normalization steps. Notice that the number of neurons correspond to the real number of neurons. S and L stand for “sigmoid” and “linear” neurons, respectively. ........................................................................................................... 21

Figure 8: Theoretical performance of the LAI retrieval. Each line refers to a land cover class and last line to all classes merged. On the left subplots, the cumulative distribution function of input data for the training (i.e. MCD15) and output retrieval. On the right subplots, scatter plot between MCD15 (x-axis) and Estimates (y-axis) are displayed. Only data for DIRECT sites (not used for training) were plotted. A, P and U corresponds to Accuracy, Precision and Uncertainty metrics (see Vermote and Kotchenova 2008 for more details). ..................................................................................... 25

Figure 9: Theoretical performance of the FAPAR retrieval. Refer to legend of Figure 8 for more details. ........................................................................................................................................................................... 26

Figure 10: Comparison of retrieval from AVHRR-NOAA-16 (N16) and AVHRR-NOAA-18 (N18) for BELMANIP-2 and DIRECT sites from 02-Jul-2005 to 31-Dec-2006. ................................................................ 27

Figure 11: In-situ validation over DIRECT sites. Ground measurement covers initially a footprint of 3x3 km. they were extrapolated to 0.05° using MCD15 products ................................................................................................................................. 28

LIST of TABLES

Table 1: Primary Sensor Data ................................................................................................................................. 12

Table 2: Ancillary Data ........................................................................................................................................... 12

Table 3: Reclassification table of the land cover classes ....................................................................................... 13
Table 4: Minimum and maximum values used for the normalization ............................................................. 20
Table 5: AVH15C1 output layers ..................................................................................................................................... 22
Table 6: Error budget based on in-situ validation ................................................................................................... 29
Table 7: QA SDS bits description. Bits are listed from the MSB (bit 7) to the LSB (bit 0) ....................... 30
1. Introduction

1.1 Purpose

The purpose of this document is to describe the algorithm submitted to the National Climatic Data Center (NCDC) by Dr. Eric Vermote and Martin Claverie of NASA/GSFC, Terrestrial Information Systems Branch, Code 619, that will be used to create the Leaf Area Index and Fraction of Absorbed Photosynthetically Active Radiation Climate Data Record (CDR) using Advanced Very High Resolution Radiometer (AVHRR) sensors onboard the NOAA 7, 9, 11, 14, 16, and 18 platforms. The actual algorithm is defined by the computer program (code) that accompanies this document, and thus the intent here is to provide a guide to understanding that algorithm, from both a scientific perspective and in order to assist a software engineer or end-user performing an evaluation of the code.

1.2 Definitions

Following is a summary of the symbols used to define the algorithm.

\[ \theta_s = \text{sun zenith angle.} \]  
\[ \theta_v = \text{view zenith angle.} \]  
\[ \Phi = \text{view-sun relative azimuth.} \]  
\[ \rho = \text{reflectance.} \]  
\[ \xi = \text{scattering angle.} \]  
\[ F_0 = \text{solar radiance.} \]  
\[ F_1 = \text{volume scattering kernel.} \]  
\[ F_2 = \text{geometric kernel.} \]  
\[ k_0, k_1, k_2 = \text{BRDF kernel coefficient.} \]

1.3 Referencing this Document

This document should be referenced as follows:

1.4 Document Maintenance

Periodic updates to the algorithm and dataset are possible to occur. This could be (for example) when improvements to the algorithm are developed. Any update will be given a new version number, and an updated version of the C-ATBD will be generated.
2. **Observing Systems Overview**

2.1 **Products Generated**

The objective of this algorithm is to retrieve the Leaf Area Index (LAI) and the Fraction of Absorbed Photosynthetically Active Radiation (FAPAR) from AVHRR sensors. The final data set contains more than 32 years of data from June 1981 to present. Products are mapped into daily 0.05°x0.05” grid, corresponding to a 3600x7200 array over the globe.

2.2 **Instrument Characteristics**

LAI and FAPAR products will be generated for each cloud-free pixel (0.05°x0.05°) observed by the NOAA-AVHRR imager channels 1 and 2 as described in the AVHRR Surface Reflectance C-ATBD (CDRP-ATBD-0495). The AVHRR Surface Reflectance CDR is also known by its legacy filename: AVH09C1. Timeline of the NOAA platform numbers is presented in Figure 1.

![Figure 1: Timeline of the NOAA platform numbers.](image)
3. **Algorithm Description**

3.1 **Algorithm Overview**

This section aims to describe the algorithm at the current level of maturity (which will be updated with each revision). The algorithm includes retrieval of LAI and FAPAR from AVHRR imagers channels 1 and 2.

3.2 **Processing Outline**

The processing outline of the AVHRR LAI / FAPAR processing is summarized in the data Flow chart of Figure 2.
Figure 2: Algorithm flowchart.
3.3 Algorithm Input

3.3.1 Primary Sensor Data

Primary Sensor Data are Bidirectional reflectance distribution function (BRDF) adjusted surface reflectance data derived from AVH09C1 products. Sun zenith angle is also used from the same products to retrieve the surface reflectance non-adjusted from the sun geometry. Data are described in Table 1.

Table 1: Primary Sensor Data

<table>
<thead>
<tr>
<th>Name</th>
<th>Type</th>
<th>Description</th>
<th>Source</th>
<th>Dimension (spatial resolution)</th>
</tr>
</thead>
<tbody>
<tr>
<td>surface reflectance (Channels 1-2)</td>
<td>Input</td>
<td>BRDF-adjusted surface reflectance</td>
<td>AVH09C1</td>
<td>0.05°</td>
</tr>
<tr>
<td>Sun zenith angle</td>
<td>Input</td>
<td>Sun zenith angle</td>
<td>AVH09C1</td>
<td>0.05°</td>
</tr>
</tbody>
</table>

3.3.2 Ancillary Data

The LAI / FAPAR algorithm requires ancillary data listed in Table 2.

Table 2: Ancillary Data

<table>
<thead>
<tr>
<th>Name</th>
<th>Type</th>
<th>Description</th>
<th>Source</th>
<th>Dimension (spatial resolution)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BRDF database</td>
<td>Ancillary</td>
<td>Pixel-based BRDF database of the VJB model</td>
<td>internal</td>
<td>0.05°</td>
</tr>
<tr>
<td>Land cover classification 1981-1994</td>
<td>Ancillary</td>
<td>IGBP* Land cover classification 1981-1994 from Hansen et al. (2000) resampled at 0.05°</td>
<td>Hansen et al. (1998)</td>
<td>0.05°</td>
</tr>
</tbody>
</table>

* International Geosphere-Biosphere Program

Land cover classification follows the IGBP (International Geosphere-Biosphere Program) labeling. The map was produced by Hansen et al. (1998) for the period 1981-1994. To avoid none consistent land cover changing, the same classification is used for the entire data set. Moreover, to reduce the number of Artificial Neuron Network (ANN) and spatial discontinuity, the number of classes were reduced from 9 to 6 (see Table 3).
Figure 3: Landcover for year 2004. Grey: water; blue: shrubland; yellow: Grasslands & Croplands & Non vegetated; light green: broadleaf forest; green: needleleaf forest; dark green: evergreen broadleaf forest.

Table 3: Reclassification table of the land cover classes

<table>
<thead>
<tr>
<th>IGBP Class Name</th>
<th>Code</th>
<th>New Class Name</th>
<th>Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>water</td>
<td>0</td>
<td>water</td>
<td>6</td>
</tr>
<tr>
<td>evergreen needleleaf forest</td>
<td>1</td>
<td>needleleaf forest</td>
<td>1</td>
</tr>
<tr>
<td>evergreen broadleaf forest</td>
<td>2</td>
<td>evergreen broadleaf forest</td>
<td>5</td>
</tr>
<tr>
<td>deciduous needleleaf forest</td>
<td>3</td>
<td>needleleaf forest</td>
<td>1</td>
</tr>
<tr>
<td>deciduous broadleaf forest</td>
<td>4</td>
<td>broadleaf forest</td>
<td>2</td>
</tr>
<tr>
<td>mixed forests</td>
<td>5</td>
<td>broadleaf forest</td>
<td>2</td>
</tr>
<tr>
<td>closed shrublands</td>
<td>6</td>
<td>shrublands</td>
<td>3</td>
</tr>
<tr>
<td>open shrublands</td>
<td>7</td>
<td>shrublands</td>
<td>3</td>
</tr>
<tr>
<td>woody savannas</td>
<td>8</td>
<td>shrublands</td>
<td>3</td>
</tr>
<tr>
<td>savannas</td>
<td>9</td>
<td>shrublands</td>
<td>3</td>
</tr>
<tr>
<td>grasslands</td>
<td>10</td>
<td>grasslands, croplands &amp; non-vegetated</td>
<td>4</td>
</tr>
<tr>
<td>permanent wetlands</td>
<td>11</td>
<td>grasslands, croplands &amp; non-vegetated</td>
<td>4</td>
</tr>
<tr>
<td>croplands</td>
<td>12</td>
<td>grasslands, croplands &amp; non-vegetated</td>
<td>4</td>
</tr>
<tr>
<td>urban and built-up</td>
<td>13</td>
<td>grasslands, croplands &amp; non-vegetated</td>
<td>4</td>
</tr>
<tr>
<td>cropland/natural vegetation mosaic</td>
<td>14</td>
<td>grasslands, croplands &amp; non-vegetated</td>
<td>4</td>
</tr>
<tr>
<td>snow and ice</td>
<td>15</td>
<td>grasslands, croplands &amp; non-vegetated</td>
<td>4</td>
</tr>
<tr>
<td>barren or sparsely vegetated</td>
<td>16</td>
<td>grasslands, croplands &amp; non-vegetated</td>
<td>4</td>
</tr>
</tbody>
</table>
3.3.3 Derived Data

<Not Applicable>

3.3.4 Forward Models

<Not Applicable>

3.4 Theoretical Description

In this section we describe the algorithms used to produce the LAI / FAPAR products.

3.4.1 Physical and Mathematical Description

3.4.1.1 VJB model

The VJB model (Vermote et al. 2009) is used to retrieve nadir-adjusted surface reflectance from BRDF-adjusted surface reflectance. The surface reflectance of the AVH09C1 product are adjusted for the sun-view geometry to a constant view zenith angle ($\Theta_v=0^\circ$) and a constant sun zenith angle ($\Theta_s=45^\circ$). However, the FAPAR is a variable that varies according to the zenith angle. It is consequently more appropriate to derive FAPAR from nadir-adjusted surface reflectance keeping the original sun zenith angle through the implementation of the VJB, for which the theoretical basis is described below.

The analysis of Parasol multidirectional data has shown that, among analytical BRDF models, the Ross–Li–Maignan model provides the best fit to the measurements (Breon et al. 2012). This model computes the reflectance as the sum of three terms:

$$\rho(\theta_s, \theta_v, \phi) = k_0 + k_1 F_1(\theta_s, \theta_v, \phi) + k_2 F_2(\theta_s, \theta_v, \phi)$$

$$= k_0 \left[ 1 + \frac{k_1}{k_0} F_1(\theta_s, \theta_v, \phi) + \frac{k_2}{k_0} F_2(\theta_s, \theta_v, \phi) \right]$$

(10)

where $F_1$ is the volume scattering kernel, based on the Ross-thick function, but corrected for the Hot-Spot process, and $F_2$ is the geometric kernel, based on the Li-sparse reciprocal function. $F_1$ and $F_2$ are fixed functions of the observation geometry, but $k_0$, $k_1$, and $k_2$ are free parameters. In the following, we will use $V$ as $k_1/k_0$ and $R$ for $k_2/k_0$. The $F_1$ and $F_2$ functions are given in (11) and (12).
where $\xi$ is the scattering angle and $\xi_0$ is a characteristic angle that can be related to the ratio of scattering element size and the canopy vertical density ($\xi_0=1.5^\circ$).

In AVH09C1 product, a correction of the directional effect is derived after transforming the measurement coordinates to standard observation geometry. The AVH09C1 standard observation geometry is for the Sun at 45° from zenith, and the observation at nadir. Therefore, the BRDF-adjusted (view and sun geometry) surface reflectance as used in AVH09C1 product is computed as eq (13), while the retrieval of the nadir-adjusted (view-geometry only) surface reflectance product is computed as eq (14).

\[
F_1(\theta_s, \theta_v, \varphi) = \frac{m}{\pi}(t - \sin t \cos t - \pi) + \frac{1 + \cos \xi}{2 \cos \theta_s \cos \theta_v}
\]

\[
\cos t = \frac{2}{m} \sqrt{\Delta^2 + (\tan \theta_s \tan \theta_v \sin \varphi)^2}
\]

\[
m = \frac{1}{\cos \theta_s} + \frac{1}{\cos \theta_v}
\]

\[
\Delta = \sqrt{\tan(\theta_s)^2 + \tan(\theta_v)^2 - 2 \times \tan(\theta_s) \times \tan(\theta_v) \cos(\phi)}
\]

\[
F_2(\theta_s, \theta_v, \varphi) = \frac{4}{3\pi} \frac{1}{\cos \theta_s + \cos \theta_v} \left[ \left( \frac{\pi}{2} - \xi \right) \cos \xi + \sin \xi \right]
\]

\[
\times \left( 1 + \left( 1 + \frac{\xi}{\xi_0} \right)^{-1} \right) - \frac{1}{3}
\]

(11)

(12)

(13)

(14)

BRDF correction is based on pre-computed coefficients of $V$ and $R$ relationship with NDVI (Vermote et al. 2009).

$V = V_{slope} \times NDVI + V_{intercept}$
R = R_slope * NDVI + R_intercep \hspace{1cm} (15)

The 4 coefficients (V_slope, V_intercep, R_slope, R_intercep) were retrieved using the 2000-2011 MODIS archives following the approach presented in Vermote et al. (2009). Figure 4 illustrate V and R global maps derived from MODIS 2000-2004.

3.4.1.2 Artificial Neuron Network building

In this section, the methodology to derive the Artificial Neuron Network (ANN) parameters used to relate nadir-adjusted surface reflectance to LAI and FAPAR is described. We used the MODIS product MCD15 as reference LAI and FAPAR values.

3.4.1.2.1 Input/output dataset for the training of the ANN

**AVH09C1 surface reflectance products**

AVH09C1 products were selected from NOAA-16 during the year 2000-2007. A full description of the products is given in the AVHRR Surface Reflectance C-ATBD (CDRP-
ATBD-0459). We retained pixel flag in the products as cloud, cloud-shadow and snow free. NDVI was derived directly from the nadir-adjusted surface reflectance.

**MCD15 LAI/FAPAR products**

The MCD15A2 (denoted MCD15 hereafter) is a global LAI and FAPAR product, composited every 8 days at 1-kilometer resolution on a Sinusoidal grid. Data were extracted for the period 2000-2007.

MODIS main algorithm is based on Look Up Tables (LUT) simulated from a three-dimensional radiative transfer model (Knyazikhin et al., 1998). Red and NIR atmospherically corrected MODIS reflectance (Vermote et al., 1997) and the corresponding illumination-view geometry are used as inputs of the LUT. The output is the mean LAI and FAPAR values computed over the set of acceptable LUT elements for which simulated and MODIS surface reflectances agree within specified level of (model and measurement) uncertainties. When the main algorithm fails, a backup algorithm based on NDVI (Normalized Difference Vegetation Index) relationships, calibrated over the same radiative transfer model simulations is used (Yang et al., 2006). We retained pixel flag as derived from the main algorithm and back-up algorithm.

FAPAR product corresponds to the instantaneous value at the time of the satellite overpass. LAI and FAPAR are first produced daily. Then, the 8 days composite corresponds to the values of the product when the maximum FAPAR value within the eight days period is observed. Note that no LAI and FAPAR values are retrieved over bare or very sparsely vegetated area, permanent ice or snow area, permanent wetland, urban area, or water bodies.

**Sites selection**

In order to limit the amount of processing data, a sample of globally distributed sites were selected. We used the BELMANIP-2 and DIRECT networks (Figure 5). Data were obtained through the On LIne Validation Exercise (OLIVE) a CEOS/LPV initiative for on line validation of global land products (http://calvalportal.ceos.org/web/olive/).

BELMANIP2 (BErnchmark Land Multisite ANalysis and Intercomparison of Products) was built using sites from existing experimental networks (FLUXNET, AERONET, VALERI, BigFoot,...) completed with selected sites from the GLC2000 land cover map. BELMANIP2 was built using vegetation land cover map. The sites selection was performed for each band of latitude (10° width) by keeping the same proportion of biome types within the selected sites as within the whole band of latitude. Attention was paid so that the sites were homogeneous over a 10x10km² area, almost flat, and with a minimum proportion of urban area and permanent water bodies. The BELMANIP2 dataset included 445 sites.

DIRECT is a collection of sites for which ground measurements are available and that have been collected (Garrigues et al, 2008) and processed according to the CEOS-LPV guidelines. There is currently 113 data sets (sites and dates of measurements) available.
Spatial and temporal consideration

We consequently extracted time series of MCD15 and AVH09C1 products over BELMANIP-2 and DIRECT sites. We focused on coincident years between AVHRR-NOAA-16 and MODIS products: 2001-2007.

MCD15 products are 1km resolution. We aggregate the product to the equivalent AVH09C1 product, i.e. 0.05°. MCD15 products are produced every 8 days. To limit the time series noise, we aggregate the product monthly. Same processing was done for AVH09C1 product.

3.4.1.2.2 Domain definition

ANN are trained over a defined area and their confidence considerably decrease out of the domain delimited by the learning dataset. The Domain definition was defined based on NOAA-16 AVHRR Surface reflectance pixels corresponding to BELMANIP-2 sites for the 2001-2007 period.

Figure 6 represents the density distribution of the learning dataset of each class and the associated domain delimited by a polygon. Polygons were defined to include 97% of the density distribution pixels (0.01 resolution for rho1 and rho2). With these selected pixels, we used a convex hull algorithm to define an envelope polygon. Finally, the envelope was simplified using a Recursive Douglas-Peucker Polyline Simplification algorithm.
3.4.1.2.3 ANN training

ANN are a connection of neurons, associated by “synaptic” weights. Each neuron transforms the sum of the weighted signal from the previous neurons according to a given transfer function and a bias. The combination of sigmoid and linear functions is capable to fit any type of function (Demuth and Beale, 1998).

The training step is divided in 2 parts: Normalization of the inputs and outputs, ANN learning and de-normalization of the outputs.

Normalization is achieved simply by scaling between the minimum and maximum values: the normalized values, Y will vary between -1 and +1, and are computed from the raw values X and the minimum $X_{\text{min}}$ and maximum $X_{\text{max}}$ values (eq 16 and 17).

$$Y = \frac{2 \times (X - X_{\text{min}})}{(X_{\text{max}} - X_{\text{min}})} - 1 \quad (16)$$

$$X = \left(\frac{1}{2} \times (X_{\text{max}} - X_{\text{min}}) \times (Y + 1)\right) + X_{\text{min}} \quad (17)$$
Minimum and maximum values of input and outputs are presented in Table 4.

<table>
<thead>
<tr>
<th>Class</th>
<th>NeedleLeaf Forest</th>
<th>BroadLeaf Forest</th>
<th>Shrublands</th>
<th>Grasslands &amp; Croplands &amp; Non vegetated</th>
<th>evergreen broadleaf forest</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Min</td>
<td>Max</td>
<td>Min</td>
<td>Max</td>
<td>Min</td>
</tr>
<tr>
<td>rho1</td>
<td>0.01</td>
<td>0.11</td>
<td>0.01</td>
<td>0.12</td>
<td>0.01</td>
</tr>
<tr>
<td>rho2</td>
<td>0.02</td>
<td>0.24</td>
<td>0.04</td>
<td>0.39</td>
<td>0.01</td>
</tr>
<tr>
<td>cos(theta)</td>
<td>0.05</td>
<td>0.85</td>
<td>0.14</td>
<td>0.88</td>
<td>0.06</td>
</tr>
<tr>
<td>NDVI</td>
<td>0.01</td>
<td>0.86</td>
<td>-0.01</td>
<td>0.87</td>
<td>-0.22</td>
</tr>
<tr>
<td>LAI</td>
<td>0.00</td>
<td>5.24</td>
<td>0.01</td>
<td>5.94</td>
<td>0.00</td>
</tr>
<tr>
<td>FAPAR</td>
<td>0.01</td>
<td>0.93</td>
<td>0.02</td>
<td>0.92</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Separate ANN were built for both output products LAI and FAPAR and each of the 5 classes as described in section 3.3.2. The ANN architecture finally retained followed the proposition of (Claverie et al. 2013) and are composed of (Figure 7):

- One input layer made of the 4 normalized input data.
- One hidden layer with 5 neurons having tangent sigmoid transfer functions.
- One output layer with a linear transfer function.
For each configuration (5 classes x 2 output variables), 10 ANN were trained, resulting to 100 ANN in total. The selection of the "optimal" ANN was based on the RMSE between the outputs and the "true" biophysical variables from sites used for validation (DIRECT network).

The learning process is mainly made of two elements: the training dataset that was described previously, and the learning rule that is now described. The Levenberg–Marquardt optimization algorithm is used to adjust the synaptic weights and neuron bias to get the best agreement between the output simulated by the network and the corresponding value of canopy biophysical variable simulated in the training database. The initial values of the weights and biases were set to a random value between -1.0 and +1.0. Two networks are trained in parallel to retrieve LAI and FAPAR, each corresponding to independent random drawing of the initial values of the synaptic weights and bias.

3.4.1.3 Class Fusion

Maps corresponding to the five land cover types are fused according to the land cover maps described in section 3.3.2. The land cover is reported in the QA SDS.

3.4.2 Data Merging Strategy

Retrieval from the various NOAA-AVHRR sensors were merged according to the timeline of Figure 1.

3.4.3 Numerical Strategy

Artificial Neural Networks can model complex non-linear and multivariate systems with very simple output models. They include a set of coefficients for sigmoid and linear function and biases. ANN are very fast in computation and do not require any specific numerical strategy.

3.4.4 Calculations

The algorithm steps (see Figure 2) are the following:

- BRDF-adjustment via VJB model
- ANN application per class
- Class fusion.

They result in one NetCDF file including the following SDS:

- LAI
• FAPAR
• QA

3.4.5 **Look-Up Table Description**

<Not Applicable>

3.4.6 **Parameterization**

Parameterization of the ANN algorithm is defined in section 3.4.1.2.2. The sets of ANN coefficients for the 5 classes and the corresponding code and ANN parameterization are reported in the delivered code.

3.4.7 **Algorithm Output**

The output of the algorithm is one NetCDF file per day containing layers listed in Table 5. An example filename is:

AVHRR-Land_v004_AVH15C1_NOAA-07_19840101_c20130813051326.nc

With the following naming convention:

<product-name> = static series name of the product with the value, “AVHRR- Land”

<product-version> = product version number, “v004”

<product-type> = NASA product type identifier with the valid domain:

“AVH15C1” = AVHRR LAI/FAPAR

<sat-id> = Source NOAA satellite ID with the valid domain:


<YYYYmmdd> = Date of the data in the file, formatted as year, month and day, with the valid range from “19810101” to present

c<processing-date> = Creation or processing date of the file identified with a ‘c’ followed by the year, month, day, hour, minute and second

.nc indicates the format (NetCDF)

**Table 5: AVH15C1 output layers**

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
<th>Units</th>
<th>Dimension</th>
</tr>
</thead>
<tbody>
<tr>
<td>FAPAR</td>
<td>long_name = Fraction of Absorbed Photosynthetically Active Radiation</td>
<td>[-]</td>
<td>3600X7200</td>
</tr>
<tr>
<td>LAI</td>
<td>long_name = Leaf Area Index</td>
<td>m².m⁻²</td>
<td>3600X7200</td>
</tr>
<tr>
<td>QA</td>
<td>long_name = Quality Assurance</td>
<td>3600X7200</td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>-------------------------------</td>
<td>-----------</td>
<td></td>
</tr>
<tr>
<td>longitude</td>
<td>long_name = longitude</td>
<td>degrees</td>
<td>7200</td>
</tr>
<tr>
<td>latitude</td>
<td>long_name = latitude</td>
<td>degrees</td>
<td>3600</td>
</tr>
<tr>
<td>time</td>
<td>long_name = time</td>
<td>days since 1981-01-01 00:00:00 UTC</td>
<td>1</td>
</tr>
</tbody>
</table>
4. Test Datasets and Outputs

4.1 Test Input Datasets

The test input dataset is 2003 of NOAA-16 AVHRR.

4.2 Test Output Analysis

4.2.1 Reproducibility

The evaluation of the reproducibility is done through the theoretical performance of the ANN. The 5 ANN related to the 5 land cover classes were trained based on MCD15 extraction over BELMANIP-2 sites. DIRECT sites were not used for the training but only for the evaluation of the ANN. Figure 8 (LAI) and 9 (FAPAR) display the theoretical performances of the ANN for the 5 classes.
Figure 8: Theoretical performance of the LAI retrieval. Each line refers to a land cover class and last line to all classes merged. On the left subplots, the cumulative distribution function of input data for the training (i.e. MCD15) and output retrieval. On the right subplots, scatter plot between MCD15 (x-axis) and Estimates (y-axis) are displayed. Only data for DIRECT sites (not used for training) were plotted. A, P and U corresponds to Accuracy, Precision and Uncertainty metrics (see Vermote and Kotchenova 2008 for more details).
Figure 9: Theoretical performance of the FAPAR retrieval. Refer to legend of Figure 8 for more details.
Another way to evaluate the reproducibility is to compare outputs from AVHRR sensors on board of 2 different platforms. We selected NOAA-16 and NOAA-18 with an overlapping period from 02-Jul-2005 to 31-Dec-2006. The analysis is carried on BELMANIP-2 and DIRECT sites. The scatterplot is displayed in Figure 10 and uncertainties are 0.35 for LAI and 0.08 for FAPAR.

Figure 10: Comparison of retrieval from AVHRR-NOAA-16 (N16) and AVHRR-NOAA-18 (N18) for BELMANIP-2 and DIRECT sites from 02-Jul-2005 to 31-Dec-2006.

4.2.2 Precision and Accuracy

The performances of the estimates were evaluated over in situ measurements from the DIRECT network; the sites were not used in the learning process. Scatter plot per class are shown in Figure 11. LAI validation scatter plots were divided among the type of measurement: Effective and true LAI, depending if the clumping factor is included or not (see Claverie et al. 2012).
Figure 11: In-situ validation over DIRECT sites. Ground measurement covers initially a footprint of 3x3 km. they were extrapolated to 0.05° using MCD15 products.

4.2.3 Error Budget

The error budget is summarized in Table 6, which includes uncertainty, precision and accuracy from the validation over DIRECT sites.
Table 6: Error budget based on in-situ validation. N correspond to the number of points used to compute the statistical metrics.

<table>
<thead>
<tr>
<th>Class</th>
<th>Effective LAI</th>
<th>True LAI</th>
<th>FAPAR</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A</td>
<td>P</td>
<td>U</td>
</tr>
<tr>
<td>NeedleLeaf Forest</td>
<td>0.29</td>
<td>0.32</td>
<td>0.37</td>
</tr>
<tr>
<td>BroadLeaf Forest</td>
<td>0.28</td>
<td>1.03</td>
<td>1.05</td>
</tr>
<tr>
<td>Shrublands</td>
<td>0.2</td>
<td>0.9</td>
<td>0.9</td>
</tr>
<tr>
<td>Grasslands &amp; Croplands &amp; Non vegetated</td>
<td>-0.08</td>
<td>0.65</td>
<td>0.65</td>
</tr>
<tr>
<td>evergreen broadleaf forest</td>
<td>0.14</td>
<td>1.35</td>
<td>1.31</td>
</tr>
<tr>
<td>All</td>
<td>0.08</td>
<td>0.89</td>
<td>0.89</td>
</tr>
</tbody>
</table>
5. Practical Considerations

5.1 Numerical Computation Considerations

No parallelization or difficulties in matrix inversions are expected. The algorithm implementation is based on ANN to increase computation speed.

5.2 Programming and Procedural Considerations

<Not Applicable>

5.3 Quality Assessment and Diagnostics

The estimate of the LAI/FAPAR is accompanied with quality assurance information. QA bits contain information stored at the product resolution. The Table 7 describes the bit ordering to use the Quality Assessment SDS.

Table 7: QA SDS bits description. Bits are listed from the MSB (bit 7) to the LSB (bit 0)

<table>
<thead>
<tr>
<th>Bits Number</th>
<th>Bits Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>6-7</td>
<td>Polygon test</td>
</tr>
<tr>
<td>00: in polygon</td>
<td></td>
</tr>
<tr>
<td>01: not in polygon</td>
<td></td>
</tr>
<tr>
<td>10: not tested (water/cloudy)</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>BRDF corrected</td>
</tr>
<tr>
<td>0: no</td>
<td></td>
</tr>
<tr>
<td>1: yes</td>
<td></td>
</tr>
<tr>
<td>2-4</td>
<td>Associated Class</td>
</tr>
<tr>
<td>001: NeedleLeaf Forest</td>
<td></td>
</tr>
<tr>
<td>010: BroadLeaf Forest</td>
<td></td>
</tr>
<tr>
<td>011: Shrublands</td>
<td></td>
</tr>
<tr>
<td>100: Grasslands &amp; Croplands &amp; Non vegetated</td>
<td></td>
</tr>
<tr>
<td>101: Evergreen broadleaf forest</td>
<td></td>
</tr>
<tr>
<td>110: Water</td>
<td></td>
</tr>
<tr>
<td>0-1</td>
<td>Quality control</td>
</tr>
<tr>
<td>00: OK</td>
<td></td>
</tr>
<tr>
<td>01: Input flag as Cloudy</td>
<td></td>
</tr>
<tr>
<td>10: Invalid input</td>
<td></td>
</tr>
<tr>
<td>11: Output out of range</td>
<td></td>
</tr>
</tbody>
</table>

5.4 Exception Handling

<Not applicable>
5.5 Algorithm Validation

The C code was compared with the original Matlab code over a random data test of 1000 sample and the full image of day 142-2003.

5.6 Processing Environment and Resources

The CDR code is run on an 8-core 2.5GHz 64-bit Xeon server, running CentOS Linux 5.9 x86_64. The code was compiled with C-compiler GCC 4.4.7. The main C-libraries were: HDF5 8.88, HDF 4.2r4, NetCDF 4.2, Zlib 1.2, szip. A 10-year data set was processed in 25 hours.
6. Assumptions and Limitations

The main limitation of the algorithm is the capacity to reproduce LAI/FAPAR dynamic over Evergreen Broadleaf forest. This is due to a saturation of AVHRR Channel 1 and 2 signals over dense vegetation cover.
7. Future Enhancements

7.1 Land Cover map

In the current version, the land cover map is based on a unique map produced for the 1981-1994 period. In the future, it will be important to implement a dynamic land cover in order to account for large scale land cover changes (e.g. deforestation).

7.2 Improvement of the algorithm for broadleaf forest

As currently used, the algorithm is not capable to reproduce a consistent spatio-temporal of LAI/FAPAR over evergreen broadleaf forest and displays some saturaturation over broadleaf forest. Future work will be focused on algorithm enhancement.
8. References


Hansen, M., R. DeFries, J.R.G. Townshend, and R. Sohlberg (1998), UMD Global Land Cover Classification, 1 Kilometer, 1.0, Department of Geography, University of Maryland, College Park, Maryland, 1981-1994.


## Appendix A. Acronyms and Abbreviations

<table>
<thead>
<tr>
<th>Acronym or Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANN</td>
<td>Artificial Neuron Network</td>
</tr>
<tr>
<td>AVHRR</td>
<td>Advanced Very High Resolution Radiometer</td>
</tr>
<tr>
<td>BELMANIP-2</td>
<td>BEnchmark Land Multisite ANalysis and Intercomparison of Products</td>
</tr>
<tr>
<td>BRDF</td>
<td>Bidirectional reflectance distribution function</td>
</tr>
<tr>
<td>C-ATBD</td>
<td>Climate Algorithm Theoretical Basis Document</td>
</tr>
<tr>
<td>CDR</td>
<td>Climate Data Record</td>
</tr>
<tr>
<td>CEOS</td>
<td>Committee on Earth Observation Satellites,</td>
</tr>
<tr>
<td>CMG</td>
<td>Climate Modelling Grid</td>
</tr>
<tr>
<td>FAPAR</td>
<td>Fraction of Absorbed Photosynthetically Active Radiation</td>
</tr>
<tr>
<td>HDF</td>
<td>Hierarchical Data Format</td>
</tr>
<tr>
<td>IEEE</td>
<td>Institute of Electrical and Electronic Engineers</td>
</tr>
<tr>
<td>IGBP</td>
<td>International Geosphere-Biosphere Program</td>
</tr>
<tr>
<td>LAI</td>
<td>Leaf Area Index</td>
</tr>
<tr>
<td>LPV</td>
<td>Land Product Validation</td>
</tr>
<tr>
<td>LSB</td>
<td>Least Significant Bit</td>
</tr>
<tr>
<td>LUT</td>
<td>Look Up Tables</td>
</tr>
<tr>
<td>MODIS</td>
<td>Moderate Resolution Imaging Spectroradiometer</td>
</tr>
<tr>
<td>MSB</td>
<td>Most Significant Bit</td>
</tr>
<tr>
<td>NCDC</td>
<td>National Climatic Data Center</td>
</tr>
<tr>
<td>NDVI</td>
<td>Normalized Difference Vegetation Index</td>
</tr>
<tr>
<td>NIR</td>
<td>Near infrared</td>
</tr>
<tr>
<td>NOAA</td>
<td>National Oceanic and Atmospheric Administration</td>
</tr>
<tr>
<td>OLIVE</td>
<td>On Line Validation Exercise</td>
</tr>
<tr>
<td>QA</td>
<td>Quality Assessment</td>
</tr>
<tr>
<td>RMSE</td>
<td>Root-mean-square error</td>
</tr>
<tr>
<td>SDS</td>
<td>Science Data Sets</td>
</tr>
<tr>
<td>VJB</td>
<td>Vermote Justice Breon</td>
</tr>
</tbody>
</table>