AVHRR-derived NDVI Metrics Product

USER MANUAL

Version 1.0

Jiang Zhu[1], Amy E. Miller[2], Parker Martyn[2], Chuck Lindsay[2], Dayne Broderson[1], Tom Heinrichs[1]

1. Geographic Information Network of Alaska, UAF
2. National Park Service, Anchorage

Sep., 18, 2014

Table of Contents

1. Introduction

Advanced Very High Resolution Radiometer (AVHRR) data are being used by the National Park Service (NPS), Alaska Region Inventory & Monitoring (I&M) Program and the University of Alaska-Geographic Information Network of Alaska (GINA) to monitor and study the interannual variability of vegetation across Alaska. GINA acquires standardized AVHRR-derived Normalized Difference Vegetation Index (NDVI) data, derives the NDVI metrics, and distributes the NDVI data and NDVI metrics data. This user manual briefly introduces how to acquire the standardized AVHRR products from existing archives at the USGS-EROS Data Center as well as how to deliver the processed results. It focuses on description of NDVI metrics algorithm and its usage.

2. AVHRR-derived NDVI product and Metrics

The U.S. Geological Survey's (USGS) Earth Resources Observation and Science (EROS) Center acquires afternoon AVHRR 1-km resolution daily observations to produce 7-day composite Normalized Difference Vegetation Index (NDVI) dataset for Alaska region. The dataset generally includes twenty-eight weekly composites for the period April through October each year since 1995. The dataset is projected in WGS84/Albers Conical Equal Area mapping grid in Geostationary Earth Orbit Tagged Image File Format (GeoTIFF). The spatial resolution of dataset is 1km x 1km square kilometers. The dataset is obtained by a series of data processes: Scene selection, satellite and solar viewing geometry, radiometric calibration, atmospheric correction, scaling the reflectance and temperature data, calculating NDVI, geometric registration, compositing, and cloud screening. We do not choose to calculate the NDVI by ourselves for two reasons: The data process is complicated and the 7-day composite NDVI dataset of USGS-EROS is widely used to do vegetation monitoring and analysis among various communities.

The dataset can be directly served in a web coverage service (WCS) to monitor the vegetation condition and can be used to calculate NDVI metrics right way. Once transferred to GINA, the data are distributed through the GINA WCS (http://ndvi.gina.alaska.edu/year/avhrr). Yearly 7-day composite NDVI data are stacked, interpolated, and smoothed, and then the yearly NDVI metrics are calculated. The yearly NDVI metrics product is also delivered through the WCS (http://ndvi.gina.alaska.edu/metrics?). Figure 2.1 describes the NDVI data processing and delivery procedure.

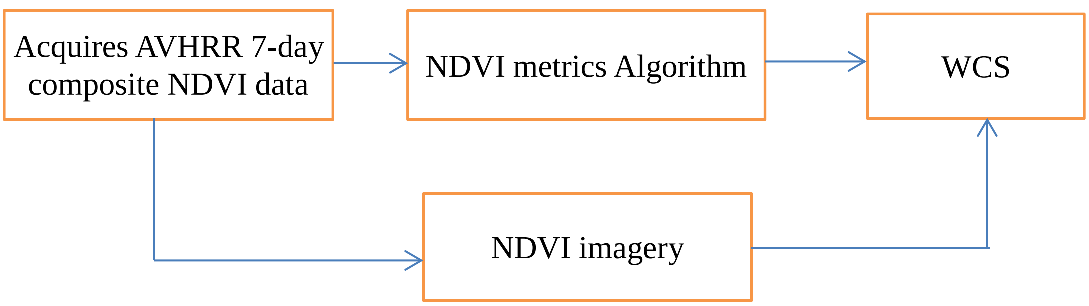

Figure 2.1. AVHRR NDVI data automatic process scheme

2.1. Source and Format of Input Data

The AVHRR 7-day composite NDVI data for Alaska generally includes twenty-eight weekly composites for the April through October each year. One composite file has the name convention of satID_regionIDyy_dddddd_tif.tar, where satID represents satellite, such as NOAA-18 (n18), regionID represents region, such as Alaska (ak), yy represents year, such as 95, 96, 11, etc. dddddd represents composite-day range from start days of year to end days of year, tif represents the file format in tar is geotiff format. For example, n18_ak11_088094_tif.tar indicates the 7-day composite data is composited only including Alaska region from 088 to 094 days of 2011 AVHRR data on board NOAA-18 satellite. Each 7-day composite AVHRR data tar file includes 14 geotiff files, where each file represents a band. One band contains the maximum NDVI value for each pixel selected from the daily overpasses. The remaining 13 bands are the data values that coincident with the observation value selected as the maximum NDVI value. They include channels 1 to 5, satellite viewing geometry data (three bands), corrected channels 1 and 2 reflectance, a quality-control band indicating the origin of the water vapor and ozone values, a band that contains a pointer to the scene identification number for each pixel selected from the same daily pass as the maximum NDVI value, and the cloud mask. The cloud mask has values from 0 to 200. The individual values are keyed to which test is ultimately used to determine if the pixel is clear or cloudy. Values less than 100 are clear, and values 100 or greater are clouds. Table 2.1 describes the bands and their possible value ranges.

Table 2.1. Band description of composite images

Band	Description	Value Range
1	Top of atmosphere Channel 1 reflectance	
2	Top of atmosphere Channel 2 reflectance	
3	Radiance	
4	Temperature	
5	Temperature	
6	NDVI	0-99, cloud, snow, water 100-200, valid NDVI
7	Satellite zenith	
8	Solar Zenith	
9	Relative Azimuth	
10	Atmospherically corrected Channel 1 surface reflectance	
11	Atmospheric corrected Channel 2 surface reflectance	
12	QA/QC	
13	Date index	
14	Cloud mask	0-99, clear, 100-200, cloudy

The dataset is available at **earthexplorer.usgs.gov**. We downloaded NDVI data from 1995 to 2012 for NDVI metrics calculation. One year of dataset includes 28 7-day composite collections of files. 2006 dataset misses two 7-day composite data collections (4/5-4/11, 4/12-4/18) and 2007 dataset misses one 7-day composite data collection (10/2-10/8).

2.2 Installation, configuration, and use of the AVHRR-derived NDVI metrics algorithm

The AVHRR-derived NDVI metrics algorithm was developed using the IDL and ENVI software package (Both IDL and ENVI are products from Exelis Visual Information Systems, Boulder, Colorado, USA). The schema of the algorithm is illustrated in Figure 2.2.

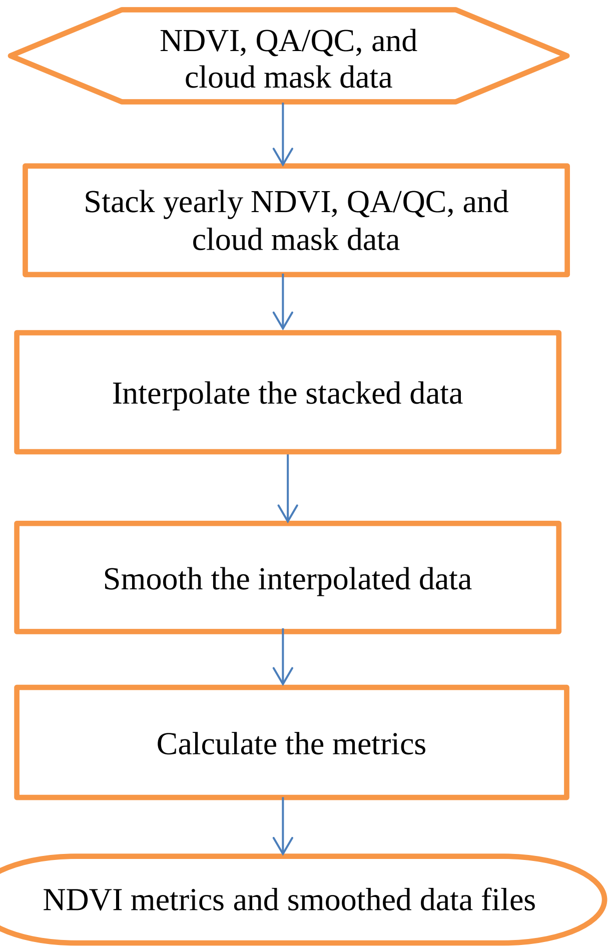

Figure 2.2. NDVI metrics algorithm schema

The yearly NDVI, QA/QC, and cloud-mask data are inputs for the NDVI metrics algorithm. The algorithm stacks yearly NDVI, QA/QC, and cloud mask data into multiple-band data sets, respectively. Then interpolates and does smooth the NAVI data with the help of information from QA/QC and cloud-mask data. Finally, the algorithm calculates the NDVI metrics and output them as well as the smoothed NDVI data.

The yearly NDVI metrics data include 12 metrics: day of onset of greenness (onp), NDVI value at onset day (onv), day of end of greenness (endp), NDVI value at the end day of greenness (endv), the duration of greenness season (durp), the day of the maximum NDVI value (maxp), maximum NDVI value (maxv), range of NDVI values (ranv), rate of green up (rtup), rate of senescence (rtdn), time-integrated NDVI (tindvi), NDVI metrics flag (mflg). Because the last 7-day composite ends on later September or early October in the USGS AVHRR 7-day composite NDVI dataset, we cannot accurately estimate the days of end of greenness for many pixels in southern part of Alaska region.

The algorithm codes were developed using the IDL programming language and ENVI software package. They work on both Linux and Windows operational environments with IDL plus ENVI installed. The instruction included here covers the installation, configuration, and use of the algorithm codes in a Linux environment.

Installation

a. Download the algorithm zip file named avhrr-ndvi-metrics-master.zip from the URL link: https://github.com/gina-alaska/avhrr-ndvi-metrics to your home directory ($HOME), and rename the zip file into "avhrr-ndvi-metrics-master.zip".

b. unzip the zip file. Files will be expanded in the "avhrr_ndvi_metrics" directory ($HOME/avhrr-ndvi-metrics-master). In this directory, there are one file (startup_nps) and several sub-directories: docs, bin, sav, codes, and scripts. The "doc" directory includes the documentation; the "~/bin" stores IDL library codes; the "~/codes" includes the IDL source codes; the "~/sav" stores the compiled IDL binary file; and the "~/scripts" includes script files used to execute the application in batch mode.

Configuration

a. Edit the startup_nps file and save to the $HOME/avhrr-ndvi-metrics directory
The file "startup_nps" includes:
ENVI, /RESTORE_BASE_SAVE_FILES
PREF_SET,'IDL_PATH', '<IDL_DEFAULT>:+$HOME/avhrr-ndvi-metrics/bin',
/COMMIT

b. edit .bashrc and add the following lines:

export ndvi_home=$HOME/avhrr-ndvi-metrics
export IDL_STARTUP=$HOME/avhrr-ndvi-metrics/startup_np

c. edit 1yr_avhrr_env.bash

export rawdata_dir="avhrr_raw_data_main_directory"
export work_dir=$HOME/avhrr-ndvi-metrics/work
export script_dir=$HOME/avhrr-ndvi-metrics/scripts
export idlprog_dir=$HOME/avhrr-ndvi-metrics/sav
export idl_dir="home directory of installed IDL"
export result_dir=$HOME/results

AVHRR raw data are stored in the "avhrr_raw_data_main_directory". Under the main directory, there are subdirectories to store different years of data. For example, year 2009 data are store in 2009 subdirectory.

Use

There are two options for running the application. The first is to run it in batch mode and the second is to run the application step by step.

a. Run the application in batch mode

```
>cd $HOME/avhrr_ndvi_metrics/scripts
>./1yr_avhrr_main_v2.bash yyyy
```

The parameter yyyy is the year that you want to process, such as 2007. It will automatically complete all steps to produce the NDVI metrics data file and smoothed data file in the $result_dir.

b. Run the application step by step

Assume the AVHRR raw data are stored in the $rawdata_dir/YYYY, where YYYY represents year (such as 2007).

Step1. produce ndvi, quality, and cloud mask file lists.

```
>./1yr_avhrr_flist.bash avhrr_raw_data_main_dir, year
```

This script produces three files. The yyyy_flist_ndvi, yyyy_flist_bq, and yyyy_flist_cldm. They include files names of ndvi, quality, and cloud mask files, respectively. For example, 2009_flist_ndvi, 2009_flist_bq, and 2009_flist_cldm.

Step2. stack the one-year data

```
>./1yr_avhrr_stack_v2.bash flist_ndvi, flist_bq, flist_cldmsk, ul_lon, ul_lat, lr_lon, lr_lat
```

The flist_ndvi, flist_bq, and flist_cldmsk are file lists which include the full path of ndvi, quality, and cloud mask file names, respectively. The ul_lon,ul_lat, lr_lon and lr_lat are upper left and low right location of the rectangular zone in longitude and latitude. They are used to subset the data. If one does not subset the data, all four location arguments are set as 0. This script produces three stacked files. They are yyyy_stacked_ndvi, yyyy_stacked_bq, and yyyy_stacked_cldm. For example, 2009_stacked_ndvi, 2009_stacked_bq, and 2009_stacked_cldm.

Step3. calculate NDVI metrics

```
>./1yr_avhrr_calmetrics.bash stacked_ndvi_file, stacked_qa_file, stacked_cldmsk_file, ver_flag
```

This script accept three stacked_file names, and user can choose ver_flg, for example, ‘ver5’. It produce a yearly metrics file and a yearly smoothed 7-day composite NDVI data file.

2. 3 Description of AVHRR-derived NDVI metrics algorithm

As described in Figure 2.2, the NDVI metrics algorithm includes stacking yearly NDVI data, interpolating the data, smoothing the data, and calculating metrics. The algorithm is composed of two parts: stack yearly data and calculating metrics.

A. stacking yearly NDVI, quality, and cloud-mask data (avhrr_stack_oneyear_data.pro)

Figure 2.3 is the flowchart of avhrr_stack_oneyear_data.pro. The stacking program takes three file lists as inputs. Three file lists are one-year-ndvi-file list, one-year-bq-file list, and the one-year-cloudmask-file list which include the full-path format file names of NDVI, quality (bq), and cloud mask files, respectively. The program goes through each group of triple files and reads them into a temporary memory space, stacks these data together, respectively. Finally, it produces three stacked data files.

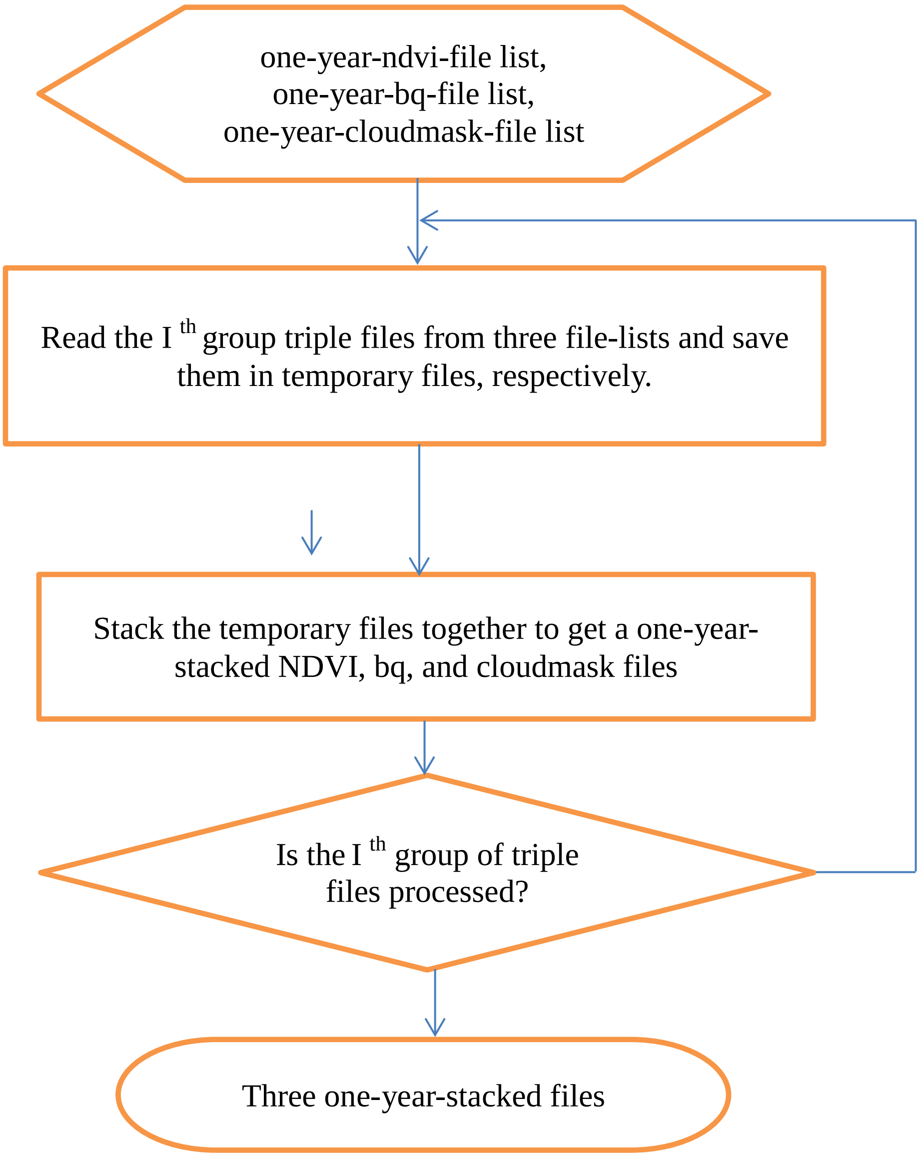

Figure 2.3. Flowchart of stacking processing

B. Calculation of NDVI metrics (avhrr_calculate_ndvi_metrics_tile.pro)

The second part of the AVHRR-derived NDVI metrics algorithm is calculation of NDVI metrics. Figure 2.4 shows the logic of how to calculate NDVI metrics. The program intakes three stacked files: stacked_ndvi_file, stacked_bq_file, and stacked_cldmsk_file. It goes through each pixel to do interpolation and smoothing of NDVI time series and calculation of metrics.

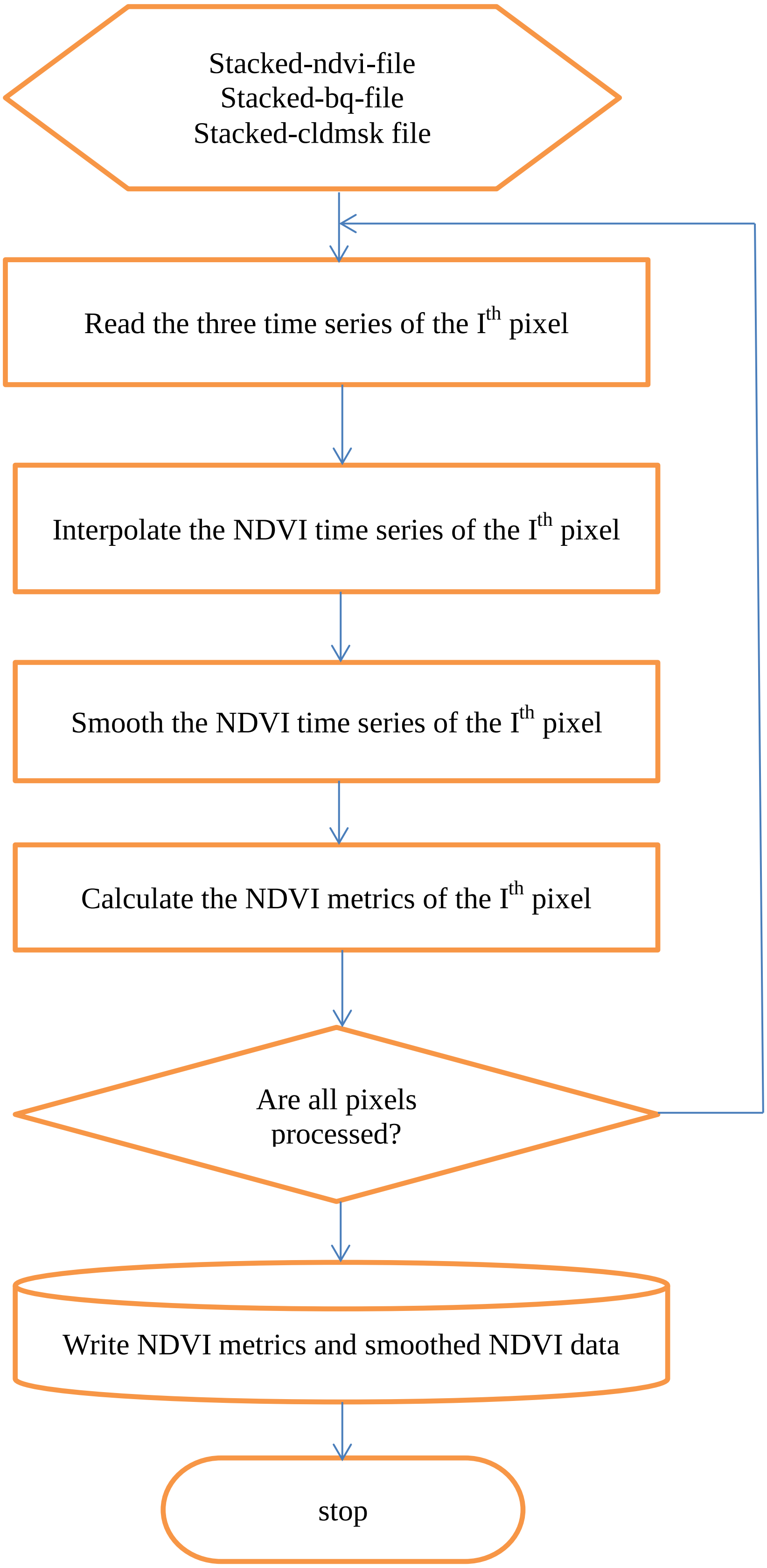

Figure 2.4 The schema of NDVI metrics calculation algorithm

a) Interpolate the stacked-data time series (avhrr_interpol_extension_1yr.pro)

The algorithm goes through each pixel to process its corresponding time series. Figure 2.5 describes the interpolation process for the time series.

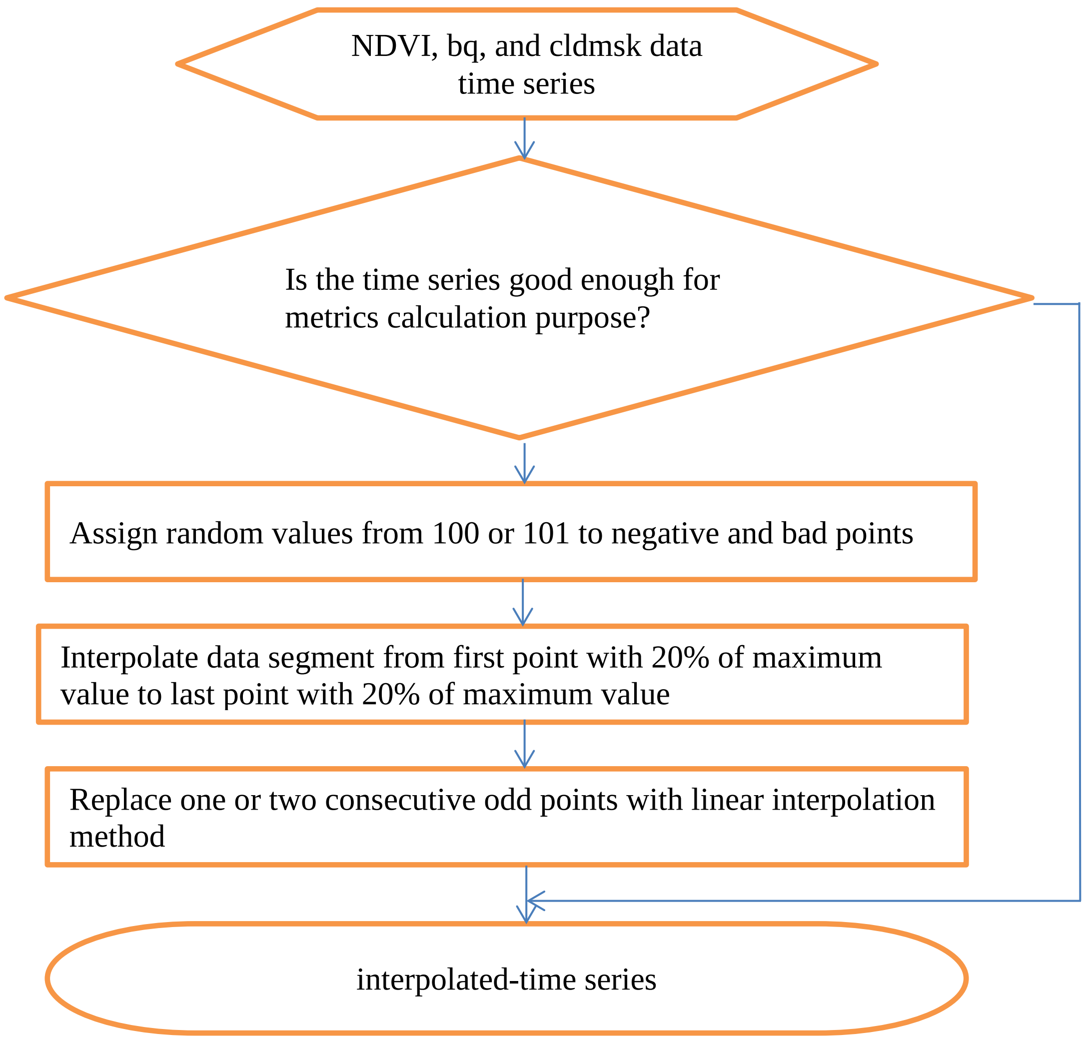

Figure 2.5 Flowchart of data interpolation for a NDVI time series

The raw NDVI time series for a pixel includes 28 weekly composite data points, starting with later March, and ending with early October. The value range of the raw NDVI time series is [0, 200], where the valid data range is [100, 200], and data in the range of [0, 100] indicate cloud, snow or water. Figure 2.6 gives an example of such a raw NDVI time series. As you can see, there are three no-valid points (value less than 100) and a few odd points in the time series. The time series has to be processed before it can be used to calculate metrics. The interpolation process targets these no-valid and odd points.

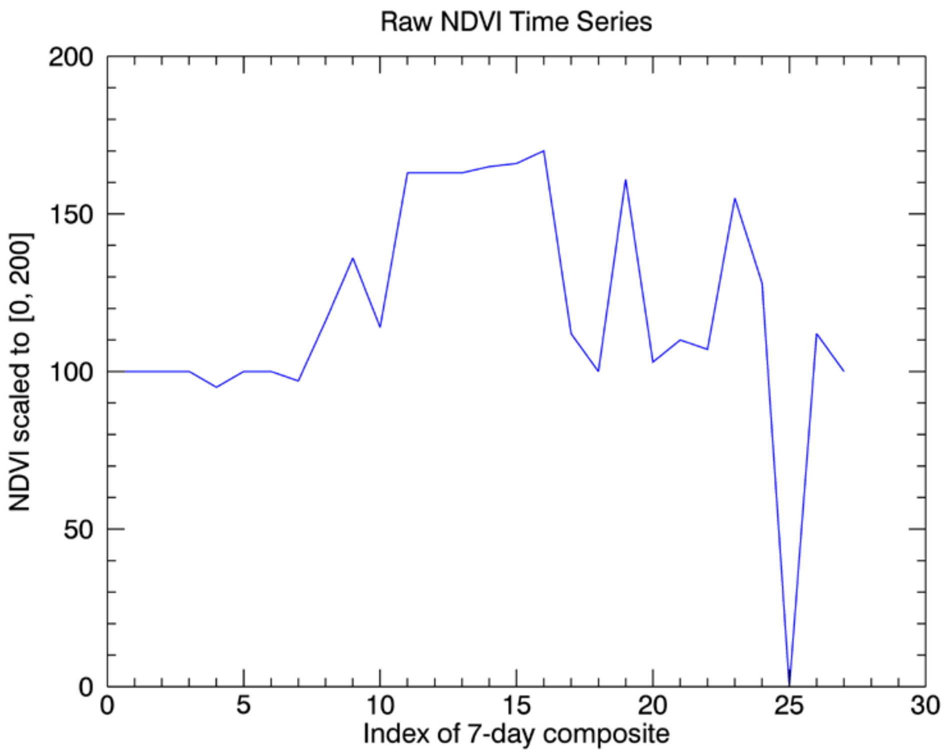

Figure 2.6 A raw NDVI time series

The interpolation subroutine first checks the raw NDVI data time series to determine if it is valid for the purpose of NDVI metrics calculation. Two conditions are applied to determine the validity of a time series: a) the time series must have at least 3 clear-sky points with NDVI value equal or greater than 25% of the valid raw data range of [100, 200]. That is to say, NDVI values must be greater than 125; b) the time series must have normal shape which defined as the maximum NDVI value point must be located between the first 20% of maximum value point and the last 20% of maximum value point. If the time series does not meet either of the conditions, the subroutine returns to its caller program. Otherwise, the subroutine continues to process the time series. Good points with value equal to or greater than 100 are kept. Other points are randomly assigned a 100 or 101 to avoid false crossovers for start of season and end of season detection. The time series is divided into three segments. The first segment is defined as from the first point of the time series to the first 20% of maximum value point-1. The second segment is from the first to the last point of 20% of maximum value points. The last segment is defined as from the last 20% of maximum value point+1 to the last point of the time series. Linear interpolation method is applied to the second segment by keeping clear-sky points with valid NDVI values and interpolating the no-valid points. The first and last points with 20% of maximum values are also used to estimate start and end of season dates (Figures 2.8 and 2.9). Then the time series is checked to determine if it includes one or two consecutive odd points which are significantly smaller than their adjacent points, and if so, these odd points are replaced with values determined by the linear interpolation method. Figure 2.7 shows the interpolated time series. As you can see, both no-valid and odd points in Figure 2.6 are linearly interpolated.

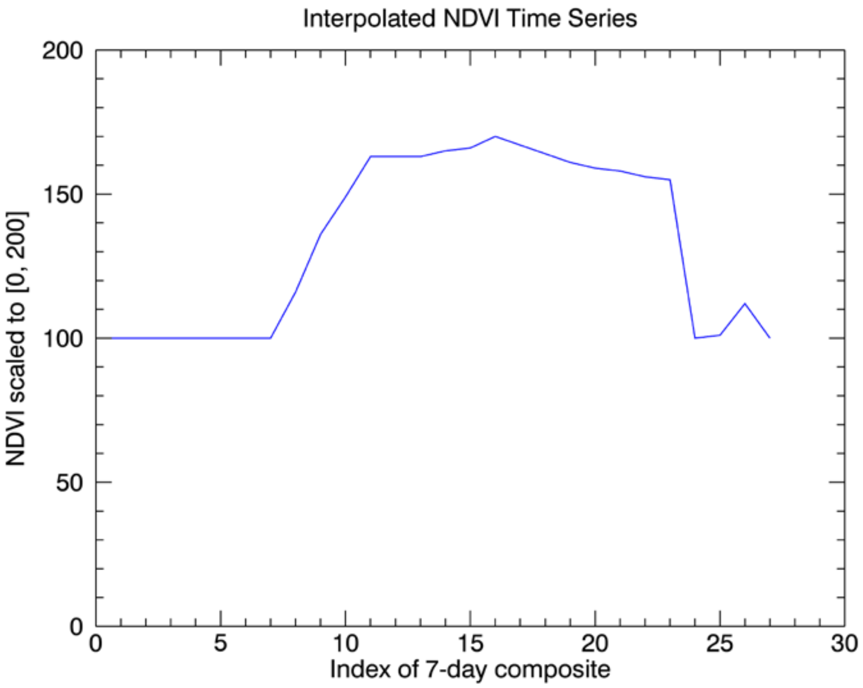

Figure 2.7. Interpolated-data time series

b) Smoothing the interpolated data time series (wls_smooth.pro)

A weighted-least-square smooth algorithm is applied to the interpolated time series. The algorithm is based on methods described by Swets (2001). Figure 2.8 shows the smoothed time series. Please notice that the value of the 24st point of the time series is increased to reflect more realistic situation.

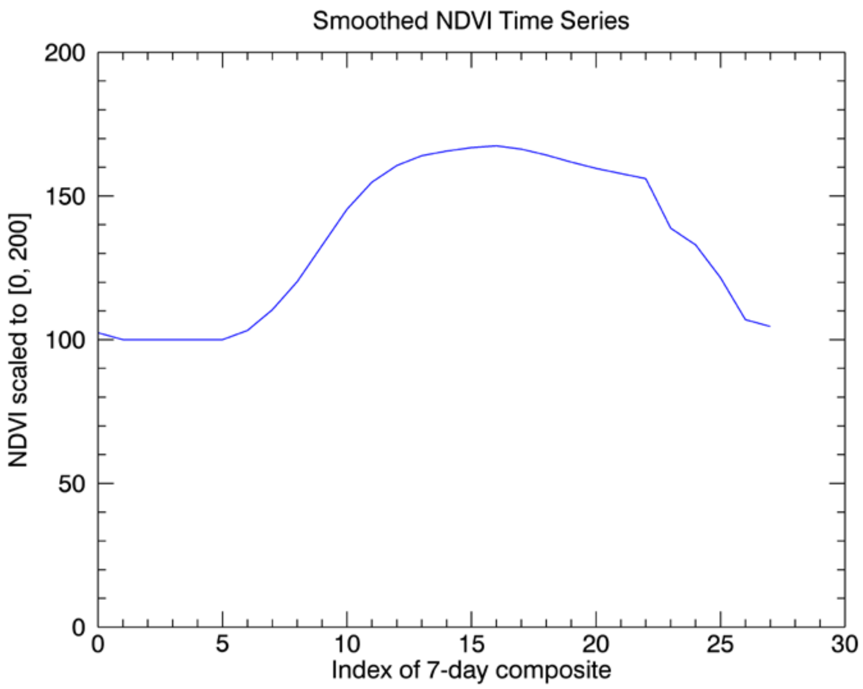

Figure 2.8 Smoothed NDVI time series

c) Calculating NDVI metrics for the smoothed-data time series

The NDVI metrics calculation method combines the delayed moving window method (Reed *et al.*, 1994) and the threshold method. Figure 2.9 describes the process of calculating the NDVI metrics from the time series.

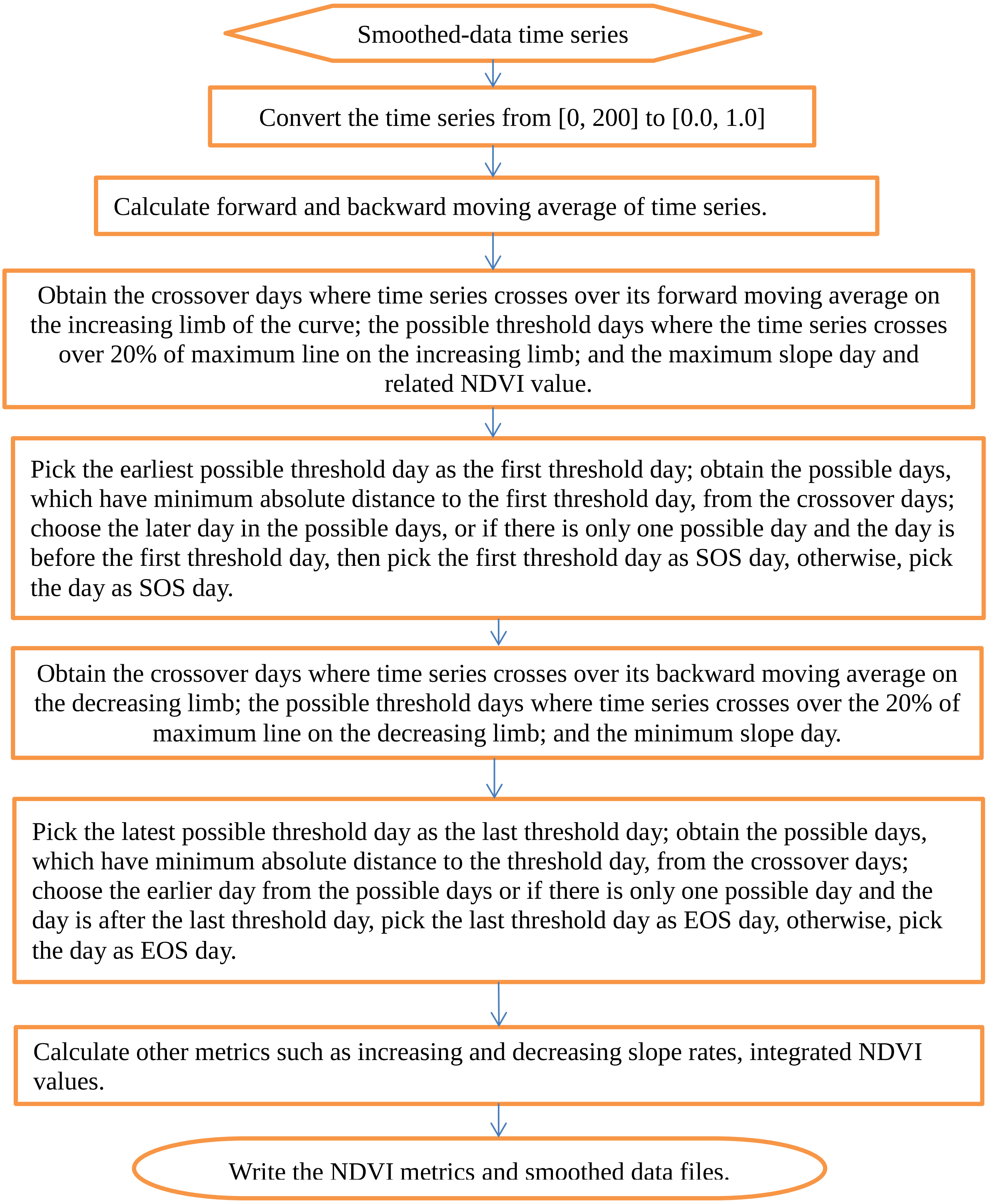

Figure 2.9. Flowchart of calculation of NDVI metrics

The calculation of NDVI metrics algorithm takes a smoothed-data time series as the input. In step 1, it converts the time series from [0, 200] to [0.0, 1.0]. In step 2, the subroutine calculates both forward and backward delay moving average of the time series. Fixed window length for the delay moving average method is chosen in the algorithm, and the default value of the window length is 16. Because one point in the time series represents 7 days, 16-point window length represents 112 days. Using a fixed-length moving window for the delayed moving average is beneficial when comparing multiple years of NDVI metrics data. The algorithm also provides a method as an option to determine dynamically the moving window length according to the possible greenness season length (described in the next section). Step 3 calculates the forward and backward moving averages of the time series. Step 4 obtains the crossover days, possible threshold days, and maximum slope days in the time series. A threshold value of 20% of maximum NDVI is used to identify possible threshold days. Step 5 chooses the earliest day from the possible threshold days as the first threshold day, and chooses the day which is the most closest to the first threshold day as the possible start of season (SOS). It then picks the later day between the possible SOS and first threshold day. Figure 2.10 gives an example to illustrate the way to determine the SOS day. The solid line is the smoothed time series and the dotted line is the forward delay moving average of the time series. The star is a possible crossover and the triangle is the first threshold day. According to the logic of the algorithm, the SOS day is equal to the crossover day, since the threshold day occurs before the crossover day.

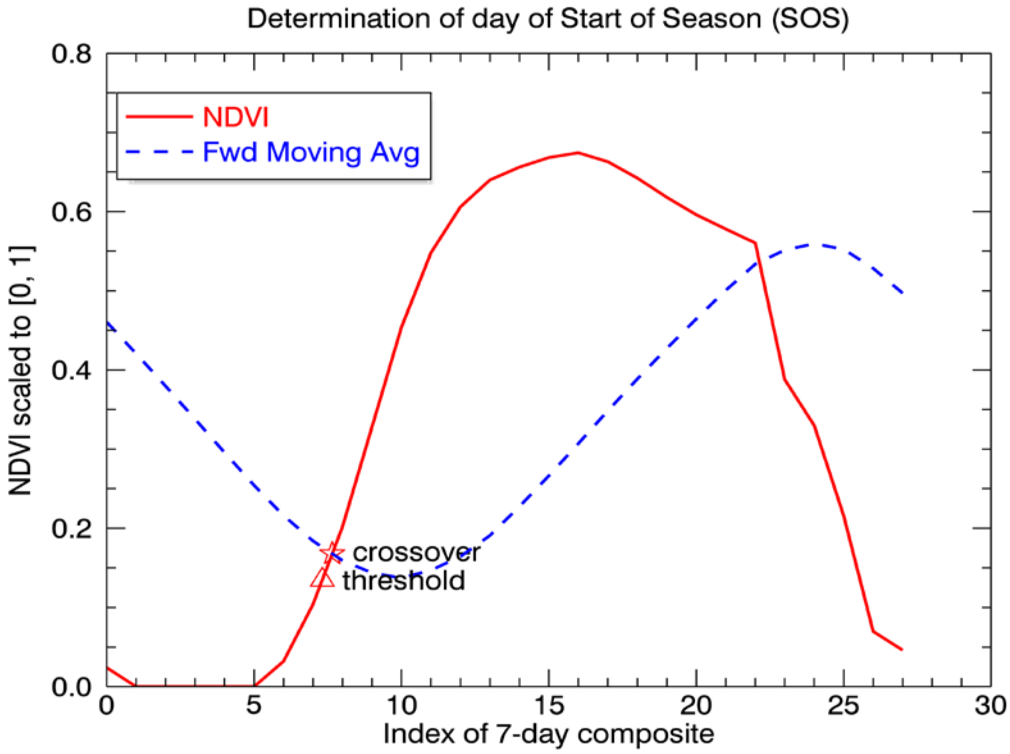

Figure 2.10 Determination of SOS day

Steps 6 and 7 determine the EOS day. They follow the similar logic as in steps 4 and 5. Figure 2.11 illustrates the process of determining the EOS day. In this case, the EOS day is equal to the crossover day, because the threshold day is later than crossover day.

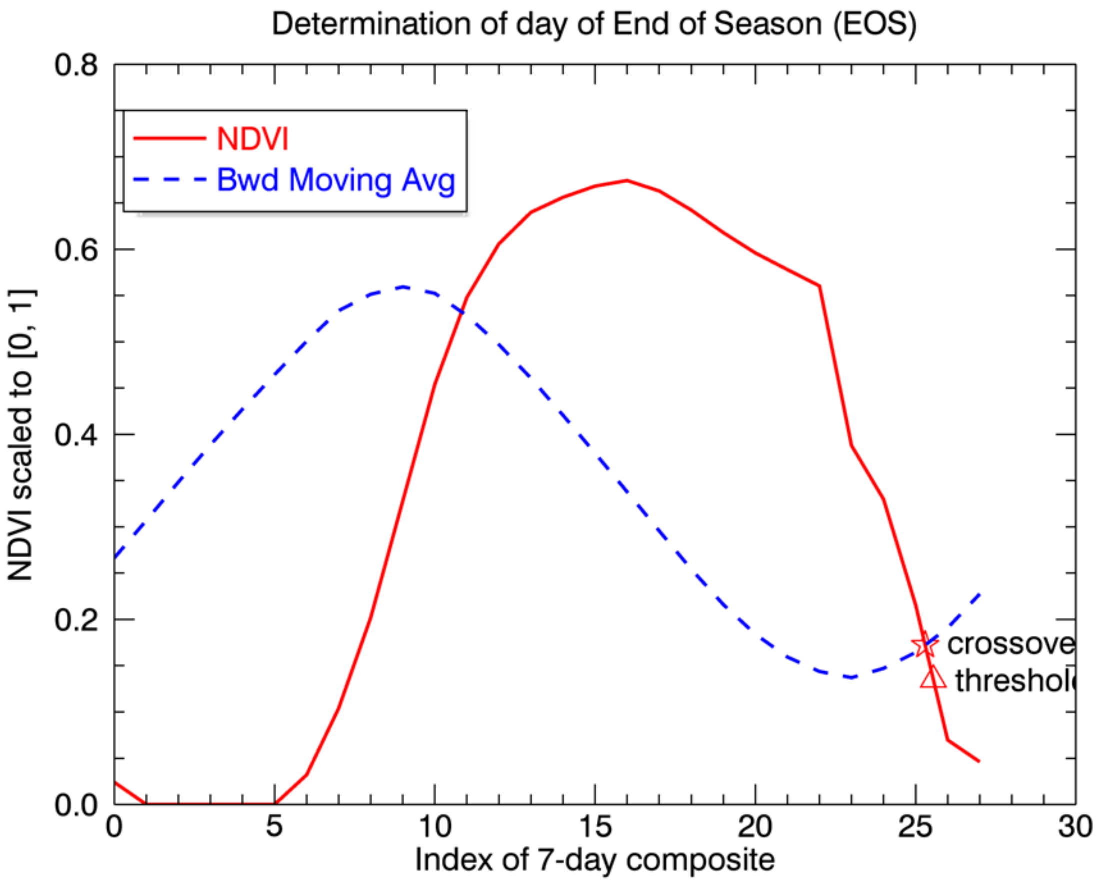

Figure 2.11 Determination of EOS day

Once the SOS day and EOS day are determined, the last step is to calculate the other metrics such as the maximum NDVI day and its NDVI value, up and down slope rates, and the integrated NDVI-day value.

Appendix A List of programs in the AVHRR-derived NDVI metrics Algorithm

1 avhrr_stack_oneyear_data.pro
1.1 start_batch.pro
1.2 read_ndvi.pro
1.2.1 subset.pro

2, avhrr_calculate_ndvi_metrics_tile.pro
2.1, start_batch.pro
2.2, avhrr_time_series_process.pro
2.2.1, avhrr_interpol_extension_1yr.pro
2.2.1.1 cutoff_interp.pro
2.2.1.1.1 interpol_line100b.pro
2.2.1.1.2 filter_2odd.pro
2.2.2 wls_smooth.pro
2.2.3 avhrr_user_metrics_by1yr.pro
2.2.3.1 ComputeMetrics_by1yr.pro
2.2.3.1.1 GetForwardMA.pro
2.2.3.1.2 GetBackwardMA.pro
2.2.3.1.3 GetCrossOver_percentage_extremeslope.pro
2.2.3.1.4 GetSOS.pro
2.2.3.1.5 GetEOS.pro
2.2.3.1.6 GetMaxNDVI.pro
2.2.3.1.7 GetTotNDVI.pro
2.2.3.1.8 GetNDVItoDate.pro
2.2.3.1.9 GetSlope.pro
2.2.3.1.10 GetRange.pro
2.2.3.2 findday.pro

Literature Cited

The conterminous United Stated and Alaska weekly and biweekly AVHRR composites, 07/08/2011

Reed B., M. Budde, P. Spencer, and A. Miller. 2006. Satellite-Derived Measures of Landscape Processes: Draft Monitoring Protocol for the Southwest Alaska I&M Network, ver. 1.0. National Park Service, Inventory & Monitoring Program, Southwest Alaska Network, Anchorage, Alaska. 30 pp.

Daniel L. Swets, Bradley C. Reed, James D. Rowland, and Shaun E. Marko. 1999. A weighted Least-Squares Approach to Temporal NDVI Smoothing. Proceedings of the 1999 ASPRS Annual Conference, Portland, Oregon, pp. 526-536.