POLISH ASSOCIATION FOR SPATIAL INFORMATION
ANNALS OF GEOMATICS 2009 ❍ VOLUME VII ❍ NUMBER 2(32)

NOISE REDUCTION OF NDVI TIME-SERIES: A ROBUST METHOD BASED ON SAVITZKY-GOLAY FILTER

REDUKCJA SZUMÓW W SZEREGACH CZASOWYCH NDVI Z ZASTOSOWANIEM FILTRA SAVITZKY-GOLAY

Jędrzej Bojanowski, Wanda Kowalik, Zbigniew Bochenek

Remote Sensing Department, Institute of Geodesy and Cartography,
Warsaw, Poland

Keywords: NDVI time series, noise reduction, Savitzky-Golay filter, AVHRR, VCI
Słowa kluczowe: szeregi czasowe NDVI, redukcja szumu, filtr Savitzky-Golay, AVHRR, VCI

Introduction

Time series data of vegetation indices derived from medium-resolution satellite data are widely used for vegetation monitoring on a global or regional level. Unquestionably, the most popular indicator is Normalized Difference Vegetation Index, which derived from NOAA/AVHRR, SPOT/VGT, TERRA/MODIS or ENVISAT/MERIS with short revisit periods, can be used as continuous daily time-series. Characterized vegetation condition can be further used for many environmental issues including yield forecasting, which is of special interest to the authors of this paper. Independently of further application, time series of NDVI have to be prepared passing all the steps of processing. Noise reduction needs to be one of those steps, but it is often wrongly omitted. Numerous researches comparing methods of noise reduction in NDVI time series prove there is no one ideal method (Hird, McDermid 2009; Van Dijk et al., 1987; Jönson, Eklundh, 2002; Chen et al., 2004; Beck et al., 2006). All of them assume that NDVI time series are related to vegetation changes and follow the annual cycle of vegetation. There is also agreement that clouds and other atmospheric conditions decrease the NDVI value, so all sudden falls in the time series can be removed.

Methods of noise removal in NDVI time series can be grouped into Fourier based fitting (Andres et al., 1994; Olsson, Eklundh, 1994), thresholds based (Viovy et al., 1992) function fitting (Jönson, Eklundh, 2002; Beck et al., 2006) and several filtering methods (Chen et al., 2004; Velleman, 1980; Ma and Veroustraete, 2006). The most recent and exhaustive comparison

(Hird, McDermid, 2009) demonstrated the general superiority of function-fitting methods. However, reliable NDVI smoothing using some parametric function fitting relies on the fact that this function fits well to all individual pixel time series, which is a pretty strong requirement. For example, Jönson and Eklundh (2002) assume that each yearly NDVI time series can be characterized by an assymetrical gaussian function, which can follow the typical phenological trend. This procedure can break down if the data does not follow the curve that belongs to the assymetric gaussian family. Additionally, function fitting methods require data from entire growing season, which is not necessary in filtering methods.

In this paper we briefly present a whole processing chain for the preparation of NDVI 10-day composites derived from NOAA/AVHRR in 1 km^2 spatial resolution and then we put strong emphasis on final noise reduction of the NDVI time series. The method of noise reduction is based on the Savitzky-Golay filter, which was introduced by Chen (2004). It increases NDVI values contaminated by unmasked clouds and atmospheric variability. Then, we test the results of noise reduction in NDVI time-series and its influence on Vegetation Condition Index (VCI).

Methods

Study area

The test site is an area of 458 700 km^2 (660 695 image pixels), covering the territory of Poland. The area is covered by diverse land use classes, but the effect of noise reduction has been tested only on agriculture lands. To derive a mask of agriculture in 1 km^2 resolution we used Corine Land Cover 2000 data, assuming two classes as agriculture land: *Non-irrigated arable land* (2.3.1) and *Complex cultivation patterns* (2.4.2). We first transformed a vector layer of CLC2000 to a raster layer with 100 m^2 resolution. Then we created a map of 1 km^2 resolution which contained a percentage of agriculture land in each 1 km × 1 km pixel. To avoid inaccuracies caused by geometric precision of the registration of NOAA/AVHRR images (approx. 500–600 meters), pixels which beyond all doubt we could assume as reliable agriculture land had to fulfill two conditions: (1) given pixel had to be agricultural in 90% and (2) it had to be surrounded by eight pixels which satisfied this 90% condition.

Data preprocessing

For this analysis, we used daily NOAA/AVHRR images from 1997 to 2008, which we radiometrically and geometrically corrected. AVHRR instruments measure the radiation in five spectral bands. The first two are centred around the red (0.58–0.68 μm) and near-infrared (0.725–1.0 μm) regions, the third one is located around 3.5 micrometer, and the last two sample the thermal radiation emitted by the Earth, around 11 and 12 micrometers, respectively. The third generation of AVHRR, first carried on the NOAA-15 platform launched in May 1998, also acquires data located at 1.6 micrometer.

Atmospheric correction of red channel 1 and near infrared channel 2 was then performed, using the 6S model (Vermote et al., 1997) and, for value of reflectance ρ, the NDVI index was calculated as:

$$NDVI = (\rho_{NIR} - \rho_{red}) / (\rho_{NIR} + \rho_{red})$$

Next, two channels in the thermal infrared (channel 4 and 5) were used to calculate surface temperature based on a split-window technique (Coll, Caselles, 1997). Then, NDVI values, reflectance in channel 1 and values of surface temperature were used as inputs to the cloud masking algorithm (Kriebel et al., 2003).

After that, 10-day composites of NDVI were calculated using maximum value composite (MVC) technique. Based on NDVI composites, Vegetation Condition Index (VCI) was computed:

$$VCI = 100(NDVI - NDVI_{min})/\ (NDVI_{max} - NDVI_{min})$$

where NDVI, $NDVI_{max}$ and $NDVI_{min}$ are decadal NDVI, multi-year maximum NDVI and multi-year minimum NDVI, respectively, for each pixel. VCI changes from 0 to 100, corresponding to changes in vegetation condition from extremely unfavourable to optimal.

We introduced VCI, because both NDVI and VCI are used widely for yield prognosis (Dąbrowska-Zielińska et al., 2002), and the influence of noise reduction in NDVI time-series on VCI is crucial for further studies. Described above NDVI 10-day composites and VCI have been used in this project. We test them for the noise reduction method.

Noise reduction

The noise reduction method we present in this paper is based on filtering, not function-fitting, as introduced by Chen (2004). It is an iterative technique which, using the Savitzky-Golay filter, preserves the upper NDVI envelope. Here, we first briefly present Savitzky-Golay filter, and then we describe the whole noise reduction method.

Savitzky-Golay filter

Savitzky and Golay (1964) introduced a particular type of low-pass filter. The formula of a simple digital filter can be written as follows:

$$g_i = \sum_{n=-m}^{m} c_n f_{i+n}$$

where each value f_i is replaced by a linear combination g_i of itself and some number of nearby neighbours (Press et al., 2007). We use only a symmetric filtering window, thus $-m$ is neighbours "to the left" and m is "to the right". For example, a simple *moving window averaging* would have constant weights $c_n = 1/(2m+1)$. To maintain higher values, Savitzky and Golay proposed filter coefficients cn given as a polynomial of certain degree. For each point of f_i, they perform a least-squares fit polynomial to all $2m+1$ points in the moving window. The matrix algebra of least-squares fitting causes that for each m and given degree of polynomial one set of $2m+1$ coefficients exists. As an example, a set of filtering coefficients (weights) c_n for 6^{th} polynomial degree and $m=4$ is:
-0.0054, 0.044, -0.15, 0.3, 0.62, 0.3, -0.15, 0.044, -0.0054.

Implementation of the method

The original method of noise removal is precisely described by Chen (2004) and according to the flowchart from this paper (Fig. 1), we briefly present the method, using as an example, time series of NDVI derived from one grid cell. Figure 2a presents NDVI time series for fourteen 10-day periods (decades) starting April 1st. There is no NDVI value for decade 19, because it was classified as cloud by the masking algorithm. As a first step, this value is linearly interpolated. We also examine and linearly interpolate relatively low values of NDVI which are preceded by a decrease of more than 0.2 and followed by an increase of more than 0.2 during 20 days. The result of this interpolation is presented on figure 2a by a dashed line.

Step 2 is very influential when it comes to the final results. In this step, *long-term change trend* fitting is performed using the Savitzky-Golay filter. After several tests for our time-series, a set of m=4 and 3rd polynomial order seems to be appropriate. Dotted line on Figure 2a presents long-term change trend.

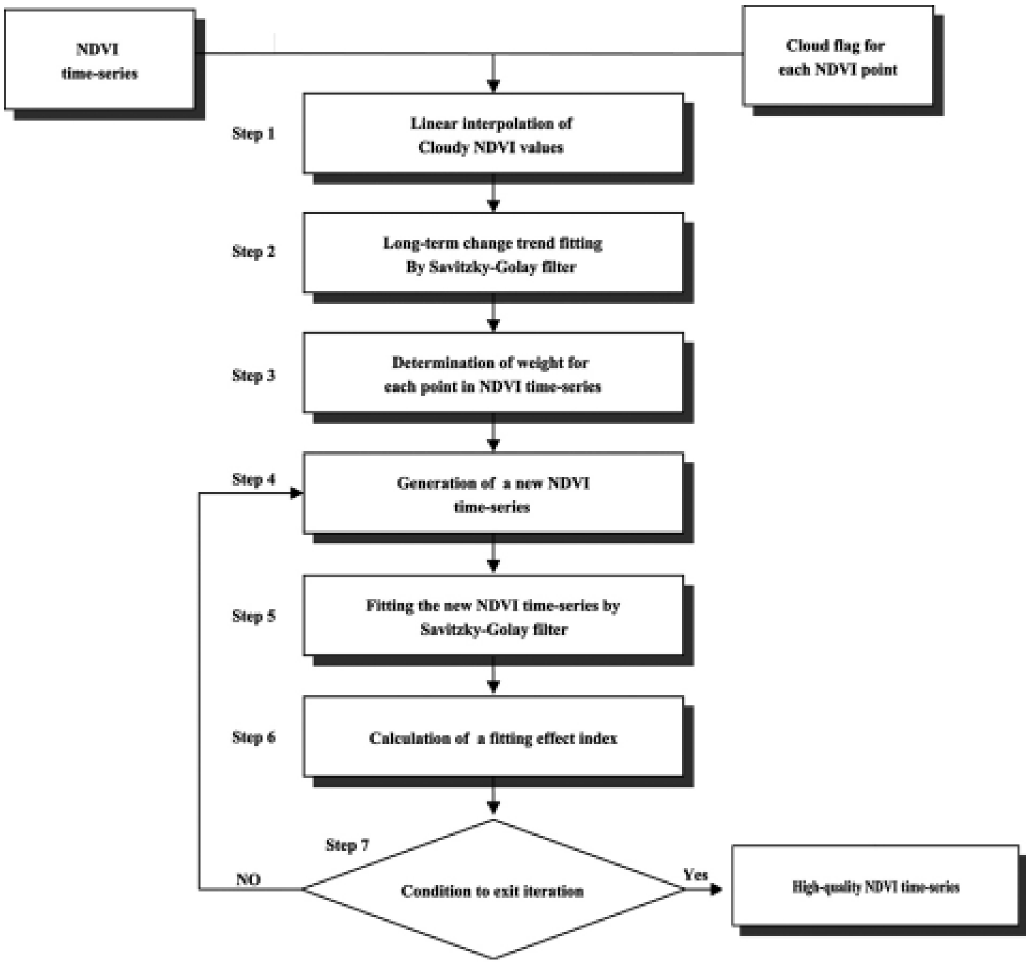

Figure 1. Flowchart of the noise reduction algorithm (Chen et al., 2004)

This new curve is further used to determine the weights of each point of NDVI time series (step 3). The idea is to signify points over the trend line by giving them high weight equalled to 1. All points which are below the trend line have lower weights (less than 1). The bigger distance between the point and the long-term change fit the smaller weight it has. We can see those weights on Figure 2a.

We assume that points below the trend line are the noisy NDVI points. Thus, in step 4, all the points of time series, which weights are lower than 1, are replaced by the values of long-term change curve (arrows on Figure 2b).

Step 4, 5 and 6 are iterative. In step 5, we aim to generate a new time-series using the Savitzky-Golay filter. Here we used a set of coefficients cn mentioned before, which are suited for m=4 and 6th polynomial order. In step 6, after filtering, we calculate a *fitting-effect index* based on the distance between newly derived time series and the original one, but we also take into account the importance of each point, which is equal to weights calculated in step 3. The formula for fitting-index for k^{th} iteration (F_k) looks as follows (Chen et al., 2004):

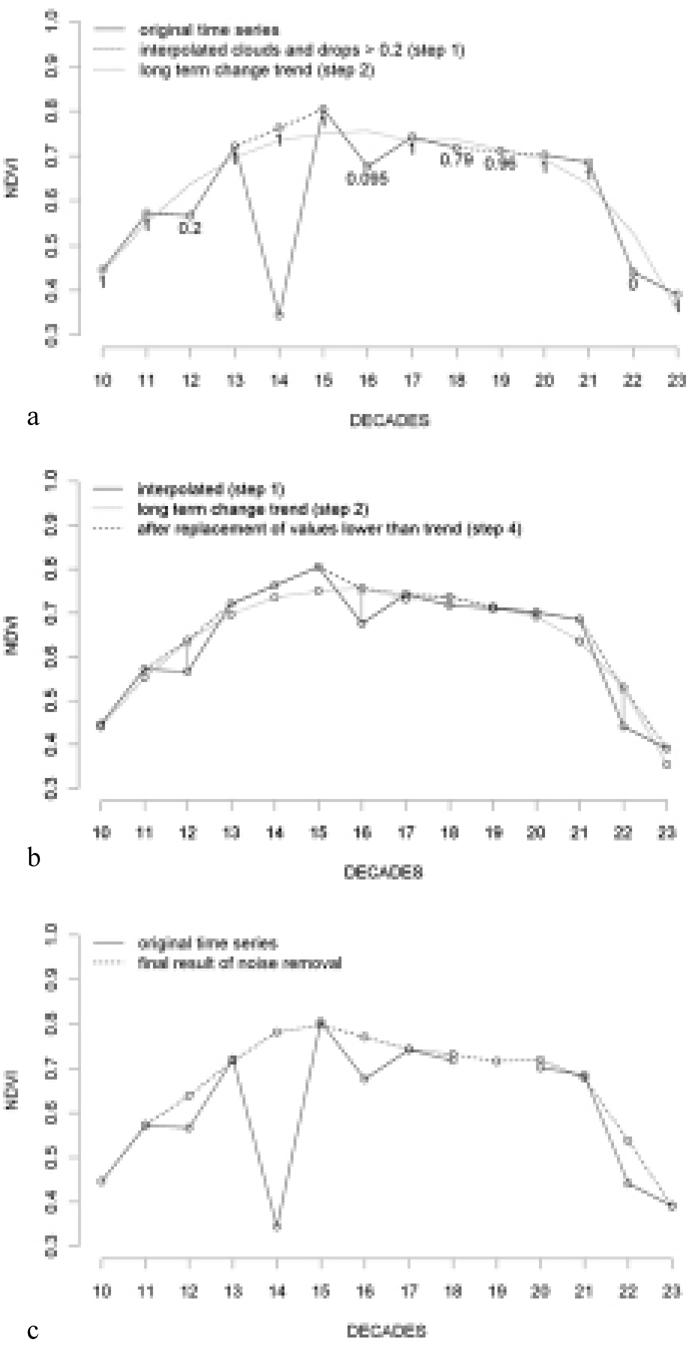

Figure 2. An example of NDVI time series for one pixel. Consecutive steps of processing are described in the text

$$F_k = \sum_{i=1}^{n} \left(\left| N_i^{k+1} - N_i^0 \right| \times W_i \right)$$

where N_i^{k+1} is i^th NDVI value of time-series derived in k^th iteration, N_i^0 is the ith NDVI value of the original NDVI time-series with interpolated cloudy NDVI values (after step 1) and W_i is the weight of i^{th} NDVI value calculated in step 5. Before next iteration, similarly to step 4, all the points of the new time series, which are below the curve of the original time series, are replaced by the values of this original time series. It effects that high original NDVI values are preserved.

Thus, for each iteration k, a new time series is obtained and their *fitting-effect indices* F_k are calculated. The analysis made by Chen (2004), then repeated on our dataset, proves that best fit (the lowest F_k) is always derived for k less than 10, usually for k between 3 and 6.

Therefore, the final time-series with removed noise is that one with the lowest F_k. It means that the number of iterations is different for each time-series derived from one pixel and one year. Figure 2c shows an example of the original time series and its final iteration.

Results and Discussion

We reduced the noise in NDVI using the method described above in more than 5 million time series, which is equal to the number of pixels multiplied by number of years. The method was implemented and run in the R environment (R Development Core Team, 2009) and the Savitzky-Golay filter was run using a signal package (Short, 2006). Working on 64-bit Windows on Pentium D 3.0 GHz with 3.24 GB of RAM, noise reduction of one time series takes 0.1 second on the average. Computation of 12 years of NDVI data for image size 660 695 pixels took seven days.

The results obtained for exemplary separate time series can be seen on Figure 3. Taking into account the theoretical curve of NDVI, which reflects annual cycle of growth and decline of vegetation, we definitely affirm that time series after noise removal has better quality. All sudden drops caused by atmospheric contamination have been removed. Simultaneously, higher values of NDVI have been preserved. It is crucial that noise reduction method does not disturb a general form of the curve, and phenological parameters like start of growing season, end of growing season, maximum green-up, etc. could be determined.

The influence of the noise reduction can be also seen in the spatial domain. Figure 4 presents an image of NDVI before and after noise removal. On the left image we can distinct pixels which are close to clouds and their NDVI values are noticeably lowed by atmospheric contamination. After noise removal this effect disappeared.

To generalize the interpretation of the results, we checked the influence of the performed method on values of NDVI averaged in administration divisions. Figure 5 shows that the differences are significant, and sudden drops have been removed while higher envelope has remained.

NDVI is often the base of computation of other indices like Vegetation Condition Index (VCI), which formula was already presented. VCI can be understood as actual NDVI scaled in relation to extreme values of NDVI which can ever appear in this area and at time of season. It makes VCI very sensitive to noise in NDVI time series. Figure 6 gives an answer that VCI calculated after noise removal in NDVI time series becomes smooth and all false variations are removed.

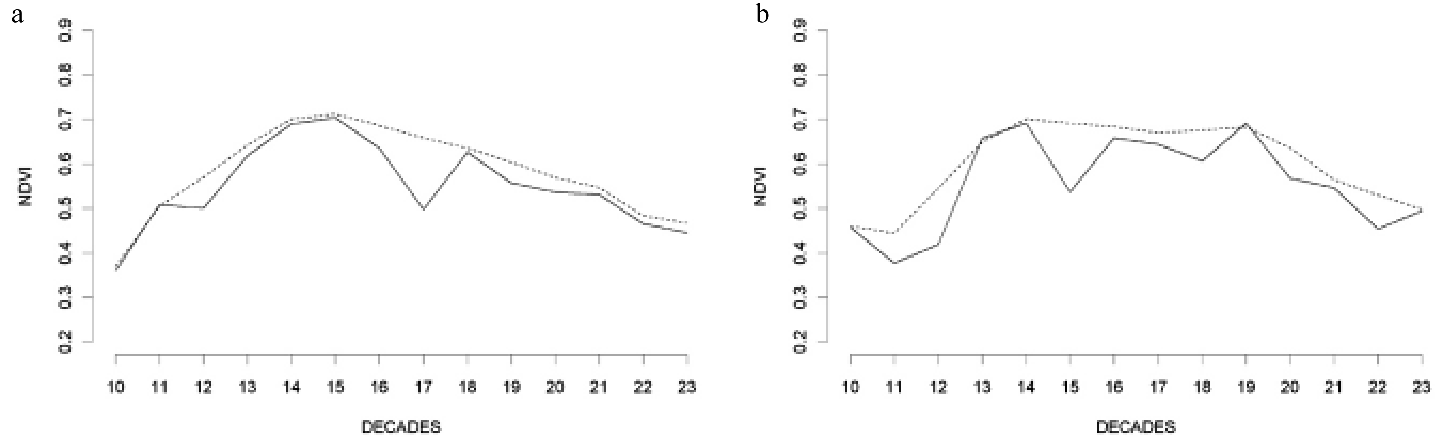

Figure 5. Examples of yearly time series of NDVI averaged in voivodships: a – Dolnośląskie 1999, b – Świętokrzyskie 2002. Solid line indicates original time series, dashed line after noise reduction

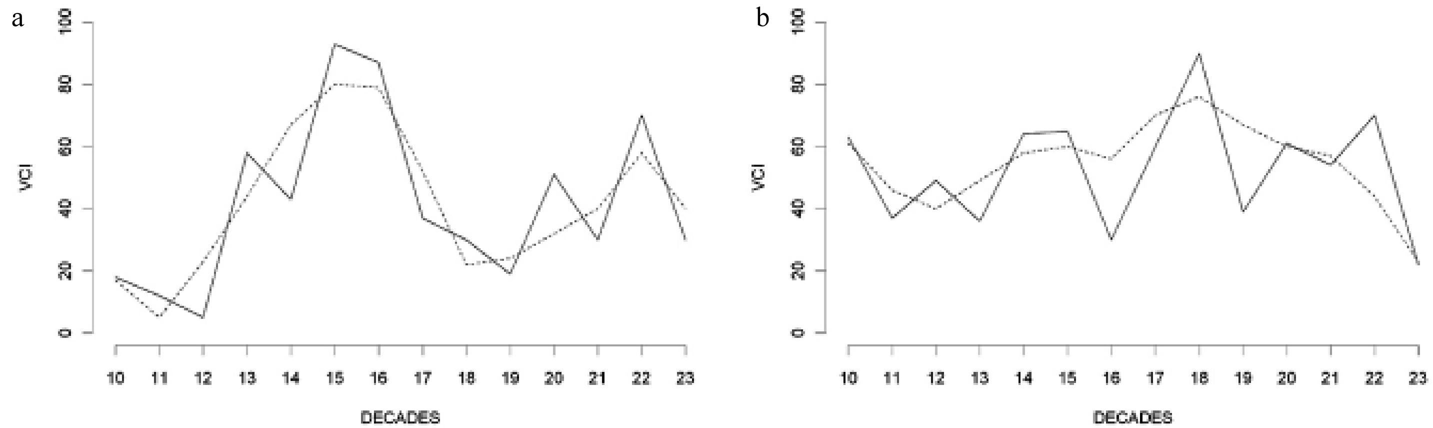

Figure 6. Examples of yearly time series of VCI averaged in voivodships: a – Łódźkie 1999, b – Dolnośląskie 2005. Solid line indicates original time series, dashed line after noise reduction.

Conclusions

The use of the Savitzky-Golay filter with a changeable number of iterations seems to be appropriate for noise removal in a NDVI time series. It preserves high values of NDVI and the general curve. It is not necessary to have NDVI values for a whole growing season, as in function-fitting methods, which gives an opportunity to perform a noise reduction before the end of season. The quality of VCI derived from noiseless NDVI is significantly higher. The applied method gives smooth NDVI and VCI time series and also delivers clean, unclouded images for visual interpretation.

References

Andres L., Salas W.A., Skole D., 1994: Fourier-analysis of multitemporal AVHRR data applied to a land-cover classification. *International Journal of Remote Sensing*, 15, No 5: 1115-1121.

Beck P.S.A., Atzberger C., Hogda K.A., Johansen B., Skidmore A.K., 2006: Improved monitoring of vegetation dynamics at very high latitudes: A new method using MODIS NDVI. *Remote Sensing of Environment*, 100, Issue 3: 321-334.

Chen J., Jönsson P., Tamura M., Gu Z., Matsushita B., Eklundh B., 2004: A simple method for reconstructing a high-quality NDVI time-series data set based on the Savitzky-Golay filter. *Remote Sensing of Environment*, 91, Issues 3-4: 332-344.

Coll C., Caselles V., 1997: A split-window algorithm for land surface temperature from advanced very high resolution radiometer data: Validation and algorithm comparison. *Journal of Geophysical Research*, 102(D14): 16, 697-16, 713.

Dąbrowska-Zielińska K., Kogan F., Ciołkosz A., Gruszczyńska M., Kowalik W., 2002: Modeling of crop growth conditions and crop yield in Poland using AVHRR based indices. *International Journal of Remote Sensing*, 23, No 6: 1109-1123.

Hird J.N., McDermid G.J., 2009: Noise reduction of NDVI time series: An empirical comparison of selected techniques. *Remote Sensing of Environment*, 113, Issue 1: 248-258.

Jönson P., Eklundh L., 2002: Seasonality Extraction by Function Fitting to Time-Series of Satellite Sensor Data. *IEEE Transactions on Geoscience nad Remote Sensing*, 40, No 8: 1824-1832.

Kriebel K.T., Gesell G., Kastner M., Mannstein H., 2003: The cloud analysis tool APOLLO: Improvements and Validation, *International Journal of Remote Sensing*, 24, No 12: 2389-2408.

Ma M., Veroustraete F., 2006: Reconstructing pathfinder AVHRR land NDVI time-series data for the Northwest of China. *Advances in Space Research*, 37: 835-840.

Olsson L., Eklundh L., 1994: Fourier-series for analysis of temporal sequences of satellite sensor imagery. *International Journal of Remote Sensing*, 15, No 18: 3735-3741.

Press W.H., Teukolsky S.A., Vetterling W.T., Flannery B.P., 2007: Numerical Recipes, 3rd edition. Cambridge University Press.

R Development Core Team, 2009: R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.r-project.org.

Savitzky A., Golay M.J.E., 1964: Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry*, 36: 1627-1639.

Short, T., 2006: Signal processing. http://cran.r-project.org/web/packages/signal, R package version 0.5.

van Dijk A., Callis S., Sakamoto C., Decker W., 1987: Smoothing vegetation index profiles: An alternative method for reducing radiometric disturbance in NOAA/AVHRR data. *Photogrammetric Engineering and Remote Sensing*, 53: 1059-1067.

Velleman P.F., 1980: Definition and Comparison of Robust Nonlinear Data Smoothing Algorithms. *Journal of the American Statistical Association*, 75 (371): 609-615.

Vermote E., Tanre D., Deuze J.L., Herman M., Morcrette J.J., 1997: Second Simulation of the Satellite Signal in the Solar Spectrum: an overview. *IEEE Transactions on Geoscience and Remote Sensing*, 35: 675-686.

Viovy N., Arino O., Belward A.S., 1992: The Best Index Slope Extraction (BISE) – A method for reducing noise in NDVI time-series. *International Journal of Remote Sensing* 13, No 8: 1585-1590.

Streszczenie

Znormalizowany wskaźnik zieleni (NDVI) otrzymywany ze średnio-rozdzielczych obrazów satelitarnych, jak NOAA/AVHRR, jest od wielu lat szeroko stosowany do monitoringu środowiska. Jednym z takich zastosowań jest modelowanie kondycji i wzrostu roślin oraz prognozowanie plonów. Wykorzystywane w tym celu szeregi czasowe wskaźnika NDVI są obarczone znaczącymi błędami wynikającymi z wpływu atmosfery i geometrii układu Słońce-Ziemia-sensor. Chmury i para wodna występujące w atmosferze pochłaniają promieniowane podczerwone, co skutkuje zaniżonymi wartościami wskaźnika NDVI. Wpływ ten można zauważyć zarówno w czasie (w szeregach czasowych NDVI), jak również w przestrzeni (na pojedynczych obrazach NDVI). Metoda redukcji szumów w szeregach czasowych NDVI bazująca na filtrze Savitzkiego i Golaya została zaprezentowana i przetestowana na dwunastoletniej bazie NDVI dla terytorium Polski. Zbadany został również wpływ redukcji szumu w NDVI na wskaźnik VCI. W wyniku zastosowanej metody znacznie poprawiła się jakość i wiarygodność szeregów czasowych NDVI i VCI, jak również otrzymano czyste obrazy przydatne przy interpretacji wizualnej.

mgr Jędrzej Bojanowski (corresponding author)
jedrzej.bojanowski@igik.edu.pl
+4822 329 19 77

mgr inż. Wanda Kowalik
wanda.kowalik@igik.edu.pl
+4822 329 19 78

dr inż. Zbigniew Bochenek
zbigniew.bochenek@igik.edu.pl
+4822 329 19 77

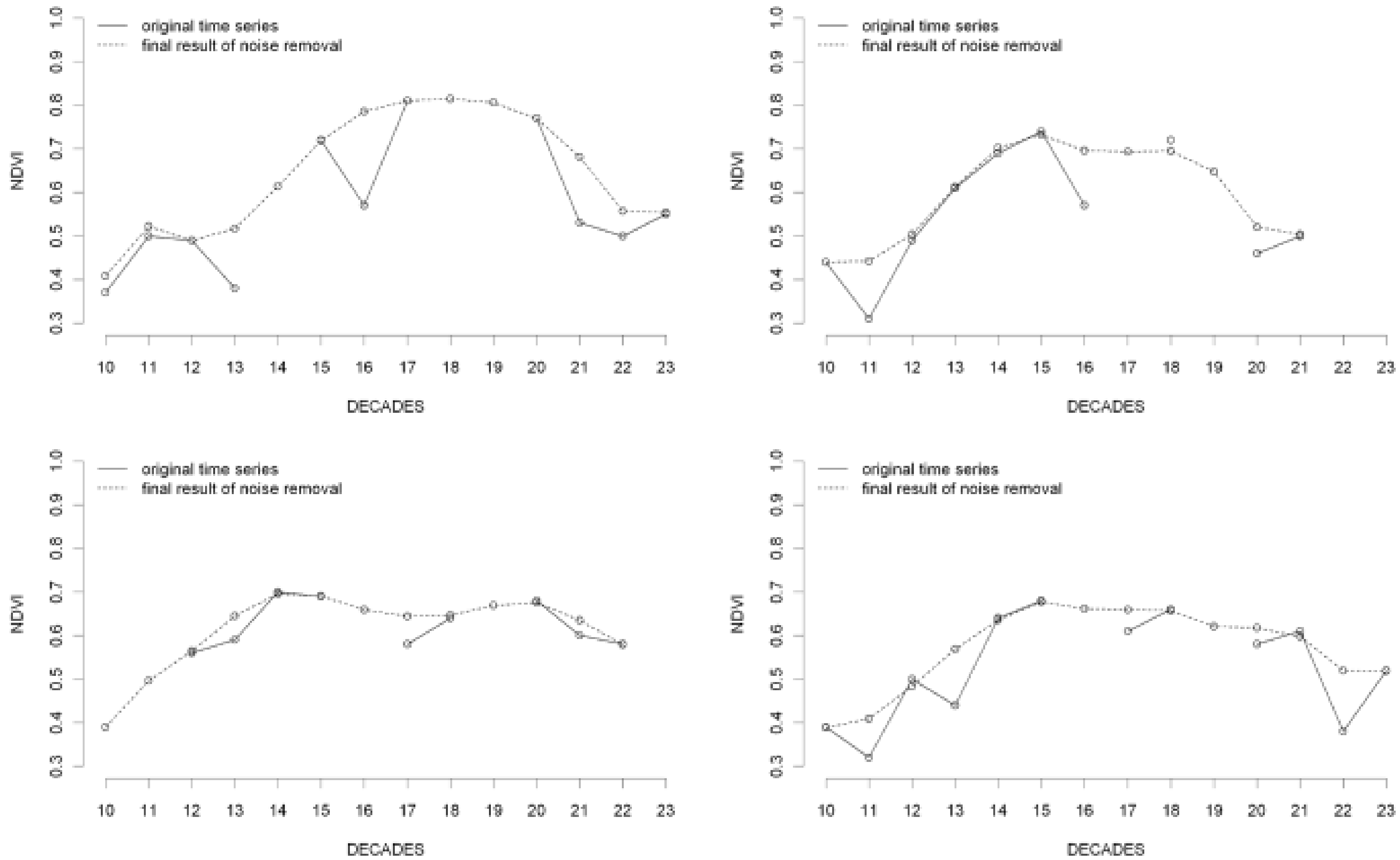

Figure 3. Examples of noise reduction method results for individual pixels

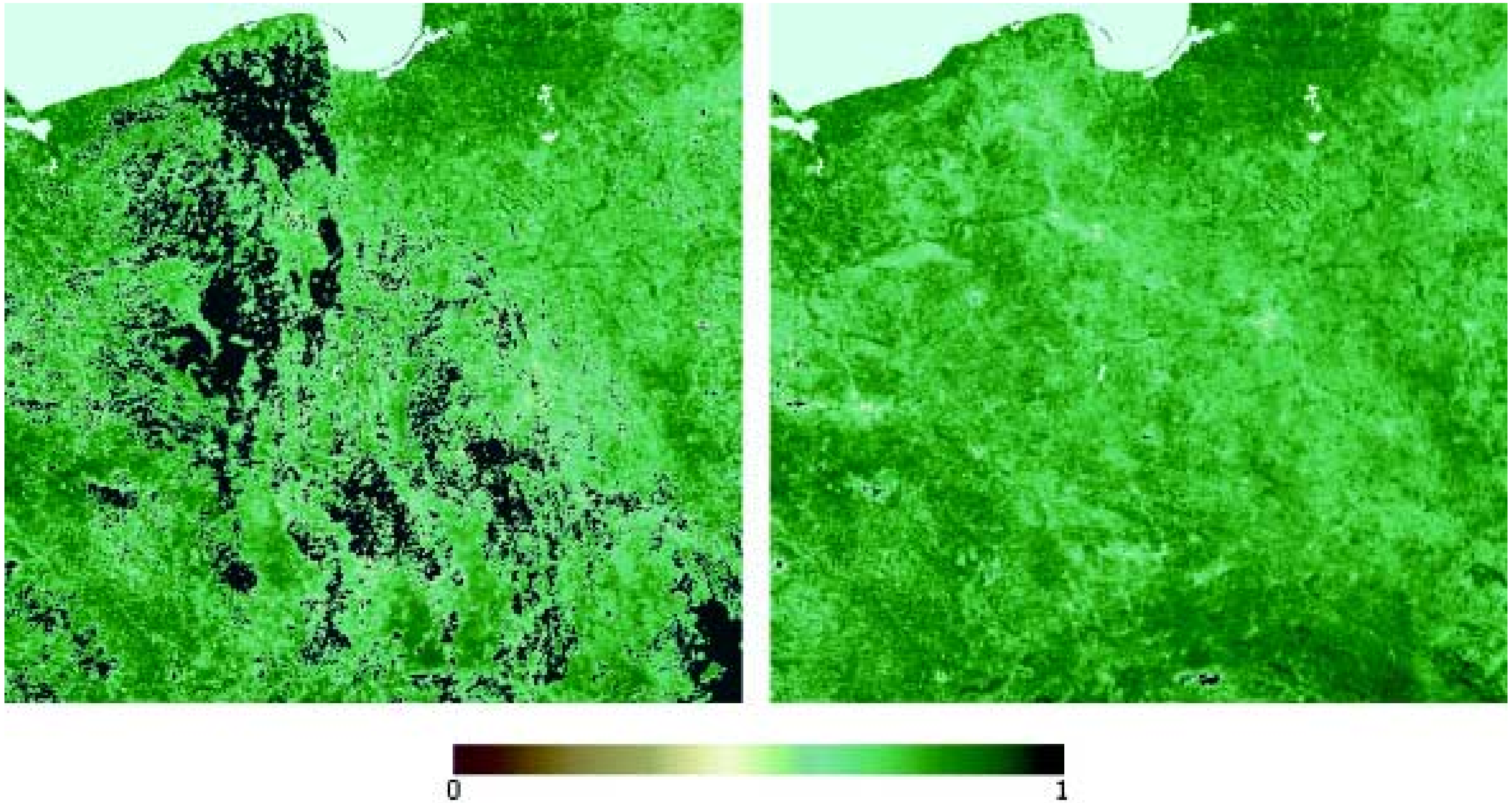

Figure 4. NDVI image of study area before (left) and after (right) noise removal