Article

Trend Change Detection in NDVI Time Series: Effects of Inter-Annual Variability and Methodology

Matthias Forkel [1],*, Nuno Carvalhais [1,2], Jan Verbesselt [3], Miguel D. Mahecha [1], Christopher S.R. Neigh [4] and Markus Reichstein [1]

[1] Department for Biogeochemical Integration, Max-Planck-Institute for Biogeochemistry, Hans-Knöll-Str. 10, 07745 Jena, Germany; E-Mails: ncarval@bgc-jena.mpg.de (N.C.); mmahecha@bgc-jena.mpg.de (M.D.M.); mreichstein@bgc-jena.mpg.de (M.R.)

[2] Faculdade de Ciências e Tecnologia, FCT, Universidade Nova de Lisboa, 2829-516 Caparica, Portugal

[3] Laboratory of Geo-Information Science and Remote Sensing, Wageningen University, Droevendaalsesteeg 3, 6708 PB Wageningen, The Netherlands; E-Mail: jan.verbesselt@wur.nl

[4] Biospheric Sciences Laboratory, NASA Goddard Space Flight Center, Greenbelt, MD 20771, USA; E-Mail: christopher.s.neigh@nasa.gov

* Author to whom correspondence should be addressed; E-Mail: mforkel@bgc-jena.mpg.de; Tel.: +49-3641-576-283; Fax: +49-3641-577-200.

Abstract: Changing trends in ecosystem productivity can be quantified using satellite observations of Normalized Difference Vegetation Index (NDVI). However, the estimation of trends from NDVI time series differs substantially depending on analyzed satellite dataset, the corresponding spatiotemporal resolution, and the applied statistical method. Here we compare the performance of a wide range of trend estimation methods and demonstrate that performance decreases with increasing inter-annual variability in the NDVI time series. Trend slope estimates based on annual aggregated time series or based on a seasonal-trend model show better performances than methods that remove the seasonal cycle of the time series. A breakpoint detection analysis reveals that an overestimation of breakpoints in NDVI trends can result in wrong or even opposite trend estimates. Based on our results, we give practical recommendations for the application of trend methods on long-term NDVI time series. Particularly, we apply and compare different methods on NDVI time series in Alaska, where both greening and browning trends have been previously observed. Here, the multi-method uncertainty of NDVI trends

is quantified through the application of the different trend estimation methods. Our results indicate that greening NDVI trends in Alaska are more spatially and temporally prevalent than browning trends. We also show that detected breakpoints in NDVI trends tend to coincide with large fires. Overall, our analyses demonstrate that seasonal trend methods need to be improved against inter-annual variability to quantify changing trends in ecosystem productivity with higher accuracy.

Keywords: greening; browning; breakpoints; seasonal cycle; season-trend model; boreal forest; tundra; fire; disturbances; Alaska

1. Introduction

Climate change will likely change biome distributions, ecosystem productivity and forest carbon stocks [1]. Such ecosystem changes can be detected and quantified using multi-temporal satellite observations of the land surface. Different states of the land surface can be measured by satellite-derived biophysical parameters [2]. The Normalized Difference Vegetation Index (NDVI) [3] is a remotely-sensed measure of vegetation greenness and is related to structural properties of plants—like leaf area index [4] and green biomass [5]—but also to properties of vegetation productivity—like absorbed photosynthetic active radiation and foliar nitrogen [5-6]. As NDVI is related to such a variety of vegetation properties, multiple explanations for a change in NDVI signals are possible. Nevertheless, the NDVI from AVHRR (Advanced Very High Resolution Radiometer) satellite observations is the only global vegetation dataset which spans a time period of three decades and thus allows the quantification and attribution of ecosystem changes as a result of ecosystem dynamics and varying climate conditions. Different ecosystem changes can be analyzed from NDVI time series. For example, annual mean or peak NDVI provides an integrated view on photosynthetic activity [7], the seasonal NDVI amplitude is related to the composition of evergreen and deciduous vegetation [8] and the length of the NDVI growing season can be related to phenological changes [9]. Thus, trend detection in NDVI time series can help to identify and quantify recent changes in ecosystem properties from a local to global scale.

Indeed, positive NDVI trends ("greening") occur in the high latitudes [10]. These greening trends were reproduced by a Dynamic Global Vegetation Model (DGVM) and attributed to increasing temperatures [11]. The temperature increase drives an expansion of shrubs in the arctic Tundra, which can be observed as greening trends [12]. The initial greening trend stalled or reversed in large parts of the boreal forest of Northern America. Negative NDVI trends ("browning") are associated with fire activity [13], increasing water vapour pressure deficit [14] or to cooling spring temperatures [15]. Regional changes in summer precipitation changed greening NDVI trends to browning trends also in Eurasian boreal forests in the late 1990s [16]. Nevertheless, browning NDVI trends are highly discussed because they differ based on the used satellite dataset [17,18]. Most studies used the GIMMS (Global Inventory, Monitoring, and Modeling Studies) NDVI dataset which was produced based on 4 km AVHRR satellite observations [19]. A comparison between the GIMMS dataset and a Canadian dataset shows weaker post-fire recovery trends and more negative NDVI trends in unburned forests in

the GIMMS dataset [20]. Other studies confirm trend estimates based on the GIMMS dataset: Despite of some regional differences in areas at very high latitudes with low vegetation cover, NDVI trends from the GIMMS dataset agree with trends from MODIS data (Moderate Resolution Imaging Spectrometer) [17,18,21]. Trends from the GIMMS dataset compare well with trends computed from Landsat imagery [22]. Changes in tree rings [23,24], temperature-induced drought stress or insect disturbances [25] were also observed in regions with browning NDVI trends. In fact, impacts of recent trends and variability of climate on ecosystems can be observed using long-term NDVI time series.

The estimation of trends depends on the length, temporal and spatial resolution of the time series, the quality of the measured data [26] and the used statistical method. Many studies calculated trends based on annual time steps from annually or seasonally aggregated values using regression analysis [27]. However, the use of linear regression analysis for estimating trends in NDVI time series violates statistical assumptions such as the independence of observations, due to temporal autocorrelation or homogeneity [28]. Accordingly, the application of temporal autocorrelation structures [13] or the use of the non-parametric Mann-Kendall test on NDVI time series was suggested to circumvent the limitations of regression analysis [28,29]. The annual aggregation of time series for trend analysis reduces the temporal resolution and time series length. The time series length is critical in determining the significance of the trend in a statistical test. On the other hand, annual aggregation supports the analysis of trends by eliminating the seasonal cycle in the NDVI time series. The seasonal cycle introduces a seasonal correlation structure that hampers trend analysis. In this context, methods were developed that make use of the full resolution time series by estimating and subtracting the seasonal cycle or by modelling the seasonal signal [30–34]. Overall, trend estimates from these different methods result in similar general spatial patterns of the major regional greening and browning trends but substantial differences in areas with weak trends [31]. In short, all trend estimation methods embed caveats that may be more or less critical depending on the application.

NDVI trends are not always monotonic but can change. A positive trend can change for example into a negative one and *vice versa*. Changes between initial greening trends in the 1980s to browning trends from the 1990s onwards in high latitude regions were detected based on the GIMMS NDVI dataset [35]. NDVI trend changes of this kind can be either gradual or abrupt [36]. For example, increasing temperatures in temperature-limited ecosystems can first support vegetation growth that results in greening NDVI trends, while a further warming can induce drought stress that slightly turns the initial greening to a browning trend (gradual change). Disturbances such as fire events can reduce the NDVI signal and initiate post-fire recovery that results in a greening trend (abrupt change). Recently, statistical methods were developed and applied to NDVI time series to detect such changes (called breakpoints) in trends. Methods like BFAST (Breaks for additive seasonal and trend) [30,33] combine trend estimation with approaches that account for breakpoints in the trend. However, the reliability of such breakpoints in NDVI trends in high latitude regions is not yet assessed.

Breakpoints in NDVI time series are related to different effects caused by inter-annual variability. Inter-annual variability of NDVI time series can be caused by (1) artefacts of a harmonized dataset from different sensors, (2) meteorological distortions like clouds or snow cover and (3) environmental processes like effects of year-to-year variations in weather conditions on plant activity or ecosystem disturbances. Inter-annual variability affects the annual mean (e.g., reduction of NDVI because of a disturbance), seasonality (e.g., longer growing season because of prolonged warmer temperatures) as

well as short-term patterns (e.g., unusual snowfall in a summer month) of NDVI time series. The aggregation of NDVI time series to mean annual values integrates these different effects which, despite the loss in temporal detail, allow us to define and quantify inter-annual variability as the standard-deviation of mean annual NDVI values.

The purpose of this study is to evaluate the performance of different methods for detecting trends and trend breakpoints in long-term NDVI time series. Previous studies have used different trend and breakpoint analysis methods on NDVI time series without or with limited demonstration of its methodological robustness [15]. By evaluating the performance of such methods, this study will enable a critical appraisal of combined trend and breakpoint detection methods for their application on NDVI time series. The methods chosen for evaluation differ on their used temporal resolution of NDVI time series, how seasonality is accounted for, and how the trend is estimated. All approaches make use of the same breakpoint detection algorithm [37]. A factorial experiment was performed based on surrogate (or "artificial") NDVI time series with different levels of trend magnitude, inter-annual and short-term variability, seasonal amplitudes and a varying number of trend changes. We tested whether the methods are able to re-detect the prescribed trend (*i.e.*, slope of the trend) and trend changes (*i.e.*, number and timing of breakpoints) in the surrogate time series. Additionally, methods were applied to real NDVI time series of Alaska and the plausibility of breakpoints was assessed in comparison to fire events and drought periods. Our results reveal a clear dependence of the method's performance on the degree of inter-annual variability.

2. Data and Methods

2.1. GIMMS NDVI3g Dataset

The GIMMS NDVI3g dataset (third generation GIMMS NDVI) is a newly available long-term NDVI dataset and was derived from NOAA AVHRR data (National Oceanic and Atmospheric Administration, Advanced Very High Resolution Radiometer) [38,39]. In comparison to a previous version of the dataset [19], it was improved for applications in high-latitude regions through calibrations to stable targets in these regions [38,39]. The dataset covers the period July 1981 until December 2011 with a 2-weekly temporal and 8 km spatial resolution. The quality of AVHRR data is affected from sensor changes between the NOAA satellites and orbital decay but it was shown that trends based on the previous NDVI dataset are not affected by these artefacts [40]. In the new GIMMS NDVI3g dataset such effects were substantially reduced [38,39].

We further pre-processed the GIMMS NDVI3g dataset for the specific requirements of our study. The year 1981 was excluded from our analysis in order to analyse only years with full data coverage. Especially in high-latitude regions NDVI observations are often affected from snow or cloud cover. Such NDVI values are flagged as "snow" or "interpolated" in the GIMMS NDVI3g dataset [38,39]. The reliability of such interpolated NDVI values under snow or cloud conditions is unclear. We addressed this fact in two ways: (1) we assumed interpolated NDVI observations under snow conditions are the best available estimate and we did not change these NDVI values (hereinafter called "all" observations). (2) NDVI values that were flagged as "snow" were excluded from the analysis (hereinafter called "ex" observations). We kept interpolated observations that were not additionally

flagged as "snow" to make sure to use enough observations throughout the growing season because three of the four assessed trend methods need seasonal observations in order to be applicable for trend detection (Section 2.3). To exclude potentially remaining effects of cloud or haze contaminations, the dataset was further aggregated to monthly temporal resolution using the monthly maximum value which is a commonly applied procedure [41].

2.2. Breakpoint Detection Algorithm

The breakpoint detection algorithm as described by Bai and Perron [37] and Zeileis *et al.* [42] was used in this study. The breakpoint detection algorithm searches for a structural change in a regression relationship, *i.e.*, for varying regression parameters before and after the breakpoint. That means a detected breakpoint splits a time series in two segments. In a first step, the ordinary-least squares moving sum (MOSUM) test is used to test for the existence of breakpoints in the time series. If the test indicates significant structural changes ($p \leq 0.05$), different numbers and locations of breakpoints are iteratively tested in the second step. For this purpose, the optimal number of breakpoints is estimated by minimizing the Bayesian Information Criterion (BIC). The optimal position of a breakpoint is estimated by minimizing the residual sum of squares of this regression [37,42].

The breakpoint detection algorithm was used based on recommendations of Bai and Perron [37]. In order to detect long-term trend changes, a minimum amount of observations between two breakpoints was defined as 48 monthly observations (or four years) and a maximum number of two allowed breakpoints were selected. Therefore, an optimized number of breakpoints between zero and two can be detected. This prevented that detected breakpoints are solely affected by year-to-year changes and supported the detection of only major breakpoints in the long-term trend. Further, detected time series segments of a length smaller than eight years were not considered as trends.

2.3. Methods for Trend Estimation

2.3.1. Trend Estimation on Annual Aggregated Time Series (Method AAT)

Method AAT estimates trends and trend changes on annual aggregated time series. The seasonal NDVI time series is first aggregated to annual values. The annual mean, growing season mean or annual peak NDVI can be calculated to aggregate the seasonal NDVI time series to annual values. Mean annual NDVI was used for the factorial experiment based on surrogate time series. Breakpoints are estimated on the annual time series using the method of Bai and Perron [37]. For each derived trend segment the slope of the trend is estimated by linear least-squares regression of the annual values against time. The significance of the trend in each time series segment is estimated by the Mann-Kendall trend test applied on the annual aggregated NDVI values [43].

2.3.2. Trend Estimation Based on a Season-Trend Model (Method STM)

The trend and breakpoint estimation in method STM (season-trend model) is based on the classical additive decomposition model and we followed the formulation used in BFAST [33,44]. The full temporal-resolution NDVI time series is explained by a piecewise linear trend and a seasonal model in a regression relationship. Thus, the NDVI value *y* at a time *t* can be expressed as:

$$y_t = \alpha_1 + \alpha_2 t + \sum_{j=1}^{k} \gamma_j \sin\left(\frac{2\pi j t}{f} + \delta_j\right) + \varepsilon_t \tag{1}$$

where α_1 is the intercept and α_2 the slope of the trend, γ are the amplitudes and δ the phases of k harmonic terms and ε is the residual error [44]. The frequency f is the number of observation per year (*i.e.*, 12 for monthly observations). Parameters α_1, α_2 are estimated using ordinary least squares (OLS) regression whereby the derived time series segments are considered as categorical interaction term with the trend slope α_2. The significance of the trend in each segment is estimated from a t-test on the interaction parameter of the regression between time series segment and α_2.

2.3.3. Trend Estimation on De-Seasonalized Time Series

Methods MAC (mean annual cycle) and SSA (annual cycle based on singular spectrum analysis) estimate trends on seasonal-adjusted time series, which is the full-resolution time series with removed seasonality. The seasonal-adjusted time series a is the difference between the original NDVI time series y and the seasonal cycle s:

$$a = y - s \tag{2}$$

The slope of the trend α_2 is estimated using OLS from the seasonal-adjusted time series:

$$a_t = \alpha_1 + \alpha_2 t + \varepsilon_t \tag{3}$$

Breakpoints are estimated on Equation (3) with different regression coefficients for each trend segment. The significance of the trend in each time series segment is estimated with the Mann-Kendall trend test applied on the seasonal-adjusted time series.

The seasonal cycle (or annual cycle) s (Equation (2)) is represented by a mean annual cycle (method MAC) and by a modulated annual cycle (method SSA). The MAC is estimated as the mean seasonal cycle from the seasonal cycles of all years. This implies that each year has the same amplitude and frequency in the seasonal cycle. However, the concept of a fixed seasonal cycle is questionable as it can change due to external forcing [34]. For example, phenological cycles might change without affecting the overall trend in a time series. Therefore, method SSA is based on a modulated annual cycle with slightly varying frequencies and amplitudes of the seasonal cycle amongst years. The modulated seasonal cycle was estimated using a one-dimensional singular spectrum analysis (SSA) as described in [45]. Singular spectrum analysis was previously used to separate remotely-sensed FAPAR (fraction of observed photosynthetically active radiation) time series into low and high frequency and seasonal time series components [32]. SSA decomposes in a first step a time series into different sub-signals with characteristic frequencies. In a second step, the sub-signals with an annual frequency were summed to build up the modulated annual cycle.

2.4. Simulation of Surrogate Time Series

2.4.1. Estimation of Inter-Annual Variability, Seasonality and Short-Term Variability from Observed Time Series

An important aspect of the experimental design was the prescription of time series properties in the surrogate (artificial) data that were observed in the NDVI datasets. In order to create surrogate time

series that mimic the full range of possible real world data, the mean, trend, inter-annual variability, seasonality and short-term variability was estimated for all observed NDVI time series of Alaska in a simple step-wise approach (Figure 1):

(1) The mean of each NDVI time series was calculated.

(2) In the second step, monthly values were averaged to annual values and the trend was calculated according to method AAT but without computing breakpoints. Hence, the slope of the annual NDVI trend over the full length of the time series was estimated.

(3) To estimate the inter-annual variability, the standard deviation and range of the annual anomalies were calculated. The mean of the time series and the derived trend component from step (2), were subtracted from the annual values to derive the trend-removed and mean-centred annual values (annual anomalies). If the trend slope was not significant ($p > 0.05$), only the mean was subtracted. The standard deviation and the range of the annual anomalies were computed as measures for the inter-annual variability of the time series.

Figure 1. Estimated time series components for a random-selected example grid cell in central Alaska (3*3 grid cells averaged around central pixel 146.424 °W, 64.762 °N). The upper panel shows the original Normalized Difference Vegetation Index (NDVI) time series with its mean value (red line). The next panels show the estimated trend, inter-annual variability (IAV) (*i.e.*, annual anomalies), seasonality (*i.e.*, mean seasonal cycle) and short-term variability (remainder component), respectively. The sum of mean, trend, IAV, seasonal and remainder component equals the original time series.
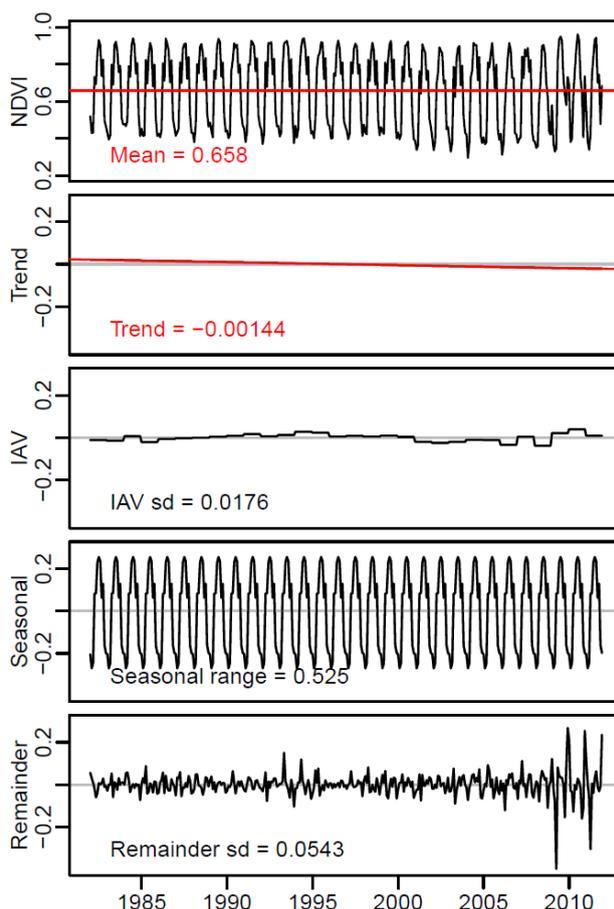
(4) In the next step, the range of the seasonal cycle was estimated. The mean, the trend component and the annual anomalies were subtracted from the original time series to calculate a detrended, centered and for annual anomalies adjusted time series. Based on this time series the seasonal cycle was estimated as the mean seasonal cycle and the range was computed.

(5) In the last step, the standard deviation and the range of the short-term anomalies were computed. Short-term anomalies were computed by subtracting the mean, the trend component, the annual anomalies and the mean seasonal cycle from the original time series. The result is the remainder time series component. The standard deviation of the remainder time series component is a measure of short-term variability.

All the described time series properties were estimated on the full NDVI dataset including all observations (*i.e.*, including snow-affected observations). Hence, we could generate a wide range of gap free surrogate time series. This procedure was applied for all NDVI time series of Alaska to estimate spatial and statistical distributions of the mean NDVI, the overall trend slope, the inter-annual variability as the standard deviation of the annual anomalies, the range of the seasonal cycle and the short-term variability as the standard deviation of the remainder time series (Figure 2).

Figure 2. Spatial and statistical distributions of NDVI time series properties in Alaska and time series components of the simulated NDVI time series. The left panel shows from top to bottom maps of the following time series properties: mean annual NDVI, slope of the annual trend (ΔNDVI/year), standard deviation of the inter-annual variability (iav), range of the seasonal cycle (seas), and the standard deviation of the remainder component (rem). The middle panel shows the statistical distribution of these properties, respectively. The right panel shows examples of the respective surrogate time series components.
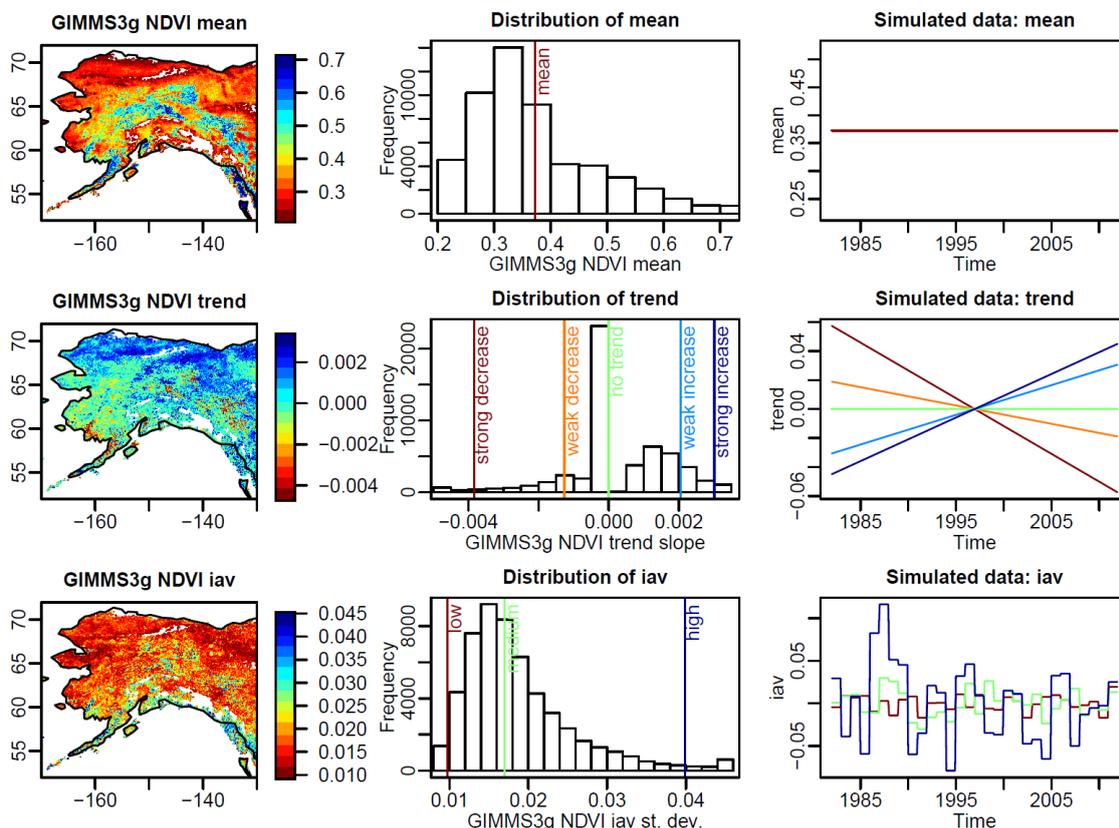
Figure 2. *Cont.*
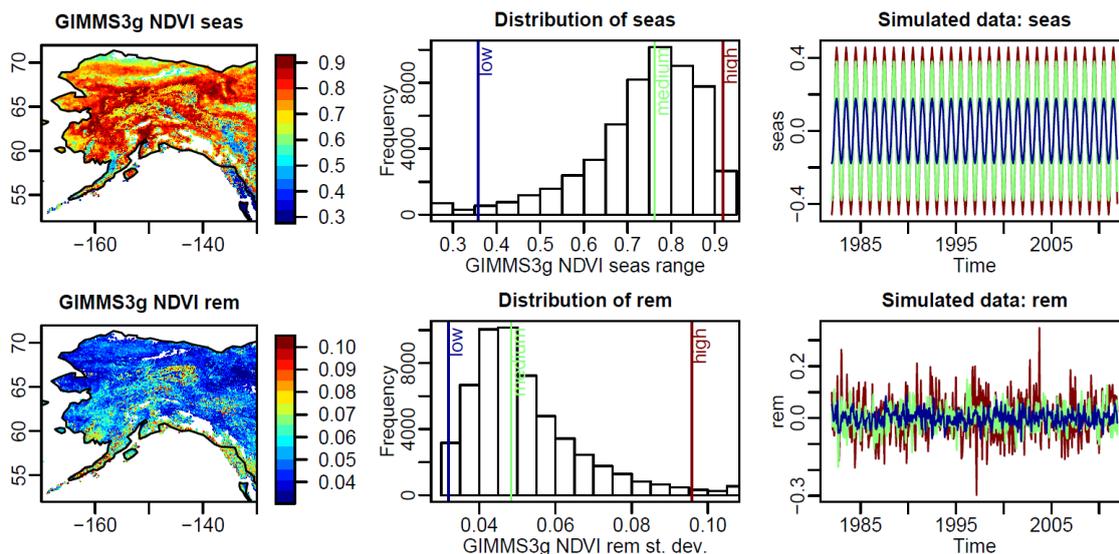
2.4.2. Surrogate Time Series and Factorial Experiment

Surrogate time series were simulated based on addition of different time series components that were estimated from observed time series properties:

$$y_t = m + T_t + I_t + S_t + R_t \tag{4}$$

where *m* is the mean, *T* is the trend component value, *I* the inter-annual variability component value, *S* the seasonal component and *R* the remainder component at time step *t*. As the estimated values of *T* and *I* have an annual temporal resolution, they were repeated for each time step t of the same year (forming a step function) to create the simulated monthly time series. The mean was taken from the mean of the observed distribution of average NDVI. We selected only one mean value for all surrogate time series because differences in mean are expressed by the intercept of the linear regression models and will not affect the trend estimate. For each of the other components, different levels were used to create surrogate time series:

(1) Trend: Time series with strong and weak positive, strong and weak negative and without a trend were created. Different magnitudes of trend slopes were derived from the 1% percentile of the observed distribution of trend slopes (strong decrease), 25% percentile (weak decrease), median (no trend), 75% percentile (weak increase) and 99% percentile (strong increase), respectively.

(2) Inter-annual variability: Time series with low, medium and high inter-annual variability were created based on normal-distributed random values with zero mean and a standard deviation according to the 1%, 50% and 99% percentiles of the observed distribution of the standard deviation of annual anomalies. Values outside the observed ranges of inter-annual variability were set to the minimum or maximum of the observed distribution, respectively.

(3) Seasonality: Seasonal cycles based on a harmonic model with low, medium, and high amplitudes were created according to the observed 1%, 50% and 99% percentiles of the distribution of seasonal ranges.

(4) Short-term variability: Different levels of short-term variability were created based on normal-distributed random values with zero mean and a standard deviation according to the 1%, 50% and 99% percentiles of the observed distribution of the standard deviation of remainder time series values.

Figure 3. Examples of simulated time series with different components of trend, IAV, seasonal and remainder referring to the simulated trend, inter-annual variability, seasonal and remainder time series components, respectively. The sum of these time series components gives the total simulated surrogate NDVI time series (upper panel). Left: time series with one breakpoint and gradual change (e.g., caused by gradual changes in environmental conditions), no trend in the first segment and decreasing trend in the second segment, medium inter-annual variability, medium seasonality and medium short-term variability. Right: Time series with one breakpoint and abrupt change (e.g., caused by a few years with exceptional favourable growing conditions), increasing trend in first segment and decreasing trend in second segment, high inter-annual variability, medium seasonality and low short-term variability.
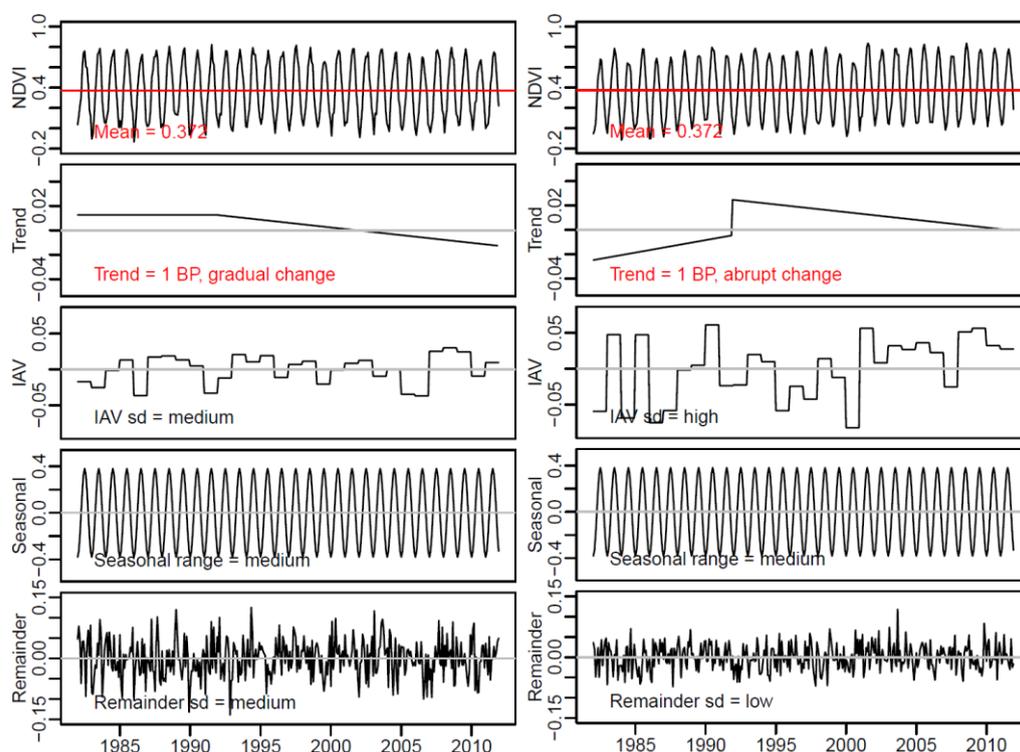
To introduce trend changes in these surrogate time series, trend components with one and two breakpoints as well as gradual or abrupt changes were created. In case of one breakpoint, the break was introduced 120 months after the beginning of the time series and in case of two breakpoints after 107 and 215 months, respectively. That means one time series can have one to three time series segments with a length of 360 months (30 years) in case of no breakpoint, 120 and 241 months in case of one breakpoint, respectively, and 107, 107 and 146 months in case of two breakpoints, respectively. The type of trend change was considered as an additional factor, whereby gradual change and abrupt changes were distinguished. A gradual change is a change between two trend segments, for which the last value of the trend component in the first segment equals the first value of the following trend

segment (Figure 3). In case of abrupt changes, trend components are disconnected between two segments (Figure 3). In summary, the following factors with different levels were considered in the experiment:

(1) Type of trend and number of breakpoints/segments (maximum 2 breakpoints = maximum 3 segments per time series with positive, negative or no trend = 27 possibilities),

(2) Trend magnitude (weak, strong),

(3) Inter-annual variability (low, medium, high),

(4) Short-term variability (low, medium, high),

(5) Type of trend change (gradual, abrupt) and

(6) Range of seasonal cycle (low, medium, high).

For each combination of these factors one surrogate time series was created. Because some combination are physically not possible (e.g., abrupt or gradual change but 0 breakpoints), in total 1377 surrogate time series were created.

2.5. Evaluation of Method Performances

2.5.1. Evaluation of Breakpoints

To evaluate the performance of the methods regarding the estimated breakpoints, the difference in number and timing of breakpoints were compared. The difference between the estimated number of breakpoints and the number of real breakpoints was calculated. The number of real breakpoints is the amount of breakpoints that was used to simulate the surrogate time series.

The timing of an estimated breakpoint was compared against the timing of the real breakpoint. For each estimated breakpoint the nearest real breakpoint was selected and the absolute temporal difference (in months) between them was calculated. If the difference is larger than five years the real breakpoint was set as undetected.

2.5.2. Evaluation of Trend Slopes and Significances

In order to evaluate the direction and significance of an estimated trend, estimated trends were compared with the real trend in a trend segment of the simulated time series. The slope and p-value of the real trend was calculated for each method based on the known real breakpoints and time series segments. To compare direction and significance of trends, trend slopes and p-values of real and estimated trends in a time series segment were classified in six trend classes:

N3: significant negative trend (slope < 0 and $p \leq 0.05$)

N2: non-significant negative trend (slope < 0 and $0.05 < p \leq 0.1$)

N1: no trend with negative tendency (slope < 0 and $p > 0.1$)

P1: no trend with positive tendency (slope > 0 and $p > 0.1$)

P2: non-significant positive trend (slope > 0 and $0.05 < p \leq 0.1$)

P3: significant positive trend (slope > 0 and $p \leq 0.05$).

Confusion matrices of estimated and prior trend classes were computed for each method in order to evaluate the accuracy of the methods for trend estimation. Confusion matrices (alternatively called

contingency table or error matrix) are standard tools to compare errors between two classifications [46]. Congalton [47] suggested to normalize confusion matrices in order to eliminate effects of different sample sizes per class and to make confusion matrices between different classifications, *i.e.*, different trend methods, comparable. Iterative Proportional Fitting Procedure (IPFP) [48] was used to normalize confusion matrices to row and column (marginal) totals of 100%. Based on normalized confusion matrices, the total normalized accuracy and the Kappa coefficient were calculated [46,47] to quantify the performance of methods for trend estimation. The total accuracy ranges between 0% (worst accuracy) and 100% (complete agreement of the two classifications) and the Kappa coefficient ranges between 0 (worst agreement) and 1 (complete agreement).

2.5.3. Evaluation of the Overall Performance for Trend and Breakpoint Estimation

The overall performance of a method was quantified by comparing the estimated with the real trend component (Equation (4)). For this purpose, the root mean square error (RMSE) between the estimated and the real trend component T was calculated for the total length n of the simulated time series:

$$RMSE = \sqrt{\frac{\sum_{t=1}^{n}(T_{est} - T_{real})^2}{n}} \tag{5}$$

This formulation involves both an effect of the estimated trend and the estimated breakpoints. An analysis-of-variance (ANOVA) was calculated for the RMSE in order to identify the factors that explain the highest fraction of the RMSE. Trend magnitude, inter-annual variability, seasonality, short-term variability, type of trend change, number of real breakpoints and method as well as their second-order interactions were used as explanatory variables in the ANOVA.

2.6. Application to Real Time Series of Alaska: Ensemble of NDVI Trends

All trend and breakpoint methods were applied to real NDVI time series of Alaska and parts of Yukon to assess differences between methods based on real data. The application of the four methods (AAT, STM, MAC, SSA) allows creating a multi-method ensemble of NDVI trend estimates. As many NDVI observations in northern regions are affected from snow, clouds or other distortions, the use of such poor quality observations causes additional uncertainties in NDVI trend estimates. To account for the effect of snow-affected observation, all methods were applied on the NDVI time series with all observations ("all") and on the NDVI time series excluding snow-affected values ("ex") (see Section 2.1). Additionally, method AAT was applied on the annual peak NDVI (defined as the annual quantile 0.9) to analyse trends (called AAT-peak) because vegetation growth in high-latitude ecosystems is usually limited to the peak production period. This setup of trend methods on the real dataset resulted in nine trend and breakpoint estimates for Alaska (AAT-all, AAT-ex, AAT-peak, STM-all, STM-ex, MAC-all, MAC-ex, SSA-all, SSA-ex). From all the nine trend estimates, the ensemble mean and standard deviation of the number of detected breakpoints, of the duration of greening and browning trends and of the trend slope were calculated. The ensemble mean NDVI slope α was calculated from the weighted mean slope α_m of a method m, weighted by the length l of the corresponding time series segment *seg* and the p-value p of the trend in the segment expressed as significance:

$$\alpha = \frac{1}{nmethod} \sum_{m=1}^{m=9} \alpha_m \qquad (6)$$

$$\alpha_m = \frac{\displaystyle\sum_{seg=1}^{seg=nseg} \alpha_{m,seg} \times (l_{seg} \times (1 - p_{seg}))}{\displaystyle\sum_{seg=1}^{seg=nseg} (l_{seg} \times (1 - p_{seg}))} \qquad (7)$$

Time series segments with a length *l* shorter than eight years were excluded from this analysis and remaining segment lengths (between eight and 30 years) were scaled to 0-1 before using them as weights in Equation (7). The uncertainty of NDVI trend slopes was calculated as the standard deviation of mean slopes α_m.

Additional datasets were used to assess the plausibility of detected breakpoints. Fire perimeter observations from the Alaskan Large Fire Database [49,50] were compared against the spatial distribution and timing of detected breakpoints. Gridded precipitation time series from the GPCC dataset (Global Precipitation Climatology Center) [51] and temperature time series from the CRU dataset [52] were used to compare breakpoints with annual temperature and precipitation anomalies (baseline 1982–2009). Additionally, photos taken in 2008 at burnt areas of the year 2004 were compared with detected breakpoints in NDVI time series to visually inspect the post-fire vegetation status in NDVI time series with breakpoints.

3. Results

3.1. Observed and Simulated Properties of NDVI Time Series

To simulate surrogate time series, statistical distributions of NDVI mean, trend, seasonality, inter-annual and short-term variability were computed from observed NDVI time series in Alaska (Figure 2). Mean NDVI ranged between 0.2 and 0.87 with an average value of *ca*. 0.37. The highest NDVI means occurred in the central Alaskan boreal forest and the lowest values in the northern Tundra regions (Figure 2). The mean of 0.37 was used as the mean value in all simulated time series. NDVI trend slopes ranged from −0.0047 to 0.0034 ΔNDVI/year, with the lowest values in some boreal forest regions and the highest values in the northern Tundra regions. NDVI trend slopes of −0.0038 (strong decrease), −0.0013 (weak decrease), 0 (no trend), 0.002 (weak increase) and 0.003 ΔNDVI/year (strong increase) were used to create trend components for the surrogate time series. The standard deviation of the annual averaged NDVI values was used as a measure for inter-annual variability. It ranged between 0.009 in some regions and 0.045 in south-western Alaska. The 0.01, 0.5, and 0.99 quantiles of 0.0097, 0.016 and 0.039 were used to create surrogate NDVI time series with low, medium and high inter-annual variability, respectively (Figure 2). The range of the mean seasonal cycle ranged from 0.27 in southern coastal regions of Alaska to 0.93 in some northern Tundra regions. Seasonal ranges of 0.34 (low), 0.76 (medium) and 0.91 (high) were used to create surrogate time series. The standard deviation of the remainder time series component was used a measure of short-term variability and ranges between 0.03 and 0.1. Values of 0.031 (low), 0.048 (medium) and 0.096 (high) were used as the standard-deviation of normal distributed random number to create

surrogate time series components of short-term variability (Figure 2). Because of the fact, that the percentiles 1% and 99% were chosen as the low and high levels for inter-annual variability, seasonality and short-term variability, simulated time series can be strongly dominated by seasonality (e.g., in case of high seasonal range but low inter-annual and short-term variability) or can show almost random behaviour (e.g., in case of low seasonality but high inter-annual and short-term variability). Consequently, the simulated NDVI time series covered not only a wide and extreme range of observed time series properties of the study region but contain time series properties that might occur under different environmental conditions.

Figure 4. Frequencies of differences between real and estimated number of breakpoints for the different methods from all experiments (blue indicates underestimation, red overestimation of the number of real breakpoints). (**a**) Performance of the methods in all experiments. (**b**) Grouped by trend magnitude. (**c**) Grouped by inter-annual variability. (**d**) Grouped by short-term variability. (**e**) Grouped by the real number of breakpoints. (**f**) Grouped by the type of trend change. (**g**) Grouped by the range of the seasonal cycle.
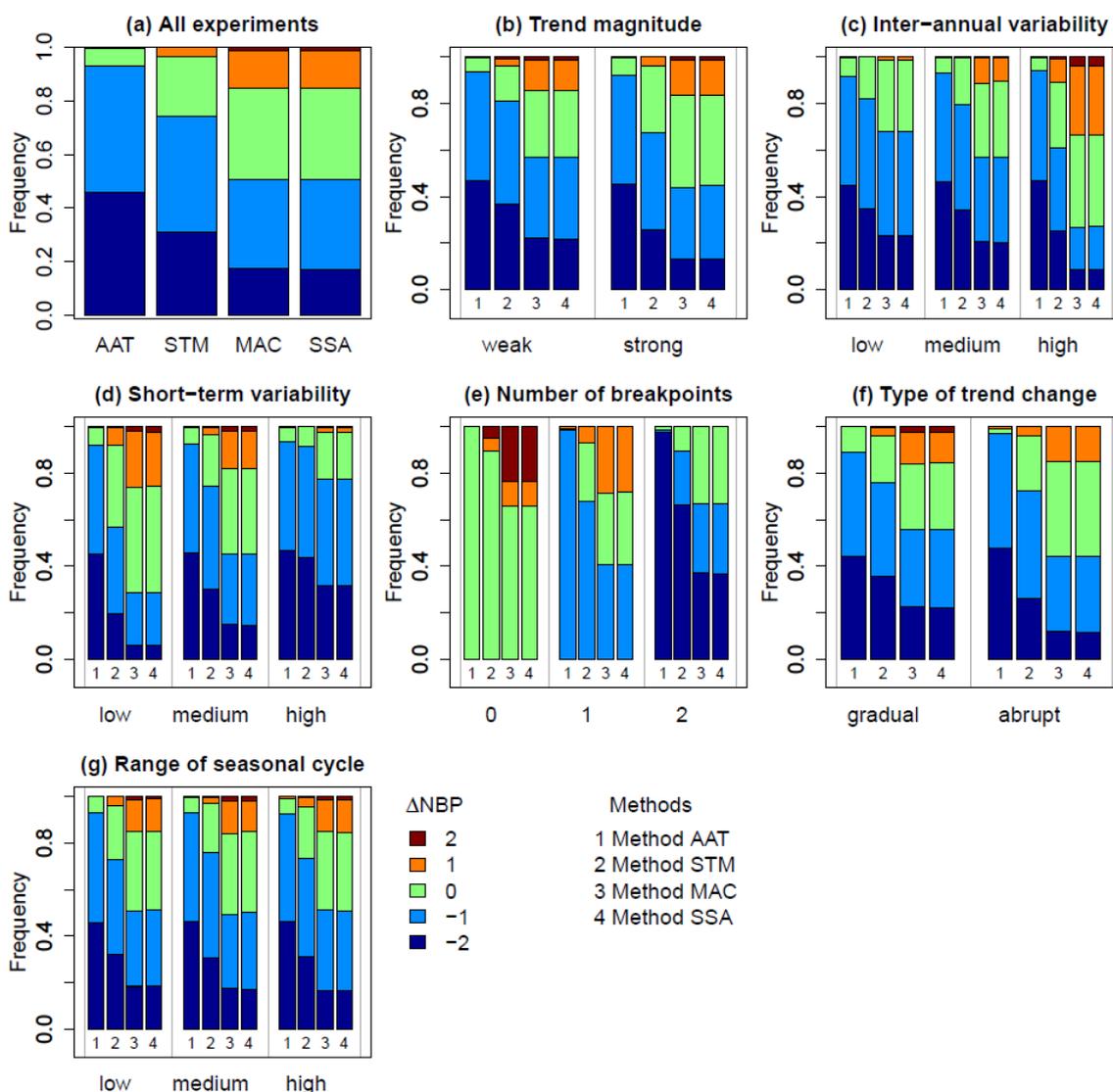
Figure 5. Distribution of the temporal absolute difference between real and estimated breakpoints. (**a**) Performance of the methods in all experiments. (**b**) Grouped by trend magnitude. (**c**) Grouped by inter-annual variability. (**d**) Grouped by short-term variability. (**e**) Grouped by the real number of breakpoints. (**f**) Grouped by the type of trend change. (**g**) Grouped by the range of the seasonal cycle. + denotes the mean of the distribution. The difference is only based on detected breakpoints. As method AAT detected fewer breakpoints, it has a much smaller sample size (n = 42) than the other methods (STM n = 732, MAC n = 1,368, SSA n = 1,380).
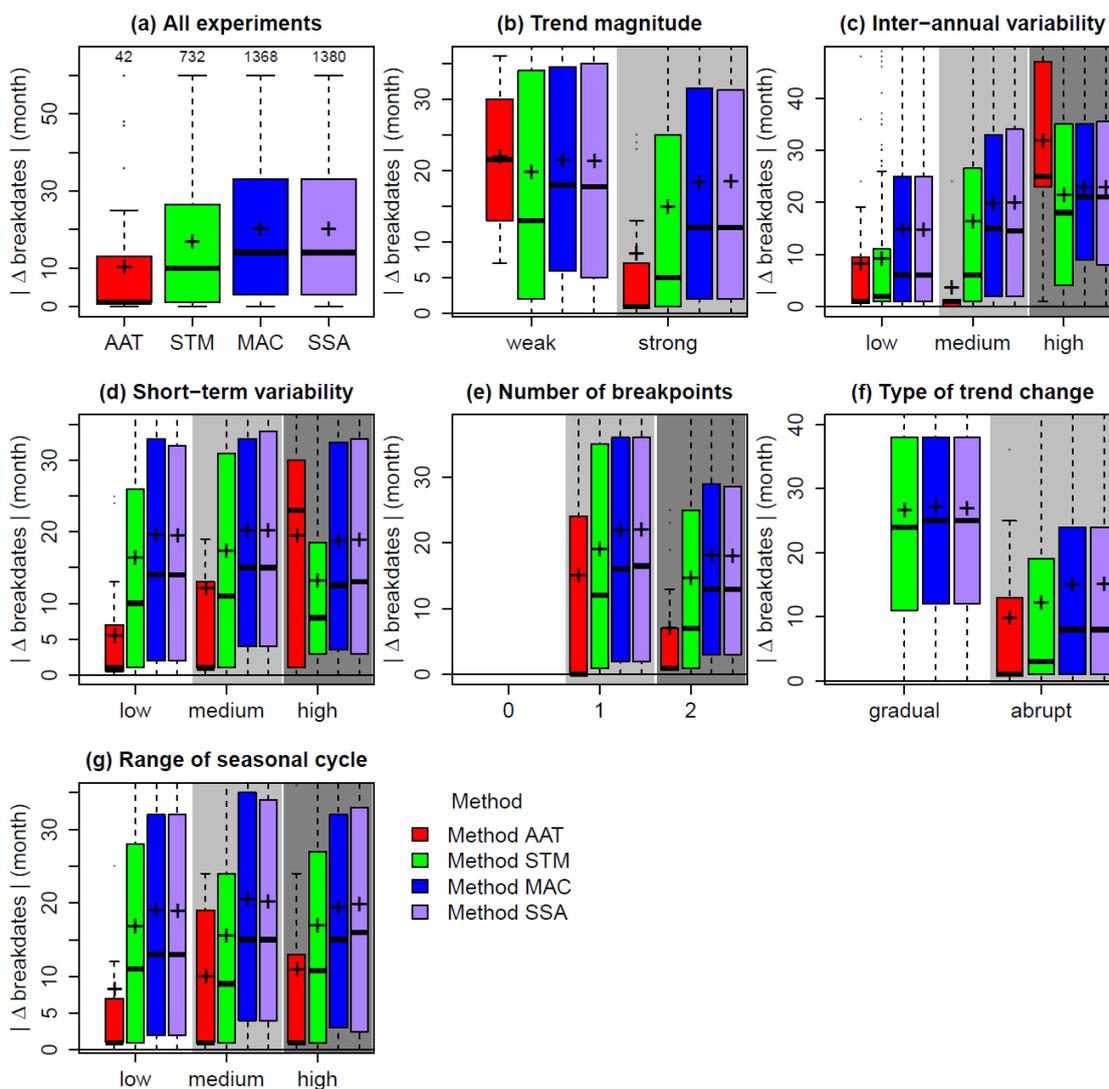
3.2. Evaluation of Estimated Breakpoints

To evaluate the performance of the methods to detect breakpoints in trends, the real and estimated breakpoints were compared. For this purpose, the difference between the estimated and real number of breakpoints was calculated and analysed grouped by methods and factors (Figure 4). All methods underestimated the number of breakpoints, *i.e.*, the number of false positive detected breakpoints was small (0.36% for method AAT, 3.8% for STM, 15.5% for MAC and 15.3% for SSA). Method AAT did not detect any breakpoint if there was no breakpoint whereas methods STM, MAC and SSA detected in up to 35% of all cases one or two breakpoints if there was no breakpoint in the surrogate

time series (Figure 4(e)). The range of the seasonal cycle had no effect on the estimation of the number of breakpoints (Figure 4(g)). The performance of all methods to estimate the correct number of breakpoints depended also on inter-annual and short-term variability. The performance of method AAT did not change with increasing inter-annual and short-term variability while methods STM, MAC and SSA had an increasing number of false positive detected breakpoints with increasing inter-annual variability and a slightly decreasing number of false positive detected breakpoints with increasing short-term variability (Figure 4(c,d)). All methods tended to perform better in case of abrupt trend changes than in case of gradual trend changes (Figure 4(f)). In short, the difference between estimated and real number of breakpoints depended mostly on the number of real breakpoints as well as short-term and inter-annual variability.

Furthermore, the temporal difference between a detected and the closest real breakpoint was calculated, to evaluate the performance of methods regarding the timing of breakpoints (Figure 5). The difference in timing was not calculated if a method did not detect a breakpoint although there were real breakpoints. In average, method AAT performed better (mean absolute difference in breakdates = 9.2 months) than method STM (16 months) and methods MAC and SSA (both 19 months) (Figure 5(a)). The error of the timing of breakpoints increased with increasing inter-annual variability (Figure 5(c)). Increasing short-term variability resulted only for method AAT in an increasing timing error (Figure 5(d)). All methods had a better timing of breakpoints in case of abrupt changes (Figure 5(f)). The difference in breakdates was lower in case of multiple breakpoints (Figure 5(e)). The range of the seasonal cycle has no effects on the timing of breakpoints (Figure 5(f)). In summary, the correct timing of breakpoints depended mostly on inter-annual variability and the type of trend change.

3.3. Evaluation of Estimated Trends

In order to evaluate if methods detect the correct trends, the direction and significance of estimated trends were compared against the direction and significance of real trends in a time series segment (Figure 6). The estimated slopes from method AAT were higher correlated with the real slopes (r = 0.74) than the estimated slopes from other methods (method STM r = 0.7, MAC r = 0.62 and SSA r = 0.61). The agreement was lower if only one of the real or estimated trend was significant (r = 0.24, AAT) or if neither the real nor the estimated trends were significant (r = 0.22, AAT). Nevertheless, estimated trend slopes from all methods were highly correlated (up to r = 0.98 between MAC and SSA, scatterplots not shown). Generally, slope estimates from method AAT were less correlated with the other methods while especially methods MAC and SSA were highly correlated. Hence, the annual aggregation approach in method AAT resulted in the most unique slope estimates compared to the other methods.

Confusion matrices were calculated to evaluate the accuracy of trend classes based on the direction and significance of a trend. Method AAT had usually higher class accuracies as well as a higher total accuracy (37.6%) and Kappa coefficient (K = 0.25) than other methods (Table 1). While method AAT correctly detected 55.24% (47.34%) of significant negative (positive) trends, methods STM, MAC and SSA correctly detected only 47.9% (47.4%), 48.1% (45.15%) and 48.1% (45.9%) of significant negative (positive) trends, respectively. Nevertheless, method AAT detected 3.1% (3.6%), method STM 5.4% (4.6%), method MAC 5.5% (6.1%) and method SSA 5.9% (6.2%) of significant negative

(positive) trends as significant positive (negative), *i.e.*, opposite, trends (Table 1). Thus, trend slopes from method AAT were higher correlated with real trend slopes and method AAT detected fewer false positive trends and more correct positive trends than methods STM, MAC and SSA.

Figure 6. Comparison of real and estimated slopes from different methods, all time series segments and all experiments. Slopes are coloured blue if both real and estimated slopes are not significant, green if only the real or estimated slope was significant and red if both slopes were significant (0.95 significance level).
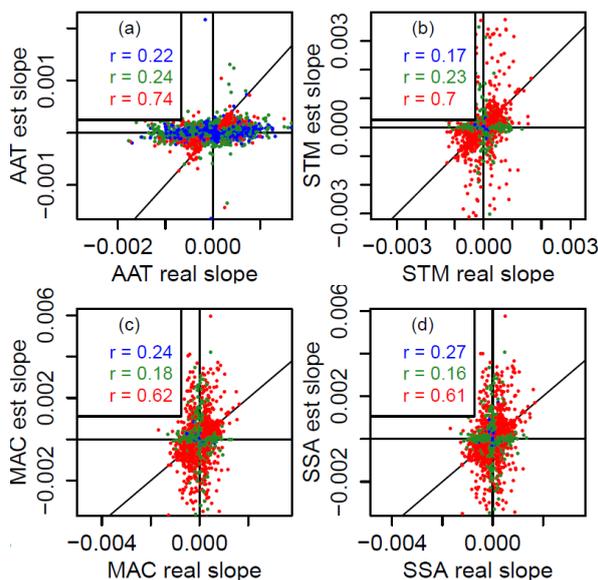
Table 1. Normalized confusion matrices of estimated and real trend classes for each method. N3: significant negative trend, N2: non-significant negative trend, N1: no trend with negative tendency, P1: no trend with positive tendency, P2: non-significant positive trend, P3: significant positive trend. ToAcc: total normalized accuracy, Kappa: Kappa coefficient.

Method AAT	Real.N3	Real.N2	Real.N1	Real.P1	Real.P2	Real.P3	Sum
Est.N3	55.24	11.18	15.57	8.44	5.95	3.62	100.00
Est.N2	12.48	43.27	26.76	11.11	0.00	6.38	100.00
Est.N1	13.27	14.55	24.55	17.46	17.74	12.42	100.00
Est.P1	10.37	10.57	15.29	24.43	24.85	14.49	100.00
Est.P2	5.54	13.72	11.98	22.01	31.01	15.74	100.00
Est.P3	3.09	6.70	5.85	16.56	20.45	47.34	100.00
Sum	100.00	100.00	100.00	100.00	100.00	100.00	600.00
ToAcc = 37.64, Kappa = 0.25							
Method STM	**Real.N3**	**Real.N2**	**Real.N1**	**Real.P1**	**Real.P2**	**Real.P3**	**Sum**
Est.N3	47.90	20.68	13.18	7.58	6.07	4.59	100.00
Est.N2	20.58	32.21	14.54	11.05	15.14	6.48	100.00
Est.N1	14.60	18.92	22.68	14.79	18.47	10.54	100.00
Est.P1	10.37	11.09	20.22	25.48	17.20	15.65	100.00
Est.P2	1.15	8.16	19.91	25.21	30.22	15.35	100.00
Est.P3	5.41	8.94	9.47	15.89	12.89	47.39	100.00
Sum	100.00	100.00	100.00	100.00	100.00	100.00	600.00
ToAcc = 34.31, Kappa = 0.21							

Table 1. *Cont.*

Method MAC	Real.N3	Real.N2	Real.N1	Real.P1	Real.P2	Real.P3	Sum
Est.N3	48.08	22.05	11.81	7.56	4.37	6.13	100.00
Est.N2	13.24	29.18	14.06	10.84	26.98	5.69	100.00
Est.N1	15.15	19.12	27.35	14.75	10.87	12.76	100.00
Est.P1	10.91	16.97	15.94	25.79	18.33	12.06	100.00
Est.P2	7.14	4.45	22.71	26.33	21.16	18.22	100.00
Est.P3	5.48	8.23	8.13	14.73	18.29	45.15	100.00
Sum	100.00	100.00	100.00	100.00	100.00	100.00	600.00
ToAcc = 32.79, Kappa = 0.19							
Method SSA	**Real.N3**	**Real.N2**	**Real.N1**	**Real.P1**	**Real.P2**	**Real.P3**	**Sum**
Est.N3	48.08	17.79	14.94	6.76	6.21	6.22	100.00
Est.N2	9.07	37.14	19.66	11.88	18.57	3.69	100.00
Est.N1	15.20	20.16	24.72	14.54	14.08	11.31	100.00
Est.P1	13.80	9.59	13.99	24.52	24.76	13.34	100.00
Est.P2	7.88	6.19	18.57	25.12	22.70	19.55	100.00
Est.P3	5.98	9.13	8.12	17.17	13.69	45.90	100.00
Sum	100.00	100.00	100.00	100.00	100.00	100.00	600.00
ToAcc = 33.84, Kappa = 0.21							

3.4. Effects on the Overall Performance of the Methods

To quantify the overall error of breakpoint and trend detection, the root mean square error (RMSE) between the estimated and real trend component was computed. The distribution of RMSE for the different experimental factors and methods is shown in Figure 7. Overall, method AAT and STM performed better than methods MAC and SSA. The error increased with increasing inter-annual variability for all methods (Figure 7(c)). For all methods the error slightly increased with increasing short-term variability (Figure 7(d)). All methods had higher errors in time series with breakpoints than in time series without breakpoints (Figure 7(e)). The range of the seasonal cycle did not affect the performance of the different methods (Figure 7(f)). The error was larger in time series with strong trends and abrupt changes than in time series with weak trends and gradual changes (Figure 7(b,f)). The higher error under strong trends was a result of the worse timing of breakpoints under these conditions (Figure 5(b)). To quantify the relative effects of the correct timing of breakpoints and inter-annual variability on the error in trend estimation, an analysis-of-variance was computed in the next step.

The contribution of different factors to the error between estimated and real trend component was analysed by an analysis-of-variance (Table 2). All considered factors explained a significant proportion of the error. Inter-annual variability explained the largest part of the error distribution (30.7%). Additionally, the type of trend change, the interaction between inter-annual variability and applied method and the trend magnitude explained large parts of the error. This is illustrated by the similar error of all methods under low and medium inter-annual variability while for methods MAC and SSA the error increased strongly under high inter-annual variability (Figure 7(c)). The seasonal range had in general only a small contribution to the overall error (0.002%). In summary, the inter-annual variability of the time series was the most important factor for the error of breakpoint and trend estimation.

Figure 7. Distribution of the root mean square error (RMSE) between real and estimated trend component for the different methods from all experiments. (**a**) Performance of the methods in all experiments. (**b**) Grouped by trend magnitude. (**c**) Grouped by inter-annual variability. (**d**) Grouped by short-term variability. (**e**) Grouped by the real number of breakpoints. (**f**) Grouped by the type of trend change. (**g**) Grouped by the range of the seasonal cycle.
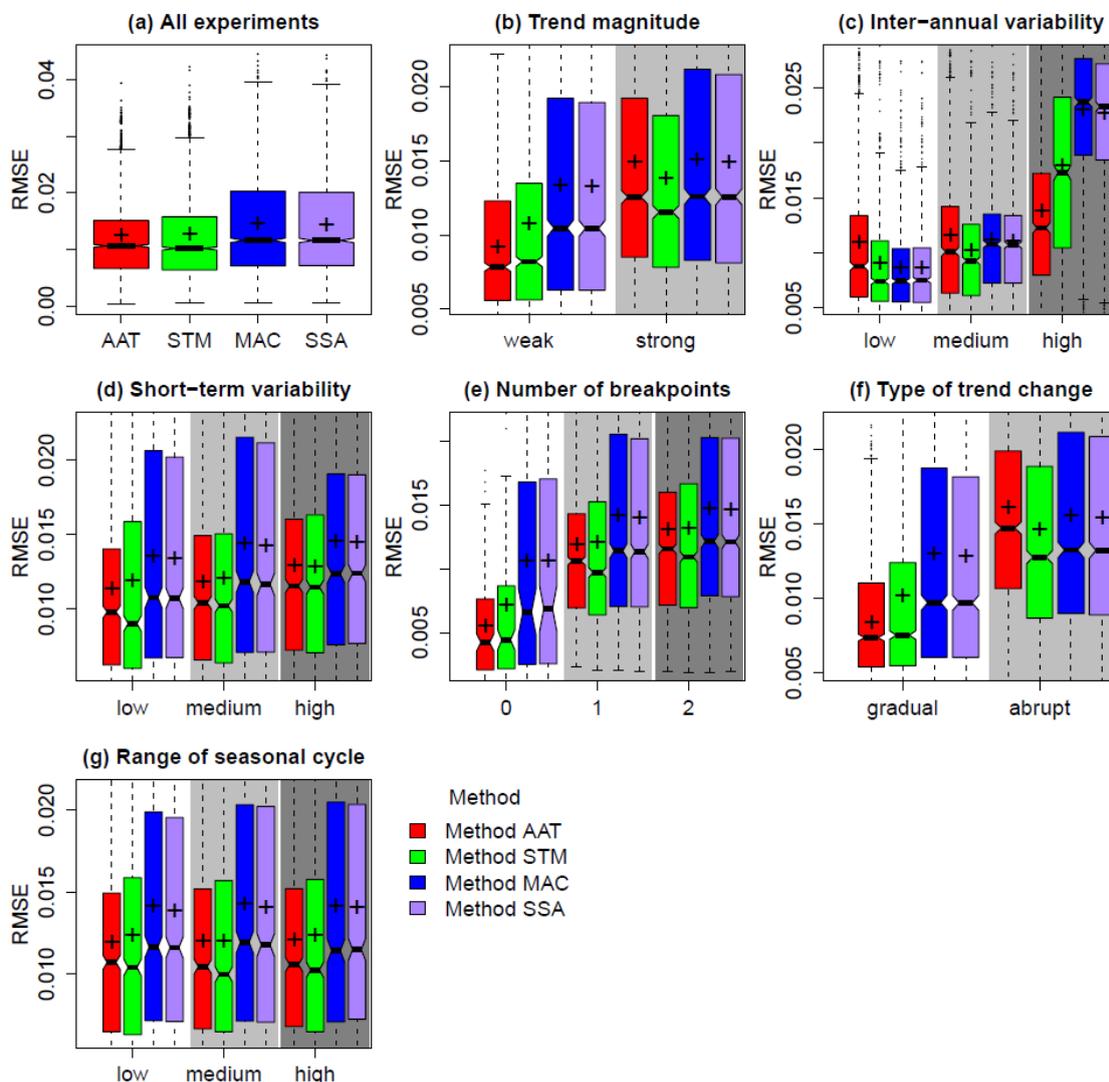

Table 2. Analysis of variance table for the RMSE between real trend and estimated trend. IAV and STV denote inter-annual and short-term variability, respectively.

Factor	Df	Sum Sq	Mean Sq	F value	P (>F)	Sum Sq/Total Sq (%)
IAV	2	0.2136	0.1068	4096.7	<2.2e-16	30.73
Type of change	1	0.0511	0.0511	1959.3	<2.2e-16	7.35
IAV * Method	6	0.0469	0.0078	299.8	<2.2e-16	6.75
Trend magnitude	1	0.0252	0.0252	966.5	<2.2e-16	3.62
IAV * STV	4	0.0153	0.0038	146.8	<2.2e-16	2.20
Type of change * Method	3	0.0121	0.0040	154.6	<2.2e-16	1.74
Method	3	0.0108	0.0036	137.8	<2.2e-16	1.55
Trend magnitude * Method	3	0.0073	0.0024	93.9	<2.2e-16	1.06

Table 2. *Cont.*

Factor	Df	Sum Sq	Mean Sq	F value	P (>F)	Sum Sq/Total Sq (%)
Number of breakpoints	2	0.0072	0.0036	137.6	<2.2e-16	1.03
STV * Type of change	2	0.0061	0.0030	116.5	<2.2e-16	0.87
STV	2	0.0024	0.0012	46.1	<2.2e-16	0.35
Trend magnitude * Type of change	1	0.0022	0.0022	86.3	<2.2e-16	0.32
Trend magnitude * Number of breakpoints	2	0.0022	0.0011	41.7	<2.2e-16	0.31
Type of change * Number of breakpoints	1	0.0022	0.0022	83.4	<2.2e-16	0.31
Trend magnitude * STV	2	0.0022	0.0011	41.3	<2.2e-16	0.31
IAV * Type of change	2	0.0005	0.0003	10.4	2.962E-05	0.08
Seasonality * Number of breakpoints	4	0.0005	0.0001	5.0	4.945E-04	0.08
STV * Number of breakpoints	4	0.0005	0.0001	4.5	1.238E-03	0.07
Trend magnitude * IAV	2	0.0004	0.0002	7.4	0.001	0.06
STV * Method	6	0.0003	0.0001	2.1	4.948E-02	0.05
Trend magnitude * Seasonality	2	0.0002	0.0001	4.6	0.010	0.03
Seasonality * Type of change	2	0.0002	0.0001	4.2	1.526E-02	0.03
IAV * Number of breakpoints	4	0.0002	0.0001	2.0	8.910E-02	0.03
Seasonality	2	0.0000	0.0000	0.2	0.799	0.00
Residuals	10,952	0.2855	0.0000			41.07

3.5. Multi-Method Ensemble of Breakpoint and Trend Estimates in Alaska

To compare breakpoint and trend estimates of the different methods under real conditions, all methods were applied to GIMMS NDVI3g time series of Alaska and the number and timing of breakpoints as well as the duration of significant greening and browning trends were calculated (Figure 8). Breakpoints were detected by multiple methods in northern Tundra regions, in some parts of eastern boreal Alaska and in south-western Alaska (Figure 8(a)). Method MAC detected in 40 % of all land grid cells one or two breakpoints if snow-affected values were excluded from the analysis (MAC-ex) (Figure 8(c)). All other methods detected less breakpoints. Especially the seasonal methods detected more breakpoints if snow-affected observation were excluded from the analysis. The lowest number of breakpoints was detected by method AAT applied to the annual peak NDVI (AAT-peak). Despite method SSA, all other methods detected breakpoints around the year 2000 with following browning trends in the northern Tundra. The detection of breakpoints controls the multi-method average duration of significant greening and browning trends. The duration of greening ranges between 20 and up to 30 years with a standard deviation between two to six years in most tundra regions. In some north-eastern parts of the study region significant greening trends last between 20 and 30 years as well but with higher uncertainties of six to 10 years (Figure 8(d,e)). Browning NDVI trends between 20 and 30 years occurred in some boreal regions of central Alaska and in south-western Alaska usually with small uncertainties of up to four years (Figure 8(g,h)). The multi-method average NDVI trend slope demonstrates that greening NDVI trends were more spatially and temporally prevalent and of higher magnitude than browning trends (Figure 8(j–l)). Greening occurred mostly in Tundra regions while browning occurred only spotted in the boreal forest. Nevertheless, greening trends occurred in some parts of the central Alaskan boreal forests too but were associated with higher uncertainties because some methods detected breakpoints with greening NDVI trends while other

methods detected no significant greening trends in these regions. All methods detected stronger greening and browning trends if snow-affected values were excluded from the analysis or only peak NDVI was used and weaker trends if all values were used (Figure 8(l)). In short, methods agreed in the estimation of the major 30 year greening and browning trends, while they had larger differences in regions where breakpoints were detected by some methods and thus trend changes are likely. Nevertheless, the treatment of snow-affected NDVI values and thus the inherent inter-annual variability of the NDVI time series caused larger differences in NDVI trend estimates than the applied trend estimation method.

Figure 8. Ensemble of breakpoint and trend estimates from all methods in Alaska. AAT, STM, MAC and SSA are the four applied trend methods. 'all' indicates that all NDVI values were used (*i.e.*, including interpolated and snow-affected observations). 'ex' snow-affected values were excluded from trend analysis. 'peak' trend was computed only on annual peak NDVI. (**a**) Mean number of detected breakpoints from all methods. (**b**) Uncertainty of the number of detected breakpoints (standard deviation of number of breakpoints from all methods). (**c**) Number of detected breakpoints grouped by method. (**d**) mean duration of greening trends (years) with associated uncertainties (**e**) and distribution of greening duration per each method (**f**). (**g**) mean duration of browning trends (years) with associated uncertainties (**h**) and distribution of browning duration per each method (**i**). (**j**) Multi-method mean trend slope (ΔNDVI/year) with associated uncertainties (**k**) and distributions per each method (**l**).
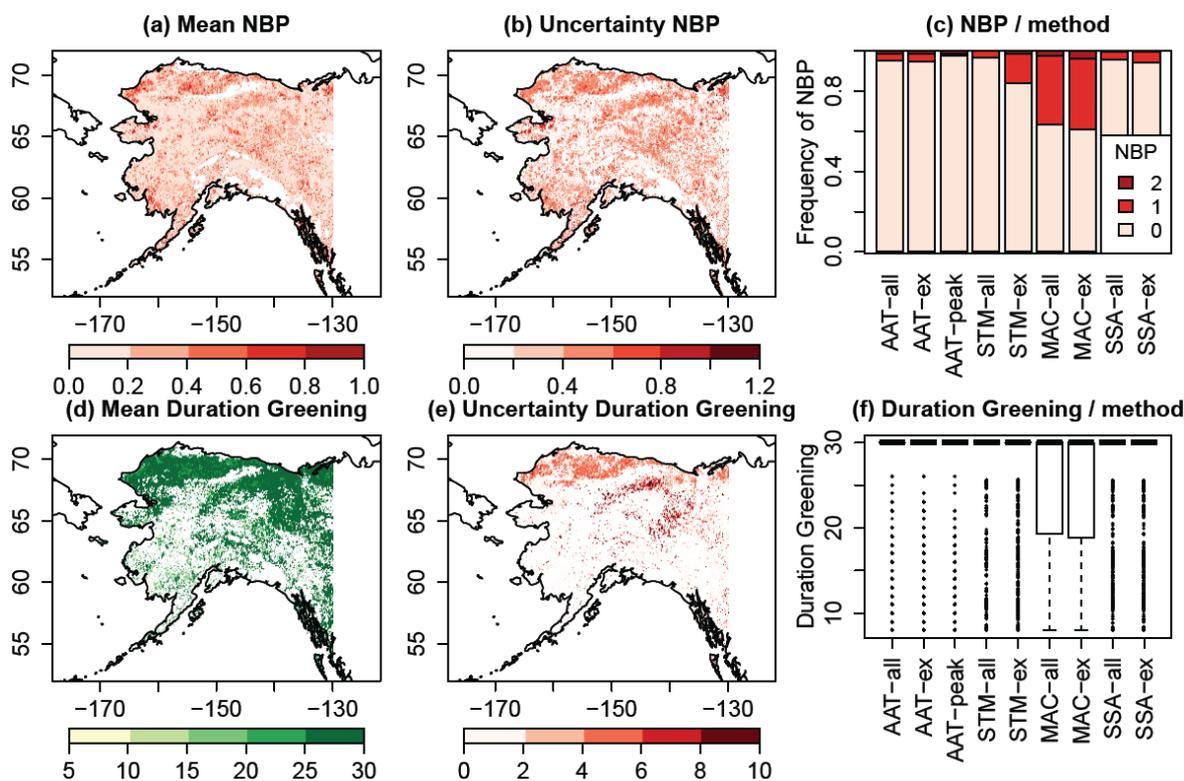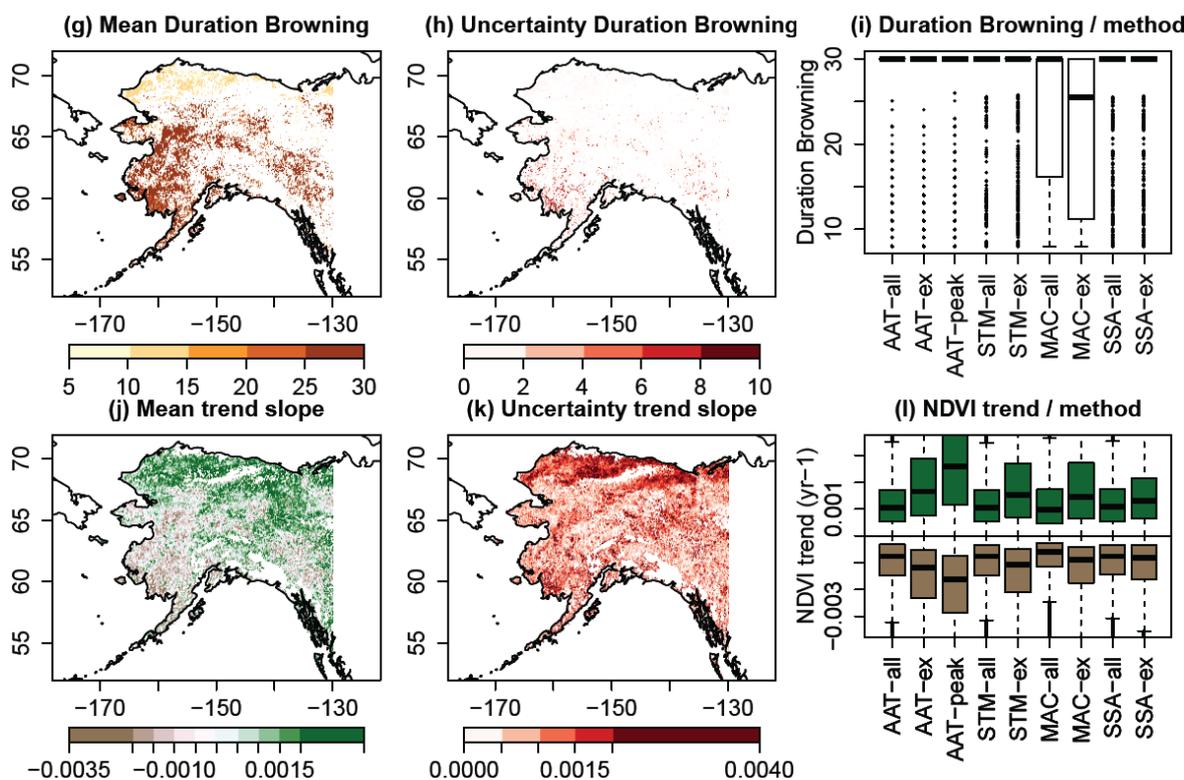
4. Discussion

4.1. Effect of Temporal Resolution on Method Performance

The correct timing of breakpoints does not depend on the temporal resolution of the time series but on how full-temporal resolution methods deal with seasonality. Two types of temporal resolution of NDVI time series were explored in this study: Method AAT makes use of a low temporal resolution based on annual aggregated data and methods STM, MAC and SSA were using a monthly temporal resolution of the time series. A more accurate detection of breakpoints was expected if a method uses the full temporal resolution than annual aggregated data. Nevertheless, the annual aggregation approach (method AAT) compared well in the timing of breakpoints like one full-temporal resolution method (method STM) although it largely under-estimated the number of breakpoints. Although the other full-temporal resolution methods (methods MAC and SSA) detected more often the right number of breakpoints, they had larger errors in the timing of these breakpoints (Figure 5). These results suggest that the estimation of the timing of breakpoints is highly sensitive on how the temporal decomposition methods account for seasonality and inter annual variability. Thus, the lower precision of the annual method limits the correct timing of within-year breakpoints, although it does not reduce the accuracy of this method in comparison to seasonal methods.

The temporal resolution of the time series affects the estimation of the trend significance. The major problem of using annual aggregated data rather than full-resolution time series is the reduction in the number of observations. This involves an underestimation of the significance of the trend [28,30]. However, this assumed limitation of the annual aggregation approach turned out to be an advantage as it decreases the risk of detecting false positive trends, *i.e.*, a significant positive trend in case of a

significant negative real trend and *vice versa*. Nevertheless, the application of all methods to real data resulted in similar trend slope estimates from different methods. Similar trend slopes with a low multi-method uncertainty were found in regions where methods usually did not detect breakpoints (Figure 8(a,j)). Also previous studies reported similarities in trend estimates from annual-aggregated and full-temporal resolution approaches as well [31]. Hence, the detection of significant trends depends rather on the estimated breakpoints and thus on time series length than on the temporal resolution of the time series that was used by the trend detection method.

4.2. Effect of Trend Changes on Method Performance

The performance of the four methods to estimate trends is lower in time series with breakpoints. All methods estimate trends with a lower error in case of simple time series with no breakpoints or in case of gradual than abrupt changes (Figure 7(e,f)). All methods present a prevalent tendency to underestimate the number of breakpoints. Method AAT underestimates the number of breakpoints in almost all cases, being the most conservative approach for estimating breakpoints, which ultimately resulted in the best overall performance for trend estimation (Figure 7(a)). On the other hand, methods MAC and SSA tended to detect breakpoints in time series without breakpoints (Figure 4(e)). In complex time series with one or multiple breakpoints or abrupt trend changes the error of the estimated underlying trend component is higher for all methods (Figure 7(e,f)). This higher error is caused by the estimated timing of the breakpoint. A larger difference between the estimated and the real breakpoint causes a higher difference between estimated and real trend slope in the time series segment before or after the breakpoint and thus a higher error in the estimated trend component. The seasonal-adjusted approaches (methods MAC and SSA) have a worse overall performance than the season-trend model approach (method STM) (Figure 7(a)). Although methods MAC and SSA detected in more cases the right number of breakpoints they had a significant number of false positive detected breakpoints (Figure 4(a)). The overestimation and/or the strong mismatch in the timing of the breakpoints cause a false detection of the underlying trend or even a detection of opposite trends, especially under high inter-annual variability (Figure 5(c)). Taken together, in order to detect the correct trends, the underestimation of breakpoints results in better trend estimates, and better estimates of the timing of major breakpoints are more important than the detection of the right number of breakpoints.

4.3. Effect of Inter-Annual Variability on Method Performance

The capability to estimate trends and breakpoints depends mostly on the robustness of a method against inter-annual variability (Figure 7, Table 2). Inter-annual variability increased the timing error of breakpoints (Figure 5(c)). The error of the estimated trend component increased for all methods with increasing inter-annual variability. Nevertheless, the error increase with increasing inter-annual variability is much stronger for methods STM, MAC and SSA than for method AAT (Figure 7(c)). Thus, the annual aggregation method is a relatively robust approach against inter-annual variability for estimating trends and trend changes. This effect is also evident from the application of the methods on real NDVI time series: The inter-annual variability of NDVI time series in Alaska is 1.6 times higher if snow-affected observations are excluded (figure not shown). This inter-annual variability caused a detection of a higher number of breakpoints if snow-affected observation were not used to compute

breakpoints in all methods (Figure 8(c)). The use or non-use of snow-affected NDVI observations for trend analysis and the associated inter-annual variability of the NDVI time series caused larger differences in breakpoint estimates and trend slopes than the choice of the trend method (Figure 8(l)).

However, we have to be cautionary in the assessment of the effect of inter-annual variability on the method performance: We assumed that the inter-annual variability as well as the short-term variability are independent, *i.e.*, temporally uncorrelated processes. Real world observations may depict long-range correlations that can be expected to induce year-to-year dependencies in the inter-annual variability. This would imply that the separation of trend and inter-annual variability is not straightforward. However, an investigation on effects of this kind is beyond the scope of our study but we assume that additional uncertainties would affect the timing of changes in trends in any methodological approach and attribution study.

4.4. Plausibility of Trend and Breakpoint Estimates in Alaska

Alaska is of special interest for the analysis of trend change detection methods because previous studies reported greening NDVI trends in tundra ecosystems of the Alaska North Slope as well as browning trends in the Alaska boreal forests [13,17,31,53]. On the other hand, it is not clear if these browning trends are monotonic trends as also different number of trend changes and a considerable amount of years with greening trends were detected [35]. No trend changes or browning trends have been reported in previous studies for the Alaskan tundra [35]. We detected breakpoints in the Alaskan tundra around the year 2000 with following browning trends across different methods and regardless if snow-affected NDVI observations or only peak NDVI were used for trend analysis or not (Figure 8(g–i)). That means that these breakpoints with following browning in the Alaskan tundra are not artefacts due to snow or cloud contaminations or other radiative surface changes outside the peak-growing season. This browning is observable due to the extended length of the GIMMS NDVI3g dataset until 2011 in comparison to previous dataset versions. It needs to be assessed if these greening-to-browning trend changes in the Alaskan tundra are due to regime shifts in climate-ecosystem interaction processes or due to inter-annual variations. Browning trends were detected by all methods in some parts of boreal Alaska. Previous studies reported negative NDVI trends especially in eastern boreal Alaska [21]. These browning trends are associated with wildfires [13], occur mainly in evergreen needle-leaf forests [17] and are related to temperature-induced drought stress and insect disturbances [21,54]. Browning trends in western boreal Alaska based on a previous version of the GIMMS NDVI dataset are more doubtful because such GIMMS NDVI had only a weak agreement with MODIS NDVI in this often cloud-affected region [21]. In contrast to previous results of [35], we detected fewer trend changes in boreal Alaska and more monotone browning trends over this 30 year period.

To evaluate the plausibility of detected breakpoints in NDVI time series of Alaska, the temporal dynamics and spatial patterns of breakpoints were compared against quality flags of the GIMMS NDVI3g dataset, temperature and precipitation anomalies as well as burnt areas from the Alaskan Large Fire Database (Figures 9 and 10). We cannot fully make sure that all detected breakpoints are only due to climate or environmental changes but might be caused by sensor changes in the GIMMS NDVI3g dataset. Sensor changes took place in 1985, 1988, 1994, 1995, 2000 and 2004 [19]. Although most breakpoints were detected in the year 2000 which had the highest number of flagged NDVI

values with reduced quality and where a sensor change took place (Figure 9), previous studies have shown that trends computed on GIMMS NDVI data are not affected by sensor changes [19,40]. Clearly, some breakpoints can be related to drought periods or fire events. Some larger fire events occurred in 1988, 1990, 1997, 2002 and 2004 and methods detected breakpoints in NDVI trends during or one year after these fire events (Figure 9(a,b)). Methods AAT, STM and MAC detected breakpoints in 1997 and 1998 in central and south-western Alaska that can be related to patterns of negative precipitation anomalies and fire activity. Alaskan fire activity is caused by droughts that are related to large-scale circulation patterns [55]. For example, El Nino causes drought and fire weather conditions in interior Alaska [55,56]. Thus it is probable, that detected breakpoints in 1997/1998 are a result of decreased NDVI signals because of drought effects after the 1997 El Nino event [57]. The largest total burnt area in Alaska occurred in 2004 under low precipitation and high temperature conditions (Figure 9(b,c)). The detection of breakpoints in NDVI time series in 2004 agreed with the spatial distribution of wildfires and *in situ* observations. Large conifer forests were burned in 2004 and were replaced by low broadleaved shrubs (dwarf birch and aspen) and grasses during post-fire succession which results in a structural change in NDVI time series (Figure 10). Especially methods AAT and MAC detected the highest number of breakpoints exactly at burn scar locations (AAT-ex n = 40, MAC n = 124) while methods STM and SSA found less breakpoints at burnt areas (STM n = 27 and SSA n = 29). Nevertheless, only a few fire-related breakpoints were detected by multiple methods and especially methods STM and MAC detected many breakpoints outside burnt areas. As method MAC had a high number of false positive detected breakpoints (Figure 4), some of these breakpoints might be false positives too. However, we cannot assess if they were caused by other environmental processes. In short, detected breakpoints can be related to environmental conditions like fire events or drought periods but need to be cross-checked against quality flags of the GIMMS NDVI3g dataset and additional environmental datasets to exclude false positive detected breakpoints.

Figure 9. Comparison of detected breakpoints with temporal fire and climate patterns in Alaska. (**a**) Time series of the total number of detected breakpoints per year for each method. (**b**) Total annual burnt area and annual anomaly of flagged GIMMS NDVI3g pixels with reduced quality. (**c**) Annual temperature and precipitation anomalies (baseline 1982–2009) averaged for Alaska.
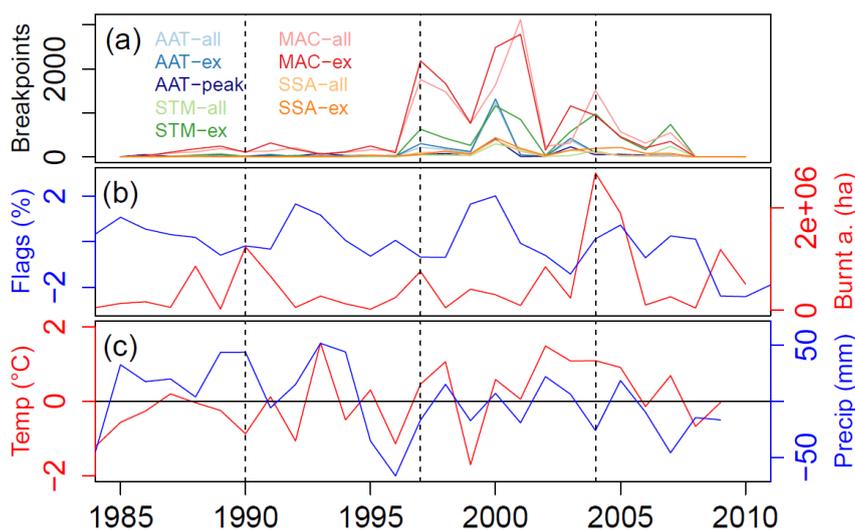
Figure 10. Comparison of detected breakpoints of the year 2004 from four different methods with 2004 burnt areas and *in situ* photos (taken in 2008). NBP is the total number of breakpoints that was detected in 2004 in this region. BPinBA denotes the percentage of breakpoints that was found inside a burnt area polygon. For methods AAT-ex and AAT-peak breakpoints for both 2003 and 2004 are shown to compensate for the lower breakpoint timing precision of these methods. All breakpoints that were found by AAT-peak were found at least also by one other method.
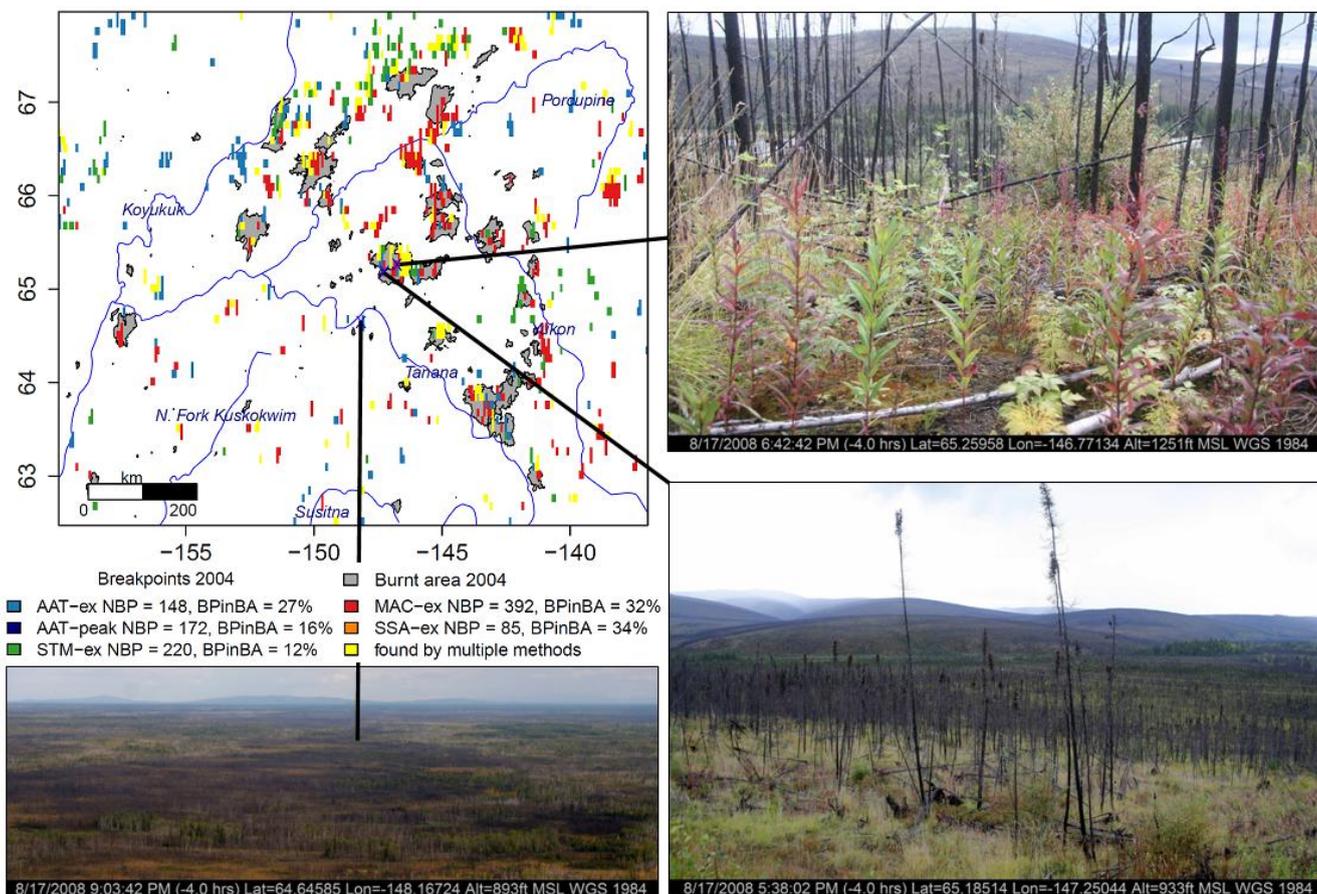
4.5. Practical Recommendations

Based on our results we want to summarize advantages and limitations of the methods and give recommendations for practical applications of trend and breakpoint detection methods on long-term NDVI time series. All tested methods offer advantages but involve also different limitations for trend and breakpoint estimation. Removing a mean annual cycle from seasonal time series (method MAC) in order to calculate trends is an often applied method. However, even if this trend analysis on seasonal-adjusted time series had the highest number of correct detected breakpoints, it involves also the highest number of false positive detected breakpoints and had a low overall performance for trend and breakpoint detection. The method that removes a modulated annual cycle as detected by singular spectrum analysis (method SSA) allows distinguishing and quantifying changes that are caused by changes in seasonality or caused by the long-term NDVI trend. Although this method resulted in a high proportion of correct detected breakpoints, it involves a high number of false positive detected breakpoints as method MAC. Additionally, the de-seasonalisation by a modulated annual cycle can remove the

inter-annual variability that is related to trends and results in a low overall performance of this method. Trend and breakpoint estimation based on a season-trend model (method STM) quantifies trends while taking into account the seasonality and noise of the NDVI time series. Thus, it allows detecting, distinguishing and quantifying changes in the phenological cycle as well as in the long-term NDVI trend [30,33,44]. Although this method under-estimated the number of breakpoints and had 5% false positive detected breakpoints, it had a good overall performance. Trend analysis on annual aggregated values (method AAT) highly underestimated the number of breakpoints but had the best overall performance of all methods considering the estimation of trends. The annual aggregation method had the lowest number of wrong detected breakpoints and detected real breakpoints were found with the highest accuracy of all methods (Figure 5). While this method reduces the temporal resolution and has a lower precision for the timing breakpoints, it offers the potential to calculate trends on specific properties of NDVI time series like annual or growing season means, as well as peak NDVI or seasonal amplitudes. The specific properties of NDVI time series in high-latitude regions (short vegetation period, snow cover, often cloud affected) limit the applicability of the seasonal methods and ultimately suggest using method AAT on peak NDVI to exclude observations affected by distortions.

Based on different advantages and limitations of all methods, we recommend using method AAT on mean growing season or peak NDVI for regions where the time series are often affected by distortions. If the seasonal NDVI time series values outside the peak NDVI period are credible, the calculation of a multi-method ensemble based on full time series could help to detect robust trends and breakpoints assuming that the agreement of multiple methods is more reliable than a single method. The later approach allows not only to detect trends but also to quantify the uncertainty of the trend estimate based on the choice of the trend method. Breakpoints can be considered as more robust if they were detected by multiple methods. Nevertheless the environmental plausibility of detected breakpoints needs to be assessed. Breakpoints with abrupt changes, *i.e.*, with higher magnitudes, were detected more accurately than breakpoints with gradual changes, *i.e.*, with low magnitudes. On the other hand false detected abrupt changes caused a low overall performance of all trend methods. Thus, we recommend to check the magnitude of changes at breakpoints and to relate these breakpoints in a driver-oriented framework [36] to potential causes of changes like land cover changes, drought or disturbances as fire or insect infestations.

As the detection of breakpoints causes additional uncertainties in trend estimates, the purpose of the breakpoint detection in analysis of NDVI trends needs to be clearly defined: Are trends or trend changes (breakpoints) in the focus of a study? Although the detection of breakpoints offer the potential to detect disturbances in NDVI time series, trend changes and trends in sub-segments of the NDVI time series, a false detection of non-existing breakpoints can result in wrong or even opposite NDVI trend estimates.

Inter-annual variability is the most important factor for the performance of methods to detect trends and breakpoints in NDVI time series. Main sources for inter-annual variability in NDVI time series are (1) contaminants like insufficient pre-processed data or insufficient harmonized multi-sensor observations, (2) meteorological distortions like clouds, dust, aerosols or snow cover and (3) environmental processes like climate variability, disturbances and land cover changes with associated changes in ecosystem structure and productivity. Usually, only the later source of inter-annual variability is of interest in NDVI time series analyses. Users of long-term NDVI datasets rely on the pre-processing

and harmonization of multi-sensor observations performed in all conscience by dataset providers. Meteorological distortions can be excluded from analyses by excluding NDVI observations that are flagged as snow or poor quality; by aggregating the bi-monthly GIMMS NDVI3g dataset to monthly temporal resolution; or by analysing only annual peak NDVI observations. As snow cover and clouds have low NDVI values, an extended snow cover can likely cause a detection of weaker NDVI greening or even the detection of browning trends. Thus, the use or non-use of snow-affected and low quality NDVI observations directly affects the inter-annual variability of the NDVI time series and such NDVI values should be excluded from trend and breakpoint analyses.

5. Conclusions

We demonstrated that increasing inter-annual variability in Normalized Difference Vegetation Index (NDVI) time series decreases the performance of methods to detect trends and trend changes in long-term NDVI time series. Trend estimation based on annual aggregated NDVI time series and the season-trend method had good overall performances. Hence, in order to detect trend changes in NDVI time series with higher precision and accuracy, one needs to improve methods that work on the full temporal resolution time series regarding the robustness against inter-annual variability. Inter-annual variability of NDVI time series can be caused by artefacts from the harmonization of a dataset from different sensors, meteorological distortions like clouds or snow and environmental processes such as climate patterns or disturbances. As a consequence, snow-affected NDVI observation or observations with poor quality need to be excluded from trend and breakpoints analyses as the performance of trend and breakpoint estimation methods decreases with increasing inter-annual variability. Methods can detect wrong or even opposite NDVI trends if they detect breakpoints in time series that have no breakpoints. Nevertheless, the detection of breakpoints offers the potential to detect trend changes or disturbances in NDVI time series.

We evaluated for the first time different methods to detect trends and trend changes in newly available 30 year GIMMS NDVI3g time series. Future studies of trends and breakpoints in long-term NDVI time series should assess the plausibility of detected breakpoints against multi-method ensemble estimates of breakpoints, quality flags of the NDVI time series and further environmental data streams, in order to prevent a detection of wrong NDVI trends.

Conflict of Interest

The authors declare no conflict of interest.

References and Notes

1. Lucht, W.; Schaphoff, S.; Erbrecht, T.; Heyder, U.; Cramer, W. Terrestrial vegetation redistribution and carbon balance under climate change. *Carbon Balance Manage.* **2006**, *1*, 6.
2. Coppin, P.; Jonckheere, I.; Nackaerts, K.; Muys, B.; Lambin, E. Digital change detection methods in ecosystem monitoring: A review. *Int. J. Remote Sens.* **2004**, *25*, 1565–1596.
3. Tucker, C.J. Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sens. Environ.* **1979**, *150*, 127–150.
4. Turner, D.P.; Cohen, W.B.; Kennedy, R.E.; Fassnacht, K.S.; Briggs, J.M. Relationships between leaf area index and Landsat TM spectral vegetation indices across three temperate zone sites. *Remote Sens. Environ.* **1999**, *70*, 52–68.
5. Gamon, J.A.; Field, C.B.; Goulden, M.L.; Griffin, K.L.; Hartley, E.; Joel, G.; Peñuelas, J.; Valentini, R. Relationships between NDVI, canopy structure, and photosynthesis in three californian vegetation types. *Ecol. Appl.* **1995**, *5*, 28–41.
6. Fensholt, R.; Sandholt, I.; Rasmussen, M.S. Evaluation of Modis LAI, FAPAR and the relation between FAPAR and NDVI in a semi-arid environment using *in situ* measurements. *Remote Sens. Environ.* **2004**, *91*, 490–507.
7. Myneni, R.B.; Hall, F.G.; Sellers, P.J.; Marshak, A.L. The interpretation of spectral vegetation indexes. *IEEE Trans. Geosci. Remote Sens.* **1995**, *33*, 481–486.
8. DeFries, R.; Hansen, M.; Townshend, J. Global discrimination of land cover types from metrics derived from AVHRR pathfinder data. *Remote Sens. Environ.* **1995**, *54*, 209–222.
9. Tucker, C.J.; Slayback, D.A.; Pinzon, J.E.; Los, S.O.; Myneni, R.B.; Taylor, M.G. Higher northern latitude normalized difference vegetation index and growing season trends from 1982 to 1999. *Int. J. Biometeorol.* **2001**, *45*, 184–190.
10. Myneni, R.B.; Keeling, C.D.; Tucker, C.J.; Asrar, G.; Nemani, R.R. Increased plant growth in the northern high latitudes from 1981 to 1991. *Nature* **1997**, *386*, 698–702.
11. Lucht, W.; Prentice, I.C.; Myneni, R.B.; Sitch, S.; Friedlingstein, P.; Cramer, W.; Bousquet, P.; Buermann, W.; Smith, B. Climatic control of the high-latitude vegetation greening trend and pinatubo effect. *Science* **2002**, *296*, 1687–1689.
12. Myers-Smith, I.H.; Forbes, B.C.; Wilmking, M.; Hallinger, M.; Lantz, T.; Blok, D.; Tape, K.D.; Macias-Fauria, M.; Sass-Klaassen, U.; Lévesque, E.; *et al.* Shrub expansion in tundra ecosystems: Dynamics, impacts and research priorities. *Environ. Res. Lett.* **2011**, *6*, 045509–045509.

13. Goetz, S.J.; Bunn, A.G.; Fiske, G.J.; Houghton, R.A. Satellite-observed photosynthetic trends across boreal north america associated with climate and fire disturbance. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 13521–13525.

14. Bunn, A.G.; Goetz, S.J.; Kimball, J.S.; Zhang, K. Northern high-latitude ecosystems respond to climate change. *Eos Trans. AGU* **2007**, *88*, 333–333.

15. Wang, X.; Piao, S.; Ciais, P.; Li, J.; Friedlingstein, P.; Koven, C.; Chen, A. Spring temperature change and its implication in the change of vegetation growth in North America from 1982 to 2006. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 1240–1245.

16. Piao, S.; Wang, X.; Ciais, P.; Zhu, B.; Wang, T.; Liu, J. Changes in satellite-derived vegetation growth trend in temperate and boreal Eurasia from 1982 to 2006. *Glob. Change Biol.* **2011**, *17*, 3228–3239.

17. Beck, P.S.A.; Goetz, S.J. Satellite observations of high northern latitude vegetation productivity changes between 1982 and 2008: Ecological variability and regional differences. *Environ. Res. Lett.* **2011**, *6*, 45501–45501.

18. Fensholt, R.; Proud, S.R. Evaluation of earth observation based global long term vegetation trends—Comparing GIMMS and MODIS global NDVI time series. *Remote Sens. Environ.* **2012**, *119*, 131–147.

19. Tucker, C.; Pinzon, J.; Brown, M.; Slayback, D.; Pak, E.; Mahoney, R.; Vermote, E.; El Saleous, N. An extended AVHRR 8-km NDVI dataset compatible with MODIS and spot vegetation NDVI data. *Int. J. Remote Sens.* **2005**, *26*, 4485–4498.

20. Alcaraz-Segura, D.; Chuvieco, E.; Epstein, H.E.; Kasischke, E.S.; Trishchenko, A. Debating the greening *vs.* Browning of the north american boreal forest: Differences between satellite datasets. *Glob. Change Biol.* **2010**, *16*, 760–770.

21. Parent, M.B.; Verbyla, D. The browning of Alaska's boreal forest. *Remote Sens.* **2010**, *2*, 2729–2747.

22. Fraser, R.H.; Olthof, I.; Carrière, M.; Deschamps, A.; Pouliot, D. Detecting long-term changes to vegetation in northern Canada using the Landsat satellite image archive. *Environ. Res. Lett.* **2011**, *6*, 045502–045502.

23. Beck, P.S.A.; Juday, G.P.; Alix, C.; Barber, V.A.; Winslow, S.E.; Sousa, E.E.; Heiser, P.; Herriges, J.D.; Goetz, S.J. Changes in forest productivity across Alaska consistent with biome shift. *Ecol. Lett.* **2011**, *14*, 373–379.

24. Berner, L.T.; Beck, P.S.A.; Bunn, A.G.; Lloyd, A.H.; Goetz, S.J. High-latitude tree growth and satellite vegetation indices: Correlations and trends in Russia and Canada (1982–2008). *J. Geophys. Res.-Biogeosci.* **2011**, *116*, doi: 10.1029/2010JG001475.

25. Verbyla, D. Browning boreal forests of western north america. *Environ. Res. Lett.* **2011**, *6*, 41003–41003.

26. Sulkava, M.; Luyssaert, S.; Rautio, P.; Janssens, I.A.; Hollmén, J. Modeling the effects of varying data quality on trend detection in environmental monitoring. *Ecol. Inform.* **2007**, *2*, 167–176.

27. Eklundh, L.; Olsson, L. Vegetation index trends for the African Sahel 1982–1999. *Geophys. Res. Lett.* **2003**, *30*, 1–4.

28. de Beurs, K.M.; Henebry, G.M. Trend analysis of the pathfinder AVHRR land (PAL) NDVI data for the deserts of central Asia. *IEEE Geosci. Remote Sens. Lett.* **2004**, *1*, 282–286.

29. de Beurs, K.M.; Henebry, G.M. A land surface phenology assessment of the northern polar regions using MODIS reflectance time series. *Can. J. Remote Sens.* **2010**, *36*, S87–S110.

30. Verbesselt, J.; Hyndman, R.; Newnham, G.; Culvenor, D. Detecting trend and seasonal changes in satellite image time series. *Remote Sens. Environ.* **2010**, *114*, 106–115.

31. de Jong, R.; de Bruin, S.; de Wit, A.; Schaepman, M.E.; Dent, D.L. Analysis of monotonic greening and browning trends from global NDVI time-series. *Remote Sens. Environ.* **2011**, *115*, 692–702.

32. Mahecha, M.D.; Fürst, L.M.; Gobron, N.; Lange, H. Identifying multiple spatiotemporal patterns: A refined view on terrestrial photosynthetic activity. *Pattern Recogn. Lett.* **2010**, *31*, 2309–2317.

33. Verbesselt, J.; Hyndman, R.; Zeileis, A.; Culvenor, D. Phenological change detection while accounting for abrupt and gradual trends in satellite image time series. *Remote Sens. Environ.* **2010**, *114*, 2970–2980.

34. Wu, Z.; Schneider, E.K.; Kirtman, B.P.; Sarachik, E.S.; Huang, N.E.; Tucker, C.J. The modulated annual cycle: An alternative reference frame for climate anomalies. *Clim. Dyn.* **2008**, *31*, 823–841.

35. de Jong, R.; Verbesselt, J.; Schaepman, M.E.; Bruin, S. Trend changes in global greening and browning: Contribution of short-term trends to longer-term change. *Glob. Change Biol.* **2011**, *18*, 642–655.

36. de Jong, R.; Verbesselt, J.; Zeileis, A.; Schaepman, M. Shifts in global vegetation activity trends. *Remote Sens.* **2013**, *5*, 1117–1133.

37. Bai, J.; Perron, P. Computation and analysis of multiple structural change models. *J. Appl. Econometr.* **2003**, *18*, 1–22.

38. Pinzon, J.; *et al.* Revisiting error, precision and uncertainty in ndvi avhrr data: Development of a consistent NDVI3g time series. *Remote Sens.* **2013**, this volume.

39. Xu, L.; Myneni, R.B.; Chapin, F.S., III; Callaghan, T.V.; Pinzon, J.E.; Tucker, C.J.; Zhu, Z.; Bi, J.; Ciais, P.; Tommervik, H.; *et al.* Temperature and vegetation seasonality diminishment over northern lands. *Nature Clim. Change* **2013**, doi:10.1038/nclimate1836.

40. Kaufmann, R.K.; Zhou, L.; Knyazikhin, Y.; Shabanov, V.; Myneni, R.B.; Tucker, C.J. Effect of orbital drift and sensor changes on the time series of avhrr vegetation index data. *IEEE Trans. Geosci. Remote Sens.* **2000**, *38*, 2584–2597.

41. Holben, B.N. Characteristics of maximum-value composite images from temporal AVHRR data. *Int. J. Remote Sens.* **1986**, *7*, 1417–1434.

42. Zeileis, A.; Kleiber, C.; Krämer, W.; Hornik, K. Testing and dating of structural changes in practice. *Computat. Statist. Data Anal.* **2003**, *44*, 109–123.

43. Mann, H.B. Nonparametric tests against trend. *Econometrica* **1945**, *13*, 245–259.

44. Verbesselt, J.; Zeileis, A.; Herold, M. Near real-time disturbance detection using satellite image time series. *Remote Sens. Environ.* **2012**, *123*, 98–108.

45. Golyandina, N.; Nekrutkin, V.V.; Zhigljavskiy, A.A. *Analysis of Time Series Structure: SSA and Related Techniques*; Chapman & Hall/CRC: Boca Raton, FL, USA, 2001; p. 305.

46. Foody, G.M. Status of land cover classification accuracy assessment. *Remote Sens. Environ.* **2002**, *80*, 185–201.

47. Congalton, R.G. A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sens. Environ.* **1991**, *37*, 35–46.

48. Deming, W.E.; Stephan, F.F. On a least squares adjustment of a sampled frequency table when the expected marginal totals are known. *Ann. Math. Statist.* **1940**, *11*, 427–444.

49. Kasischke, E.S.; Williams, D.; Barry, D. Analysis of the patterns of large fires in the boreal forest region of Alaska. *Int. J. Wildland Fire* **2002**, *11*, 131–144.

50. FRAMES: Alaska Large Fire Database. 2011.

51. Schneider, U.; Fuchs, T.; Meyer-Christoffer, A.; Rudolf, B. Global Precipitation Analysis Products of the GPCC. http://www.dwd.de/bvbw/generator/DWDWWW/Content/Oeffentlichkeit/ KU/KU4/KU42/en/Reports__Publications/GPCC__intro__products__2011,templateId=raw,prope rty=publicationFile.pdf/GPCC_intro_products_2011.pdf (accessed on 17 February 2012).

52. Mitchell, T.D.; Jones, P.D. An improved method of constructing a database of monthly climate observations and associated high-resolution grids. *Int. J. Climatol.* **2005**, *25*, 693–712.

53. Neigh, C.; Tucker, C.; Townshend, J. North american vegetation dynamics observed with multi-resolution satellite data. *Remote Sens. Environ.* **2008**, *112*, 1749–1772.

54. Verbyla, D. The greening and browning of alaska based on 1982–2003 satellite data. *Glob. Ecol. Biogeogr.* **2008**, *17*, 547–555.

55. Macias Fauria, M.; Johnson, E.A. Large-scale climatic patterns control large lightning fire occurrence in canada and alaska forest regions. *J. Geophys. Res.* **2006**, *111*, G04008.

56. Hess, J.C.; Scott, C.A.; Hufford, G.L.; Fleming, M.D. El Niño and its impact on fire weather conditions in alaska. *Int. J. Wildland Fire* **2001**, *10*, 1–13.

57. McPhaden, M.J. Genesis and evolution of the 1997–98 El Niño. *Science* **1999**, *283*, 950–954.