How to measure ecosystem stability? An evaluation of the reliability of stability metrics based on remote sensing time series across the major global ecosystems

WANDA DE KEERSMAECKER1, STEFL HERMITTE2,3, OLIVIER HONNAY4, JAMSHID FARIFTEH1, BEN SOMERS5 and POL COPPIN1

1M3-BIORES, KU Leuven, Willem de Croylaan 34, Heverlee B-3001, Belgium, 2Royal Netherlands Meteorological Institute (KNMI), Wilhelminalaan 10, De Bilt NL-3732, The Netherlands, 3Department of Earth and Environmental Sciences, KU Leuven, Celestijnenlaan 200E, Heverlee B-3001, Belgium, 4Ecologie, Evolutie en Biodiversiteitsbehoud, KU Leuven, Kasteelpark Arenberg 31, Heverlee B-3001, Belgium, 5Division of Forest, Nature and Landscape, KU Leuven, Celestijnenlaan 200E, Heverlee B-3001, Belgium

Abstract

Increasing frequency of extreme climate events is likely to impose increased stress on ecosystems and to jeopardize the services that ecosystems provide. Therefore, it is of major importance to assess the effects of extreme climate events on the temporal stability (i.e., the resistance, the resilience, and the variance) of ecosystem properties. Most time series of ecosystem properties are, however, affected by varying data characteristics, uncertainties, and noise, which complicate the comparison of ecosystem stability metrics (ESMs) between locations. Therefore, there is a strong need for a more comprehensive understanding regarding the reliability of stability metrics and how they can be used to compare ecosystem stability globally. The objective of this study was to evaluate the performance of temporal ESMs based on time series of the Moderate Resolution Imaging Spectroradiometer derived Normalized Difference Vegetation Index of 15 global land-cover types. We provide a framework (i) to assess the reliability of ESMs in function of data characteristics, uncertainties and noise and (ii) to integrate reliability estimates in future global ecosystem stability studies against climate disturbances. The performance of our framework was tested through (i) a global ecosystem comparison and (ii) a comparison of ecosystem stability in response to the 2003 drought. The results show the influence of data quality on the accuracy of ecosystem stability. White noise, biased noise, and trends have a stronger effect on the accuracy of stability metrics than the length of the time series, temporal resolution, or amount of missing values. Moreover, we demonstrate the importance of integrating reliability estimates to interpret stability metrics within confidence limits. Based on these confidence limits, other studies dealing with specific ecosystem types or locations can be put into context, and a more reliable assessment of ecosystem stability against environmental disturbances can be obtained.

Keywords: climate disturbances, ecosystem stability, Normalized Difference Vegetation Index, reliability, remote sensing, resilience, resistance, variance

Received 26 June 2013; revised version received 6 November 2013 and accepted 9 November 2013

Introduction

The occurrence of climate extremes is changing globally. Models project a substantial change in temperature extremes, a higher frequency of heavy precipitation, and more intense droughts over many areas of the globe (IPCC, 2012). These extreme climate events are likely to impose increased stress on ecosystems, and they may jeopardize the various services that ecosystems provide to man and society (Thomas et al., 2004; Philippart et al., 2011; Gosling, 2012). Therefore, it is of major importance to assess the effects of extreme climate events on ecosystems, effects that depend on both the exposure of the ecosystem and its susceptibility (IPCC, 2012). It is crucial in this context to quantify ecosystem properties and the stability of these properties following a climate-induced disturbance.

Ecosystem stability can be derived from the measured ecosystem properties through quantifying their response to climate-induced disturbances. In this context, three types of ecosystem stability metrics (ESMs) have been proposed, derived from time series that express anomalies relative to the seasonal climatology of the ecosystem property. First, resistance quantifies the direct impact of a perturbation on the ecosystem property. Consequently, resistance expresses the ability of
the ecosystem to maintain its original state following an environmental perturbation, and it can be quantified based on the magnitude of the anomaly at the moment of perturbation (Lloret et al., 2007; Van Ruijven & Berendse, 2010; Vogel et al., 2012). Second, resilience defines the rate of return to the equilibrium state after the ecosystem has been disturbed, and it can be expressed by the degree of temporal relation between observations (Telesca & Lasaponara, 2006; Telesca et al., 2008; Zaccarelli et al., 2011; Dakos et al., 2012). Finally, variance gives a more general idea of ecosystem stability. It is defined by the standard deviation or the coefficient of variation of the anomaly time series (Pimm, 1984). An ecosystem property will show larger variance when the resistance is lower and when the return to the equilibrium state is slower (Pimm, 1984; Tilman & Downing, 1996; Telesca & Lasaponara, 2006; Lloret et al., 2007; Vogel et al., 2012).

Although many studies have quantified the stability of different ecosystems from different regions (Potter et al., 1999; Fang et al., 2001; Knapp & Smith, 2001; Neigh et al., 2008; Vicente-Serrano et al., 2013), a systematic comparison of the stability of the major global ecosystems for the three ESMs is still lacking. Such a comparison is not straightforward as the reliability of the ESMs can be expected to differ due to varying data quality and properties for different locations and context. For example, noise introducing factors, such as the presence of clouds, snow or amounts of aerosols, the frequency of measurements or the performance of a measurement device, vary spatially, and may affect the performance of each ESM differently (Samanta et al., 2010, 2011, 2012a,b). Some of these factors can be corrected for (e.g., systematic errors), but others create noise in the time series that directly interferes with the stability metrics. This implies that it might not be known whether a deviation from the ecosystem’s equilibrium state is due to measurement noise or due to the response to the environmental perturbation (Hird & McDermid, 2009). Furthermore, also time series characteristics, such as measurement frequency, and the number of measurements might influence ESMs. Therefore, quantifying and comparing the stability of different ecosystems requires fundamental insight into the accuracy and reliability of ESMs.

This research aims to address the current need for a more thorough understanding of the performance of resistance, resilience, and variance ESMs and how they can be used to compare ecosystem stability globally. The specific objectives are to provide a framework (i) to assess the reliability of ESMs, where reliability is defined as accuracy in function of data characteristics, uncertainties and noise; and (ii) to integrate reliability in the comparison of the stability of the major global ecosystems. To meet these objectives, we used global land-cover (LC) types by means of global subsets of the Moderate Resolution Imaging Spectroradiometer (MODIS) Normalized Difference Vegetation Index (NDVI) time series, and we assessed the performance of our framework through (i) a global ecosystem comparison and (ii) a comparison of ecosystem stability following the 2003 European heat wave (Fink et al., 2004; Rebetez et al., 2006). As such, our study offers a general approach for implementation within studies that monitor ecosystem stability against climate disturbances.

Materials and methods

Data

We used remote sensing NDVI time series of global LC types. NDVI time series quantify the amount and greenness of vegetation (Rouse et al., 1974) and thus correlate with vegetation biomass, dynamics, and the fAPAR. Moreover, they are globally available at high temporal resolutions from various sensors. Therefore, they are very suitable as a global ecosystem health state indicator (Townshend & Justice, 1986; Tucker et al., 1986; Peruelas & Filella, 2001; Slayback et al., 2002; Jia et al., 2005; Lasaponara, 2006; Piao et al., 2011).

Global subsets of NDVI data were extracted from the MODIS Land Product based on the MODIS ASCII Subset for the Northern hemisphere (Oak Ridge National Laboratory distributed Active Archive Center (ORNL DAAC), 2010). NDVI time series for the period 2001–2006 were compiled in two steps from the Terra MODIS Land cover (MCD12Q1) and NDVI (MOD13Q1) products for 7 × 7 km areas around 1079 flux towers or field sites. The flux towers are part of the FLUXNET (Baldocchi et al., 2001), a global network of micrometeorological tower sites where exchanges of carbon dioxide, water vapor, and energy between terrestrial ecosystems and the atmosphere are measured.

Firstly, MCD12Q1 data that provide yearly estimates of the International Geosphere-Biosphere Programme Classification (IGBP) LC type (Loveland et al., 1995; Friedl et al., 2002, 2010) were used to assign each NDVI time series to its corresponding LC type. This was carried out by determining the dominant LC type (i.e., forest types, open and closed shrubland, savanna types, grasslands, permanent wetlands, cropland, snow, and barren) for each site based on the 14 × 14 0.5 km pixels around each flux tower or field site (see Fig. 1). Secondly, the NDVI time series that correspond to the dominant LC type of each site were extracted from the 28 × 28 0.25 km 16-day MOD13Q1 NDVI data. This extraction results in a set of \( n_{LC} \) NDVI time series per LC type which are assumed to represent the characteristic behavior of each LC type (Lhermitte et al., 2011a).

Methods

The framework to assess the reliability of ESMs and compare the ecosystem stability globally consisted of two steps. Firstly,
the accuracy of each ESM in function of data and noise characteristics was determined using a sensitivity experiment. Monte Carlo (MC) simulations were used in the sensitivity experiment as they allow control over the time series characteristics in combination with noise levels, which is difficult using the original time series where this is predefined. Secondly, the reliability of each ESM was calculated for the experimental time series of the global subset by relating the data and noise characteristics of the time series to the obtained accuracy from the sensitivity experiment. This subsequently allowed global comparison of the time series.

Accuracy of ESMs in function of data and noise characteristics

MC simulations. Normalized Difference Vegetation Index time series of the mean $\bar{f}^{LC}(t)$ and the standard deviation $\sigma^{LC}(t)$ per land-cover type were used as basis for the MC simulations (Lhermitte et al., 2011a). The $f^{LC}(t)$ and $\sigma^{LC}(t)$ time series were created by (i) removing all low-quality data points from the $n_{LC}$ NDVI time series (i.e., associated MODLAND quality flags larger than 1), (ii) calculating the time series of mean $\bar{f}^{LC}(t)$ and the standard deviation $\sigma^{LC}(t)$ for each LC type, and (iii) temporally interpolating $\bar{f}^{LC}(t)$ and $\sigma^{LC}(t)$ for missing observation dates. This was performed for the Northern hemisphere only (i.e., 90% of the sites) to avoid the effect of different seasons in both hemispheres. Together $\bar{f}^{LC}(t)$ and $\sigma^{LC}(t)$ time series represent the characteristic behavior of each LC type (Fig. 1) and allow regenerating time series with similar characteristics.

As nonstationarity may interfere with the ESMs (Hu et al., 2001), the $f^{LC}(t)$ and $\sigma^{LC}(t)$ were tested for stationarity before MC simulations. Firstly, nonstationary time series with a stochastic trend (i.e., time series that do not tend to evolve...
toward an equilibrium, such as for closed shrublands, permanent wetlands, and snow and ice) were removed based on a Philips Perron test (Phillips & Perron, 1988) and Augmented Dickey Fuller (ADF; Dickey & Fuller, 1979) test. Secondly, the \( \tilde{f}_{LC}^{t}(t) \) and \( \sigma_{LC}^{t}(t) \) time series were detrended by removing the deterministic trend (if significant) of the Mann-Kendall trend test, modified to take into account the autocorrelation in the time series (Sen, 1968; Hamed & Ramachandra Rao, 1998; Yue & Wang, 2004). Finally, a unit root using the ADF test (without a trend in the test equation) was performed on the detrended and deseasonalized time series to assure stationarity. Rao, 1998; Yue & Wang, 2004). Finally, a unit root using the ADF test (without a trend in the test equation) was performed on the detrended and deseasonalized time series to assure stationarity.

The simulated noise types are common in ecological time series of biomass or remote sensing–based indices. For example, white or random noise (Fig. 2a) might occur due to measurement and sensor uncertainties when measuring biomass or due to simultaneous presence of several uncorrected noise components (e.g., geometric errors or atmospheric conditions) in remote-sensing data (Jónsson & Eklundh, 2004; Hird & McDermid, 2009). Negatively or positively biased noise (Fig. 2b) may be the result of insufficient drying or calibration errors when measuring biomass (Cherubini et al., 1998) or cloud cover, snow, and shadow in remote sensing time series (Roerink et al., 2000; Jónsson & Eklundh, 2002; Chen et al., 2004). Trends (Fig. 2f) can be due to sensor degradation, or changing environments and natural succession (Fensholt & Proud, 2012; Wang et al., 2012). Furthermore, the length and temporal resolution of ecological time series can vary characteristics (further referred to as noise types) and different levels of change (further referred to as noise levels). The adapted changes allow to determine the sensitivity of ESMs to (i) white noise, (ii) biased noise, (iii) length of the time series, (iv) missing values, (v) temporal resolution, and (vi) a linear trend. An overview of the equations and parameters for the MC simulations can be found in Table 1, while a graphical overview of the types of noise can be found in Fig. 2.

The resulting stationary \( \tilde{f}_{LC}^{t}(t) \) and \( \sigma_{LC}^{t}(t) \) were used to simulate 500 sample NDVI time series per LC \( f_{LC}^{t}(t) \) using (Viovy et al., 1992):

\[
\begin{align*}
\tilde{f}_{LC}^{t}(t) &= \tilde{f}_{LC}^{t}(t) + \tilde{\sigma}_{LC}^{t}(t), \\
\sigma_{LC}^{t}(t) &= \beta_{LC}^{t} + \xi_{LC}^{t}(t), \\
\end{align*}
\]

in which \( \alpha \) is a random value given by a pseudorandom generator following a normal distribution \( N(\mu = 0, \sigma = 1) \). The use of \( \tilde{f}_{LC}^{t}(t) \) assures the removal of all noise due its mean filter effect, while the use of \( \sigma_{LC}^{t}(t) \) allows to represent the natural variability within a land-cover class.

Additionally, the 500 sample time series per LC were adapted by introducing six types of change to the data characteristics (further referred to as noise types) and different levels of change (further referred to as noise levels). The adapted changes allow to determine the sensitivity of ESMs to (i) white noise, (ii) biased noise, (iii) length of the time series, (iv) missing values, (v) temporal resolution, and (vi) a linear trend. An overview of the equations and parameters for the MC simulations can be found in Table 1, while a graphical overview of the types of noise can be found in Fig. 2.

### Table 1: Equations for the different Monte Carlo experiments

<table>
<thead>
<tr>
<th>Effect</th>
<th>Equation</th>
<th>Parameters</th>
<th>Parameters setting</th>
</tr>
</thead>
<tbody>
<tr>
<td>White noise</td>
<td>( f_{LC}^{t}(t) = \tilde{f}<em>{LC}^{t}(t) + a</em>{LC}^{t}(t) + \beta(t) )</td>
<td>( \beta(t) = \text{Pseudorandom value from normal distribution } N(\mu = 0, \sigma = \sigma_{WN}) )</td>
<td>( w_{WN} = 0.01, 0.03, \ldots, 0.19 )</td>
</tr>
<tr>
<td>Biased noise</td>
<td>( f_{LC}^{t}(t) = \tilde{f}<em>{LC}^{t}(t') + a</em>{LC}^{t}(t) )</td>
<td>( t' = \text{Time series that consists of the original time series of Eqn (10) with a number } N(t_{BN}) ) of observations replaced by noise (biased low value of 0.1)</td>
<td>( w_{BN} = 1, 2, \ldots, 10 )</td>
</tr>
<tr>
<td>Length of the time series</td>
<td>( f_{LC}^{t}(t) = \tilde{f}<em>{LC}^{t}(t') + a</em>{LC}^{t}(t') )</td>
<td>( t' = \text{Randomly shortened time series by removing all observations of } \sigma_{TR} \text{ years. The removed years were chosen randomly from a discrete uniform distribution} )</td>
<td>( w_{TR} = 1, 2, 3 )</td>
</tr>
<tr>
<td>Missing values</td>
<td>( f_{LC}^{t}(t) = \tilde{f}<em>{LC}^{t}(t') + a</em>{LC}^{t}(t') )</td>
<td>( t' = \text{Time series that consists of the original time series of Eqn (10) with a number } N(t_{MV}) ) of observations randomly set as missing value</td>
<td>( w_{MV} = 11, 17, \ldots, 71 )</td>
</tr>
<tr>
<td>Temporal resolution</td>
<td>( f_{LC}^{t}(t) = \tilde{f}<em>{LC}^{t}(t') + a</em>{LC}^{t}(t') )</td>
<td>( t' = \text{Time series that consists of the original time series of Eqn (10) with a number } N(t_{TR}) \text{ observations removed every } \sigma_{TR} \text{ observations. The start date was selected randomly out of the first } \sigma_{TR} \text{ observations} )</td>
<td>( w_{TR} = 2, 3, \ldots, 7 )</td>
</tr>
<tr>
<td>Linear trend</td>
<td>( f_{LC}^{t}(t) = \tilde{f}<em>{LC}^{t}(t) + a</em>{LC}^{t}(t) + \beta(t) )</td>
<td>( \beta(t) = \text{Linear trend with } -w_{LT} \Delta_{t} + w_{LT} t' ) ( w_{LT} = 0.0001, 0.0003, \ldots, 0.002 ) in which ( t' = t/16 ) and ( \Delta t ) the measurement period</td>
<td></td>
</tr>
</tbody>
</table>

LC, landcover.

significantly (Fig. 2e, c) with often several missing observations (Fig. 2d) as a result of the labor-intensive work or destructive impact of biomass measurements (Tackenberg, 2007). This is also the case for remote sensing data as different sensor or preprocessing procedures (Roerink et al., 2000; Jönsson & Eklundh, 2002; Chen et al., 2004; Hilker et al., 2009) in combination with irregular revisit times or cloud cover (Hilker et al., 2009; Ju et al., 2010; Lhermitte et al., 2011b; Veraverbeke et al., 2011) might result in time series with different length, temporal resolution, and/or missing values.

Ecosystem stability metrics. The 500 NDVI time series per LC, noise type, and noise level that result from the MC simulations were subsequently used to calculate the ESMs. This was performed by first removing the climatological mean (climatology, i.e., the mean NDVI for a particular day of year, across several years) from each time series and then calculating resistance, resilience, and variance from the anomaly time series as expressed in Table 2.

**Resistance** of an ecosystem expresses the ability to withstand environmental perturbations. A more resistant and stable ecosystem will show a smaller deviation from the seasonal climatology after an environmental disturbance (see Eqn 8 in Table 2; Lloret et al., 2007; Van Ruijven & Berendse, 2010; Vogel et al., 2012). As the use of anomalies in Eqn (8) does not account for the ecosystem’s capacity to change, the resistance is normalized using the climatology, i.e., the NDVI anomaly is divided by the average NDVI at that time of the year (see Eqn 9 in Table 2; Van Ruijven & Berendse, 2010; Vogel et al., 2012). Consequently, a decrease in biomass will have a stronger effect on ecosystems with a low amount of biomass than with a high amount of biomass.

**Resilience** is defined as the rate of return to its equilibrium state after a perturbation (Pimm, 1984). Ecosystems with a low resilience will return slowly to their equilibrium state, while high resilient ecosystems have a short return time (Lhermitte et al., 2010, 2011b). The temporal relation between observations serves as a metric of resilience and can be expressed by the autocorrelation at lag 1 ($r_1$), where higher $r_1$ relate to more similar subsequent anomalies and thus slower return to equilibrium. Therefore, resilience might be expressed as $1 - r_1$ (Dakos et al., 2012; see Eqn 10 in Table 2). Alternatively, two metrics have been developed to characterize resilience based on the variance of the frequency spectrum of the anomaly time series, i.e., the normalized spectral entropy ($H_{SN}$, see Eqn 13 in Table 2; Zaccarelli et al., 2011) and the spectral scaling component ($\alpha$, see Eqns 11 and 12 in Table 2). The normalized spectral entropy expresses the evenness of the distribution of variance, where higher entropies indicate more resilient ecosystems. The spectral scaling component is given by the slope of the logarithm of the spectrum upon the logarithm of the reverse frequency (Telesca & Lasaponara, 2006; Telesca et al.,

Fig. 2 Example of the types of noise added to the time series. The time series without noise are represented in blue, while the time series with noise are represented in red.
capacity (1−ρ1), the spectral scaling coefficient α and the normalized spectral entropy HSN, and (iii) variance [Standard deviation (SD) and Coefficient of Variation (CV)]. These measures are derived for a time series ts with N observations from time t$_1$ to t$_N$. The climatology of the time series is given by ts$_C$, while the time series of anomalies equals ts$_A$ = ts(t$_i$) − ts$_C$(t$_i$).

<table>
<thead>
<tr>
<th>Metric</th>
<th>Equations</th>
<th>Description measure</th>
<th>References</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resistance</td>
<td>Diff at $t_i$: $D(t_i) = ts(t_i) - ts_C(t_i)$</td>
<td></td>
<td>Vogel et al. (2012); Lloret et al. (2007); Van Ruijven &amp; Berendse (2010)</td>
<td>Low absolute (normalised) values indicate high resilience</td>
</tr>
<tr>
<td></td>
<td>ND at $t_i$: $ND(t_i) = \frac{[ts(t_i) - ts_C(t_i)]}{ts_C(t_i)}$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Resilience</td>
<td>$1 - \rho_1$ = $1 - \sum_{n=1}^{N} \frac{[ts_n(t_i) - ts_C(t_i)]}{[ts_n(t_i) - ts_C(t_i)]}$</td>
<td>$X$ stands for the mean of $X$</td>
<td>Dakos et al. (2012)</td>
<td>High $1 - \rho_1$ indicates high resilience</td>
</tr>
<tr>
<td></td>
<td>$\alpha_{ARFIL} = \log(PS) = c + \alpha_{ARFIL} \log(1/\lambda)$</td>
<td>PS represents the power spectrum, defined with the ARFIL algorithm, and $\lambda$ the frequency</td>
<td>Broersen et al. (2004)</td>
<td>Negative/zero/positive $\alpha_{ARFIL}$ or $\alpha_{DFA}$ values indicate antipersistent/white noise/persistent behavior</td>
</tr>
<tr>
<td></td>
<td>$\alpha_{DFA} = y(t_n) = \sum_{n=1}^{N} \frac{ts_n(t_i) - ts_C(t_i)}{\sum_{n=1}^{N} \frac{ts_n(t_i) - ts_C(t_i)}{}}$</td>
<td>$N_n$ represents the amount of boxes of length $n$ in the time series and $y_n(t_i)$ the fitted trend in each of the boxes</td>
<td>Peng et al. (1995)</td>
<td>Antipersistent/white noise/persistent behavior</td>
</tr>
<tr>
<td></td>
<td>$H_{SN} = H_{SN} = \frac{\sum_{k=0}^{N} \frac{p_k}{PS_k}}{\sum_{k=0}^{N} \frac{PS_k}}$</td>
<td>$K$ represents the frequency $\lambda_k$, $N_k$ the number of frequencies and $PS_k$ the power spectrum at frequency $\lambda_k$</td>
<td>Zaccarelli et al. (2011)</td>
<td>High $H_{SN}$ indicates high resilience</td>
</tr>
</tbody>
</table>

2008). An $\alpha$ equal to, larger and smaller than zero can be interpreted, respectively, as an indication of a white noise process (random anomalies), persistent behavior (i.e., positive/negative anomalies are likely followed by positive/negative anomalies) and antipersistent behavior (i.e., positive/negative anomalies are followed by negative/positive anomalies). Therefore, high-resilience ecosystems are expected to have an $\alpha$ approaching zero. A more detailed description can be found in the Supporting Information.

Variance denotes the total variability due to environmental perturbations (Eqn 14 in Table 2) and gives a more general measure of ecosystem stability. Ecosystems with a strong sensitivity to perturbations and a slow return to their equilibrium state will have a larger variability compared to ecosystems that are insensitive to perturbations and rapidly return to equilibrium (Pimm, 1984). In this context, the variance can be normalized for the average value of the ecosystem variable, giving the coefficient of variation (Pimm, 1984; see Eqn 15 in Table 2).

Sensitivity of the ESMs. Subsequently, the sensitivity of each ESM to the different noise types and noise levels was evaluated by comparing (i) the ESM derived from the adapted sample time series where noise types were introduced (Table 1) and (ii) the ESM derived from the unadapted sample time series of Eqn (1). As the latter can be considered the true ESM of the time series, the comparison gives a measure of reliability as it quantifies the accuracy or deviation from the actual stability due to changes in data and noise characteristics. In this study, the Mean Absolute Percentage Error (MAPE), i.e., the percentage error in ESM when noise is introduced, was used to quantify the reliability in function of varying noise types and noise levels:

$$\text{MAPE} = \frac{\sum_{j=1}^{n_{LCT}} \frac{\sum_{i=1}^{n_{obsLCT}} |\text{meas}_i - \text{meas}_i|}{n_{LCT}}\times 100}{\sum_{j=1}^{n_{LCT}} \frac{\sum_{i=1}^{n_{obsLCT}} \text{meas}_i}{n_{LCT}}}$$

where $n_{LCT}$ represents the number of LC types, $n_{obsLCT}$ the number of time series per LC type, meas is the ESM derived from the unadapted time series and meas$_i$ the ESM derived after adapting the time series by introducing different noise types and noise levels.

Reliability of ESMs of experimental time series

Data and noise characteristics of experimental time series. As the Monte Carlo sensitivity analysis provides the accuracy of each ESM in function of data and noise characteristics,
it can be used to estimate the reliability or MAPE of an experimental, i.e., nonsimulated time series. This implies, however, that data and noise characteristics have to be known for the experimental time series. For the global NDVI subset data, this means that the noise levels of different noise types (i.e. white noise, biased noise, missing values, trends) have to be calculated for each site, whereas the temporal resolution and time series length do not play a role as they are equal for all sites.

Prior to the estimation of the noise levels of different noise types, the NDVI subset data were preprocessed, i.e., observations with the MODIS snow and ice, shadow, adjacent cloud or mixed cloud flags equal to ‘present’ or aerosol quantity flag equal to ‘climatology’ or ‘high’ were removed. To reduce noise and missing values in the time series, all time series of the dominant LC type were averaged per site. Due to the high impact of biased noise on the ESM reliability (see Results and Discussion), we decided to replace the biased noise by missing values to increase reliability of global comparison of the ecosystem stability. Based on these mean NDVI time series per site, white noise was estimated using the power spectral estimate at the highest frequency (Fig. 3). The spectrum of a time series represents the distribution of time series variance upon frequency, where low and high frequencies are, respectively, associated with long- and short-term patterns in the time series. White noise is temporally uncorrelated and thus characterized by a flat spectrum. As higher frequencies of a time series spectrum are more likely dominated by the white noise component, its variance is estimated as the area underneath the horizontal line defined by the spectrum at the highest frequency (Halley, 1996). Biased noise was estimated as the amount of observations that are local extremes, i.e., differ more than two times the standard deviation of the anomalies from the average of the previous and next observation, whereas missing values were extracted directly from the time series. Trends were defined as the slope of the anomaly time series using a linear regression model.

**Global ecosystem stability and reliability of ESMs.** Based on mean NDVI time series per site, the resistance, resilience, and variance were calculated for each site, and the corresponding reliability was estimated using the calculated noise levels of that site. For example, if a time series has a white noise level that corresponds to 10% uncertainty in the sensitivity experiment (MAPE = 10), a biased noise level that corresponds to 5% uncertainty (MAPE = 5) and 8% uncertainty due to missing values (MAPE = 8), the total MAPE of the ESM equals 23%. Trends were not included in the MAPE but represented separately, as they might be the result of changing ecosystem characteristics and thus might provide additional information about ecosystem stability.

Finally, the ecosystem stability and MAPE results were also summarized per LC type to compare the ecosystem stability of different LC types. This was performed for LC types with at least five sites and restricted to (i) sites with MAPE’s < 30 and (ii) MAPE > 50 to assure a fair comparison between the LC types while taking into account the effect of ESM reliability.

**Results**

**Accuracy of ESMs in function of data and noise characteristics**

The results of the sensitivity experiment (Fig. 4) show that white noise, biased noise, and trends have a much stronger effect on the accuracy of resistance, resilience, and variance metrics than the length of the time series, the temporal resolution or the amount of missing values. For example, MAPE’s larger than 100 (i.e., more than 100% uncertainty about the true value of the ESM) are already reached with low noise levels (e.g., a few biased values or a white noise level of 0.05). These white noise levels are not unrealistic to be encountered in real NDVI time series as they correspond to 20% noise for a grassland where NDVI ranges from 0.15 to 0.4 or 8% noise for a forest ranging from 0.3 to 0.9. It indicates, moreover, that one biased value already can have an enormous effect on the ESM’s accuracy as it can result in uncertainties of 10–100% (MAPE = 10–100), which is equal to the uncertainty produced by removing 40–70 observations from the 115. It should, however, be noted that some residual white noise might be present in the time series used for the Monte Carlo experiment, which may cause an overestimation of the effect of noise on the ESMs. The mean white noise standard deviation of the Monte Carlo time series has been estimated as 0.01, which corresponds to a MAPE ranging from 2 to 60.

---

Fig. 3 Example of the white noise estimation based on the power spectrum of the time series. The power spectrum is represented by the blue curve, where the area underneath this curve equals the total variance. The red area underneath the horizontal line, given by the power at the highest frequency, represents the variance of the white noise component.
The sensitivity experiment (Fig. 4), moreover, shows that the effects are not similar for all ESMs. Resistance metrics are very sensitive to white noise and trends but are not affected by the length of the times series, its temporal resolution or amount of missing values, whereas variance metrics show also a strong increase in MAPE’s due to biased noise. The resilience metrics (except spectral entropy) are also very sensitive to white noise and biased noise (e.g., the effect of removing 1 year or 23 data points is smaller than the addition of one biased value) but they also show a strong sensitivity to temporal effects, like missing values, temporal resolution, and length of the time series.

The sensitivity experiment also demonstrates that the accuracy of the ESMs does not depend on the temporal pattern of missing data. For example, removing 50% of the data points from the time series by (i) reducing the length from 5 to 2.5 years, (ii) resampling at half the temporal resolution, or (iii) randomly introducing 50% missing values, results in very similar MAPE’s for the resilience and variance metrics [i.e., MAPE of (i) 3–70, (ii) 5–66, (iii) 4–52 for resilience and (i) 8, (ii) 8, (iii) 9 for variance, respectively].

**Global ecosystem stability**

Comparison of the ESMs for the global subsets (Fig. 5) shows the importance of understanding the reliability or MAPE’s when comparing global ecosystem stability. Firstly, it is clear that the MAPE’s of each ESM differ strongly. For example, more than 90% of the sites show a MAPE of 142.61 for the difference metric, whereas this is 97.39 for one minus the autocorrelation at lag 1 (1 − ρ), and 11.07 for spectral entropy. As a result, it is much more difficult to compare resistance than resilience between sites. Secondly, the MAPE’s show regional spatial patterns. For example, the resilience measures of the sites near the equator all show large MAPE’s, which makes comparison of the resilience measures of these sites more error prone.

However, if we restrict the sites to MAPE’s < 30 to assure a fair comparison between ESMs, it is clear that
the LC types differ in ecosystem stability (Fig. 6). Forests, for example, are more resistant (i.e., large percentage of small anomalies and normalized anomalies in Fig. 6a, b) than other natural LC types, such as woody savanna, savanna, and grassland, that show larger anomalies. Only barren or sparsely vegetated cover show smaller anomalies than forests. Figure 6a, b also illustrates that LC types with large intra- and interannual variability, such as grasslands or croplands, that are managed by crop rotations and irrigation, show lower resistance. This large interannual variability is also apparent when spatially comparing the difference measures for 2003–2004 with a maximum MAPE of 30 (Fig. 7): in 2003 large negative anomalies can be observed in the croplands and cropland-natural mosaic of Europe, reaching negative NDVI anomalies up to −0.10, whereas in 2004, the anomalies are mostly positive or slightly negative, i.e., most are larger than −0.02. These anomalies are in general much lower for the forest LC types.

Comparison of the $1 - \rho_1$ resilience metric shows that forests are again the most stable land-cover type. $1 - \rho_1$ approximates one, indicating high resilience and anomalies that approximate white noise (Fig. 6c). This is not the case for the other natural vegetation types (i.e., shrubland, (woody) savanna, and grassland), which have lower resilience, and the non-natural LC types (i.e., cropland (and natural mosaic) and bare soil), which show an intermediate resilience. Restricting data based on their MAPE influences the results significantly. Generally, the resilience of LCT with higher MAPE’s is lower. Furthermore, spatial differences in resilience between sites can be observed as sites in central USA, Mediterranean Europe, Mongolia, Australia, southern Africa, and Argentina show lower resilience compared with other sites (Fig. 5a, b). Also sites with higher resilience seem to have a lower reliability, which is also apparent for the spatial distribution of the different noise components (Fig. 5e–g).

Fig. 5 Spatial distribution of the stability metrics and errors, i.e., (a) resilience ($1 - \rho_1$); (b) resilience (spectral entropy); (c) variance (standard deviation); (d) variance (coefficient of variation); (e) the spatial distribution of white noise; (f) the spatial distribution of biased noise; (g) the spatial distribution of missing values. The color of the points represents the stability of the site, while their size indicates the Mean Absolute Percentage Error (MAPE) of the stability metric; (h) the spatial distribution of trends with their significance level. The slope of the trends is given by the color of the points and the significance level by their size.
Although the variation within the LC types is large, the standard deviation of forests and cropland/natural mosaic shows to be smaller compared with savanna, grassland, and cropland (Fig. 6d). Again, the restriction of the MAPE values influences the results. When only time series with a MAPE larger than 50 are taken into account, the standard deviation of the ecosystems is generally smaller compared to time series with a MAPE smaller than 30.

Discussion

A framework for ESM reliability

This study provides a framework to assess the reliability of ESMs before comparing and interpreting them. The framework is based on the sensitivity of ESMs in function of data and noise characteristics of time series of ecosystem properties. The high sensitivity of the ESMs to data and noise characteristics, stresses the importance of implementing the proposed approach before interpreting or comparing ESMs.

In the first place, our framework provides a methodology to interpret the ESMs of any time series within confidence limits dependent on the type of metric and the measured data. As such, it allows comparing ESMs measured at different locations that account for their reliability. For the NDVI subset data, for example, it is evident that the stability of forests with higher probability of snow/ice or regions with high cloud probability (e.g. tropical forests or regions located along coastlines) shows higher uncertainties and MAPE’s than regions with little noise or missing values.

Second, it provides a valuable tool to optimize the preprocessing of time series in function of the sensitivity. In the presented study on NDVI, for example, it is clear that biased observations have larger impact on the MAPE than missing values. Therefore, it suggests the replacement of biased values by missing values, as this reduces the uncertainty or MAPE for one observation with approximately 95, 60, and 10 for the variance, resilience, and resistance metrics, respectively.

Third, we demonstrate the strong effects of trends on the MAPE’s of the stability measures. Therefore, it is important to be aware of trends in the data before interpreting any stability measure as a false trend not related to ecosystem properties (e.g., due to sensor drift) might falsely indicate low stability, whereas a real trend (e.g., biomass loss after continuous thinning) might correctly be associated with real changes in stability.
Stability of the major global ecosystems

The comparison of global land-cover types based on their ESMs while accounting for their reliability might help to interpret results of other studies for specific ecosystem types or locations. In this way, a more conscious assessment of ecosystem stability against environmental disturbances can be obtained.

The calculated MAPE’s in our study point at very high uncertainties in the calculated ESMs for the Amazon forest, which indicates that it is very difficult to draw accurate conclusions on the stability of vegetation indices such as NDVI and EVI over the Amazon, due to the high amount of noise in the data. This was also concluded by Samanta et al. (2010, 2012a,b) who illustrated that the greening up shown by Huete et al. (2006) and Saleska et al. (2007) was due to inclusion of atmosphere-corrupted data, such as data contaminated with aerosols, in their analysis. Nevertheless, if only sites with high reliability are included, forests are the most stable ecosystems with respect to resistance. Only barren or sparsely vegetated cover show smaller anomalies than forests, but this might be explained by the large fraction of soil which buffers the fluctuations of the vegetation present (Somers et al., 2011). The high resistance of forests is also apparent when looking at the vegetation response to the European heat wave of 2003 (Fink et al., 2004; Rebetez et al., 2006). Our results of the 2003 resistance metric are consistent with the work of Ciais et al. (2005), who estimated a 30% reduction in gross primary production in 2003, but with much larger anomalies for crops and grassland (Reichstein et al., 2007), and only a moderate response for broadleaf forest (Lobo & Maisongrande, 2006). The NDVI anomalies over Europe in 2004 also show that most of the sites recovered 1 year after the drought, which was confirmed by Gobron et al. (2005). However, the results are also dependent on data quality. For example, the resistance of croplands is larger in case data with a MAPE smaller than 30 are selected, compared to the results using data points with a MAPE larger than 50.

Our results, further, indicate that the general trend in stability of the LCT differs for the resilience as well as for the variance metrics using different restrictions on the MAPE. In general, time series with a higher MAPE showed a higher resilience and lower standard deviation, which might be due to both the geographical location of high- vs. low-quality data as well as the quality of the data itself.
of the data. Higher levels of white noise, for example, might reduce the autocorrelation of the data and thus increase the resilience. The lower standard deviation of data with a higher MAPE might be explained by the geographical location of these data. Data with a MAPE larger than 50 are more likely to be situated in northern regions, tropical regions or deserts, areas with a lower standard deviation of their anomalies. Furthermore, the trends in stability of LCT for data with a MAPE smaller than 30 are less pronounced compared with data with a MAPE larger than 50.

To conclude, we have proposed a framework to assess the reliability of ESMs through quantifying the sensitivity of the resistance, resilience, and variance metrics to data and noise characteristics. With the application of this framework to 2001–2006 global subsets of NDVI data, we have proposed a methodology to interpret ESMs within confidence limits, and to optimize the preprocessing of the data to avoid spurious observations. Moreover, we have clearly shown the importance of reliability estimates to put the interpretation and comparison of ecosystem stability measures into context, contributing to a reliable future monitoring of ecosystem stability against global climate change and its increased frequency of disturbances.

Acknowledgements

We would like to thank the reviewers for their valuable comments and suggestions, which improved the quality of our manuscript. Funding support for this project has been provided by KU Leuven. S. Lhermitte was supported as postdoctoral researcher for Fonds Wetenschappelijk Onderzoek – Vlaanderen and the contribution of Ben Somers was supported by the Belgian Science Policy Office in the framework of the STEREO II Programme – Project REMEDY (SR/67/164).

References


Supporting Information

Additional Supporting Information may be found in the online version of this article:

The resilience metrics, i.e. the autocorrelation at lag 1, normalised spectral entropy and the spectral scaling component, are explained more thoroughly based example time series of a stable and unstable ecosystem.

Data S1. Understanding the stability metrics.

Figure S1. Example of the stability metrics: (a) The time series (ts) of a stable ecosystem X (red dashed line), the clima-tology (tsc), black) and the anomalies (tsa, red); (b) The time series of a stable ecosystem Y (red dashed line), the clima-tology (black) and the anomalies (blue); (c) The histogram of the anomalies of time series X (red) and Y (blue); (d) The sample autocorrelation of time series X (red) and Y (blue), (e) The spectrum upon the frequency and the logarithmic relationship between the spectrum and 1/frequency for time series X (red) and Y (blue).