

## ***Interactive comment on “Time series of vegetation indices and the modifiable temporal unit problem” by R. de Jong and S. de Bruin***

**Anonymous Referee #1**

Received and published: 13 October 2011

The authors correctly point out several possible pitfalls for those who fit time trends to time series of vegetation indices. In summary, the authors argue that time trends estimated using ordinary least squares can be deceiving if (1) the time series starts and ends at different phases of the seasonal cycle, (2) the time series is ‘binned’ at frequencies that do not correspond to an entire seasonal cycle, (3) the time series are aggregated over periods longer than one growing season. While many of these pitfalls are correct, the authors fail to show that these difficulties plague previous results. Furthermore, the issues raised by the authors probably are not the greatest problem associated with efforts to fit time trends to time series of vegetation indices. I do have some questions regarding the authors’ results. One issue concerns the statistical significance of results. I assume that the authors use t-tests that are implied by the text on lines 17–20 on page 8549. But the results of the t-tests are not reported in any of the

C3565

tables. For example, do all of the results reported in Table 1 reject the null hypothesis that the coefficient associated with time is statistically different from zero? And if so, at what significance level?

Another issue concerns the degree to which the results are robust. The authors analyze a function that has a peak-to-peak amplitude that is consistent with a temperate non-forest environment. If the amplitude is increased or decreased, what happens to the results? My guess is that the importance of the distortions reported by the authors declines as the amplitude is decreased. Ad absurdum, a seasonal cycle with no amplitude would not generate the types of distortions described by the authors. So, the authors may want to identify the amplitude at which these distortions appear and the type of ecosystem that is consistent with this ‘threshold amplitude.’

Furthermore, I am not sure about the relevance of some of the authors’ cautions. To date, most analyses of vegetation indices focus on mid and high latitudes (in part because cloud cover limits observations at low latitudes). At these mid and high latitudes, there is only one growing season per year, and so concerns about time series with different start and end dates are largely moot. As such, this largely alleviates the concern expressed by the authors on page 8551, lines 11–12 “but in reality calendar years may not fit the periodicity of VI time series because of shifts in vegetation phenology and variations in growing season length. In the Northern Hemisphere, the growing season at mid and high latitudes fits within a single growing season.”

Indeed, I think the paper would gain added relevance if the authors could identify some analyses that they think suffer from the distortions, which they describe. For example, the Zhou et al (2001) paper referenced by the authors (the earliest paper and hence most likely to suffer from such problems) analyzes trends over time in which NDVI for single growing seasons or a specific month are dependent variables. Although I am not an expert in this area, I don’t know of any published analyses in which the authors simply fit a time trend to a continuous series of monthly or biweekly values.

C3566

Finally, I think that the authors miss the most important issue associated with fitting time trends to time series of vegetation indices. These indices partially reflect biomass. As such, an increase in biomass persists from one period to the next. Similarly, a loss of biomass, such as a tree fall, persists from one period to the next. As such, it may be reasonable to represent the vegetation index using the following data generating process:

$$VI_t = VI_{t-1} + \mu_t$$

In which  $VI$  is the vegetation index and  $\mu$  is a disturbance to vegetation. As such, the vegetation index contains a stochastic trend. This implies that the time series contain a unit root or near unit root. Even if  $\mu$  is iid, the  $VI$  time series will meander around its initial value.

Under these conditions, fitting a time trend to the  $VI$  time series (e.g.  $VI_t = \alpha + \beta \text{Time} + \varepsilon_t$ ) will generate very misleading results. Specifically, Monte Carlo simulations indicate that the coefficient associated with Time ( $\beta$ ) will be statistically different from zero about 80 percent of the time if the data generating process that creates  $VI$  looks like equation 1. Such results are termed 'spurious regressions'. As such, spurious regression results may be a more important pitfall for efforts to fit time trends to vegetation indices.

---

Interactive comment on Biogeosciences Discuss., 8, 8545, 2011.