Biogeosciences Discuss., 12, C7328–C7333, 2015
www.biogeosciences-discuss.net/12/C7328/2015/

Biogeosciences
Discussions
Open Access

Interactive comment on "Modelling anomalies in the spring and autumn land surface phenology of the European forest" *by* V. F. Rodriguez-Galiano et al.

Anonymous Referee #2

Received and published: 3 November 2015

General comments

1. This ms offers an account of using random forests to characterize land surface phenology (LSP) anomalies over "the European Forest". Unfortunately, the authors do not go further to define the spatial domain of their analysis and there are no maps either in the main article nor in the supplement. This lack of spatial information about the domain is disconcerting, but not as disturbing as the complete lack of discussion or visual display of spatial residuals for a method that the authors repeatedly tell us is able to cope with spatial nonstationarity. When describing the modeling phenomena in space and time, it is incumbent upon the authors to show, not just tell, the readers about model performance. Aspatial, atemporal goodness of fit metrics, like the pseudo coef-

ficient of determination tells us about just one aspect of the model – its overall ability to explain the appearances. But how well it explains in a particular location or specific period will depend on the selected random forest models. That we should expect the model fits equally well everywhere would be to overlook the purported strength of the modeling approach – its flexibility. That the authors need to map out the model results and analyze the spatial and temporal patterns of residuals; otherwise the reader is left to conclude that the model approach is not as good as the hype.

2. The authors have chosen to focus on modeling LSP anomalies rather than the LSPs directly. OK, that is fine to explore potential drivers of LSP variation and change, but the authors have chosen to produce the anomalies in a non-standard way that shrouds the underlying linear decomposition idea of anomalies. The authors state the following on 11839 l25f: "The value of the targeted year was excluded in the computation to enhance the interannual variation." While the authors cite Saleska et al. 2007 as precedent, that doesn't mean such exclusion was a good idea. First, the long-term (here 11 years) average is now contingent depending on year, so there are 11 long-term averages, each based on 10 years. Furthermore, if the targeted year was excluded in the calculation of the average, then it should have also been excluded in the calculation of the standard deviation, so now there are also several standard deviations. Whether 10 or 11, there are too few years to get a stable estimation of the standard deviation and thus the significance of the various anomalies are called into question. Finally, it is indeed difficult to reconstitute the observational field from the anomaly field due to the exclusion of the targeted years. A simple temporal median and its deviation would be a better approach to maintain decomposability and transparency.

3. Perhaps more important to the intended message of this ms is that modeling LSP anomalies is an activity distinct from modeling LSPs. What is the object of the modeling? The former explains the appearances to gain insight in what kinds of things drive deviations from a simple temporal average. The latter aims both to explain the past temporal patterns and to provide a means for forecasting LSP. This former object is

what the authors try to do by sifting through variable importance. But note, the authors do not explore in this ms the spatially explicit pattern of variables. Instead, all that is provided are global evaluations of the relative importance of each variable to explaining the overall variation of the dataset. Most of the literature on LSP seeks to do the latter and thus it is misleading or at least inaccurate to cite papers doing the latter as inadequately doing the former. Different objectives requires different methods.

4. While the authors' enthusiasm for RF is understandable, their line of argument frequently devolves into boosterism rather than a serious comparison of linear methods to modeling LSP or LSP anomalies. As an apparent afterthought, in the Discussion section starting on p 11847 l 12f, there is a mention that the authors also did multivariate linear regression on the same variables and achieved coefficients of determination of 0.36 and 0.26 for spring and fall anomalies across the spatial domain. Unfortunately the details of what they did are too few and there are no direct comparisons spatially or temporally so it is hard to know just how poorly a "typical" modeling approach worked. The comparison of the predicted versus observed phenology anomalies for RF and the ill-described linear regression models relies on highly leveraged coefficients of determination rather than RMSEs and biases.

5. Certainly there is not sufficient evidence presented in the ms to conclude as stated in the end of the abstract: "This research, thus, shows clearly the inadequacy of the hitherto applied linear regression approaches for modeling LSP and paves the way for a new set of scientific investigation based on machine learning methods." Hardly! If the authors want to achieve that end, then they need to construct a very different kind of inter-comparison study.

Specific comments

P 11835 l 13 what was the precision of the spring index method used in the Schwartz et al. 2006 paper? What was the precision of the extended Spring Indices published in papers earlier this year by Schwartz and his colleagues? The SI and SI-x use daily

meteorological data as input.

p 11835 l 12 are these relative errors referring to the day of the year?

P 11836 l 4 no, not to reflect all land covers, which are a human concept, but rather that remote sensing of phenology relies fundamentally on reflected shortwave radiation upwelling from mixtures of things at the surface. Moreover there are no well-defined phenophases for land covers as there are for specific species, and we have well defined phenophases for on a fraction of the dominant or economically important plant species.

P 11836 l 12 "relatively unstudied" is hardly accurate. There is a robust line of papers looking at climate mode influence on LSPs. Some older examples include: Tucker et al. 2001 IJBM, Buermann et al. 2003 JGR, Gong & Shi 2003 IJRS, Potter et al. 2003 GCB, Zhang et al. 2004 GCB, Cook et al. 2005 GCB, de Beurs & Henebry 2008 JClim. And the field phenology community has long known this as well, e.g., Post & Stenseth 1999 Ecology, Menzel et al. 2005 GCB.

P 11837 l 15f testing for a legacy effect in the LSPs is an interesting idea, but scaling issues would suggest that any signal would be swamped.

P 11838 l 17 by focusing on "climate-driven anomalies in phenology" changes in LSPs that arise from disturbances such as fire or flooding and changes in land cover or land use or harvesting in agroforestry plantations will all be overlooked and simple contribute to the noise in the system.

P 11838 l 21 "monthly average values"!? Below it is specified that daily data were used. Which is it?

P 11838 l 25f "rather than fixed calendar dates" this is a strawman argument. Who uses "fixed calendar dates"?

P 11839 what are the units of measurement in the DAL and SIS?

P 11839 l 20-21 "using the methodology described in Dash et al. (2010)." Succinctly describe the method.

P 11839 l 25 “long-term mean” is a stretch for 11 years of data. How about using “multi-year mean”, if the suggested approach of the median is not adopted.

P 11840 l 2 resampling the 1km and 0.05 degree data to 0.25 degrees will greatly degrade the LSP signal. Perhaps a map of the residuals from the median will help to understand the spatial performance of the RFs.

P 11840 l 4 change “with less than 50” to “with fewer than 50” since the quantity is countable.

P 11840 eschew “Julian date” and use “day of year (DOY)” instead.

P 11840 l 17f Clarify: “Relative differences of the climatologies. . ..” What climatologies?

P 11840 l 21f put this information with units into a table

P 11841 l 13 “and may vary spatially” but not temporally?

P 11842 15 decode “oob”

P 11842 l 17 “from 1 to 9” but figures 2 & 3 show 12 and 14 predictors

P 11842 l 25-26 “rather than human-imposed temporal scales” but assigning 30 or 90 day windows is a “human-imposed” scale, it is not? , P 11844 l 3 “absolute fixed dates of the calendar months” there is that strawman argument again!

P 11845 l 16 “climatic predictors” the predictors used relate to weather, not climate.

P 11845 l 25 “improving the temporal matching between LSP anomalies and the preceding climatic anomalies” the ms does not show this.

P 11846 l 3 “a consistent divergent effect was observed between spring and autumn” what is this? Not clear.

P 11846 ll10-12 that spring is in the mid to high latitudes is driven by the timing of

the thaw is not new information, nor is the observation that the arrival of autumn is a complicated, contingent process depending on the growing season’s trajectory coming into the fall season.

P 11846 l 29 “our results support this hypothesis” what hypothesis exactly? Restate it here succinctly.

Figure 1 has 3 typographical errors in the spelling of anomalies and explanative should be explanatory. Where is the legend to decode the significance of the shapes?

Figure 2 what is the scale of “relative error”? should it read 0-200%?

Figure 3 has several shortcomings. First is the poor contrast in pseudo-r2. Move these important numbers to the top of the bars. (Note that r2 is called the coefficient of determination, not the “square correlation coefficient”.) Second, it is not clear why there are 12 and 14 climatic drivers shown when the pseudo-r2 is stable after 8 or fewer. How to determine when the model is sufficiently rich?

Figure 4 is a classic case of a highly leveraged regressions. Not one of these patterns show good performance. What are the biases and RMSEs of these models?

Interactive comment on Biogeosciences Discuss., 12, 11833, 2015.