Responses to Anonymous Referee #1

Xu and coauthors used two different NDVI phenology metrics to determent the dynamics of spring green-up dates, and studied the correlation between green-up and preseason T and precipitations. Large-scale phenology study has been a hot pot in the global change ecology study, and comparison studies in the RS-based phenology dates have been investigated several times recent years, but did not find consistent results that single method could be perfectly used to extract phenology data from the RS data series and therefore multiple methods that were used to extract phenology transition dates were recommended. This study focused on the RS based phenology and try to understand the difference in the RS-based phenology dates between two NDVI metrics, it would be a good contribution to understand the RS phenology, but I have several major comments that hopefully can help to improve this analysis.

Authors: We sincerely thank the reviewer for constructive criticisms and valuable comments. We have edited and rewritten parts of this manuscript. Our responses and detailed edits are indicated below in the point-to-point responses.

1) In this study, the authors focused on the AVHRR and MODIS, and found that, over the period 2001-2013, difference in magnitude and sign in spring phenology dates between these two dataset. Even, over the long period, i.e. 1983-2014 using AVHRR, globally delayed spring phenology dates were reported, which is different from the in situ data, as well

as many regional phenology studies. Considering large variation in spring phenology, a 10yrs trends may holds large uncertainty in trend analysis. Furthermore, only one single methods, i.e. piecewise logistic method, might be also generate large uncertainty in the green-up extraction. Therefore, I would suggest applying multiple methods to extract green-up dates. Since large uncertainty in the gridded climate data, validated study using another climate dataset would be suggested, and the results could be put in the appendix.

Authors: This study was inspired by the different spring phenology trend between AVHRR and MODIS, and between AVHRR and in situ observation. In 1980s and 1990s, AVHRR showed a consistent advanced trend with the in-situ observations, even though the advanced magnitude differs. The main difference was found after 2000, which is the period overlapping with MODIS product for comparison. We agree with the reviewer that a trend analysis over a 13 years is a bit short, however, we compared the estimated mean spring phenology and interannual variability between AVHRR and MODIS NDVI derived SGs. Furthermore, our analysis of spring phenology sensitivity to temperature and precipitation is the interannual response to climate.

We add a paragraph in section 2.4 to elucidate why we use a single method to smooth NDVI time series and derive greenup-date other than multi-methods. The common used regression methods to reconstruct NDVI time series and derive SG include Savitzky-Golay fitting method, spline smoothing, asymmetric Gaussian functions, double logistic function, and harmonic analysis of times series. These methods are valid in fitting NDVI gaps and reducing noise (Hird and McDermid, 2009; Cai et al. 2017), however, can make differences in estimating phonological stages (Cong et al., 2013). It is hard to justify which method is better. In order to reduce the mixed uncertainty of reconstruction methods and NDVI products, here we used one regression method to reconstruct the NDVI series. The double logistic method uses least-square fitting to half growing season (Zhang et al., 2003). It requires no smoothing parameter and is more robust than other methods in estimating the vegetation seasonal dynamics, when there is no local calibration (Cai et al., 2013).

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

Cai, Z., P. Jönsson, H. Jin, L. Eklundh (2017). Performance of smoothing method for reconstructing NDVItime-series and estimating vegetation phenology from MODIS data. Remote Sensing, 9, 1271, doi:10.3390/rs9121271.

Cong, N., T. Wang, H. Nan, Y. Ma, X. Wang, R. B. Myneni, S. Piao (2013). Changes in satellite-derived spring vegetation green-up date and its linkage to climate in China from 1982 to 2010: a multi-method analysis. Global Change Biology, 19, 881-891, doi:10.1111/gcb.12077.

Zhang, X., Friedl, M.A., Schaaf, C.B., Strahler, A.H., Hodges, J.C.F., Gao, F., Reed, B.C., Huete, A. (2003). Monitoring vegetation phenology using MODIS. Remote Sensing of Environment, 84, 471–475.

2) Wrong estimation in preseason and T/Precipitation sensitivity.

About the preseason issues, from the figure 3, very large difference preseaon-T were found between MODIS and AVHRR, but using same climate dataset and very similar green-up data, except Tibet and Polar regions, is it possible the large difference related to the statistics method? i.e. the DF is too small, i.e. 2001-2014, and could not be perfectly used to determine the preseason length. Or other climatic issues might affect the phenology process, and thus the effect of precipitation and radiation should be excluded from the preseason estimation. Anyway, the large difference in the preseason estimation is wired, and it would be substantially affect the estimation in the sensitivity estimation.

Authors: We checked our analysis of preseason length for both MODIS and GIMMS and added more discussion about this discrepancy in section 4.2. We calculated correlations between the time of SG and mean temperature in a period proceeding mean SG from 15 to 120 days with an increment of 3 days and identify preseason as the period (15-120) in which mean temperature is best correlated with SG. Although the mean SG patterns look similar for MODIS and GIMMS (Figure 1a and 1b), the difference can be larger than 20 days in some regions (Figure S1b). Therefore, the difference in preseason length for temperature is not only propagated from the differences in mean SG but also the interannual variability of SG. The preseason length inferred from MODIS (41±31days) is very close to the preseason length inferred from GIMMS in an earlier longer period over 1982-2005 (43±30 days). The consistent preseason length inferred from MODIS and GIMMS over two different period, and the stronger MODIS SG-temperature coupling makes it more confident to use MODIS NDVI in the available period and GIMMS NDVI data in the earlier period.

In the temperature/precipitation sensitivity of phenology, only significant relationships were recorded and mapped, but the percentage should also be provided. From the results and figure4, it seems quiet large percentage pixels were removed. if 90% were insignificant and removed, then a mean values across the 10% in T-sensitivity would be nonsense. I would suggest providing all data, both insignificant and sig correlations, and calculating the mean values and providing the percentage of sigs.

Authors: We replaced the Figure 4 to show all the pixels where the valid preseason and preseason climate-SG correlation were calculated. We marked the pixels with 90% confidence level with black dots. The percentages of sensitivity at 90% confidence level are given in section 3.3. About 43% of the land fraction shows significant sensitivity of D_{SG}^{M} to T_m ($p < 0.1$) compared with 13% of the land fraction with significant sensitivity of D_{SG}^{G} to T_m. About 11% of the land fraction shows significant sensitivity of D_{SG}^{M} to P_c ($p < 0.1$) as compared with 3% of the land fraction with significant sensitivity of D_{SG}^{G} to P_c. We add the mean sensitivity of SG to temperature and precipitation in section 3.3 for all the pixels. For the biome-scale sensitivity of SG to temperature, we found that filtering with ($p<0.1$) criteria tends to exclude pixels with lower sensitivity. But the stronger sensitivity inferred by MODS SG than that inferred by GIMMS SG remains with/without p-value filtering. The biome-scale sensitivity of SG to temperature for MODIS versus GIMMS for all p-value and p-value<0.1 are plotted below (Figure R1).

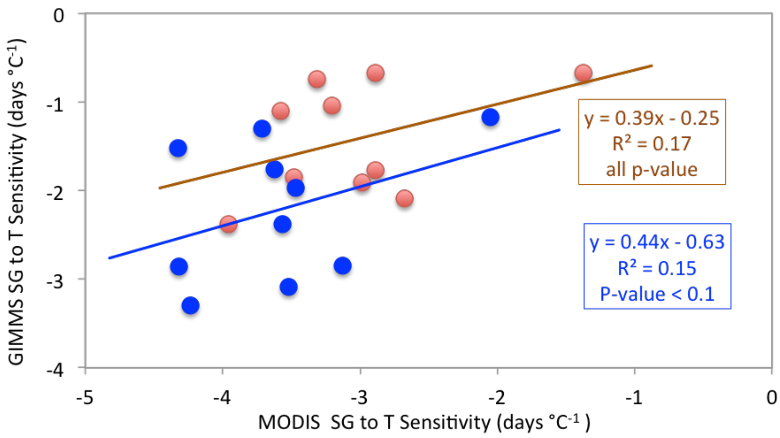

Figure R1 The biome-scale sensitivity of SG to temperature for MODIS versus GIMMS

3) The reference should be provided in many arguments, such as L37, what's means of several changes? Need references; L114, need references, also why moisture? Generally by T and photoperiod, rather air or soil moisture.

Authors: We rephrased these sentences and references are amended.

We agree with the reviewer that temperature and photoperiod are very important controller of phonological cycle in temperate and boreal vegetation. In the regions with permanent or periodic water stress, water availability mediates the phonological cycle. For example, Delayed spring phenology in response to increased air temperature has been reported in East Asia semiarid regions due to reduced precipitation (Shen et al., 2011; Yu et al., 2003). In a Mediterranean forest and in a mediterranean shrubland, rainfall pattern plays an important role in regulating the phonological change (Peñuelas et al., 2004).

Shen, M., Y. Tang, J. Chen, X. Zhu, Y. Zheng (2011). Influences of temperature and precipitation before the growing season on spring phenology in grasslands of the central and eastern Qinghai-Tibetan Plateau. Agricultural and Forest Meteorology, 151, 1711–1722. https://doi.org/10.1016/j.agrformet.2011.07.003

Yu, F., Price, K. P., Ellis, J., & Shi, P. (2003). Response of seasonal vegetation development to climatic variations in eastern central Asia. Remote Sensing of Environment, 87, 42–54. https://doi.org/10.1016/S0034-4257(03)00144-5

Peñuelas, J., I. Filella, X. Zhang, L. Llorens, R. Ogaya, F. Lloret, P. Comas, M. Estiarte, J. Terradas (2004). Complex spatiotemporal phonological shifts as a responses to rainfall changes. New Phytologist, 161, 837-846.