The study evaluated the plant phenology simulated by CRESCENDO land surface models using satellite observational LAI products. Specially, the 4GST method was applied to extract the times of start and end of growing season based on the simulated and remote sensing monthly LAI values. Then, the growing season types, variability of growing season start and end, latitudinal variability, and reginal variability were com- pared between the model simulations and satellite observations. Recommendations were also given for future model improvements. In general, the manuscript was written well, organized well, and the results were summarized clearly and interesting. So, I think the manuscript can be accepted for publication on the journal.

We thank the reviewer for her/his positive feedbacks and useful comments.

Only one main remark is that the description of the phenology schemes of the models. As we know, the phenology schemes in the models are quite different, in terms of their parameterizations of solar radiation, day-length, temperature, and soil moisture conditions. In section 2.2, the description of phenology schemes makes me a little bit hard to follow the differences among these models. So, I encourage the authors to summarize the similarities and differences of the processes of the schemes, according to some standards such as how to parameterize the effects of soil moisture, how to parameterize the effects of soil temperature etc. This summary will help us understanding the differences of the model and simulated results more clearly (e.g., Page 12, 349).

We thank the reviewer. In the revised version of the manuscript we will summarize and re-organize the section 2.2 in order to make the phenology scheme description of each land surface model more concise and easier to compare. Besides, we update table 1 to contain further details on the different phenology parameterization adding information on root zone depth, temperature and moisture thresholds, and variables used in detecting the phenology phases:

Table 1. Grid spatial resolution used for each land surface model (LSM) and brief summary of their main features. PFT stands for Plant Functional Type and CFT stands for Crop Functional Type. The second part gives further details on the phenology schemes.

LSM	Original Resolution	PFT	Soil level	CFT	Phenology scheme	Phenology drivers	Root zone
CLM 4.5	1.25° x 0.9375°	15	15	1 C3	evergreen; seasonal-deciduous; stress-deciduous	Soil temperature; soil moisture; day-length	*Zeng* (2001)
CLM 5.0	0.5° x 0.5°	16	20	2 C3	evergreen; seasonal-deciduous; stress-deciduous	Soil temperature; moisture day-length; precipitation	*Jackson et al.* (1996)
JULES-ES	1.875° x 1.25°	13	4	1 C3, 1 C4	Deciduous trees	Surface temperature	*Wiltshire et al.* (2020a)
JSBACH	1.9° x 1.9°	12	5	1 C3, 1 C4	evergreen; summergreen; raingreen; grasses; tropical crops; extra-tropical crops	air temperature; soil temperature; soil moisture; NPP	*Kleidon* (2004)
LPJ-GUESS	0.5° x 0.5°	25	2	3 C3, 2 C4	evergreen; seasonal-deciduous; stress-deciduous	Soil temperature; soil moisture	Root in top soil layer†: Herbaceous PFTs 90%, Woody PFTs 60%
ORCHIDEE	0.5° x 0.5°	15	11	1 C3, 1 C4	deciduous; dry and semi-arid; grasses and crops grasses and crops	Air temperature; soil moisture soil moisture	Exponential profile within the 2m soil *Krinner et al.* (2005)
ISBA-CTRIP	1° x 1°	16	14	1 C3, 1 C4	leaf biomass	Leaf biomass	*Canadell et al.* (1996)

LSM	Temperature variable	Temperature threshold	Moisture variable	Moisture Threshold	Reference
CLM 4.5	Third soil layer* temperature	0 °C	Third soil layer* water potential	-2 MPa	*Oleson et al.* (2013)
CLM 5.0	Third soil layer** temperature	0 °C	Third soil layer** water potential	-0.6 MPa	*Lawrence et al.* (2018)
JULES-ES	mean daily surface temperature	5 °C	None	N/A	*Clark et al.* (2011)
JSBACH	depending on the phenology: air, pseudo-soil temperature	depending on the phenology: 4°C, 10°C	soil moisture in the root zone	wilting point of 0.35 m/m	*Mauritsen et al.* (2019)
LPJ-GUESS	Top soil layer† temperature	5 °C	water stress scalar (ω)	minimum of ω ($\omega_{\min}$)	*Smith et al.* (2014)
ORCHIDEE	mean daily air temperature	Sum above -5 °C	soil moisture in root zone	Increase of moisture over 5 days	*Botta et al.* (2000)
ISBA-CTRIP	no phenology model: LAI is deduced from leaf biomass through Specific Leaf Area, which depends on nitrogen content				*Delire et al.* (2020)

* about 6cm depth. ** 9cm depth. † 25cm depth.

Meanwhile, in the results section, more direct comparisons among the model simulations should be made towards the differences of processes.

We increase the discussion and reference to the difference among models in the result sections of the revised version of the manuscript, especially in sections 3.2, 3.3, 3.4, and 3.5.

In addition, the comparisons were based on monthly LAI values (Page 4, line 89 and Page 9 Line 268). However, the temporal scale may cover up the real phenology characteristics. For example, based on the 8-day LAI data, Zhang et al., (2019) demonstrated that the CLM simulated growing season type is TGS in a temperate grassland, but the MODIS LAI-based type was SGS-S. It seems like that this discrepancy was not found in the study (Fig. 2b). There- fore, the monthly LAI mean output from the models may cause uncertainties on the model evaluation. Moreover, as mentioned by the authors (Page 16 Line 479), double cropping cropland can not be easily detected by the monthly LAI data, for example, a large area of winter wheat-summer maize double cropping system in the North China Plain was not detected by the method based on MODIS LAI (Fig 1 a). So, the uncertainty from the monthly LAI output from the models should be also discusses. I have no other remarks.

We thank the reviewer.
The use of monthly data is another limitation in our methodology, indeed.
Following the reviewer's suggestion, we add this point to the discussion in Section 4.3:

"Another limitation of the present evaluation is the monthly temporal frequency. Data at a higher frequency, indeed, might lead to a more detailed bias assessment. The use of a different temporal frequency may also influence phenology type detection. For example, Peano et al. (2019), that uses 15-day LAI data, detect a slightly different distribution of CLM4.5 SGS-D and TGS types in Australia, Horn of Africa, and Brazil. Similarly, Zhang et al. (2019), which analyses CLM4.5 in Northeast China with 8-day LAI data, obtain TGS type in areas recognized as SGS-S in the present analysis."