LETTER

Pine caterpillar occurrence modeling using satellite spring phenology and meteorological variables

Hao Hua1,2, Chaoyang Wu1,2,∗, Rachhpal S Jassal3, Jixia Huang4, Ronggao Liu1,2,5 and Yue Wang5

1 The Key Laboratory of Land Surface Pattern and Simulation, Institute of Geographical Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, People’s Republic of China
2 University of the Chinese Academy of Sciences, Beijing 100049, People’s Republic of China
3 University of British Columbia, 2357 Main Mall, Vancouver, BC V6T 1Z4, Canada
4 Precision Forestry Key Laboratory of Beijing, Beijing Forestry University, Beijing 100083, People’s Republic of China
5 Center for Biological Disaster Prevention and Control, National Forestry and Grassland Administration, Beijing 100013,
People’s Republic of China

∗ Author to whom any correspondence should be addressed.
E-mail: wucy@igsnrr.ac.cn

Keywords: day of insect occurrence (DIO), evapotranspiration, larval feeding, overwinter period, spring phenology

Abstract

Outbreaks of leaf-feeding Lepidopteran insects substantially weaken the quality of forest trees and strongly affect the ecosystem functions of plant photosynthesis and carbon uptake. The narrow phenological time window of leaf out about ten days, during which Lepidopteran larvae feed on high nutrient newly flushed leaves, may change the insect community and outbreak dynamics by determining the survival rate of larvae. The Chinese pine Caterpillar (Dendrolimus tabulaeformis Tsai et Liu) infestation of the northern Chinese pine (Pinus tabulaeformis) forest in China is a major concern, and accurately modeling the day of insect occurrence (DIO) in the spring remains challenging. With continuous in-situ observed insect activities of 20 plots and satellite and meteorological observations from 1983 to 2014, we found a strong synchronization \( r = 0.54, \ p = 0.001 \) between the satellite-based vegetation spring phenology, i.e. the green-up day (GUD), and DIO of the pine caterpillar over time. We used partial least squares regression and ridge regression models, and identified that monthly preseason air temperature, wind speed, specific humidity, and downward radiation were key environmental cues that awakened the overwintering pine caterpillars. After removing the collinearity of multiple variables, we showed that the dimensionality reduction-based regression models substantially improved the accuracy of DIO modeling than commonly used models, such as interval and degree-day models. In particular, including GUD significantly enhanced the predictive strength of the models increasing the coefficient of determination \( R^2 \) by 17.1% and consequently a decrease of 16.5% in the root mean square error. We further showed that evapotranspiration changed the environmental moisture content, which indirectly affected the activities of insects. Our results revealed a useful linkage between spring leaf development and insect occurrence, and therefore are of great importance for the large-scale monitoring of pest outbreaks with future remote sensing observations.

1. Introduction

Forest insects and related diseases reduce the quality and vitality of forest trees, increase ecological vulnerability, and are important drivers of global forest ecosystem change (Kurz et al 2008, Kautz et al 2018, Seidl et al 2018). Biotic disturbances cause extensive loss of trees of specific species within the landscape with serious consequences for tree biodiversity and carbon sink functions (Clark et al 2010, Bao et al 2020). Therefore, examining the mechanisms and conditions of forest insects and disease occurrence is necessary to conserve, manage, and develop forest resources.
The pine caterpillar (*Dendrolimus tabulaeformis Tsai et Liu*) has been the most serious destructive defoliator in northern China (Fan et al 2013, Zhang et al 2014). The infestation might occur each year and causes the trees to lose leaves, resulting in changes in the host trees’ physiology (Seidl et al 2017, Bao et al 2019). Once an outbreak occurs, the number of larvae per tree might exceed 400, which would have a significant effect on the tree's growth and even result in tree mortality (Fan et al 2013, Zhang et al 2018). The day of insect occurrence (DIO) refers to the timing when the pine caterpillar larvae first appear in the spring and travel up to trees for feeding, which occurs after overwintering in the thin layer around the roots of the host, under the rocks, the roots of the grass, and the covered pits. As the defensive compounds of young leaves tend to increase gradually early in the season after 30° days (Tikkkanen and Julkunen-Titto 2003, Falk et al 2018), Lepidopteran larvae might benefit from feeding on the leaves during a narrow time window before the leaves develop enough defensive compounds (Ekholm et al 2020), suggesting that larvae feeding with leaf flush are more likely to develop into outbreaks (Thomson et al 1984, Forkner et al 2008). The ability to predict DIO dynamics has important consequences for the management of destructive forest pests (Chen et al 2003), as well as the deployment of phenologically appropriate, timed insecticide treatments (Waldstein and Reissig 2001).

Temperature and photoperiod are the two major environmental factors that influence folivore and host tree phenology in the spring (Cannell and Smith 1983, van Asch and Visser 2007, Chuine et al 2016). Traditional researches analyze insect activities and population dynamics with ground meteorological data using the temperature-based degree-day (DDC) models (Mussey and Potter 1997, Wang et al 2013, Frank 2021). These models often use a baseline temperature above a particular threshold, below which no insect development is expected, and accordingly could provide a decent general description of larval occurrence (Buse and Good 1996, Bergant et al 2005). Due to the assumption of a simple linear relationship between temperature and insect developmental rate in DDC models, these models often tended to overestimate the development rate at high temperatures. As the synchrony of larval feeding with leaf flush exists, host plant phenology can reflect insect activity to some extent, although the underlying mechanism may be different for larvae and their host plants (van Asch and Visser 2007, Forkner et al 2008, Ekholm et al 2020). Regionally, larval feeding is often accompanied by the expansion of *Populus × Canadensis* buds and *Salix matsudana* buds, and the opening of *Salix babylonica* buds. However, the regional long-term records of host plant phenology (i.e. budburst, leaf expansion) are rare and hard to obtain.

During the past few decades, the increased use of remote sensing has greatly expanded the observation of traditional plant phenology (Richardson et al 2013, Keenan et al 2014, Fu et al 2015, Senf et al 2017). Landscape-scale vegetation spring phenology, that is, green-up date (GUD), depends on detecting the timing of phenological events in the temporal profile of greennis-related vegetation indices, such as the normalized difference vegetation index (NDVI) and the enhanced vegetation index (Piao et al 2019). But only in a study area with strong homogeneity, satellite-derived phenology can reflect to some extent the response of ecosystems to climate change. And it has the potential for insect occurrence modeling.

Based on 32 years of continuous field work in-situ observation data from fixed monitoring stations of pine caterpillars in Qingsongling, China from 1983 to 2014, our objectives are (a) to fill the knowledge gap of main drivers of insect activities (*Dendrolimus tabulaeformis Tsai et Liu*) in early spring in the study area, (b) to explore the synchrony of larval feeding with satellite-derived phenology and improve the predictive power of DIO models combining remote sensing (NDVI) with biotic variables (vegetation phenology), and finally (c) to explain the underlying mechanism of the synchrony between GUD and DIO.

### 2. Materials and methods

#### 2.1. Study area

The study area, Qingsongling, is located in eastern Jianping County of Liaoning Province, China (41.72°–41.86° N, 119.80°–119.99° E), which has a temperate continental monsoon climate (figure 1). The mean annual air temperature is approximately 6.9 °C (ranging from −21.3 °C to 28.2 °C in a year) and the total precipitation sum reaches approximately 440 mm (He et al 2012). The State Forestry Administration of China reports that, *Pinus tabuliformis* appeared on 327.8 km² in Qingsongling, accounting for 93.1% of the total area covered by forest and the rest is mixed forest, scrub and other vegetation. Therefore, *Pinus tabuliformis* is the dominant tree species in the study area.

#### 2.2. Insect activities data

We set up 20 plots (50 m × 40 m) in the study area, and selected 20 *Pinus tabulaeformis* trees each for observation according to the ‘Z’ method (figure 2(a)). We used the following strategies to arrange the plots: (a) as evenly and uniformly distributed and homogeneously as is physically possible; (b) there are no artificially constructed buildings or roads in the neighborhood; (c) there are trees that are similar in age, height, and planting distance. The plots were not at regular intervals, within about 2 km. Before the larvae traveled up the tree in early March, we used a special sticky plastic skirt to wrap around the tree’s breast...
Figure 1. The location of the study area in eastern Jianping County of Liaoning Province (b), China (a). (c) The red triangles represent the sampling points and the dotted line represent the boundary of the study area.

Figure 2. (a) Schematic diagram of the ‘Z’ method to sample 20 trees in plots (50 m × 40 m); (b) the synchronization of the satellite-derived phenology (GUD: green-up day; LSD: leaf senescence day) of the host plant extracted from normalized difference vegetation index (NDVI) and the development stages of the pine caterpillar.
height, and regularly checked and recorded the number of larvae in the plastic skirt every day until no larvae were detected on the tree for three consecutive days. The DIO within a plot was the date when the first larva has been observed on any tree. The site-wide DIO was defined when all plots have recorded at least one larva sighting. We calculated DIOs by setting different thresholds and the interannual variations between them were highly correlated (supplementary figure S7). DIO data was continuously recorded from 1983 to 2014.

2.3. Meteorological and land surface data
The China Meteorological Forcing Dataset (CMFD) is a gridded near-surface meteorological dataset specially developed for the study of land surface processes in China (http://data.tpdc.accn/en/data/8028b944-daaa-4511-8769-965612652c49/), with a spatial resolution of 0.1° (Yang and He 2019). We used 2 m air temperature (Temp), specific humidity (Shum), 10 m wind speed (Wind), downward shortwave radiation (Srad), and precipitation rate (Prec) as the pre-season predictor variables (Fu et al. 2015, Wang et al. 2019) based on possible effects on larvae (van Asch and Visser 2007) from 1983 to 2014.

As the larvae overwinter is under ground cover, land skin parameters should also be taken into consideration. We used the ECMWF Reanalysis v5 (ERA5)-Land reanalysis dataset that provides a consistent view of the evolution of land variables over several decades at an enhanced resolution compared to ERA5 (www.ecmwf.int/en/era5-land) (Muñoz-Sabater et al. 2021). These data was gridded to 0.1° × 0.1° with monthly mean averages provided. We used skin reservoir content and skin temperature to reflect the land skin moisture and temperature from 1983 to 2014.

We also used the eight day moderate resolution imaging spectroradiometer (MODIS) Evapotranspiration/Latent Heat Flux product (MOD16A2, 500 m, version 6) from 2001 to 2014 (Running et al. 2017, Li and Xiao 2019). To reduce study uncertainties, we evaluated the evapotranspiration (ET) from MODIS and ERA5-Land (supplementary figure S5) in the study area from 2001 to 2014, which signaled a strong positive correlation ($r = 0.95, p < 0.001$). As the ET from MODIS has a higher resolution and is independent of other predictors, we used the MODIS data to explore the underlying mechanism. We interpolated linearly it to a daily scale and evaluated the most optimal window size (WS) before and after GUD to explore the relationships between ET and DIO. We defined a WS ranging from 0 to 50 days, and the duration was WS days before and after GUD (GUD ± WS days) occurred.

2.4. Satellite NDVI dataset
NDVI is a proxy for vegetation greenness (Jeong et al. 2017). To coincide with the years of the site-based DIO, we used the NDVI GIMMS third-generation data (https://data.tpdc.ac.cn/en/data/9775f2b4-7370-4e5e-a537-3482c9a83d88/) (Tucker et al. 2005) from 1983 to 2014. With a 1/12° spatial resolution, the fortnightly NDVI dataset has long-term observational data since 1982. As the longest NDVI record, the GIMMS3g NDVI data have been extensively applied for spring phenological research (Jeong et al. 2011). We used seven NDVI pixels in the study area to extract vegetation phenology after area-weighted averaging.

2.5. Determination of spring phenology
Based on the time range of NDVI3g (1982–2015) and DIO (1983–2014), our study focused on the period of 1983–2014. After area averaging and data preprocessing, we first applied a modified Savitzky–Golay filter to exclude abnormal NDVI data and take the envelope (Shen et al. 2014, Cao et al. 2018). Then we applied a double logistic function (Elmore et al. 2012, Gonsamo et al. 2013) to fit the annual NDVI curve (supplementary figure S1):

$$f(t) = \alpha_1 + \frac{\alpha_2}{1 + e^{(\alpha_3 - t)/\alpha_4}} - \frac{\alpha_5}{1 + e^{(\alpha_6 - t)/\alpha_7}}$$  \hspace{1cm} (1)

where $t$ is time in days, $f(t)$ is the NDVI value at time $t$, and $\alpha_1, ..., \alpha_7$ are fitting parameters. Specifically, $\alpha_1$ is the background NDVI, represents the mean values of NDVI in the non-growing season of vegetation; $\alpha_2$ and $\alpha_3$ represent the difference between the background and the amplitude of the late summer and autumn plateau NDVI units; $\alpha_4$ and $\alpha_6$ are the midpoints of the transitions for green-up and senescence/abscission, respectively; $\alpha_5$ and $\alpha_7$ are the transitions curvature parameters (normalized slope coefficients). All seven optimal parameters in the study area were presented in the supplementary material (supplementary table S4). GUD was calculated as the time when the change rate of fitted NDVI reached its first local maximum value in spring (Gonsamo et al. 2013, Wang and Liu 2021).

2.6. Commonly used models to predict DIO
We applied two traditional models commonly used to predict DIO (Li 1959, Chai et al. 2002, Li et al. 2002, Zhou et al. 2012). The first was the interval method, which predicts the DIO based on the average interval of a multi-year overwintering period:

$$F = H_i + X$$  \hspace{1cm} (2)

where $F$ is the date of the current developmental stage starts, $H_i$ is the date of the previous developmental stage starts, $X$ is the average interval obtained from field observations. In this study, $F$ is DIO, and $H_i$, acting as an input, is the specific date when the larvae begin overwintering in previous year.

The second was the DDC model which follows the idea of accumulated forcing temperature (van Asch and Visser 2007, Liu et al. 2018). Using the baseline
temperature, accumulated forcing temperature, and the local daily temperature during growth of the pine caterpillar, the occurrence period was predicted as:

\[ DD(d) = \max\left(T_{\text{mean}}(d) - T_b, 0\right) \]  
\[ DD_{\text{threshold}} = \sum_{d=d_0}^{DIO} DD(d) \]

where \(DD(d)\) is the growing degree on date \(d\), \(T_b\) is the baseline temperature, \(T_{\text{mean}}(d)\) is the daily mean temperature on date \(d\), \(DD_{\text{threshold}}\) is the accumulated forcing temperature from \(d_0\) to DIO and \(d_0\) is the first day of accumulation, set as 1 January of the year. DIO was defined as the date that \(DD(d)\) first exceeded the multiyear mean \(DD_{\text{threshold}}\). When predicting DIO of overwintering larvae rather than egg hatching, the average value of the multi-year winter minimum temperature was used as the baseline in this study (van Asch and Visser 2007).

### 2.7. Statistical analysis strategy

To remove covariate effects of environmental variables on DIO, we applied a partial correlation analysis to determine the optimal length of the preseason for each variable in the study area (supplementary table S3), as the most relevant period for DIO ranged several months before it, and the length differed among environmental cues. The optimal preseason for each variable was defined as the period (with one month steps) for which the partial correlation coefficient between DIO and variables was highest during 1983–2014 (Fu et al 2015).

We used partial correlation analyses, excluding interference between variables, to examine the responses of several environmental cues to insect activities in terms of DIO (Guo et al 2021). We applied two regression models for key variables selection. We used the partial least squares regression (PLS-R) and ridge regression (Ridge-R) to solve the variable collinearity issue. For the PLS-R, the variable importance in projection (VIP) was applied (Tran et al 2014, Farrés et al 2015), with the absolute value representing the driver’s influence on DIO. The PLS-R model has a ‘greater than one’ criteria (Guo et al 2021), which states that a VIP value bigger than one shows that the driver significantly affects DIO. Ridge-R works directly with the predictors and adds a penalty related to the size of the coefficients to remove the variable collinearity (Hoerl and Kennard 1970) which is being increasingly adopted for variable selection and sensitivity analysis due to its reliable model performance (Graham 2003, Chen et al 2020). For Ridge-R, the sign of the model coefficient (MC) showed the positive or negative impacts of the driving factors and the \(t\)-test was used to calculate the significance.

When using the models, all predictor variables have been standardized and centered to remove the influence of units. Leave-one-out and K-fold cross-validation methods, built into the functions in R, were also used to ensure the credibility of the results. R packages ‘pls’ (Wehrens et al 2021) and ‘ridge’ (Cule and De Iorio 2013) were used for applying and testing the models (R Core Team 2020). All statistical significance was set at \(p < 0.05\).

A detailed data analysis procedure was presented in figure 3. Before performing the analysis, we resampled CMFD, ERA5-Land, and MODIS datasets into 1/12° to match GUD derived from NDVI3g data. And we used the same method to average them. We summarized all the datasets and models used in this study in the supplementary material (supplementary tables S1 and S2).

### 3. Results

#### 3.1. Responses of DIO to predictor variables

After eliminating the variable covariance, DIO showed significantly positive correlations with wind speed \((r = 0.57, p < 0.01)\) and GUD \((r = 0.45, p < 0.05)\). However, significantly negative correlations were observed for air temperature \((r = -0.42, p < 0.05)\), specific humidity \((r = -0.47, p < 0.05)\), and downward shortwave radiation \((r = -0.52, p < 0.01)\). We found that the precipitation, skin reservoir content, and skin temperature showed non-significant impacts on DIO (figure 4(a)).

We used the VIP of PLSR and MCs of Ridge-R to represent the contribution of each predictor to DIO. GUD and Wind also showed a positive effect on DIO, opposite to Temp, Shum, and Srad (figures 4(b) and (c)), which supported the partial correlation results. We also applied the ordinary least squares linear regression and similar results were obtained (supplementary figure S4). Comprehensive considering the results of significance test and the ‘greater than one’ criterion, we selected Green up day (GUD), air temperature (Temp), specific humidity (Shum), 10 m wind speed (Wind), and downward shortwave radiation (Srad) as key predictor variables for the modelling of DIO.

#### 3.2. DIO’s synchronization with green-up data

Figure 5(a) shows that DIO and GUD were synchronized with a significant positive correlation \((r = 0.54, p = 0.001)\). The pixel-based GUD and plots-based DIO shows similar synchronization (supplementary figure S6). The difference (in days) between the DIO and the GUD of the larvae’s specific host tree determined the degree of synchrony. The degree of synchrony can vary strongly between years with a mean value of 4.22 days (figure 5(b)).

#### 3.3. Predicting DIO by different models

Among the two commonly used prediction models, the DDC model outperformed the simple interval
method (figures 6(a) and (b)), but the explanatory power was still low ($R^2 = 0.20$, root mean square error (RMSE) = 6.99 days). The multivariate regression methods based on dimensionality reduction, when only considering environmental variables (Temp, Wind, Shum, Srad), showed a significantly stronger predictive strength than commonly used methods (figures 6(c) and (d)). PLS-R ($R^2 = 0.64$, RMSE = 3.33 days) performed similar with Ridge-R ($R^2 = 0.64$, RMSE = 3.39 days) in the term of RMSE. After combining both biotic and abiotic factors, GUD was added to PLS-R ($R^2 = 0.75$, RMSE = 2.78 days) and Ridge-R ($R^2 = 0.75$, RMSE = 2.80 days), and the models performed better than only considering environmental variables (figures 6(e) and (f)). Overall, GUD enhanced the predictive strength of the models increasing $R^2$ by 17.1% and consequently a decrease of 16.5% in RMSE.

### 3.4. Underlying mechanism including spring GUD for DIO modeling

To better understand the reasons for a better performance of modeling strategy with GUD, we further explored the possible changes over the spring leaf developmental stage that may influence caterpillar activities. We found a significant positive correlation between average ET and DIO with the WS between 18 and 32. At the optimal WS of GUD ± 26 days, ET showed the strongest impacts on DIO with $r = 0.55$ ($p < 0.05$) (figures 7(a) and (b)).

### 4. Discussion

#### 4.1. Complex mechanism of larvae overwintering and occurrence

Previous studies have shown that meteorological conditions are the driving factors for the activity
Figure 4. (a) Partial correlation coefficients (PCOR) between DIO and environmental variables. Ridge regression’s model coefficients (b) and partial least squares regression’s variable importance in projection (VIP) values (c) acted as the sensitivity of DIO to environmental variables. We used five near-surface environment variables and two land skin variables, including 2 m air temperature (Temp), specific humidity (Shum), 10 m wind speed (Wind), downward shortwave radiation (Srad), precipitation rate (Prec), skin reservoir content (Src) and skin temperature (Skt). Asterisks indicate level of significance ($p < 0.05$: ‘*’; $p < 0.01$: ‘**’; $p < 0.001$: ‘***’) of t-test.

dynamics of forest insects (Frank 2021), especially photoperiod and temperature (van Asch and Visser 2007). However, reasons for the relationships between wind speed and air humidity and larval emergence are limited. Wind speed generally obstructs larvae awakening as it can change local short-term weather conditions, either by directly blowing down tree larvae or by destroying their overwintering environment (He et al 2012).

As a result of the sluggish development and low survival of pine caterpillars caused by dehydration, ET was found to be substantially linked with DIO (Han et al 2008). The higher environmental moisture content is favorable to their survival and development (Wu et al 2021). Freeze-tolerant insects withstand internal ice formation and maintain a high supercooling point through the production of ice nucleators (Zeng et al 2008). The environmental moisture
content can therefore impact the water balance of insects both directly and indirectly with implications on larvae’s metabolic activities. ET is the sum of water evaporation and transpiration from a surface area (such as soil, canopy interception, water bodies, and plants) to the atmosphere (Wang and Dickinson 2012). In spring, plants need water for photosynthesis and transpiration pull promotes the absorption of soil water (Tang et al. 2021). The stronger the ET during GUD, the more local cooling and reduction of environmental moisture content will be caused, which is not conducive to the survival of insects and leading to a much later DIO.

4.2. Synchronization of vegetation phenology and insect activities

The synchronization of host plant phenology and insect activities has long been noticed by predecessors, as vegetation and insect phenology both are regulated by regional environmental characteristics (van Asch and Visser 2007, Forkner et al. 2008, Ekholm et al. 2020), although the underlying mechanism maybe different for herbivores and their host plants. The pattern of synchrony may vary under climate change (Schwartzberg et al. 2014), which may consequently poses challenges for our model development. Therefore, we suggest that combining this synchronization with environmental variables in the model will help reduce the uncertainty.

This synchronization also makes vegetation phenology, a more readily observable land surface variable, somewhat of a proxy for insect activity. Future vegetation phenology changes can now be accurately predicted with the development of numerous spring phenology models (Liu et al. 2018), such as the one-phase growing degree days, spring warming models and two-phase sequential, parallel models. Our analysis indicates that the spring phenology (GUD) will be a crucial variable for predicting DIO in addition to meteorological data. We therefore expect that future higher spatial-temporal resolution, and even ‘real-time’ GUD monitoring, will undoubtedly provide a finer description of the synchronization between plant phenology and insects activities, even limitations existed in the current phenology data and model developments. The predictive power of insect activity models will be substantially increased in future due to the advancements in mechanism interpretation and data quality control.

4.3. Scale issues in remote sensing

Numerous factors, such as the homogeneity of the study area landscape, which makes it more challenging to identify mechanisms than ground-based observation, hinder the proper use of remote sensing data in ecological studies (Wu and Li 2009). Because the homogeneity of the study area could amplify the characteristics of individual tree species, the phenology of Pinus tabuliformis can actually be detectable by satellites because of the seasonal needle drops (Wang et al. 2017), i.e. evergreen needleleaf forests do regularly shed older needles as newer needles fill in (Vilhar et al. 2013). The needle drop is mostly noticeable when several trees start to lose their needles at the same time (supplementary figure S3), which is a seasonal process and therefore is detectable from remote sensing observations.

Spatial mismatch is another concern for the proper use of remote sensing data. The GIMMS3g NDVI dataset has been widely used to monitor terrestrial vegetation change over the Northern terrestrial ecosystems (Piao et al. 2020). This dataset has the advantage of a very long-term observation from 1982 to 2015, and its biweekly observation can support the monitoring of vegetation cycle changes in a year (Piao et al. 2019). Since the first Landsat satellite
launched in 1972, Landsat data can also support long-term monitoring with a higher spatial resolution (Wulder et al 2019). However, the persistent presence of atmospheric contamination hinders the reconstruction of the vegetation index time series within one year. Therefore, the GIMMS3g NDVI dataset was a better choice to retrieve the land surface phenology of the study area after time series reconstruction. Future sensor performance with a higher accuracy to enable more accurate extraction of land surface phenology will be of great help to support modeling of various terrestrial variables including DIO.
5. Conclusion

While previous works on insect activity models mostly considered abiotic factors such as temperature and photoperiod, we attempted to model the DIO in spring using GUD, a biotic factor of the host phenology based on NDVI. Combined with remote sensing and reanalysis data, we screened the key environmental cues that affect insect occurrence and built predictive models for DIO. The dimensionality reduction-based regression methods greatly enhanced the explanatory power compared with commonly used models. We also found that biotic and abiotic factors have a joint effect on DIO, and satellite-derived spring phenology has great potential in predicting DIO, which can significantly enhance the predictive strength of the models increasing the coefficient of determination ($R^2$) by 17.1% and consequently a decrease of 16.5% in the RMSE. Environmental moisture content changes induced from ET before and after GUD may be a key factor in awakening the overwintering larvae. Our results are therefore of great significance for studying insect activities from a landscape-scale perspective and contribute to forestry management, and insect prevention modeling.

Data availability statement

The data generated and/or analysed during the current study are not publicly available for legal/ethical reasons but are available from the corresponding author on reasonable request.

Acknowledgments

This work was funded by the National Key R&D program of China (2019YFA0606603), the National Natural Science Foundation of China (42125101) and by the CAS Interdisciplinary Innovation Team (JCTD-2020-05).

ORCID iDs

Hao Hua https://orcid.org/0000-0002-1679-412X
Ronggao Liu https://orcid.org/0000-0003-2819-7178

References

Bao Y, Na L, Han A, Guna A, Wang F, Liu X, Zhang J, Wang C, Tong S and Bao Y 2020 Drought drives the pine caterpillars (Dendrolimus spp.) outbreaks and their prediction under different RCPs scenarios: a case study of Shandong Province, China For. Ecol. Manage. 475 118446
Bao Y, Wang F, Tong S, Na L, Han A, Zhang J, Bao Y, Han Y and Zhang Q 2019 Effect of drought on outbreaks of major forest pests, pine caterpillars (Dendrolimus spp.), in Shandong Province, China Forests 10 1–16
Bergant K, Trdan S, Žnidarič D, Crepinšek Z and Kajfež-Bogataj L 2005 Impact of climate change on developmental dynamics of Thrips tabaci (Thysanoptera: thripidae): can it be quantified? Environ. Entomol. 34 755–66
Buse A and Good J E G 1996 Synchronization of larval emergence in winter moth (Operophtera brumata L.) and budburst in pedunculate oak (Quercus robur L.) under simulated climate change Ecol. Entomol. 21 335–43
Cannell M G R and Smith R I 1983 Thermal time, chill days and prediction of budburst in Picea sitchensis J. Appl. Ecol. 20 951–63
Cao R, Chen Y, Shen M, Chen J, Zhou J, Wang C and Yang W 2018 A simple method to improve the quality of NDVI time-series data by integrating spatiotemporal information with the Savitzky–Golay filter Remote Sens. Environ. 217 244–57
Chai S, Xu G, Xie K, Lu N and Li G 2002 Forecast of occurrence stage and harm extent of dendrolimus houi lajonquière J. Northeast For. Univ. 17 54–7
Chen L, Hänninen H, Rossi S, Smith N G, Pau S, Liu Z, Feng G, Gao J and Liu J 2020 Leaf senescence exhibits stronger climatic responses during warm than during cold autumns Nat. Clim. Change 10 777–80
Chen Z, Clancy K M and Kolb T E 2003 Variation in budburst phenology of douglas-fir related to western spruce
budworm (Lepidoptera: Tortricidae) fitness J. Econ. Entomol. 96 377–87
Chuine I, Bonhomme M, Legave J-M, Garcia de Cortazar-atauri I, Charrier G, Lacointe A and Améglio T 2016 Can phenological models predict tree phenology accurately in the future? The unrevealed hurdle of endodormancy break Glob. Change Biol. 22 3444–60
Clark K L, Skowronski N and Horn J 2010 Invasive insects impact forest carbon dynamics Glob. Change Biol. 16 88–101
Cale E and De Iorio M 2013 Ridge regression in prediction problems: automatic choice of the ridge parameter Genet. Epidemiol. 37 704–14
Ekholm A, Tack A J M, Pulkkinen P and Roslin T 2020 Host plant phenology, insect outbreaks and herbivore communities—the importance of timing J. Anim. Ecol. 89 829–41
Elmore A J, Guinn S M, Minsley B J and Richardson A D 2012 Landscape controls on the timing of spring, autumn, and growing season length in mid-Atlantic forests Glob. Change Biol. 18 656–74
Falk M A, Lindroth R L, Keefer-Ring K and Raffa K F 2018 Genetic variation in aspen phytochemical patterns structures windows of opportunity for gypsy moth larvae Oecologia 187 471–82
Fan J, Xie Y, Xue J, Zhang Y, Yang Q and Bengoechea J 2013 Cellular apoptosis of hemocytes from dendrolimus tabulaeformis tsai et liu larvae induced with the secondary metabolites of Beauveria brongnartii (Sacc.) Petch PLoS One 8 e71600
Farré R, Platikanov S, Tsakovski S and Tauter R 2015 Comparison of the variable importance in projection (VIP) and of the selectivity ratio (SR) methods for variable selection and interpretation J. Chemom. 29 528–36
Forrner R E, Marquis R J, Lill J T and Le C J 2008 Timing is everything! Phenological synchrony and population variability in leaf-chewing herbivores of Quercus Ecol. Entomol. 33 276–85
Frank S D 2021 Review of the direct and indirect effects of warming and drought on scale insect pests of forest systems Forestry 94 167–80
Fu Y H et al 2013 Declining global warming effects on the phenology of spring leaf unfolding Nature 526 104–7
Gonsamo A, Chen J M and D’Odorico P 2013 Deriving land environmental humidity on the survival and development of the larva of Dendrolimus tabulaeformis Chin. J. Appl. Entomol. 1231–42
Graham M H 2003 Confronting multicollinearity in ecological multiple regression Ecology 84 2009–15
Guo M, Wu C, Peng J, Liu L and Li S 2021 Identifying contributions of climatic and atmospheric changes to autumn phenology over mid-high latitudes of Northern Hemisphere Glob. Planet. Change 197 103396
Han R-D, Parajulaee M, He Z and Ge F 2008 Effects of environmental humidity on the survival and development of pine caterpillars, Dendrolimus tabulaeformis (Lepidoptera: Lasiocampidae) Insect Sci. 15 147–52
He S-Y, Iao Y-Q, Wen J-B, Zhao Y-X and Zong S-X 2012 Influence of climate warming on overwintering behaviour of the larva of Dendrolimus tabulaeformis Chin. J. Appl. Entomol. 49 1231–42
Hoefl E A and Kennard R W 1970 Ridge regression: biased estimation for nonorthogonal problems Technometrics 12 55–67
Jeong S-J, Ho C-H, Gim H-J and Brown M E 2011 Phenology shifts at start vs. end of growing season in temperate vegetation over the Northern Hemisphere for the period 1982–2008 Glob. Change Biol. 17 2345–99
Jeong S-J, Schimel D, Frankenberg C, Drewry D T, Fisher J B, Verma M, Berry J A, Lee J-E and Joiner J 2017 Application of satellite solar-induced chlorophyll fluorescence to understanding large-scale variations in vegetation phenology and function over northern high latitude forests Remote Sens. Environ. 190 178–87
Kautz M, Anthone P, Meddens A J H, Pugh T A M and Arneth A 2018 Simulating the recent impacts of multiple biotic disturbances on forest carbon cycling across the United States Glob. Change Biol. 24 2079–92
Keenan T F et al 2014 Net carbon uptake has increased through warming-induced changes in temperate forest phenology Nat. Clim. Change 4 598–604
Kurz W A, Dyndom C C, Siminson G, Rampley G J, Neilson E T, Carroll A L, Etaba T and Sarsanyik L 2008 Mountain pine beetle and forest carbon feedback to climate change Nature 452 987–90
Li B 1959 Preliminary study of the life history of the oil pine caterpillar (Dendrolimus spectabilis Butl.) in Shantung Province Acta Entomol. Sin. 9 316–24
Li L I and Xiao A 2019 A global, 0.05-degree product of solar-induced chlorophyll fluorescence derived from OCO-2, MODIS, and reanalysis data Remote Sens. 11 2563
Li R, Chai S, Xu G, Duan Z, Xie K and Li G 2002 A preliminary study on forecast of the occurrence period of Dendrolimus kikuchii J. Southwest For. Univ. 29 49–52
Liu Q, Fu Y H, Liu Y, Janssens I A and Piao S 2018 Simulating the onset of spring vegetation growth across the Northern Hemisphere Glob. Change Biol. 24 1342–56
Muñoz-Saléter J et al 2021 ERAS-Land: a state-of-the-art global reanalysis dataset for land applications Earth Syst. Sci. Data 13 4349–83
Mussey G J and Potter D A 1997 Phenological correlations between flowering plants and activity of urban landscape pests in Kentucky J. Econ. Entomol. 90 1613–27
Piao S et al 2020 Characteristics, drivers and feedbacks of global greening Nat. Rev. Earth Environ. 1 14–27
Piao S, Liu Q, Chen A, Janssens I A, Fu Y, Dai J, Liu L, Lian X, Shen M and Zhu X 2019 Plant phenology and global climate change: current progresses and challenges Glob. Change Biol. 25 1922–40
R Core Team 2020 A language and environment for statistical computing R Found. Stat. Comput. 2 (available at: www.R-project.org)
Richardson A D, Keenan T F, Migliavacca M, Ryu Y, Sonnett O and Toomey M 2013 Climate change, phenology, and phenological control of vegetation feedbacks to the climate system Agric. For. Meteorol. 169 156–73
Running S, Mu Q, Zhao M and Moreno A 2017 MOD16A2 MODIS/terra net evapotranspiration 8-day L4 global 500m SIN grid V006 NASA EOSDIS L. Process. IAAC L5 3.54 (available at: https://scholarworks.umt.edu/nts_pubs/2680/https://scholarworks.umt.edu/nts_pubs/2680/)
Schwartzberg E G, Jamieson M A, Raffa K F, Reich P B, Montgomery R A and Lindroth R L 2014 Simulated climate warming alters phenological synchrony between an outbreak insect herbivore and host trees Oecologia 175 1041–9
Seidl R et al 2017 Forest disturbances under climate change Nat. Clim. Change 7 395–402
Seidl R, Klonner G, Rammer W, Eisl F, Moreno A, Neumann M and Dullinger S 2018 Invasive alien pests threaten the carbon stored in Europe’s forests Nat. Commun. 9 1–11
Sent C, Seidl R and Hostert P 2017 Remote sensing of forest insect disturbances: current state and future directions Int. J. Appl. Earth Obs. Geoinf. 60 49–60
Shen M, Zhang G, Cong N, Wang S, Kong W and Piao S 2014 Increasing altitudinal gradient of spring vegetation phenology during the last decade on the Qinghai–Tibetan Plateau Agric. For. Meteorol. 189–190 71–80
Tang Y Y, Chen J P, Zhang F and Yuan S S 2021 Spatiotemporal analysis of evapotranspiration and effects of water and heat on water use efficiency Water 13 1–17
Thomson A J, Shepherd R F, Harris J W E and Siversides R H 1984 Relating weather to outbreaks of western spruce budworm, Choristoneura occidentalis (Lepidoptera: Tortricidae), in British Columbia Can. Entomol. 116 375–81
Tikkanen O-P and Julkunen-Tiitto R 2003 Phenological variation as protection against defoliating insects: the case of Quercus robur and Operophtera brumata Oecologia 136 244–51

Tran T N, Afanador N L, Buydens L M C and Blanchet L 2014 Interpretation of variable importance in partial least squares with significance multivariate correlation (sMC) Chemom. Intell. Lab. Syst. 138 153–60

Tucker C J, Pinzon J E, Brown M E, Slayback D A, Pak E W, Mahoney R, Vermote E F and El Salloum N 2005 An extended AVHRR 8-km NDVI dataset compatible with MODIS and SPOT vegetation NDVI data Int. J. Remote Sens. 26 4485–98

van Asch M and Visser M E 2007 Phenology of forest caterpillars and their host trees: the importance of synchrony Annu. Rev. Entomol. 52 37–55

Vilhar U, Beuker E, Mizunuma T, Skudnik M, Lebourgeois F, Soudani K and Wilkinson M 2013 Tree phenology forest monitoring Developments in Environmental Science vol 12, ed M Ferretti and R B Fischer (Amsterdam: Elsevier) ch 9, pp 169–82 (available at: www.sciencedirect.com/science/article/pii/B9780080982229000091)

Waldstein D E and Reissig W H 2001 Apple damage, pest phenology, and factors influencing the efficacy of tebufenozide for control of obliquebanded leafroller (Lepidoptera: Tortricidae) J. Econ. Entomol. 94 673–9

Wang J and Liu D 2021 Vegetation green-up date is more sensitive to permafrost degradation than climate change in spring across the northern permafrost region Glob. Change Biol. 28 1–14

Wang J, Wu C, Wang X and Zhang X 2019 A new algorithm for the estimation of leaf unfolding date using MODIS data over China’s terrestrial ecosystems ISPRS J. Photogramm. Remote Sens. 149 77–90

Wang K and Dickinson R E 2012 A review of global terrestrial evapotranspiration: observation, modeling, climatology, and climatic variability Rev. Geophys. 50

Wang R, Chen J M, Liu Z and Arain A 2017 Evaluation of seasonal variations of remotely sensed leaf area index over five evergreen coniferous forests ISPRS J. Photogramm. Remote Sens. 130 187–201

Wang X Y, Zhao C Y and Jia Q Y 2013 Impacts of climate change on forest ecosystems in Northeast China Adv. Clim. Change Res. 4 230–41

Wehrens R, Hiemstra P, Kristian M and Liland H 2021 pls: partial least squares and principal component regression R package version 2.8-0

Wu H and Li Z-L 2009 Scale issues in remote sensing: a review on analysis, processing and modeling Sensors 9 1768–93

Wu S J, Zhu T H, Qiao T M, Li S J and Shan H 2021 Prediction of the potential distribution of dendrolimus houi lajonquiere in Sichuan of China based on the species distribution model Appl. Ecol. Environ. Res. 19 2227–40

Wulder M A et al 2019 Current status of Landsat program, science, and applications Remote Sens. Environ. 225 127–47

Yang K and He J 2019 China meteorological forcing dataset (1979–2018) National Tibetan Plateau Data Center (https://doi.org/10.11888/AtmosphericPhysics.tpcc.249369.file)

Zeng J-P, Ge F, Su J-W and Wang Y 2008 The effect of temperature on the diapause and cold hardness of Dendrolimus tabulaeformis (Lepidoptera: Lasiocampidae) Eur. J. Entomol. 105 599–606

Zhang N, Zhang X, Yang G, Zhu C, Huo L and Feng H 2018 Assessment of defoliation during the Dendrolimus tabulaeformis Tsai et Liu disaster outbreak using UAV-based hyperspectral images Remote Sens. Environ. 217 323–39

Zhang S, Zhang Z, Wang H and Kong X 2014 Molecular characterization, expression pattern, and ligand-binding property of three odorant binding protein genes from Dendrolimus tabulaeformis J. Chem. Ecol. 40 396–406

Zhou G, Zhang G and Liang Q 2012 Impact of meteorological conditions on Dendrolimus tabulaeformis and establishment of its prediction model J. Northwest For. Univ. 40 131–4