Risk Modeling for the Emergence of the Primary Outbreak Area of the Siberian Moth *Dendrolimus sibiricus* Tschetv. in Coniferous Forests of Central Siberia

Andrey A. Goroshko [1], Svetlana M. Sultson [1], Evgenii I. Ponomarev [2], Denis A. Demidko [1], Olga A. Slinkina [1], Pavel V. Mikhaylov [1,*], Andrey I. Tatarintsev [1], Nadezhda N. Kulakova [1] and Natalia P. Khizhniak [1]

1   Scientific Laboratory of Forest Health, Reshetnev Siberian State University of Science and Technology, 660037 Krasnoyarsk, Russia; utrom3@gmail.com (A.A.G.); sultson2011@yandex.ru (S.M.S.); sawer_beetle@mail.ru (D.A.D.); sloa@mail.ru (O.A.S.); tatarintsevai@sibsau.ru (A.I.T.); nadezha21@mail.ru (N.N.K.); natalia-mx@mail.ru (N.P.K.)
2   Federal Research Center "Krasnoyarsk Science Center, Siberian Branch, Russian Academy of Sciences", 660036 Krasnoyarsk, Russia; evg@ksc.krasn.ru
*   Correspondence: mihaylov.p.v@mail.ru

Abstract: In the southern taiga of Siberia, periodic outbreaks of the Siberian moth *Dendrolimus sibrircus* Tschetv. have been observed. The outbreaks result in the defoliation of Siberian fir *Abies sibirica* Ledeb. and Siberian pine *Pinus sibirica* Du Tour. stands across approximately one million hectares, leading to dieback of the affected forests. This is largely attributable to the inability to promptly identify the onset of the pest population growth in a timely manner, particularly in the context of expansive forest areas with limited accessibility. It is feasible to enhance the efficacy of monitoring Siberian moth populations by discerning stands with the highest propensity for damage and concentrating efforts on these areas. To achieve this, we employed machine learning techniques, specifically gradient boosting, support vector machines, and decision trees, training models on two sets of predictors. One of the datasets was obtained through a field study conducted in forest stands during the previous outbreak of the Siberian moth (2015–2018), while the other was derived from the analysis of remote sensing data during the same period. In both 2015 and 2016, the defoliation was most accurately predicted using gradient boosting (XGB algorithm), with ROC-AUC values reaching 0.89–0.94. The most significant predictors derived from the ground data were the proportions of Siberian fir, Siberian spruce *Picea obovata* Ledeb., and Scots pine *Pinus sylvestris* L., phytosociological data, tree age, and site quality. Among the predictors obtained from the analysis of remote sensing data, the distance to disturbed forest stands was identified as the most significant, while the proportion of dark coniferous species (*A. sibirica*, *P. sibirica*, or *Picea obovata* Ledeb.), the influx of solar radiation (estimated through the CHILI index), and the position in the relief (mTPI index) were also determined to be important.

Keywords: machine learning; forecasting; taiga; defoliation; *Dendrolimus sibiricus*; remote sensing; forest characteristics

1. Introduction

Of the various ecological groups of forest insects capable of outbreaks, the greatest danger is posed by those that are defoliators. In some cases, the damage they cause leads to the death of a significant proportion of a forest stand, up to the death of almost all their host trees [1–6]. The Siberian moth *Dendrolimus sibiricus* Tschetverikov, 1908 (Lepidoptera,

Lasiocampidae), is one such insect species. It has been observed that 50% defoliation of so-called dark coniferous species (including Siberian pine *Pinus sibirica* Du Tour and Siberian fir *Abies sibirica* Ledeb.) by its caterpillars often results in the death of several tens of percent of the tree layer. In the event of 75% or more defoliation, the forest stand will typically undergo a complete dieback [7–11]. The outbreaks of *D. sibiricus* in the taiga zone of Krasnoyarsk Krai alone affect approximately one million hectares of dark coniferous forests [12], i.e., forests dominated by *P. sibirica*, *A. sibirica*, or *Picea obovata* Ledeb., and have the potential to result in their complete disappearance over areas spanning several hundred thousand hectares [10].

In regard to the mitigation of the detrimental effects of defoliation by the Siberian moth, the advantage of prompt identification of affected sites is evident. Nevertheless, this is a challenging undertaking in Siberia. In the regions of Siberia that suffer the most from damage to Siberian pine and Siberian fir forests by *D. sibiricus* (Krasnoyarsk Krai and Tomsk Oblast), the areas of Siberian fir or Siberian pine reach 13.72 and 4.21 million hectares, respectively [13]. The road density in these regions is 0.0137 and 0.0360 km^{-2}, respectively [14]. Notwithstanding the uneven distribution of both damaged forests and the transport network, these data indicate that access to potential outbreak spots is extremely difficult. This, in turn, presents a challenge to the timely detection of an increase in the pest's numbers. This situation results in an inadequate level of defoliating pest management in these forests [15], thereby complicating the planning of management measures [16].

The issue can be partially addressed by identifying, in accordance with specific criteria, those forest stands where the likelihood of a Siberian moth outbreak is the greatest. Previously published theories of the population dynamics demonstrate the fundamental possibility of this. In the primary outbreak areas (these are the areas where the population of the pest increases most rapidly at the beginning of the outbreak and where damage occurs first [8,17]) of the Siberian moth, which are characterized by the most favorable habitat conditions, the population density of the pest increases due to the migration of individuals from adjacent areas [8,10]. The primary outbreak areas are smaller than those that *D. sibiricus* defoliates throughout the outbreak [12,18]. This makes the timely detection of the beginning of an increase in numbers more challenging. However, if a risk assessment procedure for identifying such areas is developed, it will be possible to survey a relatively small area to confirm the threat from the Siberian moth, which can be done in a short time and at the appropriate juncture.

Previously, efforts have been made to study the landscape-ecological confinement of defoliators outbreak areas, with a view to understanding the influence of spatial heterogeneity of landscapes on their population dynamics [19]. For different insect species and regions, the dependence of damage intensity on the confinement of outbreaks to different relief elements [16,20–22], soil characteristics [16,21], and forest characteristics [20–22] has been substantiated. This issue has also been studied for the Siberian moth [8,10,12,18,23–28]. However, these studies did not contribute to the development of a risk analysis system.

Among the features that could prove useful for identifying a potential primary outbreak area of *D. sibiricus*, the majority of authors cite the forest type. The identification of these types is based on the methodology established by Braun-Blanquet [29] and Sukachev [30], with the classification determined by the prevailing understory plant community [8,10,18,24–26,28], independent of dominant tree species. In the context of characterizing outbreak areas in mountainous regions, the height above sea level [8,12,18,25,27], steepness [12,18,27], slope aspect [8,10,12,18,24,27], and the confinement of the forest stand to a certain part of the slope [8,10,23,24] are of particular significance. The soil drainage is related to the forest type (i.e., plant community) and the location of the forest site on the slope. Some authors [8,23] have indicated a direct connection between these factors and the

probability of damage caused by the Siberian moth. In addition to tree species composition, which is addressed in one form or another in the aforementioned works, the density and age of the forest stand were found to influence the probability and extent of damage caused by *D. sibiricus* [8,10,24–26,28]. Ultimately, a number of authors posit that the probability of an outbreak is heightened in forest stands that have been subjected to disturbance (such as logging, previous defoliation, or fire) or in proximity to such stands [23,24,26].

Despite the comprehensive description of the landscape characteristics preferred by the Siberian moth, including those of the taiga of Krasnoyarsk Krai [10,25,26,28], the localization of its primary outbreak areas presents a significant challenge. The first reason is that these characteristics were considered in isolation from one another. Thus far, no attempts have been made to construct a risk assessment model based on these characteristics, similar to the models developed for bark beetles [31–33]. Another reason for these difficulties is that forest stand characteristics, such as forest type and density, are commonly employed to assess the risk of an outbreak in a given area [8,10,26]. Reliable data of this kind can typically be obtained from a comprehensive survey of relatively limited areas [31]. Attempts to utilize forest inventory data encompassing the entire study region to identify potential primary outbreak areas of *D. sibiricus* have frequently resulted in inaccuracies (V.V. Soldatov, personal communication).

As the most practical solution, we propose a transition from the utilization of data that can only be obtained through comprehensive field studies to the incorporation of remote sensing (RS) data as predictors [31]. The principal objective of this study was to evaluate the precision of risk assessment models constructed using machine learning to predict the location of primary outbreak areas of the Siberian moth based on RS data. In order to ascertain whether the constructed models would be applicable across different outbreaks or whether they were specific to a particular outbreak, it was necessary to compare the patterns of spatial distribution of the Siberian moth outbreak with the results of previous studies. To this end, analogous models were constructed utilizing ground survey data as predictors, and the impact of specific forest site characteristics on the emergence of primary outbreak areas was investigated.

2. Materials and Methods

2.1. Study Area

The study focuses on dark coniferous forests that have been partially affected by the outbreak of the Siberian moth between 2015 and 2018 (Figure 1). These forests are situated within the middle taiga subzone of the eastern part of the West Siberian Plain. From an administrative perspective, the study area is located within the Yeniseysky District of Krasnoyarsk Krai.

The study area is characterized by a continental climate, with low winter temperatures and the stagnation of cold air in river valleys and basins. The absolute minimum air temperature is $-57\,°C$, while the absolute maximum is $+37\,°C$. The period during which the average daily temperature is below $-5\,°C$ lasts for approximately five months (from November to March), while the period during which the average daily temperature is below $0\,°C$ lasts for approximately half a year. The frost-free period lasts for 103 days, with the first frost observed as early as the beginning of September [34]. The average annual temperature is between -1.5 and $-0.6\,°C$, with a growing season average temperature of $13.5–14.5\,°C$ [35,36]. The average annual precipitation is 460 mm [34].
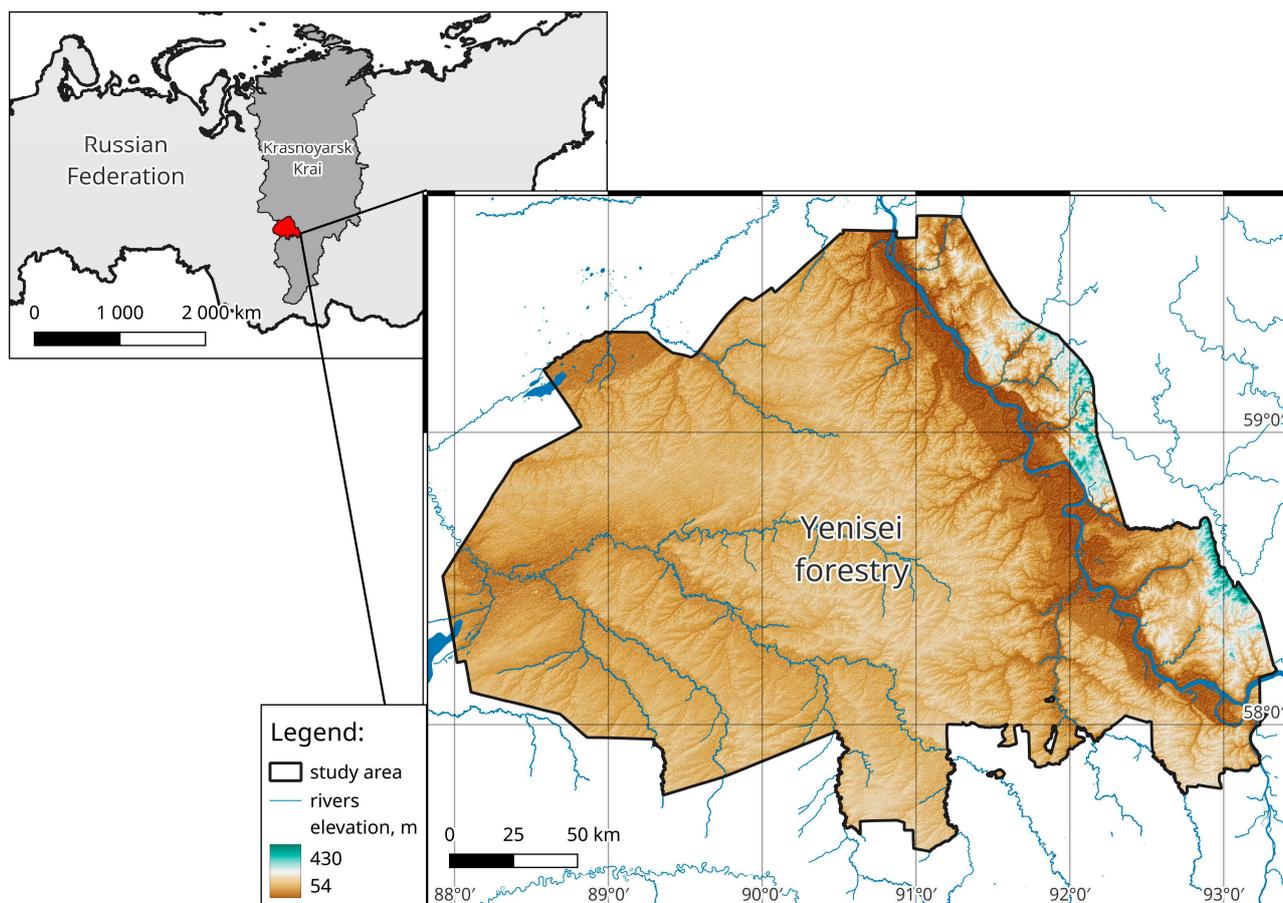
Figure 1. Study area (Yeniseyskoye forest district, Krasnoyarsk Krai).

The zonal vegetation type is taiga. Dark coniferous forests prefer loamy soils as they are more demanding in terms of air humidity and constant moderate soil moisture. Coniferous stands, dominated by Scots pine *Pinus sylvestris* L., occupy sandy and swampy soils. Large areas are covered by deciduous forests (mainly dominated by birch *Betula* ssp.), which usually replaced coniferous forests [37,38]. The outbreak of the Siberian moth developed in forest stands dominated by Siberian fir and Siberian pine. The first damage was recorded during the analysis of the RS data in 2015 [39].

2.2. Species Background

The Siberian moth *D. sibiricus* is one of the most destructive defoliators in Eurasia. The modern area of this species within Russia encompasses the southern and central taiga as well as mountain forests, extending from the Middle Volga Region to Sakhalin. Additionally, it occupies the northern regions of Kazakhstan, Mongolia, China, and Japan [40,41]. Its outbreaks have been documented in the southern part of the range [41]. In the study area, pest outbreaks occur approximately once every 15 years [10,42].

Imagoes of this species are observed from late June to early August, with a peak in numbers occurring in mid-July. The period of egg hatching begins in late July, after which the caterpillars commence feeding [7,9,41]. The preferred food of the caterpillars is the needles of various species of *Abies*, *Larix*, and some five-needle pines, including *P. sibirica* [43,44]. However, in the study area, Siberian fir and Siberian pine forests incur the majority of damage [10,44]. Typically, this insect species requires two years to complete one generation [7,9,41]; however, under favorable conditions, the majority of the population becomes monovoltine [10,41]. In October, the caterpillars embark on their annual search for

sheltered overwintering sites, which they typically locate beneath a layer of moss or litter. In some years, the migration to the crowns commences as early as April, with the majority of individuals arriving in mid-May, coinciding with the melting of the snow. Pupation occurs in the crowns between June and early July [7,41].

In studies conducted in the mid-twentieth century, the authors highlighted the influence of topography on the formation of Siberian moth outbreak areas [7,23]. The prevailing view is that the most favorable conditions for caterpillars wintering on the soil surface under plant cover or litter are those of good drainage. This is particularly the case in the upper parts of hills [7,9,10,41] or slopes [9,10,23,41]. The situation for wintering caterpillars is significantly more adverse on relief elements where water run-off is challenging [23], which can be ascribed to their inability to withstand low temperatures in humid conditions [45].

Notwithstanding their cold resistance, the caterpillars of the Siberian moth are markedly thermophilic. A number of researchers [23,24,26] have highlighted the proximity of the primary outbreak areas of this species to forest areas that have previously been disturbed by other factors. This phenomenon may be attributed to a favorable alteration in the microclimate, which has become warmer and drier [8].

2.3. Obtaining and Preparing Forest Inventory, Forest Cover and Orographic Data

The first set of predictors for modeling was based on ground data, comprising stand characteristics within the study area. These were represented by a vector layer at the level of forest compartments, which were defined as forest areas exhibiting homogeneous main characteristics. The area of these compartments ranged from 1 to 506 ha, with a median of 18 ha. From this, compartments covered by forest were selected for modeling. A total of fourteen stand characteristics were identified for inclusion in the modeling procedure (Table 1).

Table 1. The characteristics of the forest compartments, as determined by ground data.

Characteristic	What Does Characteristic Specify	Unit of Measurement	Scale of Measurement
forest compartment area	The area of forest compartments	ha	ratio
age	Average age of dominant tree species	year	ratio
relative stocking	Ratio of the basal area of a stand to the basal area of a 'normal' stand		ratio
site quality	Index of potential site productivity expressed by average height of dominant tree species compared with 'normal' stand		ordinal
soil moisture	Index of long-term moisture conditions		ordinal
group of forest types	Dominance of some ecological group of understory plant species		nominal
share of tree species	Share of the tree species (fir, spruce, Siberian pine, Scots pine, larch (*Larix sibirica* Ledeb.), birch, aspen (*Populus tremula* L.) or willow (*Salix* ssp.)) in the stand's growing stock	'unit'; each unit ≈ 10% of the total growing stock	ratio

The groups of forest types were identified based on the species composition of the understory plant community [29,46,47], which indirectly characterizes a complex of soil conditions. The group of forest types was a categorical variable and included eight groups (Table 2).

Table 2. A brief description of groups of forest types. The predominant size of soil grains is specified in accordance with ISO 14688-1:2017 [48].

Group of Forest Types	Plants Dominated in Unederstory Layers	Soil Fertility	Most Typical Soil Humidification Regime and Grain Size
feather moss	Hylocomiaceae	poor	moderately wet, coarse or medium slit
tallgrass	some tall grasses, like species of *Heracleum*, *Aconitum*, *Veratrum* and others	very rich	moderately wet, medium slit
shrub	*Vaccínium* ssp.	poor	moderately wet, from sand to medium slit
lichen	*Cladonia* and *Cetraria* species	extremely poor	dry, sand
sedge	*Carex macroura* Meinsh.	rich	moderately wet, coarse or medium slit
mixed grass	a variety of typical mesophilic forest herbs without explicit dominants	very rich	moderately wet, coarse or medium slit
sphagnum	*Sphagnum* ssp.	extremely poor	extremely stagnant wet, from medium slit to fine slit
grass-swamp	a variety of typical hydrophilic herbs without explicit dominants	extremely poor	extremely flowing wet, from medium slit to clay

The site quality and soil moisture were ordered factors, with six classes ranging from the most productive to the least productive and five moisture classes, from the driest to the wettest soil. The proportion of tree species within the tree layer was assigned integer values between 0 (indicating the absence of species) and 10 (representing absolute dominance) for each of the eight forest-forming species.

The data were processed in the R 4.0.2 statistical computing environment [49] with the RStudio 2022.07.1-554 graphical interface [50]. The preliminary processing of the data on stand characteristics was conducted using the dplyr 1.0.9 [51] and sf 1.0-8 [52] packages.

The second set of predictors comprised derivatives of RS data normalized to a spatial resolution of 270 m per pixel (EPSG:4326, WGS 84). Two indices were employed to account for orographic conditions. The first index, the Multi-Scale Topographic Position Index (mTPI) [53], enables the differentiation of depressions (negative values) and watersheds (positive values). The second index, Continuous Heat-Insolation Load (CHILI) [54], estimates the radiation balance of sites and takes values from 0 (representing very cold habitats) to 255 (representing very warm habitats). The aforementioned indices were obtained from the Google Earth Engine website [55], with a spatial resolution of 270 (mTPI) and 90 (CHILI) m per pixel. The CHILI data were interpolated to a spatial resolution of 270 m per pixel using a linear approach.

The sites with high amounts of the Siberian moth host tree species were identified by remote sensing data analysis as forest stands dominated or co-dominated by dark coniferous species (Figure 2). This was achieved by utilizing the data from vegetation maps of Russia [56] with a spatial resolution of 230 m per pixel, which was then refined to a spatial resolution of 270 m per pixel by the nearest neighbor interpolation method. The woody vegetation classes that may be regarded as potential habitats for *D. sibiricus* were identified as classes 1 ($\geq$80% of the crown area is comprised of spruce, fir, or Siberian pine) and 10 (the crown area of coniferous and deciduous tree species is represented in approximately equal proportions). In contrast to the ground data, forests comprising other coniferous and deciduous species were initially excluded from the processing.
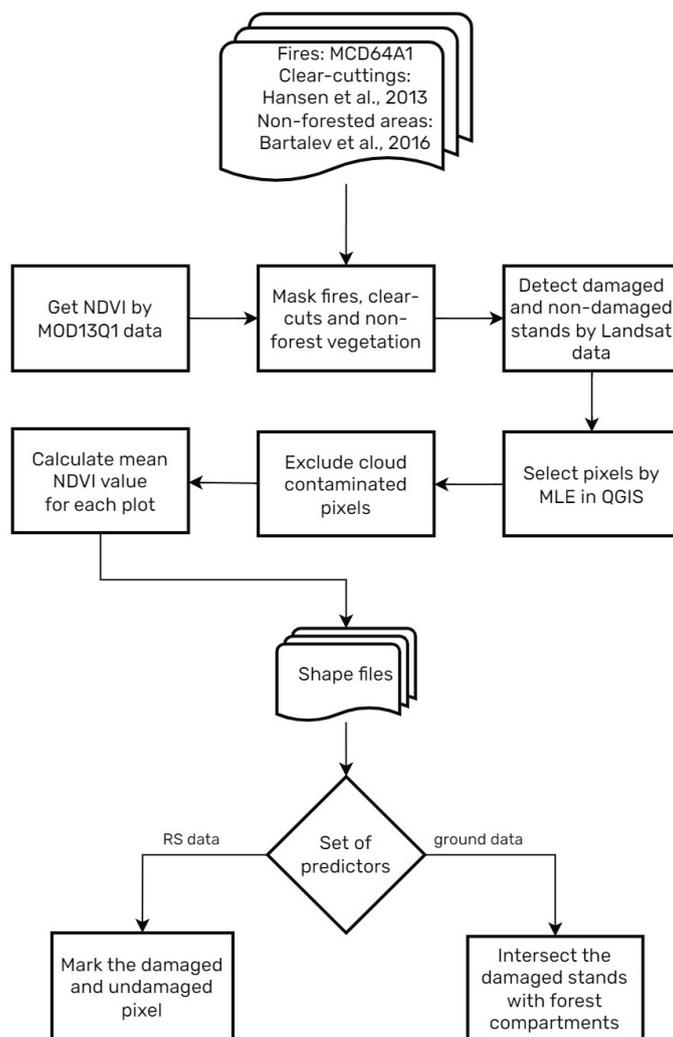
Figure 2. Schematic representation of the processing of RS data on vegetation cover [56,57].

The areas of forest disturbance from 2005 to 2014 were identified (Figure 2) using the Hansen map of global deforestation [57], which has a spatial resolution of 30.92 m per pixel. The data were divided into two sets: one comprising damage occurring five years prior to the outbreak (since 2010) and the other comprising damage occurring between five and ten years prior to the outbreak (before 2010). Both data sets were normalized to 270 m per pixel using nearest neighbor interpolation, with the objective of matching the rest of the data and excluding small areas of damage from the analysis. Furthermore, a window filter was applied to the 270 m per pixel data in order to exclude single pixels within a 3×3 cell neighborhood. Based on this data, raster layers were constructed in which each pixel contained the distance to the nearest disturbed forest area (for the period from 2010 to 2014 and for the period from 2005 to 2009).

All forms of interpolation were conducted utilizing the distance function from the Terra 1.6-47 package [58]. The geodata layers were limited to the territory of the Yeniseyskoye forest district, and within this territory, to the dark coniferous and mixed dark coniferous-deciduous stands. For the purposes of this study, stands dominated by Scots pine, larch (*Larix sbirica* Ledeb.), or deciduous species were excluded from consideration when remote sensing data were analyzed.

2.4. Detecting Damaged Stands Using Remote Sensing Data

The dependent variable was the presence of disturbance caused by the Siberian moth at the commencement of the peak phase of its outbreak. Models were constructed for the end of 2015 and the end of 2016.

To identify forest areas damaged by the Siberian moth, we employed the use of the NDVI index, which was calculated using MODIS data (Figure 2). The NDVI values were extracted from the MOD13Q1 vegetation index product [59] with a spatial resolution of 250 m per pixel. The data set encompassed the period of highest photosynthetic activity of vegetation (from June to September) in 2014–2018. To identify and exclude from the analysis any changes in NDVI associated with fires, the MCD64A burned area product [60], also formed using MODIS data, was used (Figure 2). Thematic MODIS products were loaded using the LAADS (Level-1 and Atmosphere Archive & Distribution System) service. A comparable issue pertaining to clear cuts was addressed through the utilization of the global forest cover loss map [57]. The lands with non-forest vegetation were identified using a vegetation map developed at the Russian Space Research Institute of the Russian Academy of Sciences [56] (Figure 2). The RS data were divided into training and test samples using Landsat images (spatial resolution of 30 m) obtained from the US Geological Survey website [61]. Visual interpretation was used to identify areas covered with dark coniferous stands, damaged and undamaged by the Siberian moth (Figure 2).

A comparison of the NDVI values derived from low-resolution (MODIS) and moderate-resolution (Landsat) satellite imagery was conducted on an individual basis for forest network plots. To perform such a comparison, the following procedure was undertaken: (1) the subsets of MODIS/Landsat pixels located within each forest network plot were selected using GIS tools (Quantum Geographic Information System, version 3.16.3) with maximum likelihood estimation; (2) low-quality pixels (contaminated by cloud cover) were excluded; and (3) the mean NDVI values for the forest network plot were calculated (Figure 2). Consequently, composites were constructed comprising data from MODIS channels 1 (620–670 nm) and 2 (841–876 nm), in addition to NDVI index values.

The results were exported in the form of polygonal shapefiles (Figure 2). The predictors were represented as polygons (forest compartments) based on the stand characteristics. In this set of predictors, the damaged forest compartments were identified by calculating the area of intersection between the damage and the forest compartment outline for a specific date. A forest compartment was designated as damaged if RS data indicated a defoliation percentage of 10% or higher. The defoliated forest compartments were assigned a label of 1, while all others were assigned a label of 0. For the RS-based dataset, the actual defoliation data was reduced to a resolution of 270 m per pixel using nearest neighbor interpolation and labeled similarly (Figure 2).

2.5. Application of Machine Learning Algorithms

The machine learning was conducted using the algorithms embedded in the mlr3 0.14.1 [62] and DALEX 2.4.2 [63] packages. The ggplot2 3.3.6 package was employed for the graphical analysis of the results [64].

The classification task was solved using three machine learning algorithms: decision tree (DT), support vector machine (SVM), and extreme gradient boosting (XGB). The DT algorithm is relatively straightforward to comprehend, whereas SVM and XGB algorithms have demonstrated efficacy in addressing a diverse array of challenges, including those pertaining to land use and language classification [65]. A forest insect damage risk assessment model based on the random forest algorithm demonstrated comparable accuracy to that of the XGB [31]. However, our preliminary study indicated that the random forest

model requires inordinate computational resources, which ultimately led to the decision to discontinue its testing.

The data regarding the number of observations for each year and damage class in the ground data and RS data are presented in Table 3. To optimize the hyperparameters of the models, this set of observations was utilized in its entirety, with the exception of SVM, for which the data volume was reduced by half due to the high computational demands of this algorithm. The original data set was divided into a training set and a testing set in a ratio of 80:20. To mitigate the imbalance of classes in the training set, observations were randomly selected from undamaged stands at a rate 10 times greater than the number of damaged stands. Subsequently, the number of damaged stands was augmented through repeated extraction of observations until it reached 50% of the number of undamaged stands. Consequently, the training set comprised two times less data for damaged stands than for undamaged ones. The ratio of classes in the test set remained unaltered (Table 3).

Table 3. The data size and class ratio in the datasets prior to upsampling.

Data	Year	Damage	All Data	Train Set	Test Set
ground data	2015	0	116,545	5340	23,309
ground data	2015	1	668	2670	134
ground data	2016	0	113,615	28,780	22,723
ground data	2016	1	3598	14,390	720
RS data	2015	0	304,535	47,780	60,907
RS data	2015	1	5972	23,890	1194
RS data	2016	0	289,326	169,450	57,865
RS data	2016	1	21,181	84,725	4236

In order to facilitate the analysis, the factors were transformed by one-hot encoder and the ordered factors were converted into integer values.

The selection of the hyperparameters of the models was conducted on the training sample using the 4-fold cross-validation scheme. The optimization criterion was the area under the curve (AUC). This criterion is not contingent on the cutoff threshold, thereby facilitating a more objective comparison of the accuracy of the algorithms. The optimal values of the hyperparameters were selected within the specified limits (Table 4) by maximizing the AUC value calculated on the test sample.

The selection of SVM and XGB hyperparameters to optimize computational efficiency was conducted in two stages. Initially, the kernel for SVM and the number of trees (nrounds) for XGB were identified. Subsequently, the remaining hyperparameters were tuned.

All classification algorithms utilized were of the soft variety, predicting the probability of an observation belonging to a specific class. To transition from probability to class prediction, a cutpoint was utilized, calculated based on the objective of maximizing the Youden J-index, as determined by the cutpointr 1.1.2 package [66].

The importance of individual predictors was evaluated through the use of the variable permutation method (feature_importance function from the DALEX package) [63], which was applied 50 times for each variable (for detailed description of the procedure see [67]). Consequently, the greater the decline in the average AUC value, the more pronounced the impact of the predictor on the accuracy of the model's prediction [63].

Table 4. Optimized boundary values for the hyperparameters of three algorithms.

Hyperparameter	What Does Hyperparameter Specify	DT	SVM	XGB
Cp form	The measure of minimal increasing of prediction accuracy after splitting	0.001–1		
Maxdepth, max_depth	The maximum depth of the tree	3–10		3–10
Minbucket	Smallest number of observations in a terminal node	1–100		
Minsplit	Smallest number of observations in the parent node	1–100		
Kernel	Specific algorithm of pattern analysis		radial, sigmoid, polinomial (degree 1 to 4)	
Cost	The measure of classification hardness		10^{-5}–10^{5} (log-scaled)	
Gamma	The measure of sample point influence on classification		10^{-5}–10^{5} (log-scaled)	
Nrounds	The number of trees			10–600
Min_child_weight	The minimum sum of weights of observations in a child node			1–10
Subsample	The fraction of observations sampled for each tree			0.5–0.8
Colsample_bytree	The subsample ratio of columns when constructing each tree			0.5–0.9
Eta	Degree of feature's weight shrinkage to prevent overfitting			0.1–0.6

3. Results

3.1. Best Models

The XGB algorithm-based models yielded the optimal results according to the ROC-AUC maximization criterion. For the test data set, the AUC value ranged from 0.89 to 0.94 (Figures 3 and 4). The hyperparameter values of the best models are presented in Table 5.
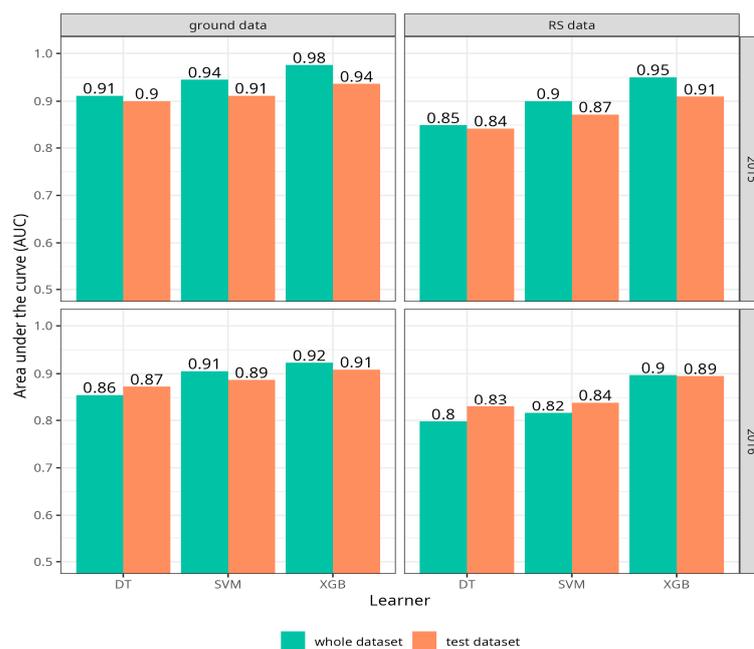

Figure 3. Classification quality of ground data and RS data in 2015 and 2016, employing a range of classification algorithms.
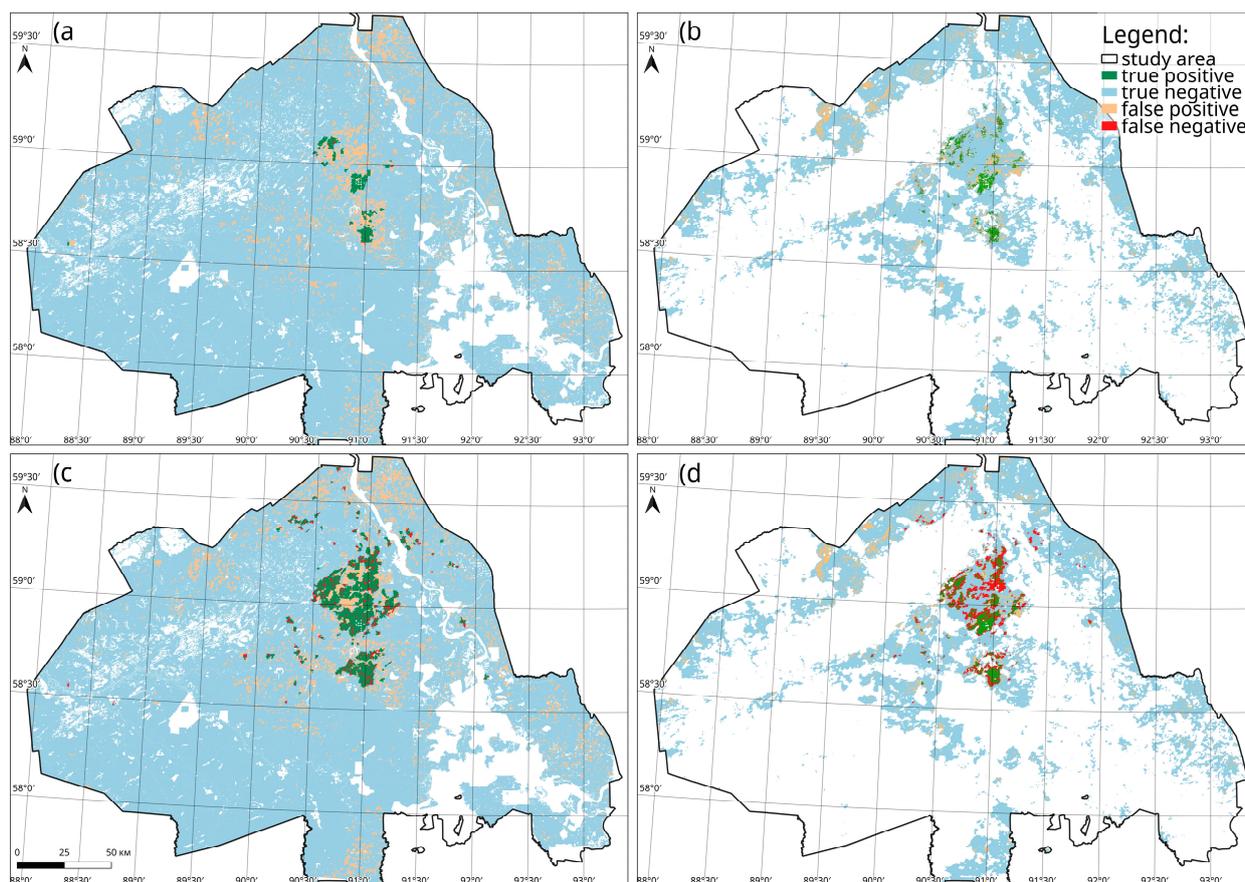
Figure 4. The results of the forecasting of defoliation of forest stands by the Siberian moth *Dendrolumus sibiricus* Tschetv. using the XGB algorithm: (**a**,**b**) based on stand characteristics for 2015 and 2016, respectively; (**c**,**d**) based on RS data for the same years.

Table 5. The best hyperparameter values for gradient boosting in the classification of forest areas damaged by the Siberian moth *Dendrolumus sibiricus* Tschetv.

Hyperparameter	2015, Ground Data	2015, RS Data	2016, Ground Data	2016, RS Data
Max_depth	10	10	8	10
Nrounds	50	200	60	300
Min_child_weight	1.712	6.953	1.252	3.363
Subsample	0.6603	0.6782	0.6053	0.748
Colsample_bytree	0.6254	0.8419	0.8426	0.8021
Eta	0.1453	0.1513	0.1463	0.1063

3.2. Importance of Predictors

When both ground data and RS data are employed in a predictive model, the contributions of the predictors in 2015 and 2016 are distributed in a similar manner (Figure 5).

The most significant predictors, as indicated by the ground data, were the share of spruce, fir, and Scots pine in the stand's growing stock, age, site quality, and group of forest types. In the model trained on 2016 data, a notable contribution was also observed for the share of aspen (*Populus tremula* L.) and birch, relative stocking, and forest compartment area (Figure 5A,B).
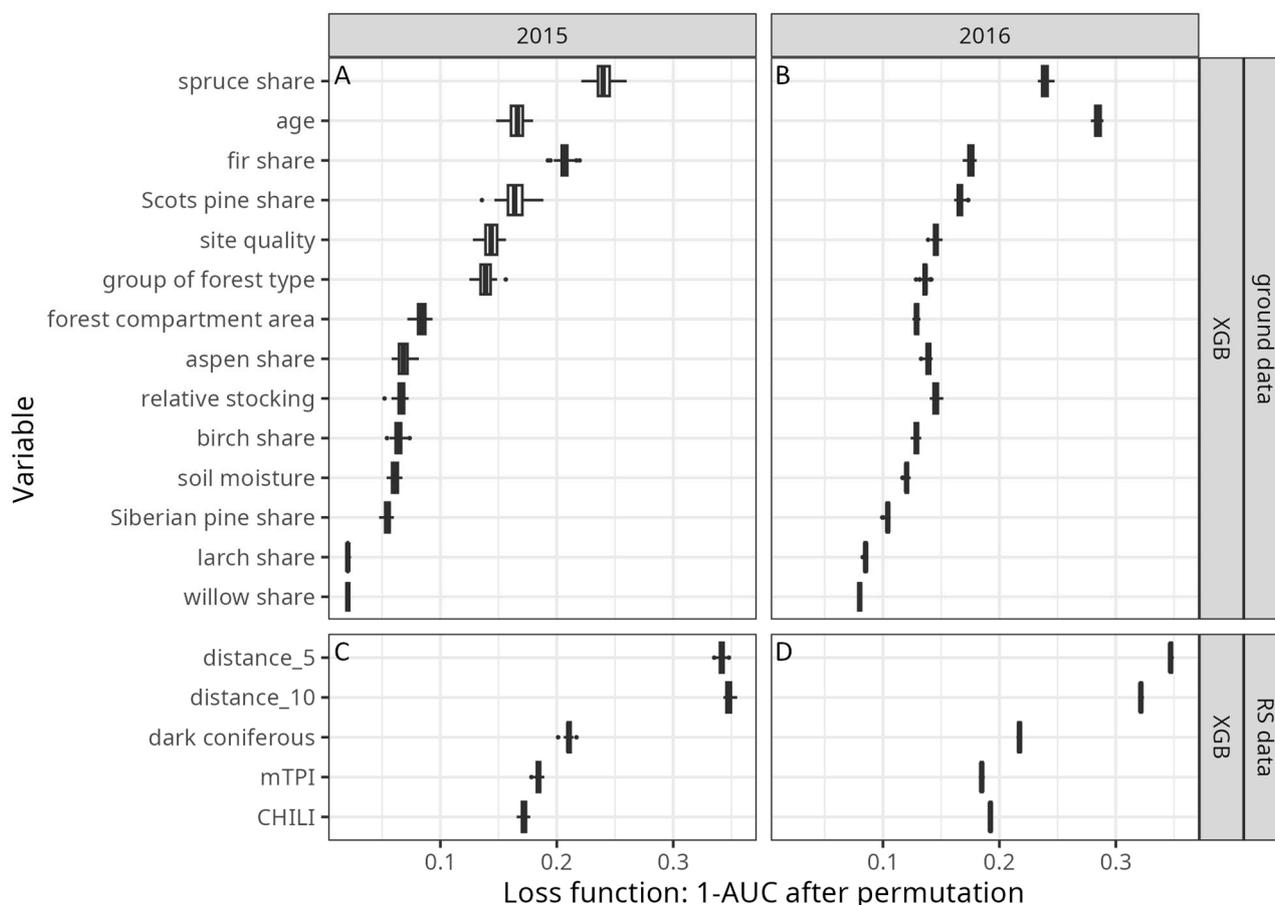
Figure 5. The role of predictors based on ground data (**A**,**B**) and RS data (**C**,**D**) in forecasting the emergence of primary outbreak areas for the Siberian moth *Dendrolumus sibiricus* Tschetv. In the RS data subfigures, distance_5 and distance_10 denote distances to forests that were disturbed from 2010 to 2014 and from 2005 to 2009, respectively.

In models constructed using RS data, the distance from disturbed stands has been identified as the most significant predictor. The role of other predictors, including the woody vegetation class (the proxy of host tree species proportion), mTPI, and CHILI indices (Figure 5C,D), has also been determined to be noteworthy.

3.3. Distributions of Predictor Values According to the Predicted Classes

The degree of dependence of a forest stand belonging to a specific class (defoliated or not defoliated) is contingent upon its species composition (Figure 6). Forest stands comprising a minimum of one unit (equivalent to 10% of the stand's growing stock) of Scots pine are rarely classified as damaged (Figure 6D). Similarly, the proportion of aspen and birch exceeding one unit is also an unlikely indicator of damage to the forest stand. Conversely, an increase in the proportion of fir in a forest stand comprising two or more units elevates the likelihood of the model classifying the forest stand as damaged (Figure 6C). The interrelation between the classification of a forest stand as defoliated and the proportion of spruce is close to a binomial distribution, with a maximum of approximately four units (Figure 6B). The probability of a forest stand being classified as damaged is demonstrably higher within the age range of 100–250 years (Figure 6A). The classification outcome is clearly dependent on stand productivity. Damage is most often predicted for stands of the third site quality class, and to a lesser extent for the second and fourth (Figure 6E). Notable differences are observed in the forecast results for different forest type groups between 2015 and 2016 (Figure 6F). If, at the outset of the outbreak,

damage is predicted almost exclusively for feather moss and sedge stands, then in the following season, the probability of damage caused by the Siberian moth increases for forests of tallgrass, mixed grass, and grass-swamp groups of types.
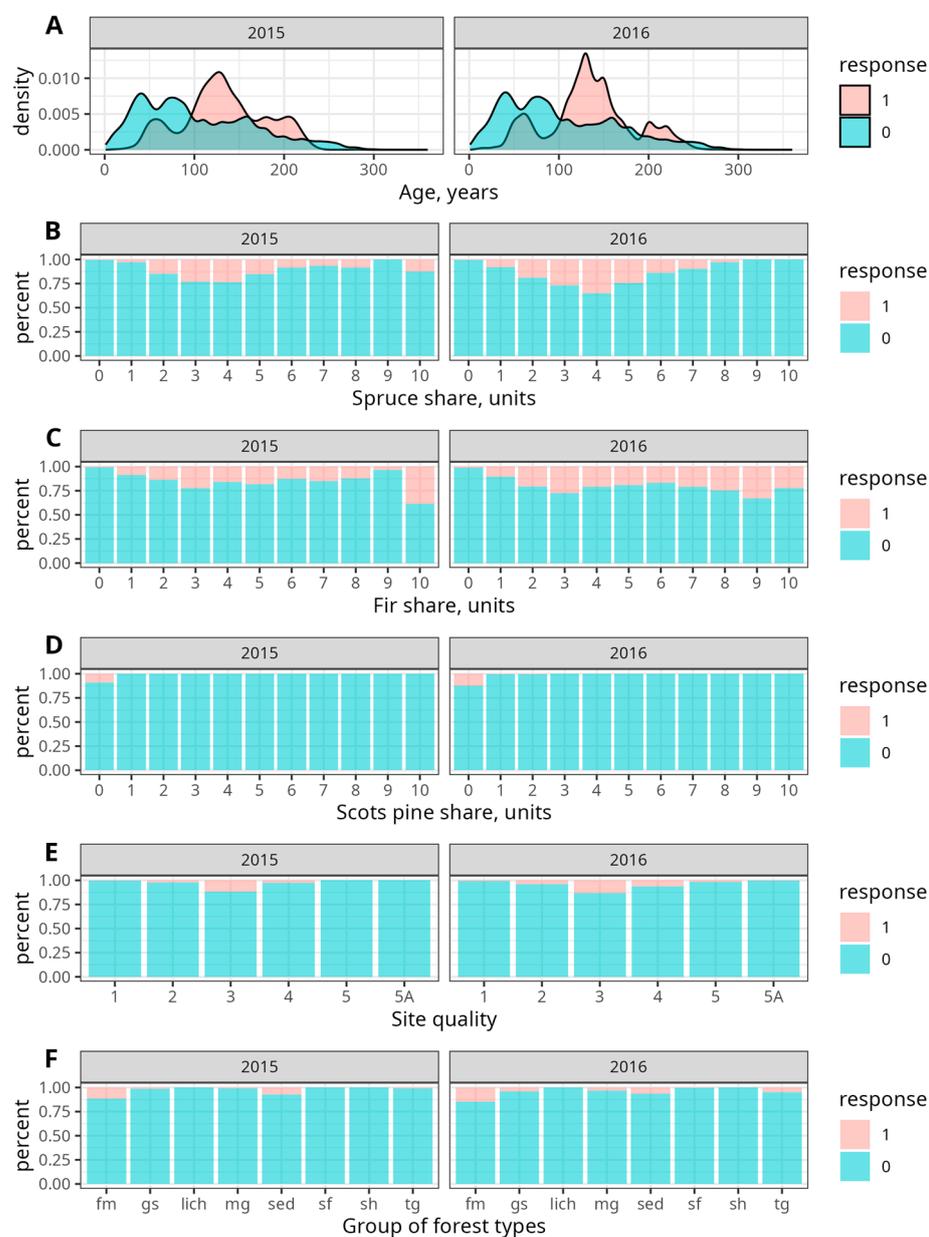

Figure 6. Distribution of predictor values obtained during the field study in relation to the model forecast results. (**A**) mean age of the forest stand, (**B**) share of spruce *Picea obovata* Ledeb., (**C**) share of fir *Abies sibirica* Ledeb., (**D**) share of Scots pine *Pinus sylvestris* L., (**E**) site quality, (**F**) forest type group.

The analysis of RS data indicates that damage is predicted with greater frequency for areas classified as dark coniferous forests (class 1) than for mixed forest (class 10) (Figure 7A). The dependence of the forecast outcome on the distance to the nearest disturbed forest stand is clearly bimodal, particularly in 2016 (Figure 7D). The initial probability peak occurs at a distance of approximately 2–3 km. The subsequent interval at which the probability of damage caused by the Siberian moth exceeds the probability of its absence commences at a distance of approximately 10 km. Additionally, forest sites for which the damage forecast was positive tend to be situated in moderately warmed areas (CHILI index) on or in close proximity to watersheds (mTPI index) (Figure 7B,C).
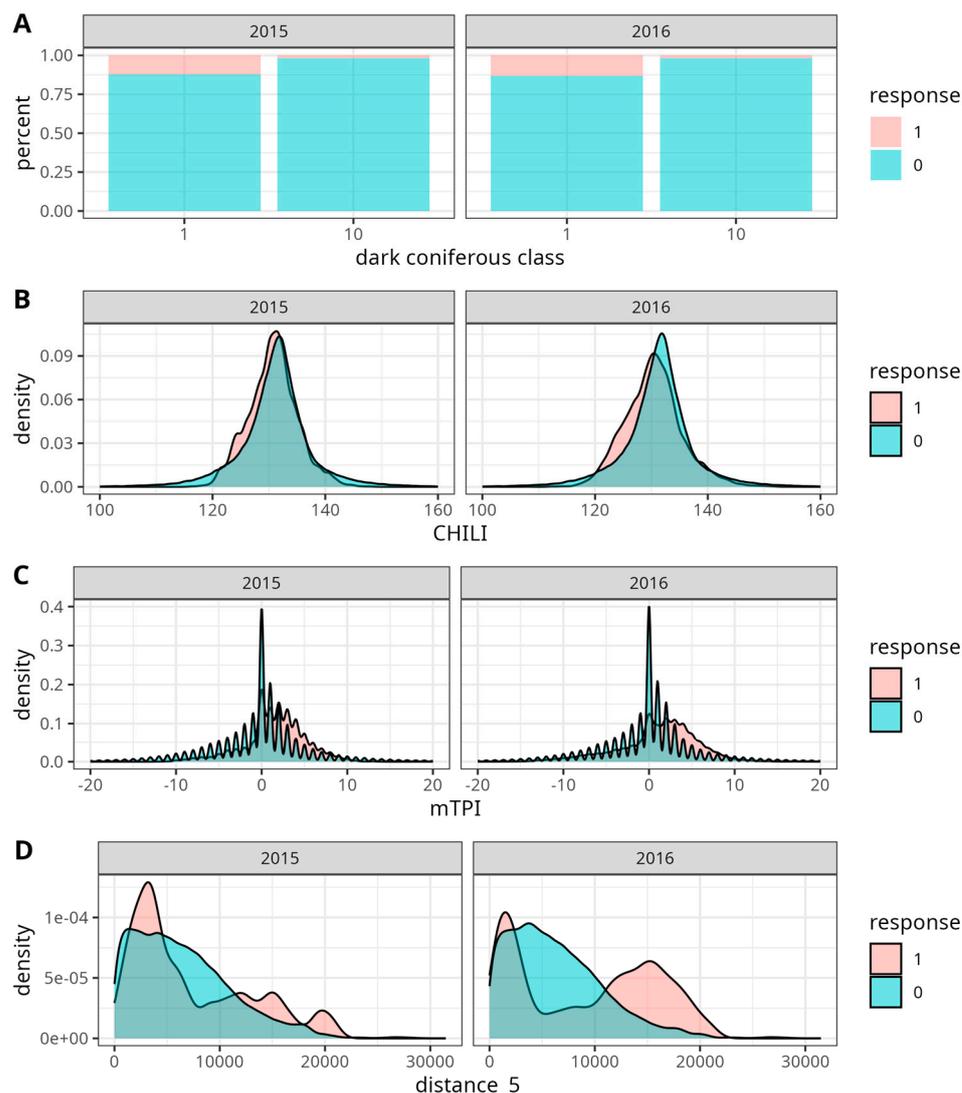
Figure 7. Distribution of predictor values obtained using RS data in relation to the outcomes of the model forecast (**A**) the proportion of preferred tree species in the stand composition, where class 1 corresponds to stands dominated by dark coniferous species and class 10 to a near-equal proportion of dark coniferous and deciduous tree species, (**B**) the CHILI index, (**C**) the mTPI index, and (**D**) the distance to the nearest forest stand that was disturbed five or less years before the Siberian moth *Dendrolumus sibiricus* Tschetv. outbreak).

4. Discussion

4.1. Results of Machine Learning Procedure

The machine learning algorithms employed enabled the accurate prediction of damage to forest stands dominated by dark coniferous species by the Siberian moth (Figures 3 and 4). The trained models demonstrated efficacy when utilizing both ground data (vector data) and RS data (raster data). Tests indicated that gradient boosting, which had previously exhibited favorable outcomes in addressing a range of other issues, yielded the most precise classification. Models constructed on this foundation can be employed to identify forest sites at risk in advance of an outbreak.

4.2. Food Availability Variables as Predictors

The most evident aspect pertains to the interrelation between the forecast outcomes and the species composition of the stands. The probability of classifying a stand as damaged

based on the proportion of fir in 2015 exhibits a distinct bimodal distribution, with maximal risk of the damage at approximately three to ten (pure fir stands) units. In the subsequent year, the distribution becomes more uniform, and the second maximum shifts to nine units (Figure 6C). This phenomenon can be attributed to a complex interplay of environmental and historical factors. As the host tree species proportion increases [21,26,68], so too does the population density of the pest, reflecting common environmental patterns [69]. Siberian fir is one of the preferred host species for the Siberian moth [45], and an increase in its proportion within the stand composition results in a growth of food resource amount for this defoliator. The results obtained are in accordance with the established patterns for this region [26].

The aforementioned findings are in alignment with the results of the RS data predictor analysis, which contributed to the model results. The greatest risk of an outbreak is associated with forest stands dominated by fir, spruce, and Siberian pine in varying proportions (class 1). Conversely, the probability of classifying mixed dark coniferous-deciduous stands (class 10) as damaged is significantly lower (Figures 5 and 7A).

The role of the shares of Siberian pine and larch, two other host species of the Siberian moth, as a predictor is unexpectedly low (Figure 4). This can be attributed to the relatively small share of Siberian pine (Figure 8G) and, in particular, larch (Figure 8D) within the forest stands of the study area when compared to fir (Figure 8B). The more common species primarily determines the amount of food resources.

One impediment to the development of the outbreak is the high proportion of non-host species [22]. The impact of birch and aspen (Figure 9B,C) on the probability of damage is particularly evident in 2016 (Figure 4). In addition to being the least favorable host among the conifers of the study area [45], Scots pine also exhibits significant differences in its soil requirements when compared to fir and Siberian pine (see Study Area). In the Yeniseysky District, Scots pine often occurs as the dominant species in pure stands, where the emergence of Siberian moth outbreaks is precluded (Figure 4).

It is noteworthy that the proportion of spruce, another non-host species, serves as an effective predictor of damage caused by *D. sibiricus* (Figure 4), with stands that contain 3–5 units of this species exhibiting the highest likelihood of defoliation (Figure 6B). This finding is at odds with the results of previous studies [26], which indicated that the probability of damage decreases significantly with the presence of more than one unit of spruce in the stand. A potential causal relationship exists between the current situation and previous Siberian moth outbreaks (see below).

The role of relative stocking as a predictor is relatively minor, yet discernible in 2016 (Figure 5). The probability of damage increases in a nearly linear manner from 0.3 to 0.9 but reaches zero at 1.0 (Figure 9H). This distribution can be attributed to the observed increase in needle mass with increasing density. The absence of damage to forest compartments with a relative stocking of 1.0 is associated with their very small number (Figure 8J). The number and area of such stands account for less than 1% of the number and area of forest stands with sufficient food resources for the Siberian moth (the share of fir is greater than or equal to two units).

There is a considerable discrepancy in the age of stands classified as damaged and the age of stands that were not affected by defoliation (Figure 6A). The probability of defoliation is elevated for stands aged 90–100 years and older that have a greater needle biomass. This is corroborated by the findings from the previous analysis conducted during the outbreak peak stage [26]; younger stands are primarily damaged by the Siberian moth during the decline phase of the outbreak [28], but we assessed the risk of damage during the rise phase.
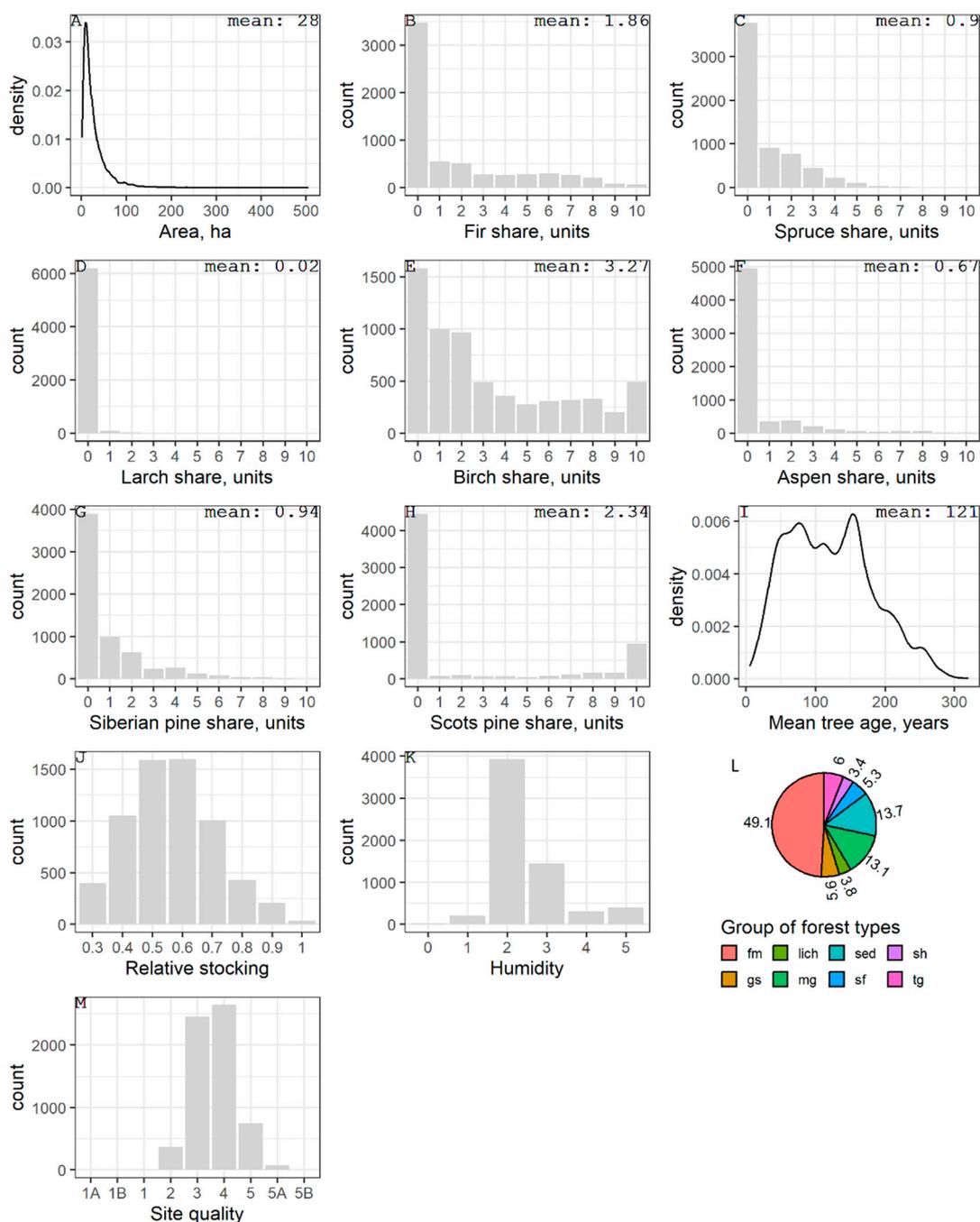
Figure 8. Empirical distributions of the predictors for the area under study. For the predictors measured on a ratio scale, the mean values are also provided. (**A**) mean areas of the forest compartments; (**B**) share of Siberian fir *Abies sibirica* Ledeb. in the stands; (**C**) the same for spruce *Picea obovata* Ledeb.; (**D**) the same for larch *Larix sibirica* Ledeb.; (**E**) the same for birch *Betula* ssp.; (**F**) the same for aspen *Populus tremula* L.; (**G**) the same for Siberian pine *Pinus sibirica* Du Tour; (**H**) the same for Scots pine *Pinus sylvestris* L.; (**I**) mean tree ages; (**J**) relative stockings; (**K**) soil humidities; (**L**) groups of forest types proportions (percents); (**M**) site qualities. The codes used to identify group of forest types: fm—feather moss, gs—grass-swamp, lich—lichen, mg—mixed grass, sed—sedge, sf—sfagnum, sh—shrub, tg—tall-grass.
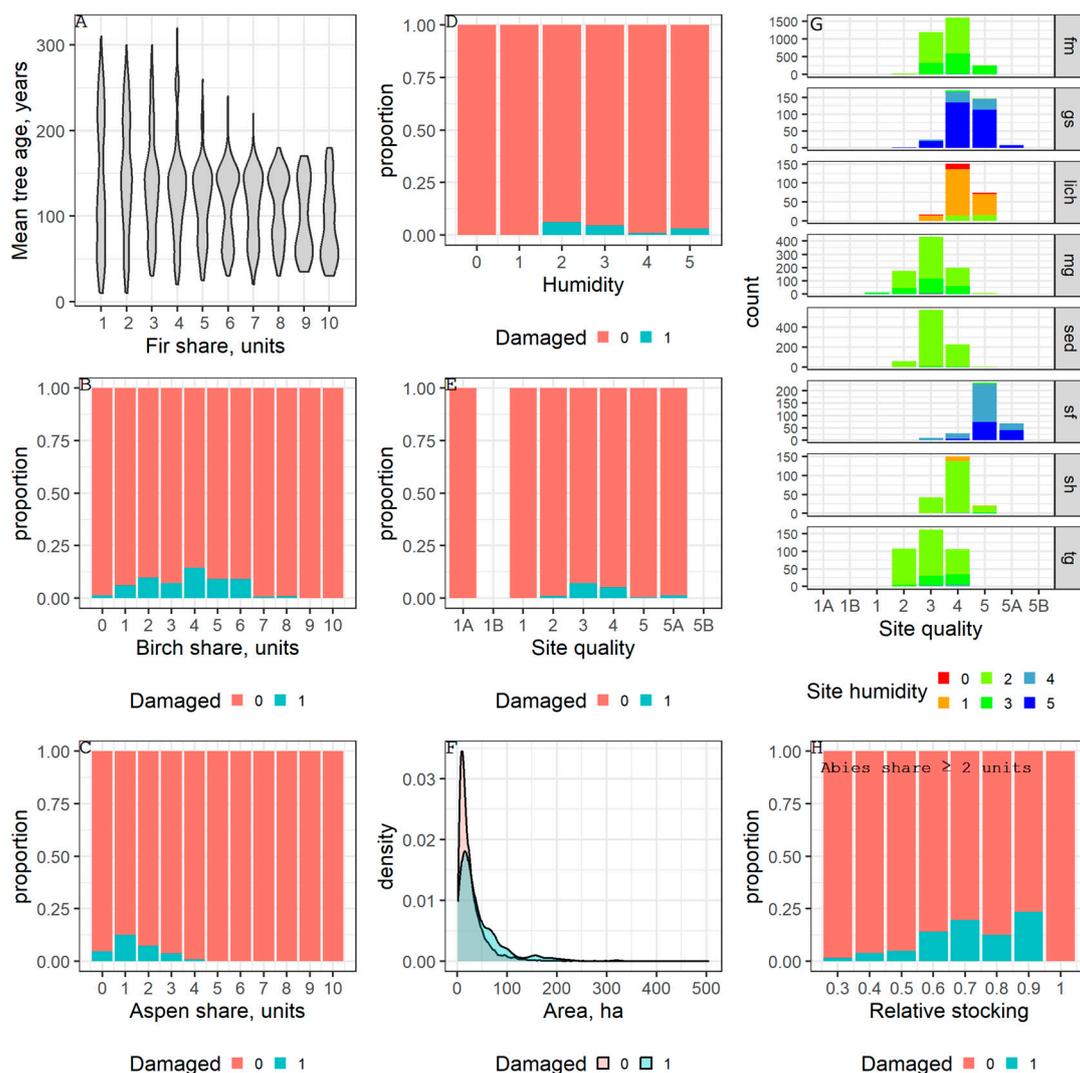
Figure 9. Relationships between some variables. (**A**) mean age and share of fir *Abies sibirica* Ledeb.; (**B**) share of birch *Betula* ssp. and defoliation; (**C**) share of aspen *Populus tremula* L. and defoliation; (**D**) humidity and defoliation; (**E**) site quality and defoliation; (**F**) forest compartment area and defoliation; (**G**) site quality, humidity and group of forest type; (**H**) relative stocking of stands with share of fir ≥ 2 units and defoliation. Group of forest types: fm—feather moss, gs—grass-swamp, lich—lichen, mg—mixed grass, sed—sedge, sf—sfagnum, sh—shrub, tg—tall-grass.

4.3. Possible Role of Historical Circumstances for Risk Assessment

We assume that bimodal distribution of the share of Siberian fir in the damaged stands (Figure 6C) is associated with the presence of stands that were destroyed by the damage caused by the Siberian moth in the 1950s and 1990s [10,42]. The occurrence of *D. sibiricus* outbreaks is often associated with stands that have previously undergone some degree of disturbance [23,24,26]. A notable proportion of the damaged stands are in aged forests, where fir is represented by a few surviving trees that withstood previous defoliation events, and the remainder of the stand is comprised of non-host species that emerged during the succession process. Other damage events were identified in younger, pure fir forests that had not previously been defoliated by the Siberian moth. This assumption is predicated on two factors. First, the higher average age of stands in which the share of fir is comparatively small. Second, the distinctly bimodal distribution of fir ages in a number of cases (Figure 9A).

A considerable proportion of spruce in the stands that were classified as damaged (Figure 6B) necessitates interpretation. It is postulated that the results can also be explained by previous outbreaks of the Siberian moth, which resulted in non-host and less defoliated spruce partially replacing dead Siberian pine and fir. In the 2015–2016 period, *D. sibiricus* defoliated retained or freshly emerged fir trees in these stands for a second time [28].

The risk of damage is significantly influenced by the size of the forest compartment (Figure 5), as evidenced by the tendency for outbreaks to gravitate towards larger forest compartments (Figure 9F). In other words, for the Siberian moth, optimal environmental conditions are created in habitats that are more homogeneous. A similar conclusion was previously reached for other species of eruptive defoliators – *Choristoneura fumiferana* (Clemens) [70] and *C. freemani* Razowski (=*occidentalis* Freeman) [71].

The interpretation of the relationship between distance to previously disturbed stands and defoliation (Figures 4 and 7D) is analogous to that of the forest compartment area. The influence of these two predictors on the probability of damage caused by *D. sibiricus* is partially contingent on the history of stand formation. As a consequence of mortality during outbreaks and subsequent successions, fragmentation of stands increases [26,72]. On the one hand, conditions conducive to a growth of the Siberian moth populations are created in more homogeneous forest sites located at a distance from former outbreak areas with fragmented forest cover. A comparable pattern of damage caused by *D. sibiricus* to previously undisturbed forests was observed in the Baikal region [7]. Conversely, fragmented stands situated in close proximity to previous outbreak areas and other disturbed sites are conducive to the emergence of new outbreak areas, where a favorable microclimate for the Siberian moth develops [26].

4.4. Relief- and Soil-Based Variables as Risk Factor

The group of forest types, soil moisture, and site quality index are closely related. The feather moss and sedge forest type groups, which are most susceptible to damage caused by the Siberian moth (Figure 6F), exhibit a near-optimal moisture regime (scores 2 or 3) and average or low productivity (site quality index 3–4) (Figure 9G). While site quality is indirectly correlated with the habitat suitability for the Siberian moth, the moisture conditions and ground cover directly influence the overwintering success of its caterpillars. The optimal conditions for the moth are found in areas with minimal excess moisture [45] and a thick moss cover [41], which is characteristic of the aforementioned forest type groups [46,73]. This establishes the relatively high predictive value of these three characteristics (Figure 5). Additionally, it elucidates the correlation between classification results and mTPI values (Figure 5). The distribution of mTPI values indicates that the risk of damage caused by the Siberian moth is heightened on elevated, drained relief elements (Figure 7C).

The contribution of the CHILI index, which characterizes incoming solar radiation, is relatively minor in comparison to other predictors derived from RS data and is only comparable to mTPI. Nevertheless, its role as a predictor is noteworthy (Figure 5). The distribution of CHILI values within stands classified as damaged and undamaged was analyzed, and it was found that the probability of defoliation is somewhat higher at CHILI values of approximately 120–130. In regions with lower (CHILI < 120) and higher (CHILI > 130) CHILI values, the risk of damage is lower (Figure 7B). This result is in accordance with the findings of the modeling of the dependence of the occurrence of *D. sibiricus* outbreaks on weather conditions. The analysis of this model demonstrated that both excess and deficiency of heat act as impediments to the occurrence of outbreaks [74].

5. Conclusions

Of the machine learning methods employed, gradient boosting (XGB algorithm) demonstrated the greatest efficacy in predicting the spatial location of stands defoliated by the Siberian moth. The ROC-AUC values of the trained model reached 0.89–0.94, depending on the year and data set.

The work of the models is well interpreted from the perspective of the ecology of *D. sibiricus*. Among the most significant are predictors that delineate the feeding grounds. In the case of ground data, this encompasses the proportion of host and non-host tree species (e.g., Siberian fir and Scots pine). In the case of RS data, this pertains to the vegetation class. Other significant predictors are observed to form local conditions (position in the relief, estimated using the mTPI index) or reflect them. Among the latter, the most significant are the group of forest types (a simplified characteristic of the plant community) and the site quality, which are indirectly related to the wintering conditions of the Siberian moth caterpillars, the CHILI index, and, to a lesser extent, the distance to disturbed stands, which characterize the microclimate. Finally, a number of predictors offer insight into the impact of successional processes on local conditions. Among these, those associated with the restoration of forest stands following previous damage stand out as particularly significant. These include the proportion of spruce in a stand composition, the age of the forest stand, and the distance to the nearest disturbed forest site.

The output of the models is a forecast map of forest stands defoliated by the Siberian moth. The forecast results facilitate an optimized approach to *D. sibiricus* population monitoring.

Author Contributions: Conceptualization, D.A.D.; Methodology, A.A.G., S.M.S., D.A.D. and P.V.M.; Validation, A.A.G. and D.A.D.; Investigation A.A.G., S.M.S., D.A.D., A.I.T. and P.V.M.; Data Curation, A.A.G., D.A.D., P.V.M. and N.N.K.; Writing—Original Draft Preparation, A.A.G., S.M.S., D.A.D. and A.I.T.; Writing—Review and Editing, A.A.G., S.M.S., D.A.D., A.I.T. and N.P.K.; Visualization, A.A.G., D.A.D., E.I.P. and O.A.S.; Project Administration, S.M.S.; Funding Acquisition, P.V.M. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: Data are contained within the article.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Prozorov, S.S. *Small Engrailed Moth Boarmia Bistortata Goeze as a Mass Pest of Siberian Fir*; Siberian Forest Engineering Institute: Krasnoyarsk, Russia, 1955; Collection XI; Issue III, pp. 55–136.
2. Cedervind, J.; Långström, B. Tree Mortality, Foliage Recovery and Top-kill in Stands of Scots Pine (*Pinus sylvestris*) Subsequent to Defoliation by the Pine Looper (*Bupalus piniaria*). *Scand. J. For. Res.* **2003**, *18*, 505–513. [CrossRef]
3. Ierusalimov, E.N. *Zoogenic Defoliation and Forest Community*; KMK Publishing House: Moscow, Russia, 2004.
4. Iqbal, J.; MacLean, D.A.; Kershaw, J.A., Jr. Balsam fir sawfly defoliation effects on survival and growth quantified from permantnt plots and dendrochronology. *Forestry* **2011**, *84*, 349–362. [CrossRef]
5. Moulinier, J.; Lorenzetti, F.; Bergeron, Y. Gap dynamics in aspen stands of the clay belt of northwestern Quebec following a forest tent caterpillar outbreak. *Can. J. For. Res.* **2011**, *41*, 1606–1617. [CrossRef]
6. Chen, C.; Weiskittel, A.; Bataineh, M.; MacLean, D.A. Evaluating the influence of varying levels of spruce budworm defoliation on annualized individual tree growth and mortality in Maine, USA and New Brunswick, Canada. *For. Ecol. Manag.* **2017**, *396*, 184–194. [CrossRef]
7. Florov, D.N. *Pest of Siberian Forests (Siberian Moth)*; OGIZ Irkutsk Book Publishing House: Irkutsk, Russia, 1948.
8. Rozhkov, A.S. *Siberian Silk Moth Outbreaks and Their Control*; Nauka: Moskow, Russia, 1965.

9. Prozorov, S.S. *Siberian Moth Dendrolimus sibiricus Tshtv in the Fir Forests of Siberia*; Siberian Forest Engineering Institute: Krasnoyarsk, Russia, 1953; Collection VII; Issue III, pp. 93–132.

10. Kondakov, Y.P. Regularities of the Siberian moth outbreaks. In *Ecology of Forest Animal Populations in Siberia*; Nauka: Novosibirsk, Russia, 1974; pp. 206–265.

11. Grodnitsky, D.L.; Raznobarsky, V.G.; Soldatov, V.V.; Remarchuk, N.P. Degradation of taiga forests disturbed by the Siberian silkmoth. *Contemp. Probl. Ecol.* **2002**, *1*, 3–12.

12. Kharuk, V.I.; Im, S.T.; Soldatov, V.V. Siberian silkmoth outbreaks surpassed geoclimatic barrier in Siberian Mountains. *J. Mt. Sci.* **2020**, *17*, 1891–1900. [CrossRef]

13. Alexeyev, V.A.; Svyazeva, O.A. *Woody Plants of Russian Forests. A List of Species and the State Account of Biodiversity of Forest Resources*; SB RAS, Sukachev Institute of Forest: Krasnoyarsk, Russia, 2009; ISBN 978-5-94668-063-9.

14. Federal State Statistics Service. Transport. Available online: https://rosstat.gov.ru/statistics/transport (accessed on 3 September 2024).

15. Mason, R.R.; Wickman, B.E. Integrated pest management of the Douglas-fir tussock moth. *For. Ecol. Manag.* **1991**, *39*, 119–130. [CrossRef]

16. Weseloh, R.M. Developing and Validating a Model for Predicting Gypsy Moth (Lepidoptera: Lymantriidae) Defoliation in Connecticut. *J. Econ. Entomol.* **1996**, *89*, 1546–1555. [CrossRef]

17. Isaev, A.S.; Khlebopros, R.G.; Nedorezov, L.V.; Kondakov, Y.P.; Kiselev, V.V.; Sukhovolsky, V.G. *Population Dynamics of Forest Insects*; Nauka: Moscow, Russia, 2001.

18. Sultson, S.M.; Goroshko, A.A.; Verkhovets, S.V.; Mikhaylov, P.V.; Ivanov, V.A.; Demidko, D.A.; Kulakov, S.S. Orographic Factors as a Predictor of the Spread of the Siberian Silk Moth Outbreak in the Mountainous Southern Taiga Forests of Siberia. *Land* **2021**, *10*, 115. [CrossRef]

19. Cooke, B.J.; Nealis, V.G.; Régnière, J. Insect defoliators as periodic disturbances in northern forest ecosystems. In *Plant Disturbance Ecology*; Academic Press: Cambridge, MA, USA, 2021; pp. 423–461.

20. Williams, C.B.; Wenz, J.M.; Dahlsten, D.L.; Norick, N.X. Relation of Forest Site and Stand Characteristics to Douglas-Fir Tussock Moth (Lep. Lymantriidae) Outbreaks in California. *Mitt. Schweiz. Entomol. Ges.* **1979**, *52*, 297–307.

21. Stoszek, K.J.; Mika, P.G.; Moore, J.A.; Osborne, H.L. Relationships of Douglas-Fir Tussock Moth Defoliation to Site and Stand Characteristics in Northern Idaho. *For. Sci.* **1981**, *27*, 431–442.

22. Zhang, B.; MacLean, D.; Johns, R.; Eveleigh, E. Effects of Hardwood Content on Balsam Fir Defoliation during the Building Phase of a Spruce Budworm Outbreak. *Forests* **2018**, *9*, 530. [CrossRef]

23. Kostin, I.A. *On the Siberian Moth Outbreak in the Mountain Forests of Eastern Kazakhstan*; Institute of Zoology of the Academy of Sciences of the Kazakh SSR: Alma Ata, Kazakhstan, 1958; Volume VIII, pp. 122–126.

24. Galkin, G.I. Some issues of the formation of reservations and primary outbreak spots of the Siberian silkmoth in Krasnoyarsk Krai forests. In *The Problem of Siberian Silk Moth (Workshop Reports)*; USSR Academy of Science; USSR Academy of Science Publishing House: Novosibirsk, Russia, 1960; pp. 21–33.

25. Kondakov, Y.P. The Siberian moth outbreaks in the forests of Krasnoyarsk Krai. In *Entomological Researches in Siberia*; KF RES: Krasnoyarsk, Russia, 2002; Volume 2, pp. 25–74.

26. Isaev, A.S.; Ryapolov, V.Y. Analysis of landscape-ecological confinement of the Siberian moth outbreak areas using aerospace photography. In *Study of Taiga Landscapes by Remote Methods*; Nauka: Novosibirsk, Russia, 1979; pp. 152–167.

27. Kharuk, V.I.; Demidko, D.A.; Fedotova, E.V.; Dvinskaya, M.L.; Budnik, U.A. Spatial and Temporal Dynamics of Siberian Silk Moth Large-Scale Outbreak in Dark-Needle Coniferous Tree Stands in Altai. *Contemp. Probl. Ecol.* **2016**, *9*, 711–720. [CrossRef]

28. Demidko, D.A.; Goroshko, A.A.; Slinkina, O.A.; Mikhaylov, P.V.; Sultson, S.M. The Role of Forest Stands Characteristics on Formation of Exterior Migratory Outbreak Spots by the Siberian Silk Moth *Dendrolimus sibiricus* (Tschetv.) during Population Collapse. *Forests* **2023**, *14*, 1078. [CrossRef]

29. Braun-Blanquet, J. *Plant Sociology; the Study of Plant Communities*, 1st ed.; McGraw-Hill Book Company: New York, NY, USA; London, UK, 1932.

30. Sukachev, V.N. *Fundamentals of Forest Typology and Forest Biogeocenology*; Selected Works; Nauka Publ.: Leningrad, Russia, 1972.

31. Fernández-Carrillo, Á.; Franco-Nieto, A.; Yagüe-Ballester, M.J.; Gómez-Giménez, M. Predictive Model for Bark Beetle Outbreaks in European Forests. *Forests* **2024**, *15*, 1114. [CrossRef]

32. Munro, H.M.; Montes, C.R.; Gandhi, K.J.K. A new approach to evaluate the risk of bark beetle outbreaks using multi-step machine learning methods. *For. Ecol. Manag.* **2022**, *520*, 120347. [CrossRef]

33. Rammer, W.; Seidl, R. Harnessing Deep Learning in Ecology: An Example Predicting Bark Beetle Outbreaks. *Front. Plant Sci.* **2019**, *10*, 1327. [CrossRef]

34. *Atlas of Krasnoyarsk Krai and the Republic of Khakassia*; Isaev, A.S., Ed.; Roskartografiya: Moscow, Russia, 1994.

35. Gorbatenko, V.P.; Ippolitov, I.I.; Kabanov, M.V.; Loginov, S.V.; Podnebesnych, N.V.; Kharyutkina, E.V. Effect of atmospheric circulation on temperature variations in Siberia. *Atmos. Ocean. Opt.* **2011**, *24*, 15–21.

36. Ippolitov, I.I.; Kabanov, M.V.; Loginov, S.V.; Podnebesnykh, N.V.; Kharyutkina, E.V.; Gorbatenko, V.P. Influence of atmospheric circulation on the temperature regime of Siberia. *Opt. Atmos. Ocean* **2011**, *24*, 15–21.

37. Kamanin, L.G.; Likhanov, B.N. (Eds.) *Central Siberia*; Nauka: Moscow, Russia, 1964.

38. Forests of the Urals, Siberia and the Russian Far East. In *Forests of the USSR*; Zhukov, A.B., Ed.; Nauka: Moscow, Russia, 1969; Volume 4.

39. Pavlov, I.N.; Litovka, Y.A.; Golubev, D.V.; Astapenko, S.A.; Khromogin, P.V. New outbreak of *Dendrolimus sibiricus* Tschetv. in Siberia (2012–2017): Patterns of development and prospects for biological control. *Cont. Probl. Ecol.* **2018**, *4*, 462–478.

40. Kononov, A.; Ustyantsev, K.; Wang, B.; Mastro, V.C.; Fet, V.; Blinov, A.; Baranchikov, Y. Genetic Diversity among Eight Dendrolimus Species in Eurasia (Lepidoptera: Lasiocampidae) Inferred from Mitochondrial COI and COII, and Nuclear ITS2 Markers. *BMC Genet.* **2016**, *17*, 157. [CrossRef] [PubMed]

41. Rozhkov, A.S. *Siberian Silk Moth*; Nauka: Moskow, Russia, 1963.

42. Kharuk, V.I.; Im, S.T.; Ranson, K.J.; Yagunov, M.N. Climate-Induced Northerly Expansion of Siberian Silkmoth Range. *Forests* **2017**, *8*, 301. [CrossRef]

43. Kirichenko, N.I.; Baranchikov, Y.N. Feeding norms for the Siberian moth caterpillars on coniferous tree species in Siberia. *Contemp. Probl. Ecol.* **2008**, *5*, 709–716.

44. Okunev, P.P. Geographical position and hazardous areas of the Siberian moth. In *Geographical Collection. V. Geographical Issues of Forestry*; USSR Academy of Science Publishing House: Moskow/Leningrad, Russia, 1955; pp. 210–222.

45. Konikov, A.S.; Platonova-Chernysheva, L.V.; Kondakov, Y.P.; Zaitseva, A.I. Adaptation of the Siberian moth to environmental conditions. To the characteristics of the factors determining the number of the Siberian moth. In *Scientific Notes of the Krasnoyarsk Pedagogical Institute*; Krasnoyarsk State Pedagogical Institute: Krasnoyarsk, Russia, 1959; Volume XV, pp. 145–175.

46. Rysin, L.P. *Five-Needle Pines Forests of Russia*; KMK Publishing House: Moscow, Russia, 2011; ISBN 978-5-87317-771-4.

47. Rysin, L.P.; Savel'eva, L.I. *Scots Pine-Dominated Forests of Russia*; KMK Publishing House: Moscow, Russia, 2008; ISBN 978-5-87317-512-3.

48. *14688-1:2017*; Geotechnical Investigation and Testing—Identification and Classification of Soil—Part 1: Identification and Description. ISO: Geneva, Switzerland, 2017.

49. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2022.

50. RStudio Team. *RStudio: Integrated Development Environment for R*; RStudio, PBC: Boston, MA, USA, 2022.

51. Wickham, H.; François, R.; Henry, L.; Müller, K. Dplyr: A Grammar of Data Manipulation. Available online: https://dplyr.tidyverse.org (accessed on 5 March 2022).

52. Pebesma, E. Simple Features for R: Standardized Support for Spatial Vector Data. *R J.* **2018**, *10*, 439–446. [CrossRef]

53. Zhao, Q.; Yu, L.; Li, X.; Peng, D.; Zhang, Y.; Gong, P. Progress and trends in the application of Google Earth and Google Earth Engine. *Remote Sens.* **2021**, *13*, 3778. [CrossRef]

54. Theobald, D.M.; Harrison-Atlas, D.; Monahan, W.B.; Albano, C.M. Ecologically-Relevant Maps of Landforms and Physiographic Diversity for Climate Adaptation Planning. *PLoS ONE* **2015**, *10*, e0143619. [CrossRef]

55. Gorelick, N.; Hancher, M.; Dixon, M.; Ilyushchenko, S.; Thau, D.; Moore, R. Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sens. Environ.* **2017**, *202*, 18–27. [CrossRef]

56. Bartalev, S.A.; Egorov, V.A.; Zharko, V.O.; Lupyan, E.A.; Plotnikov, D.E.; Khvostikov, S.A.; Shabanov, N.V. *Satellite Mapping of the Vegetation Cover of Russia*; ISR RAS: Moscow, Russia, 2016.

57. Hansen, M.C.; Potapov, P.V.; Moore, R.; Hancher, M.; Turubanova, S.A.; Tyukavina, A.; Thau, D.; Stehman, S.V.; Goetz, S.J.; Loveland, T.R.; et al. High-Resolution Global Maps of 21st-Century Forest Cover Change. *Science* **2013**, *342*, 850–853. [CrossRef]

58. Hijmans, R.J. Terra: Spatial Data Analysis. Available online: https://rspatial.github.io/terra/ (accessed on 10 March 2022).

59. Didan, K.; Munoz, A.B.; Solano, R.; Huete, A. *MODIS Vegetation Index User's Guide (MOD13 Series)*; Version 3.0; University of Arizona: Tucson, AZ, USA, 2015.

60. Giglio, L.; Boschetti, L.; Roy, D.P.; Humber, M.L.; Justice, C.O. The Collection 6 MODIS burned area mapping algorithm and product. *Remote Sens. Environ.* **2018**, *217*, 72–85. [CrossRef]

61. United States Geological Survey. Available online: https://earthexplorer.usgs.gov/ (accessed on 11 October 2022).

62. Lang, M.; Binder, M.; Richter, J.; Schratz, P.; Pfisterer, F.; Coors, S.; Au, Q.; Casalicchio, G.; Kotthoff, L.; Bischl, B. Mlr3: A Modern Object-Oriented Machine Learning Framework in R. J. *Open Source Softw.* **2019**, *4*, 1903. [CrossRef]

63. Biecek, P. DALEX: Explainers for Complex Predictive Models in R. *J. Mach. Learn. Res.* **2018**, *19*, 1–5.

64. Wickham, H. *Ggplot2: Elegant Graphics for Data Analysis*; Springer: New York, NY, USA, 2016; ISBN 978-3-319-24277-4.

65. Talukdar, S.; Singha, P.; Mahato, S.; Shahfahad; Pal, S.; Liou, Y.-A.; Rahman, A. Land-Use Land-Cover Classification by Machine Learning Classifiers for Satellite Observations—A Review. *Remote Sens.* **2020**, *12*, 1135. [CrossRef]

66. Thiele, C.; Hirschfeld, G. cutpointr: Improved Estimation and Validation of Optimal Cutpoints in R. *J. Stat. Softw.* **2021**, *98*, 1–27. [CrossRef]

67. Biecek, P.; Burzykowski, T. *Explanatory Model Analysis: Explore, Explain, and Examine Predictive Models*; Chapman & Hall/CRC: New York, NY, USA, 2021; ISBN 978-0367135591.

68. Lemaire, J.; Vennetier, M.; Prévosto, B.; Cailleret, M. Interactive effects of abiotic factors and biotic agents on Scots pine dieback: A multivariate modeling approach in southeast France. *For. Ecol. Manag.* **2022**, *526*, 120543. [CrossRef]

69. Nicholson, A.J. An outline of the dynamics of animal populations. *Aust. J. Zool.* **1954**, *2*, 9–65. [CrossRef]

70. Robert, L.-E.; Sturtevant, B.R.; Cooke, B.J.; James, P.M.A.; Fortin, M.-J.; Townsend, P.A.; Wolter, P.T.; Kneeshaw, D. Landscape host abundance and configuration regulate periodic outbreak behavior in spruce budworm *Choristoneura fumiferana*. *Ecography* **2018**, *41*, 1556–1571. [CrossRef]

71. Ellis, T.M.; Flower, A. A multicentury dendrochronological reconstruction of western spruce budworm outbreaks in the Okanogan Highlands, northeastern Washington. *Can. J. For. Res.* **2017**, *47*, 1266–1277. [CrossRef]

72. Coops, N.C.; Gillanders, S.N.; Wulder, M.A.; Gergel, S.E.; Nelson, T.; Goodwin, N.R. Assessing changes in forest fragmentation following infestation using time series Landsat imagery. *For. Ecol. Manag.* **2010**, *259*, 2355–2365. [CrossRef]

73. Pologova, N.N.; Chernova, N.A.; Klimova, N.W.; Duykarev, A.G. Diversity of Siberian pine forests related to their habitat. *Russ. J. For. Sci.* **2013**, *4*, 32–42.

74. Demidko, D.A.; Goroshko, A.A.; Sultson, S.M.; Kulakova, N.N.; Mikhaylov, P.V. Weather Data-Based Prediction of the Siberian Moth *Dendrolimus sibiricus* Tschetv.: A Case Study. *Contemp. Probl. Ecol.* **2024**, *17*, 379–392. [CrossRef]