NASA DEVELOP National Program
California – Ames

Spring 2022

Vermont & New Hampshire Ecological Forecasting

Monitoring Trends in Tree Defoliation Due to *Lymantria dispar* Outbreaks to Predict Future Hardwood Tree Mortality and Health Impacts

DEVELOP Technical Report

Final – March 31st, 2022

Seamore Zhu (Project Lead)
Katie Caruso
Morgan Dean
Jacob Orser

Advisors:
Juan Torres-Pérez, NASA Ames Research Center (Science Advisor)
Valerie Pasquarella, Boston University (Science Advisor)

1. Abstract

The invasive, herbivorous insect *Lymantria dispar* is a major defoliator of hardwood trees in the northeastern United States. Established populations of *L. dispar* typically rest at low levels but undergo recurring outbreaks that cause tree mortality if they occur in quick succession or are combined with other stressors on tree health. Widespread defoliation events disrupt local wildlife, economies, and livelihoods. Accurate monitoring of defoliation events is necessary to implement effective land management practices that support tree health. There are challenges to accurately monitoring *L. dispar* outbreaks that include the ephemeral character of defoliation disturbances and the difficulties of conducting large-scale surveys of *L. dispar* populations using existing aerial and ground-based data collection methods. To better monitor the impact of *L. dispar* on forests in Vermont and New Hampshire, the NASA DEVELOP team partnered with organizations responsible for supporting land and invasive species management including the Forest Ecosystem Monitoring Cooperative, the Vermont Agency of Agriculture, Food and Markets, the University of New Hampshire Cooperative Extension, and the New Hampshire Division of Forests and Lands, Forest Health Program. The team used NASA Earth observations collected by the Terra, Soil Moisture Active Passive (SMAP), Shuttle Radar Topography Mission (SRTM), Landsat 7, and Landsat 8 satellites along with ancillary datasets to map historical tree defoliation from 2012 to 2021. In support of partners' future land management efforts, the team created a Google Earth Engine tool that displays annual defoliation extent.

Key Terms

remote sensing, Landsat, ecological forecasting, invasive insect, Lymantria dispar, Google Earth Engine

2. Introduction

2.1 Background

Lymantria dispar, commonly known as the spongy moth, is an insect native to Europe and Asia that was introduced to the Northeast United States in the late 1860s. *Lymantria dispar* is a generalist herbivorous feeder that prefers hardwood trees but feeds on other plant species if their preferred hosts are defoliated (Swier et al., 2016). Due to the slow spread of *L. dispar*, which occurs across large distances primarily through wind dispersal of larvae suspended on silk threads, only 40 percent of the potential range of *L. dispar* in the U.S. has experienced defoliation events. (Elkinton & Liebhold, 1990; Liebhold et al., 2000). *Lymantria dispar* populations are held in check by various factors. One primary control is a non-native fungus *Entomophaga maimaiga*, which attacks moth larvae. Other factors include forest predators such as small mammals, other fungi, and parasites that infect moths (Hajek et al., 1995). In recent years, Vermont (VT) and New Hampshire (NH) have experienced multiple events of extreme defoliation. In New Hampshire, state officials estimated 2021 total defoliation in the range of 50,000 acres (Cherry, 2021).

Outbreak events occur in periodic cycles. The reasons behind outbreak events are studied but are relatively difficult to track because pre-outbreak populations can remain at low levels for years. An outbreak is defined as over 5000 egg masses of *L. dispar* per hectare (Liebhold et al., 2000). The concern in VT and NH is that outbreaks reduce annual tree growth, lower timber values, decrease fruit production, and decrease habitat (Liebhold et al., 2000). Consecutive years of outbreaks, particularly when combined with unsuitable weather, poor soil, and low tree vigor, may cause tree mortality. Mast failure and tree mortality lead to cascading mortality within the ecosystem.

The ephemeral character of these disturbance events, combined with difficulties of accurately surveying *L. dispar* populations, poses challenges to monitoring defoliation caused by *L. dispar* outbreaks. Remote sensing is a cost-effective tool with the potential to improve monitoring efforts in VT and NH forests. Monitoring *L. dispar* outbreaks using remote sensing can support invasive species management efforts by revealing outbreak dynamics and improving predictive models of defoliator abundance and risk to forests (Pasquarella et al., 2021). However, a limitation of previous remote sensing models is their reliance on a single index to measure and monitor defoliation events. Prior remote sensing models rely on measuring change in greenness to

identify potential defoliation. The goal of this project was to improve defoliation identification by using a Random Forest model, a supervised classification method, that allows the input of multiple predictor variables to classify potential defoliation events.

For this feasibility project, the study area was limited to VT and NH (Figure 1). This region is defined by its mixed forests of broadleaf, deciduous, and boreal tree species. Mixed hardwood forests of maple, birch, and beech, as well as northeastern spruce-fir forests, dominate VT and NH forests. We worked with partners to reveal trends before, during, and after large defoliation events during the selected 2012 to 2021 study period.

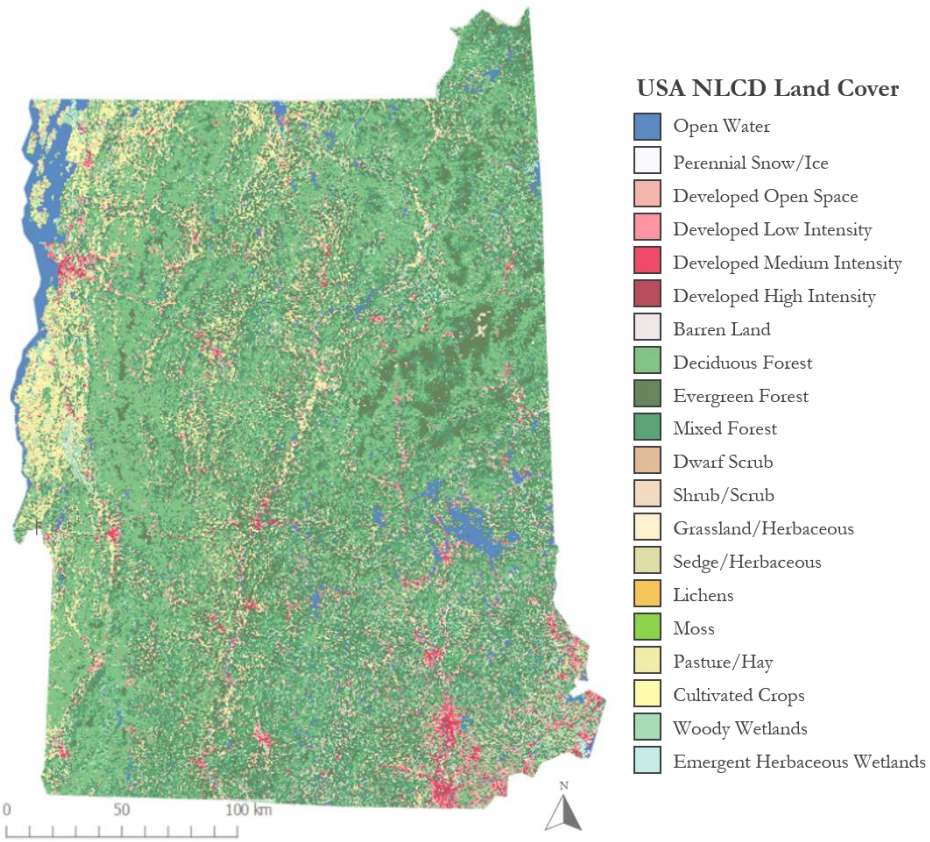

Figure 1. Vermont and New Hampshire deciduous forests, mixed deciduous-evergreen forests, scrubland, and woody wetlands.

2.2 Project Partners & Objectives

Our DEVELOP team partnered with the Forest Ecosystem Monitoring Cooperative (FEMC), the Vermont Agency of Agriculture, Food and Markets (VAAFM), the University of New Hampshire Cooperative Extension (UNH Extension), and the New Hampshire Division of Forests and Lands, Forest Health Program in our effort to monitor trends in defoliation by *L. dispar* in VT and NH. FEMC monitors northeastern temperate forest regions to track long-term trends in ecosystem health. VAAFM is involved with protecting VT's agriculture from threats including invasive insects like *L. dispar*. The UNH Extension protects NH's natural resources by providing communities with education on forest health and invasive pest management. The Forest Health Program with the NH Division of Forests and Lands addresses forest pest challenges by monitoring damage using aerial surveys and ground observations as well as working alongside landowners.

The partner organizations have used aerial surveys to locate *L. dispar* outbreaks and subsequent defoliation. However, they are interested in other forms of modeling that accurately identify defoliation extent in VT and NH, particularly following low-severity defoliation events which are less likely to be captured using aerial survey methods. Quantitative assessments of annual change in defoliation over this period can support partners in understanding defoliation trends as well as the environmental and geographical factors that may be correlated with outbreak events. The objectives of this project were to provide partners with potential defoliation extent maps for 2012–2021 and a Google Earth Engine (GEE) tool for comparing defoliation extent across the ten-year study period. We also created an ArcGIS StoryMap to communicate the project's results to the partner organizations and the general public in support of project partners' outreach and education efforts.

3. Methodology

3.1 Data Acquisition

To generate annual defoliation extent maps for 2012–2021, we classified defoliation severity across the VT and NH study area using Random Forest classification in GEE. The Topographically Integrated Geographic Encoding and Referencing (TIGER) U.S. Census States 2018 dataset was used to delineate the VT and NH study area geometry in GEE. Next, in the absence of ground data on *L. dispar* defoliation events during our study period, we gathered the following datasets to generate 7337 points of defoliation events that occurred in our study area and period to train our Random Forest model to identify defoliation. First, we utilized aerial surveys conducted in 2017, 2018, and 2021 by the Forest Health Program with the NH Division of Forests and Lands to measure defoliation caused by *L. dispar,* which were contained in shapefiles with geometries indicating observed defoliation and their associated severity rating (light, moderate, or severe defoliation). The landcover band of the 2016 National Land Cover Database (NLCD) was used to mask out regions of our study area that were not deciduous forests, mixed deciduous-evergreen forests, scrubland, and woody wetlands and, therefore, unlikely to experience defoliation during *L. dispar* outbreaks. For the survey years, we used harmonic condition-monitoring scripts authored by Dr. Valerie Pasquarella to create layers that identified decreases in greenness in our study area, which indicate additional areas of potential defoliation due to *L. dispar*.

Once we generated the training points necessary to create a training dataset, we gathered environmental predictor variables that may contribute to *L. dispar* outbreaks for every year from 2012 to 2021 to improve the accuracy of the Random Forest defoliation classifier. The vegetation indices that functioned as predictor variables were derived from imagery collected by the Landsat 8 Operational Land Imager (OLI) and Landsat 7 Enhanced Thematic Mapper Plus (ETM+). We refined the time period during which we measured summer indices from June 1 to September 1 to improve the likelihood of capturing defoliation caused by *L. dispar*, which peaks mid-June through late July, while generating enough images to cover our study area after cloud filtering and masking. Spring indices from April 1 to June 1 were also collected to help the classifier differentiate winter logging from summer defoliation. Surface temperature data, which were acquired from the Landsat 8 Thermal Infrared Sensor (TIRS) and Landsat 7 ETM+, may play a role in limiting the extent of *L. dispar* and, therefore, defoliation (Reilly et al., 2014). Elevation, slope, and aspect impact *L. dispar* population growth (Walter et al., 2015) and were gathered from Version 4 of the Shuttle Radar Topography Mission (SRTM) dataset. Spring precipitation, drought severity, humidity, and soil moisture all have implications for *E. maimaiga*, a fungal agent that prefers a moist environment and regulates populations of *L. dispar* (Pasquarella et al., 2017; Reilly et al., 2014). We obtained precipitation data from the Climate Hazards Group InfraRed Precipitation with Station (CHIRPS) and the Global Precipitation Measurement (GPM) Integrated Multi-satellitE Retrievals for GPM (IMERG). We sourced the Palmer Drought Severity Index (PDSI) and specific humidity from the University of Idaho's Gridded Surface Meteorological (gridMET) dataset, and soil moisture measurements from the Soil Moisture Active Passive (SMAP) mission. Data on snow cover, which insulates egg masses from low temperature and thereby increases winter survival of egg masses at ground level (Andresen et al., 2001), were retrieved from the Terra Moderate Resolution Imaging

Spectroradiometer (MODIS). The aforementioned Earth observations of precipitation, drought severity, snow cover, and humidity are available in Google Earth Engine. Soil nitrogen concentrations are also negatively correlated with *L. dispar* defoliation and potential tree mortality (Conrad-Rooney et al., 2021). As the prediction precision of soil nitrogen content decreases with soil depth (Wang et al., 2017), we selected mean nitrogen content at a depth of 5-15cm from the International Soil Reference and Information Centre (ISRIC) World Soil Information dataset.

Next, we gathered all predictor variables by year, creating one set to describe conditions for each of the 10 years—so, 10 sets in total. Climate data including precipitation, drought severity, humidity, and soil moisture were filtered annually from May 1 to August 1 to capture the spring season, a period when moisture affects the survival and summer abundance of *E. maimaiga* (Pasquarella et al., 2017; Reilly et al., 2014). Snow cover was filtered annually to winter months from December 1 to March 1, when snow and cold temperatures typically occur throughout the two states. The Existing Vegetation Type (EVT) classification from the U.S. Geological Survey provided a detailed distinction between vegetation types in our study area, which may correlate to defoliation due to *L. dispar*'s host preferences. Finally, the mean Tasseled Cap Greenness (TCG) change scores derived from Dr. Valerie Pasquarella's harmonic condition monitoring scripts were brought in as a predictor variable because greenness decrease is very closely tied to defoliation.

Table 1
Predictor variables acquired for classifying defoliation

Index	Predictor Variable	Source	Native Resolution	Image Dates
1	Tasseled Cap Greenness (TCG)	Landsat 8 OLI Landsat 7 ETM+	30m	4/1–6/1 (2012–2021)
2				6/1–9/1 (2012–2021)
3	Normalized Difference Vegetation Index (NDVI)	Landsat 8 OLI Landsat 7 ETM+	30m	4/1–6/1 (2012–2021)
4				6/1–9/1 (2012–2021)
5	NDVI Spring to Summer Difference	Landsat 8 OLI Landsat 7 ETM+	30m	4/1–6/1 (2012–2021) 6/1–9/1 (2012–2021)
6	NIR/Red (Simple Ratio)	Landsat 8 OLI Landsat 7 ETM+	30m	6/1–9/1 (2012–2021)
7	Enhanced Vegetation Index (EVI)	Landsat 8 OLI Landsat 7 ETM+	30m	6/1–9/1 (2012–2021)
8	Surface Temperature	Landsat 8 OLI Landsat 7 ETM+	30m	6/1–9/1 (2012–2021)
9	Elevation	SRTM	30m	N/A
10	Slope	SRTM	30m	N/A
11	Aspect	SRTM	30m	N/A
12	Precipitation	CHIRPS	5566m	5/1–8/1 (2012–2021)
13	PDSI	gridMET	4638.3m	5/1–8/1 (2012–2021)
14	Specific Humidity	gridMET	4638.3m	5/1–8/1 (2012–2021)
15	Snow Cover	Terra MODIS	250m	12/1–3/1 (2011–2021)
16	Surface Soil Moisture	SMAP	10000m	5/1–8/1 (2012–2021)
17	Nitrogen Content	ISRIC World Soil Information	250m	N/A

18	Existing Vegetation Type	USGS	30m	N/A
19	Mean TCG Change Score	Harmonic Condition Monitoring Project (Contact: Valerie Pasquarella)	30m	6/1–9/1 (2012–2021)

3.2 Data Processing

To process vegetation indices for each year, we began by removing images with greater than 80 percent cloud cover and applied a bitmask to mask out clouds, cloud shadows, and snow from images that remained. Using data contained in spectral bands, we calculated vegetation indices to incorporate them into the Random Forest model. Due to cloud cover in this area, we had to create mosaics for each of the indices described in Table 2.

To determine the health of vegetation in the study area, the Normalized Difference Vegetation Index (NDVI) was calculated in GEE as the ratio of the difference between portions of the red and near-infrared spectrum over the sum of portions of the red and near-infrared spectrum (Table 2, Vermote et al., 2016). We further evaluated the properties of green vegetation by calculating Tasseled Cap Greenness (TCG; Table 2; Baig et al., 2014). Changes in vegetation water content were calculated with the Normalized Difference Water Index (NDWI; Table 2; Gao, 1996). We also calculated the Enhanced Vegetation Index (EVI) and Simple Ratio, two other vegetation indices (Pasquarella et al., 2021; Vermote et al., 2016). To reduce the likelihood that the Random Forest model would attribute changes in greenness to logging rather than defoliation, we calculated the difference between spring and summer NDVI for each year (Table 2; Choi et al., 2021).

Table 2
Equations for deriving indices used as predictor variables for Random Forest classification.

Predictor Variable	**Equation**
Tasseled Cap Greenness (TCG; Baig et al., 2014)	$(-0.2848 \times BLUE) - (0.2435 \times GREEN) - (0.5436 \times RED) + (0.7243 \times NIR) + (0.0840 \times SWIR1) - (0.1800 \times SWIR2)$
Normalized Differenced Vegetation Index (NDVI; Vermote et al., 2016)	$\frac{NIR - RED}{NIR + RED}$
NDVI Spring to Summer Difference (Choi et al., 2021)	$R_i - G_i$ *Where R is the mean NDVI value of pixel i in the summer, And G is the mean NDVI value of pixel i in the spring*
Normalized Difference Water Index (NDWI; Gao, 1996)	$\frac{GREEN - NIR}{GREEN + NIR}$
Simple Ratio (NIR/Red) (Pasquarella et al., 2021)	$\frac{NIR}{RED}$
Enhanced Vegetation Index (EVI; Vermote et al., 2016)	$2.5 \times \frac{NIR - RED}{NIR + 6 \times RED - 7.5 \times BLUE + 1}$

Calculations were required for the environmental condition variables such as nitrogen content, soil moisture, snow cover, specific humidity, PDSI, precipitation, surface temperature, and mean TCG. Relatively stable topography variables such as existing vegetation type, slope, aspect, and elevation did not require processing because of minimal variation from year to year.

To create the training set used to classify defoliation in the Random Forest model, we generated random points within the intersections of the aerial survey geometries and deciduous masked areas for each year from 2012 to 2021. We defined classes of defoliation severity by mean TCG change scores where a lower, more negative value corresponds to less greenness or more defoliation. 'Severe' was between -4500 and -3500, 'Moderate' was between -3500 and -2500, and 'Light' was between -2500 and -1500. Points in deciduous areas with scores outside of these ranges were defined as 'No Defoliation' points. Logging areas from 2017, 2018, and 2021 were located using June 1 through September 1 true color imagery from the Sentinel-2 MultiSpectral Instrument (MSI). Random points from these areas were also included as points to train the classifier to differentiate between defoliation and logging. The training dataset from all three years totaled 7337 points throughout the two states.

Predictor variables from 2017, 2018, and 2021 were assigned to corresponding points from our training set so that each point in each year included information on predictor variables values for that specific location and year. These points were then used to train the Random Forest classifier.

3.3 Data Analysis

3.3.1 Random Forest Classification

To classify defoliation severity over our study area, we assigned corresponding predictor variable values to all our training points by year. Then, we trained a Random Forest classifier with ten trees on our training sample data. We then applied the classification model to the images containing the predictor variables for each year from 2012 to 2021 to generate the final annual defoliation extent maps. In the absence of ground-truth data on observed defoliation during our study years, we were not able to conduct a formal validation process to test the accuracy of our predicted defoliation classifications. Instead, we produced a confusion matrix representing performance in classifying our training points correctly (Table 3). We generated an accuracy score of 0.989 based on the number of correctly classified points over the total. The Random Forest classification function within Google Earth Engine also calculates scores representing the predictor variable importance[1]. Our predictor variables in order of importance for classification were the mean TCG change score, spring NDVI, spring TCG, summer NDVI, slope, surface temperature, spring to summer NDVI difference, elevation, nitrogen content, EVI, aspect, Simple Ratio, summer TCG, precipitation, snow cover, specific humidity, PDSI, soil moisture, and EVI.

Table 3
Random Forest Classification Confusion Matrix

Actual \ Classified	**No Defoliation**	**Light**	**Moderate**	**Severe**
No Defoliation	4254	6	1	0
Light	16	601	22	0
Moderate	2	12	953	0
Severe	1	0	4	42

*3.3.2 Condition Assessment Tool to Estimate Patterns of the Invasive Lymantria dispa*R

To improve defoliation monitoring in VT and NH, our NASA DEVELOP team created the Condition Assessment Tool to Estimate Patterns of the Invasive Lymantria dispaR (CATEPILR) on the GEE Interface API. CATEPILR consists of an interactive graphical user interface (GUI) that enables users to view and compare changes in forest conditions between any two years of the study period. Within CATEPILR, users can create Random Forest prediction maps based on a selection by year. Further functionality enables the user to look at the differences in Random Forest predictor variables over time.

4. Results & Discussion

4.1 Analysis of Results

To measure trends in potential defoliation due to *L. dispar* across VT and NH, we mapped the total extent of potential defoliation from 2012 to 2021 (Figure 2) as well as the annual extent of defoliation across the ten-year study period.

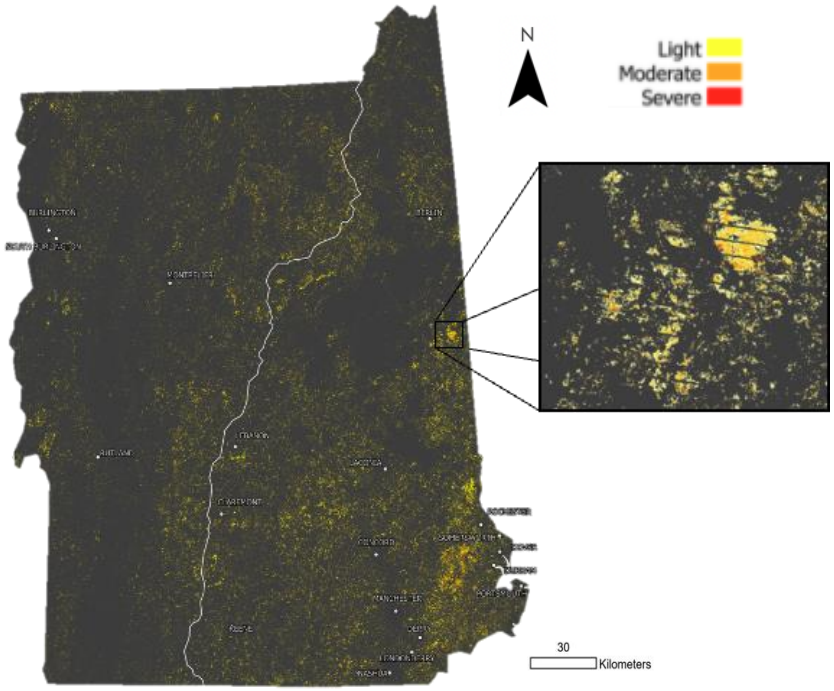

Figure 2. Total potential defoliation extent from 2012 to 2021 across the VT and NH study area, binned into three severity classes.

Most of the areas (69.57 percent) where potential defoliation occurred from 2012 to 2021 were categorized as light severity. In contrast, a minority of areas (<1 percent) of the total extent of defoliation experienced severe potential defoliation (Table A1). Among the years where potential defoliation was measured, 2021 experienced the greatest extent of potential defoliation, with nearly 800 km² affected across the total 49,114 km² VT and NH study area (Table 4).

Table 4
Annual historical potential defoliation extent between 2012–2021.

Time Period	**Defoliation Extent by Severity Class (km²)**			**Total Defoliation (km²)**
	Light	*Moderate*	*Severe*	
2012	279	95	0.08	375
2013	217	187	0.09	405
2014	42	47	13	102
2015	147	52	0.3	199
2016	143	33	0.2	176
2017	16	21	6.3	43
2018	230	45	0.2	275
2019	155	61	0.02	216
2020	408	142	0.6	551
2021	548	249	1.3	799

We also calculated total defoliation on state and federal protected lands in 2021. These layers were sourced from the USGS Protected Areas Database of the United States (PAD-US; Gap Analysis Project Status Codes 1-3). We found that in Vermont, 12.8 km² of the total 165.5 km² potential tree cover defoliation occurred on protected lands, consisting of 7.73 percent of total potential defoliation in the state. In New Hampshire, 47.3 km² of the total 633.0 km² potential defoliation fell within protected lands, constituting 7.47 percent of total defoliation in the state. Total defoliation on protected lands for both states totaled 60.1 km² or 7.52 percent of total potential defoliation in the two states (Table 5).

Table 5
Total potential defoliation on state and federal protected lands in 2021

Potential Defoliation Extent	Vermont	New Hampshire	Both States
Total defoliation extent (km²)	165.5	633.0	798.5
Overlap with protected lands (km²)	12.8	47.3	60.1
Overlap with protected lands (%)	7.73	7.47	7.52

These areas represent protected land in danger of degradation due to defoliation. Repeat outbreaks may threaten the health and viability of deciduous trees in these areas and have repercussions for the entire ecosystem. As some of these protected lands are managed for limited multiple uses including logging, *L. dispar* damage may decrease the quality and abundance of lumber, which could lead to negative economic impacts. Viewing the 2021 potential defoliation map may help in the assessment of management strategies on protected lands.

One of the community concerns is forest health. Tree mortality may occur after multiple defoliation events or an intensive defoliation event in combination with other stressors. To understand the dynamics of forest health, we performed analysis of repeated defoliation to identify patterns of *L. dispar* outbreaks in VT and NH. We calculated the sum of potential defoliation events per pixel to gain a sense of overall defoliation frequency in the past decade (Figure 3).

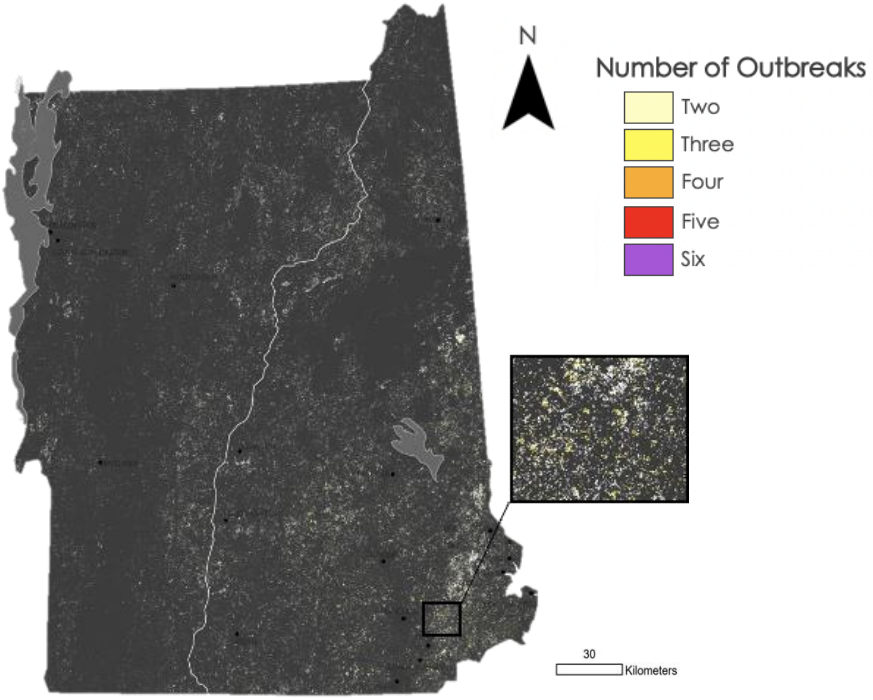

Figure 3. Total potential defoliation years per pixel across the study period.

We can identify that the spongy moth spreads over time. 1615.911 km² were affected by potential defoliation between 2012 and 2021. This corresponds to just under 400,000 acres. Figure 3 helps to illustrate the vast area that *L. dispar* outbreaks can affect within a 10-year time frame. There are areas of repeated defoliation events, which occupy less than 250 km², or roughly 62,000 acres (Figure A1). Figure 3 and Table 6 help to illustrate the repetition of *L. dispar* outbreaks.

Table 6
Extent of repeated historical potential outbreaks in the same area

Number of Outbreaks	**Area Affected (km²)**
One	2616
Two	220
Three	25
Four	3.1
Five	0.31
Six	0.017

We also identified the extent to which outbreaks impact particular areas. This analysis was conducted by summing severity across our study area and period to identify areas impacted heavily by defoliation. In VT and NH from 2012 to 2021, most areas experienced a single light-severity defoliation event, but some areas experienced heavy repeated defoliation. Severity scores are some combination of defoliation events. For example, a severity score of one corresponds to one light defoliation event where a severity score of two could either be one moderate defoliation event or two light defoliation events. A severity score of ten is a combination of multiple defoliation events at different severities; this could be 10 years of light defoliation, 3 years of severe defoliation and one year of light defoliation, or another combination of yearly severity scores.

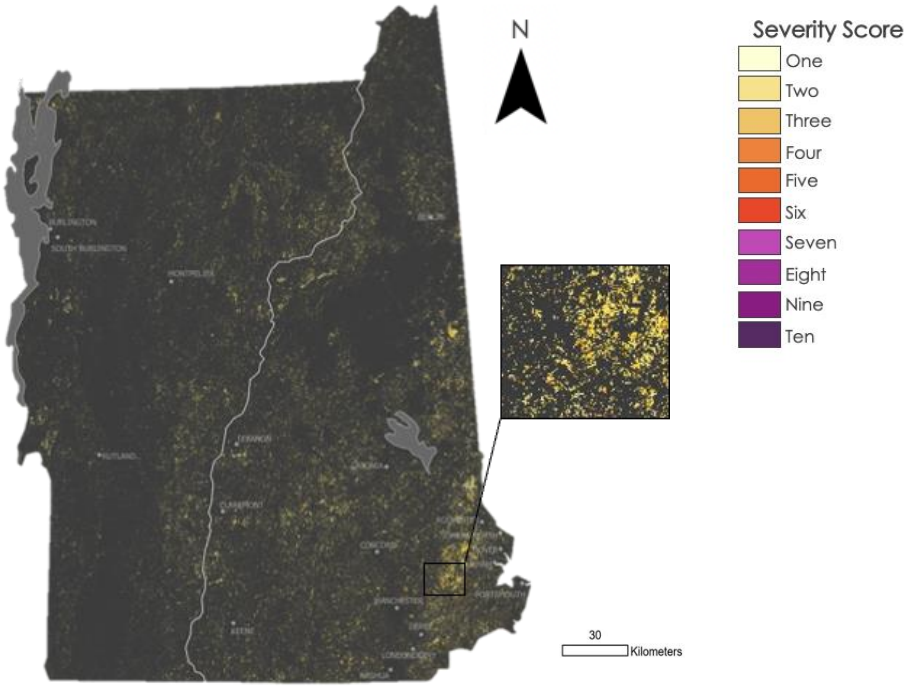

Figure 4. Sum of potential defoliation severity from 2012 to 2021.

We found that 1371.27 km² of the VT and NH study area experienced defoliation at a 10-year severity level of 1 (Figures 5 & 6). This corresponds to around 34,0000 acres affected in the past decade by a single light defoliation event. Further, we calculated that 223.81 km² were affected by defoliation with a severity ranking

of 4 or more. This correlates to either multiple years of severity of 1 or a combination of more intense severity seen in fewer years. We found 75 km² was impacted at a combined severity of over 6 in the past decade. The analysis behind this number is that areas impacted experienced multiple years of intense defoliation, a minimum of 2 years of severe defoliation or up to 6 plus years of light defoliation.

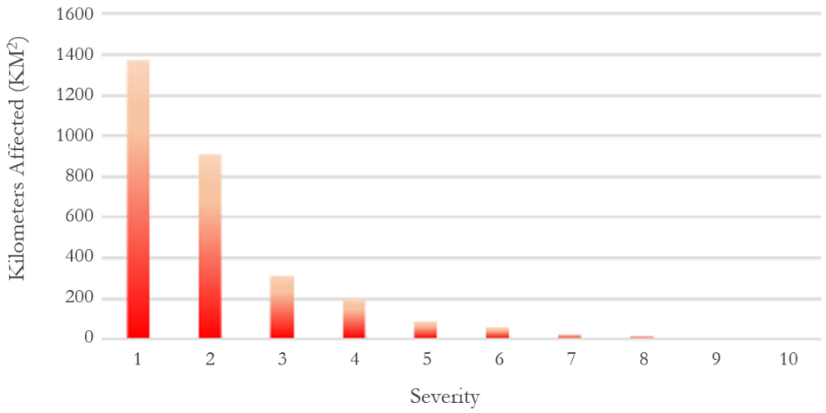

Figure 5. Sum of kilometers affected by each 10-year severity score.

One source of error in this analysis may be the misrepresentation of data. If an area receives a score of 6, that could mean six years of light severity or two years of heavy severity. While these receive the same objective score, the impact is vastly different, and management is different as well. However, when we combine Figures 3 and 4, we can see patterns arise. Areas with 4, 5, and 6 years of outbreaks are also areas with combined severity scores of 6 to 10.

To understand this, we zoomed into an area in southern New Hampshire. This area experienced repeated potential defoliation events in the years 2012, 2017, 2018, and 2019 with ranging intensities. The map in the middle is the project study area with the inset on the left being a cut out of the southern region of interest. To examine each year, the maps on the right and bottom show the sum scores and the stated year score.

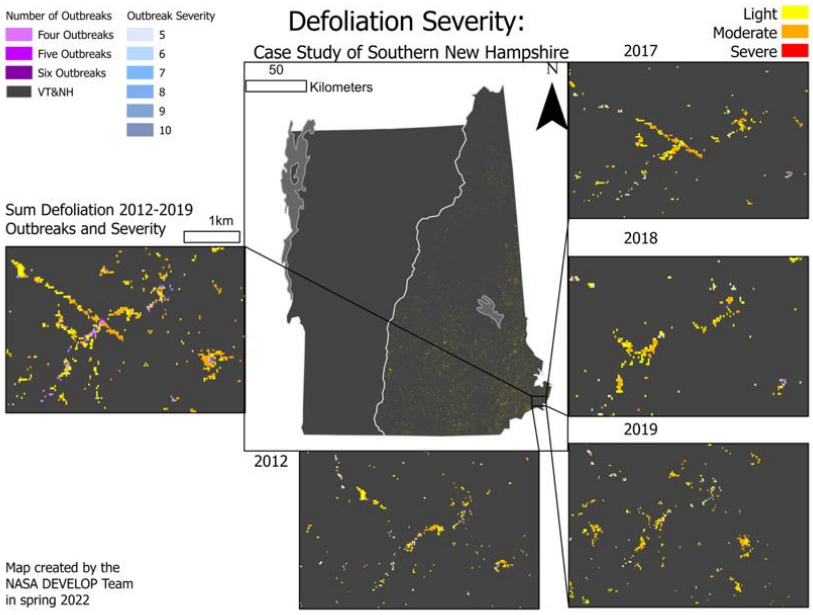

Figure 6. Defoliation repetition in southern New Hampshire.

This map shows that in 2012, a potential defoliation event occurred in this area of interest. Next, looking on the top-right we can see that in 2017 this area flared up again with defoliation events affecting the forest. Not only does this support the idea that *L. dispar* experience outbreak events but also that they can remain undetected in an area for many years prior to outbreak events. The importance of reviewing the repetition of defoliation lies in the implication of these analyses for land management practices. These areas that are highlighted, which experienced multiple years of impact, could be studied in greater depth to understand the effects of outbreak events on forest health.

4.2 Future Work

While our Random Forest classification may be predictive of future *L. dispar* driven tree defoliation and mortality, the lack of ground-based observations makes it difficult to quantify the accuracy of our algorithm. Validation was restricted to three years (2017, 2018, and 2021) of data randomly sampled in areas classified by tree cover type. We derived our accuracy metrics from these samples, and the most immediate future work should be focused on conducting fieldwork to obtain ground truth samples for comprehensive model validation.

Using our methodology, the future incorporation of remotely sensed imagery with higher spatial, spectral, and temporal resolution can provide more accurate classifications of defoliation severity. These data are also necessary for forecasting tree mortality given that the drivers of tree health are scale-dependent and time-dependent (Swier et al., 2016). While defoliation events occur across the Northeast United States and into Canada, the study area in this project was limited to Vermont and New Hampshire; thus, future research can expand on this study area. Altogether, the inclusion of higher resolution data within our model may yield accurate, near real time results that can better inform the decision making of land managers.

5. Conclusions

In this study, we modeled the extent of potential defoliation due to *L. dispar* across VT and NH from 2012 to 2021. Of the years we modeled, the potential defoliation extent in 2021 was the greatest; that year accounted for 25 percent of the total extent of defoliation across the 10-year study period. This finding agrees with the

project partners' field observations of defoliation (J. Rosovsky, personal communication, March 28, 2022). Since there is a large area where defoliation events may occur, efforts that increase monitoring extent can improve management practices that mitigate damage caused by *L. dispar*. This project can support land management efforts by highlighting areas of potential historical defoliation and providing a tool that helps land managers understand the mechanics and patterns of outbreak events.

Mortality typically occurs among healthy hardwood trees after two or more years of heavy defoliation and can lead to a loss of ecosystem services (Swier et al., 2016). We measured where and how often outbreaks were repeated during the study period to identify areas where trees may have greater risk of mortality. Most areas of potential defoliation (91 percent, 2616 km^2) were limited to a single year, and most defoliation events were of low severity (Figures 2 & 4). The majority (88 percent, 220 km^2) of the areas where repeated potential defoliation occurred experienced only two outbreaks. Therefore, mortality is unlikely to have occurred among host trees in most of the areas that experienced potential defoliation unless those defoliation events were combined with another stressor of tree health such as drought.

This project is centered around the communities that are affected by *L. dispar* outbreaks. Thus, we prioritized our partners' concerns in this research. First, our partners described evidence of the resilience of forests following *L. dispar* outbreaks. Most areas where we identified historical defoliation only experienced one defoliation event, and trees have shown resilience to single defoliation events. Further, our partners highlighted the need for a cost-effective monitoring tool. Defoliation monitoring based on remote sensing data has been found to more accurately estimate retrospective *L. dispar* abundance than some ground-based abundance assessment methods can (Pasquarella et al., 2021). Thus, we created the graphical user interface CATEPILR hosted in GEE to increase project partners' defoliation monitoring capacity. Through the creation of CATEPILR, our DEVELOP team provided partners with a tool that identifies potential defoliation events as new Landsat scenes become available. This tool highlights variables correlated to outbreak events that can direct management efforts and may be used in future research. Finally, our project partners requested an outreach tool for the public. Through the creation of an Esri StoryMap, research and results were effectively communicated to the public to assist in education and the understanding of *L. dispar*'s habits in Vermont and New Hampshire.

6. Acknowledgments

The Vermont and New Hampshire Ecological Forecasting Team thanks our project partners from the Forest Ecosystem Monitoring Cooperative (FEMC), the Vermont Agency of Agriculture, Food and Markets (VAAFM), the University of New Hampshire Cooperative Extension (UNH Extension), and the New Hampshire Division of Forests and Lands, Forest Health Program. We are grateful to Drs. Juan Torres-Pérez and Valerie Pasquarella for advising this study, Britnay Beaudry for offering guidance, support, and feedback throughout this project, and Hayley Pippin, Dr. Kenton Ross, and Tasos Tentoglou for providing technical assistance.

This material contains modified Copernicus Sentinel data (2017, 2018, and 2021), processed by ESA.

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Aeronautics and Space Administration.

This material is based upon work supported by NASA through contract NNL16AA05C.

7. Glossary

CATEPILR – Condition Assessment Tool to Estimate Patterns of the Invasion of Lymantria dispaR. A Google Earth Engine tool to estimate defoliation due to *L. dispar* that updates with newly ingested satellite imagery.

Defoliation – The loss of leaves on a plant. The severity of leaf loss be detected through remotely sensed imagery.
Earth observations – Satellites and sensors that collect information about the Earth's physical, chemical, and biological systems over space and time.
EVI – Enhanced Vegetation Index. A metric of vegetation greenness using atmospheric corrected surface reflectance.
MODIS – Moderate Resolution Imaging Spectroradiometer. Hyperspectral satellite-based sensor with 36 bands.
NDVI – Normalized Difference Vegetation Index. Measures the difference between near-infrared and red light as a quantifier of vegetation greenness.
Random Forest – A supervised machine learning classification algorithm. This algorithm uses training data as inputs in a multitude of decision-trees that each assign a prediction. These predictions are then averaged for the final classification.
TCG – Tasseled Cap Greenness. A weighted sum of bands representing the degree of greenness of a pixel.

8. References

Abatzoglou, J. (1979). *GRIDMET: University of Idaho Gridded Surface Meteorological Dataset*. University of California Merced. https://www.climatologylab.org/gridmet.html

Abatzoglou, J. (1980). *GRIDMET DROUGHT: CONUS Drought Indices*. University of California Merced. https://www.climatologylab.org/gridmet.html

Andresen, J. A., McCullough, D. G., Potter, B. E., Koller, C. N., Bauer, L. S., Lusch, D. P., & Ramm, C. W. (2001). Effects of winter temperatures on gypsy moth egg masses in the Great Lakes region of the United States. *Agricultural and Forest Meteorology*, *100*, 85–100. http://doi.org/10.1016/S0168-1923(01)00282-9

Baig, M. H. A., Zhang, L., Shuai, T., & Tong, Q. (2014). Derivation of a tasseled cap transformation based on Landsat 8 at-satellite reflectance. *Remote Sensing Letters*, *5*(5), 423–431. https://doi/full/10.1080/2150704X.2014.915434

Dewitz, J. (2019). *National Land Cover Database (NLCD) 2016 Products* (ver. 2.0, July 2020) [Data set]. U.S. Geological Survey data release. *US Geological Survey*. https://doi.org/10.5066/P96HHBIE

Cherry, M. (2021, June 21). *Outbreak of gypsy moth caterpillars in north central New Hampshire*. WMUR. https://www.wmur.com/article/gypsy-moth-caterpillar-outbreak-nh/36793109

Choi, C. T. H., Bakke, A., Posen, A., Rock, M., & Vannest, N. (2021). Colorado Ecological Forecasting: Monitoring post-fire cheatgrass (*Bromus tectorum*) distribution to inform management planning. [Unpublished manuscript]. NASA DEVELOP National Program, Colorado – Fort Collins.

Conrad-Rooney, E., Barker Plotkin, A., Pasquarella, V. J., Elkinton, J., Chandler, J. L., & Matthes, J. H. (2021). Defoliation severity is positively related to soil solution nitrogen availability and negatively related to soil nitrogen concentrations following a multi-year invasive insect interruption. *AoB PLANTS*, *12*(6). https://doi.org/10.1093/aobpla/plaa059

Elkinton, J. S., & Liebhold, A. M. (1990). Population dynamics of gypsy moth in North America. *Annual Review of Entomology*, *35*, 571–596. https://doi.org/10.1146/annurev.en.35.010190.003035

Funk, C. (2015). *CHIRPS Daily: Climate Hazards Group InfraRed Precipitation With Station Data* (Version 2.0 Final). Climate Hazards Group. https://doi.org/10.15780/G2RP4Q

Gao, B. (1996). NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sensing of Environment*, *58*(3), 257–266. https://doi.org/10.1016/S0034-4257(96)00067-3

Hall, D. K. and G. A. Riggs. (2016). *MODIS/Terra Snow Cover Daily L3 Global 500m SIN Grid* (Version 6). NASA National Snow and Ice Data Center Distributed Active Archive Center. https://doi.org/10.5067/MODIS/MOD10A1.006

LANDFIRE. (2014). *LANDFIRE EVT (Existing Vegetation Type)* (v1.4.0). U.S. Department of Agriculture, Forest Service, U.S. Department of the Interior, U.S. Geological Survey, and The Nature Conservancy. https://landfire.gov/evt.php

Liebhold, A., Elkinton, J., Williams, D., & Muzika, R. (2000). What causes outbreaks of the gypsy moth in North America? *Population Ecology*, *42*(3), 257–266. https://doi.org/10.1007/PL00012004

O'Neill, P. E., Chan, S., Njoku, E. G., Jackson, T., & Bindlish, R.. (2018). *SMAP L3 Radiometer Global Daily 36 km EASE-Grid Soil Moisture* (Version 5). NASA National Snow and Ice Data Center Distributed Active Archive Center. https://doi.org/10.5067/ZX7YX2Y2LHEB

Pasquarella, V. P., Bradley, B. A., & Woodstock, C. E. (2017). Near-real-time monitoring of insect defoliation using Landsat time series. *Forests*, *8*(8), 275–292. https://doi.org/10.3390/f8080275

Pasquarella, V. (2021). *valpasq/condition_monitoring: v0.2-alpha* (v0.2-alpha). Zenodo. https://doi.org/10.5281/zenodo.4565535

Pasquarella, V. P., Mickley, J. G., Plotkin, A. B., MacLean, R. G., Anderson, R. M., Brown, L. M., Wagner, D. L., Singer, M. S., & Bagchi, R. (2021). Predicting defoliator abundance and defoliation measurements using Landsat-based condition scores. *Remote Sensing in Ecology and Conservation*, *7*(4), 592–609. https://doi.org/10.1002/rse2.211

Poggio, L. (2020). *SoilGrids250m 2.0 - Total nitrogen*. ISRIC - World Soil Information. https://doi.org/10.17027/isric-soilgrids.f0797a68-1692-11ea-a7c0-a0481ca9e724

Reilly, J. R., Hajek, A. E., Liebhold, A. M., & Plymale, R. (2014). Impact of *Entomophaga maimaiga* (Entomophthorales: Entomophthoraceae) on cutbreak gypsy moth populations (Lepidoptera: Erebidae): The role of weather. *Environmental Entomology*, *43*(3), 632–641. https://doi.org/10.1603/EN13194

State of New Hampshire Division of Forests & Lands Forest Health Program (2021). Forests and Lands Maps [Data Set]. New Hampshire Division of Forests & Lands. https://www.nh.gov/nhdfl/reports/maps.htm

Swier, S. R., Eaton, A. T., Maccini, R., Gladders, D., & Bennett, K. (2016). *Gypsy Moth, Pest Fact Sheet 70*. University of New Hampshire Cooperative Extension. https://extension.unh.edu/resource/gypsy-moth-fact-sheet

US Census Bureau. (2018). *TIGER: US Census States 2018*. US Census Bureau. https://www.census.gov/geographies/mapping-files/time-series/geo/tiger-line-file.2018.html

U.S. Geological Survey Earth Resources Observation and Science Center. (2018). *Digital Elevation - Shuttle Radar Topography Mission (SRTM) Void Filled* [Data set]. US Geological Survey. https://doi.org/10.5066/F7F76B1X

U.S. Geological Survey Earth Resources Observation and Science Center. (2020). *Landsat 7 ETM+* [Data set]. US Geological Survey. https://doi.org/10.5066/P9C7I13B

U.S. Geological Survey Earth Resources Observation and Science Center. (2020). *Landsat 8 OLI/TIRS* [Data set]. US Geological Survey. https://doi.org/10.5066/f78s4mzj

U.S. Geological Survey Gap Analysis Project (GAP). (2020). Protected Areas Database of the United States (PAD-US) 2.1 [Data set]. *U.S. Geological Survey*. https://doi.org/10.5066/P92QM3NT

Vermote, E., Justice, C., Claverie, M., & Franch, B. (2016). Preliminary analysis of the performance of the Landsat 8/OLI land surface reflectance product. *Remote Sensing of Environment*, *185*(2), 46–56. https://doi.org/10.1016/j.rse.2016.04.008

Walter, J. A., Meixler, M. S., Mueller, T., Fagan, W. F., Tobin, P. C., & Haynes, K. J. (2015). How topography induces reproductive asynchrony and alters gypsy moth invasion dynamics. *Journal of Animal Ecology*, *84*(1), 188–198. https://doi.org/10.1111/1365-2656.12272

Wang, S., Zhuang, Q., Wang, Q., Jin, X., & Han, C. (2017). Mapping stocks of soil organic carbon and soil total nitrogen in Liaoning Province of China. *Geoderma*, *305*, 250–263. https://doi.org/10.1016/j.geoderma.2017.05.048

9. Appendices

Appendix A

Table A1
Total historical potential defoliation extent

	Defoliation Extent			
	Light	*Moderate*	*Severe*	*Total*
Total defoliation (km²)	2185.08	933.12	22.55	3140.74
Percentage of total defoliation	69.57	29.71	0.72	100

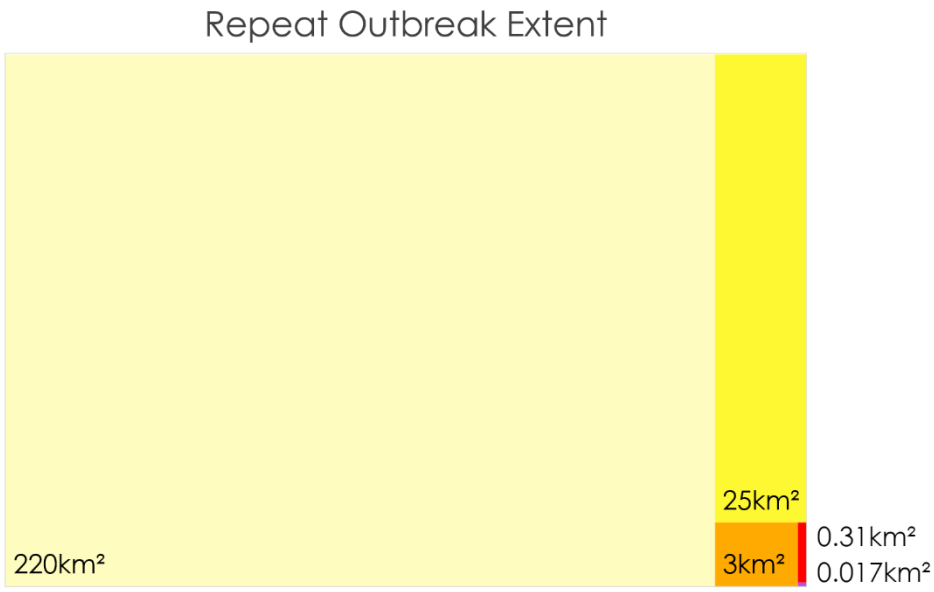

Figure A1. Repeated potential defoliation events between 2012 and 2021, ranging from two events (lightest yellow) to six (purple).