dfoliatR: An R package for detection and analysis of insect defoliation signals in tree rings

Christopher H. Guiterman[*,a,b], Ann M. Lynch[a,c], Jodi N. Axelson[c]

[a]*Laboratory of Tree-Ring Research, University of Arizona, 1215 E Lowell St. Box 210045, Tucson, AZ, 85721*
[b]*Three Pines Forest Research, LLC, PO Box 225, Etna, NH, 03750*
[c]*U.S. Forest Service, Rocky Mountain Research Station, 1215 E Lowell St. Box 210045, Tucson, AZ, 85721*
[d]*Dept of Environmental Science, Policy & Management, University of California, Berkeley, Berkeley, CA 94720*

Abstract

We present a new R package to provide dendroecologists with tools to infer, quantify, analyze, and visualize growth suppression events in tree rings caused by insect defoliation, and to reconstruct defoliator outbreak chronologies. 'dfoliatR' is based on the OUTBREAK program and builds on existing resources in the R and dplR computing environments. It performs an indexing procedure to remove climatic signals, represented in non-host chronologies or other reconstructed climate series, from the host-tree series. It then infers defoliation events in individual trees based on user-specified thresholds. Site-level analyses identify outbreak events that synchronously affect user-defined numbers or proportions of involved host trees. Functions are provided for summary statistics and graphics of tree- and site-level series. 'dfoliatR' expands on OUTBREAK to provide greater control of suppression thresholds, additional output tables, and high-quality graphics. We provide two example data sets and script to enable users to gain familiarity with the package and its capabilities. The source code is available in the Comprehensive R Archive Network (CRAN) and on GitHub.

Key words: Dendroecology, spruce budworm, Choristoneura, pandora moth, Coloradia pandora Blake, larch-bud-moth

---

[*]Corresponding Author
Email addresses: `chguiterman@arizona.edu` (Christopher H. Guiterman), `ann.m.lynch@usda.gov` (Ann M. Lynch), `jodi.axelson@berkeley.edu` (Jodi N. Axelson)

1. Introduction

Variation in the width and morphology of annual radial growth rings in trees permits dating and quantification of past forest insect defoliator outbreaks. Defoliation can be distinguished from climate- and other disturbance-related influences by comparing ring-width or other annually-resolved features in the wood of host species to that of non-host species or annually-resolved climate records. The effect of defoliation on radial growth of trees has been recognized since the 1860s, and used to reconstruct outbreak regimes since the 1950s (Blais 1954; Lynch 2012). It was not until the 1980s, however, that precise dendrochronological techniques were applied for inferring defoliation events and reconstructing defoliator outbreak regimes (Swetnam, Thompson, and Sutherland 1985; Speer 2010; Lynch 2012). The first studies (Swetnam, Thompson, and Sutherland 1985; Swetnam and Lynch 1989, 1993) focused on developing historical reconstructions of western spruce budworm (WSBW; *Choristoneura freemani* (Lepidoptera: Tortricidae); previously known as *C. occidentalis*) outbreak regimes. The methodology has since been successfully applied to a wide range of defoliator species, most of which are conifer herbivores, and has evolved in sophistication for a wide range of ecosystem situations (Lynch 2012).

The main dendrochronological tool for inferring, dating, and characterizing defoliator outbreaks from tree-ring records has been the software routine OUT-BREAK (Swetnam, Thompson, and Sutherland 1985; Holmes and Swetnam 1986; Swetnam and Lynch 1989). OUTBREAK computes indices (described later in detail) of suppressed growth by subtracting a detrended and standardized climate series (a "control" chronology) from individual host-tree detrended and standardized radial growth series after the host and non-host series have been brought to a common variance. The non-host chronology usually consists of a site chronology developed from non-host tree species growing on a climate-sensitive site, but a gridded climate data point series, like the North American Drought Atlas (Cook and Krusic 2004) also suffices. If the host and non-host species respond similarly to climate (which can and should be tested), the de-

2

rived series retains variability that the host and non-host series do not have in common, generally the insect signal and some unexplained variability (noise). The user defines a rule base specifying the magnitude and duration that a period of indexed growth suppression must meet or surpass for a period of suppressed growth to be inferred as a defoliation event at the tree level.

Though powerful, OUTBREAK is outdated and increasingly difficult to use in modern computing environments. It was written in FORTRAN V with inherently severe restrictions, as RAM and disk space were limited at that time (256 kb and 10 MB, respectively) and FORTRAN conventions imposed very strict formatting, file naming, and output conventions. The program lacks a graphical interface or capabilities, forcing users to import generated text files into spreadsheets or other software to assess results and performing analyses. Furthermore, OUTBREAK can only handle one test at a time, creating barriers to batch operation and a large burden for researchers with a datasets including multiple sites. We developed `dfoliatR` as an `R`- and `dplR`-based library to overcome these issues.

`dfoliatR` adds to a growing suite of dendrochronology packages in the `R` computing environment (R Core Team 2019). Stemming from the `dplR` library (Bunn 2008) that enables `R` users to read and write an array of tree-ring data formats, standardize ring width series, build and evaluate chronologies, and perform quality control (to name a few), one can now also measure ring widths from scanned images of prepared samples (Shi et al. 2019; Lara, Bravo, and Sierra 2015), conduct and check crossdating (Bunn 2010), analyze sub-annual anotomical features (Campelo et al. 2016; Rathgeber et al. 2011), and perform many analytical tests (Jevšenak and Levanič 2018; Zang and Biondi 2015). Tools for assessing stand dynamics and disturbance analyses are under rapid development, with new packages for assessing growth and release events (`TRADER`: Altman et al. 2014), metrics of growth resilience (`pointRes`: van der Maaten-Theunissen, van der Maaten, and Bouriaud 2015), and fire history (`burnr`: Malevich, Guiterman, and Margolis 2018). The key objective of `dfoliatR` is to provide tools to identify and analyze insect defoliation and outbreak events by building on

3

the methods employed by OUTBREAK. It capitalizes on the robust software already available in `R` by using `dplR` data formats for incoming tree-ring series and providing output data formats embodied by the `tidyverse` (Wickham et al. 2019) that include efficient data manipulation (Wickham et al. 2020) and graphics (Wickham 2016).

In this paper, we describe the statistical methods employed by `dfoliatR`, compare results to those produced by OUTBREAK, and present an example analysis inlcuding test data sets and script. Users need not have much experience in `R` to replicate the analyses and graphics as presented. The `R` code below is executable in an `R` session once the required libraries are installed and loaded. Support documentation in addition to this paper is provided within the package via standard help menus and on the package website (https://chguiterman.github.io/dfoliatR/), which includes up-to-date vignettes that describe various routines. Code to generate a preprint of this manuscript, including the `R` scripts and tubular and graphical output is available from https://github.com/chguiterman/dfoliatR_paper.

2. Overview of the software

The `dfoliatR` library requires two sets of tree-ring data to infer defoliation and outbreak events:

- Standardized ring-width series for individual trees of the host species
- A standardized tree-ring chronology from a local non-host species, or a climate reconstruction

Users can develop these data sets in the software of their choosing, such as `dplR` or ARSTAN (Cook and Holmes 1996). It is important that the host-tree data include only one tree-ring series per tree. Both `dplR` and ARSTAN have options for averaging multiple sample series into a tree-level series.

At the heart of `dfoliatR` lies two functions: `defoliate_trees()` and `outbreak()`. These identify defoliation events on individual trees (Figure 1) and then composite across multiple trees to identify outbreak events (Figure 2).

4

2.1. Identifying Defoliation of Trees

The `defoliate_trees()` function is the point of entry to the `dfoliatR` library. It performs two processes, removing climate-related growth signals from the host-tree series and identifying tree-level defoliation events. The climatic or non-defoliation signals in each host-tree series are characterized by a non-host chronology or climate reconstruction. `dfoliatR` removes the non-defoliation signal by subtracting the non-host series from each host-tree series, which generates a residual index. In OUTBREAK, this residual index was termed the "corrected index." We call it the "growth suppression index" (GSI). The GSI is calculated the same as in OUTBREAK for each host tree as

$$\text{GSI}_i = \text{H}_i - \left(\text{NH}_i - \overline{\text{NH}}\right) \frac{\sigma_{\text{H}}}{\sigma_{\text{NH}}} \tag{1}$$

where H and NH are the host-tree series and the non-host chronology, in year i, respectively. Only the common period between the host-tree series and the non-host chronology are used in Equation 1. The host and non-host chronologies are brought to common variance by scaling the non-host chronology by its mean ($\overline{\text{NH}}$) and multiplying by the ratio of host and non-host standard deviations ($\frac{\sigma_{\text{H}}}{\sigma_{\text{NH}}}$), which approximates the variance of the host tree series.

Negative departures in the normalized GSI (NGSI; GSI converted to z-scores) that surpass user-specified thresholds in duration and magnitude are defined as *defoliation events.* As in OUTBREAK, NGSI must reach the magnitude threshold, assessed for each year within the departure sequence. The default setting is -1.28 (NGSI is in units of standard deviation), which was previously determined to be representative of WSBW effects (Swetnam and Lynch, 1989) and is commonly used for other species. Failure to reach the specified threshold results in sequences not being inferred as defoliation events.

Event duration is assessed by examining sequences of negative NGSI (that exceed the magnitude threshold) before and after the year of maximum departure. Each defoliation event is allowed one single-year positive excursion on each side of the year of maximum departure. Duration is computed across the

entire negative sequence that allows these two positive excursions. As in OUT-BREAK, the user specifies a duration threshold (minimum number of years) for a departure sequence to be inferred as a defoliation event. The default threshold is eight years, as is commonly used in WSBW studies (Swetnam and Lynch, 1989). Failure to reach the specified duration threshold results in sequences not being inferred as defoliation events, regardless of NGSI values (i.e., both thresholds must be met). Researchers can, and should, adjust the duration and magnitude parameters accordingly and critically evaluate the results, as insect species vary in the length of their outbreaks and the degree to which they can suppress tree growth.

Like OUTBREAK, users are provided an option to suspend the duration threshold at the recent end of the series in cases where an outbreak event is known to be ongoing. This should be used if the user has direct knowledge of defoliation at the site during the sampling campaign. The advantage of allowing potentially short, series-end events is that it allows the current event to be included in return-interval estimates, and can aid in identifying the start-year for the current defoliation event or outbreak.

Diverging from OUTBREAK, `dfoliatR` allows users to extend defoliation events on individual trees by bridging between sequential events. In cases where two defoliation events are separated by a single year, bridging will link them into a single event. This option was added to `dfoliatR` during the testing phase of development, when we realized that OUTBREAK deliberately omits sequential, or back-to-back events, even when these events surpass thresholds in duration and magnitude. Instead, OUTBREAK will select the one sequential event with the greatest negative departure year. In every case we assessed (described below) we felt that the OUTBREAK-omitted defoliation events should have been maintained and recorded. Due in large part to reconstructions using OUTBREAK (see papers cited by Lynch 2012), we now know considerably more about forest defoliator outbreak regimes than we did in the 1980s when OUTBREAK was written. We think that two or more prolonged events separated by a single year should in some situations be considered a single event.

6

This is particularly relevant to WSBW and spruce budworm (*C. fumiferana*), for which regimes have been reconstructed for several geographic areas and for which researchers and forest health experts have gained considerable knowledge (Schmitt, Grimble, and Searcy 1984; Sanders et al. 1985; Brookes et al. 1982, and many later publications). In many cases, the greatest growth suppression often occurs late in the outbreak due to cumulative effects accrued on a tree's resources (Swetnam and Lynch 1989, 1993). We urge caution in using the bridging option, however, because it may not be appropriate for all studied insects, such as in situations where impacted stands barely recover from one outbreak before another begins, as with pine processionary caterpillars (*Thaumetopoea pityocampa* (Lepidoptera: Thaumetopoeidae)) (Carus 2004, 2009).

2.2. Inferring Outbreak Events

Defoliation of one or a few trees does not constitute an outbreak. To determine when defoliation becomes an *outbreak event*, `dfoliatR` composites the individual tree defoliation series into a site-level chronology with the `outbreak()` function. Users have options to define the number and/or the proportion of trees required for an event to be considered an outbreak. Three parameters control whether a defoliation event constitutes an outbreak: the minimum number of trees available, the minimum number of trees recording defoliation, and the percent of trees recording defoliation. The first allows the researcher to make a judgement call as to the confidence ascribed to reduced sample depth toward the ends of their chronologies, thus compensating for the "fading record problem" (Swetnam, Allen, and Betancourt 1999). The second two parameters adjust the scale of defoliation considered to be an outbreak. Absolute numbers of trees and percentages can be applied separately or in conjunction, following filtering conventions in tree-ring fire history studies (Malevich, Guiterman, and Margolis 2018). We urge users to carefully consider the choice of absolute numbers in situations where the number of trees represented in the series varies with time, or the choice of percentages when sample size is small.

3. Evaluation

3.1. Approach

We tested `dfoliatR` against OUTBREAK by comparing NGSI to OUT-BREAK's normalized corrected indices for individual trees and years, defoliation status for individual trees and years, and percentage of trees recording outbreaks at the site level, using standardized ring-width data from eight sites in British Columbia, Colorado, and New Mexico, and author-provided non-host site chronologies.

We detrended host data for both `dfoliatR` and OUTBREAK using AR-STAN (v6.1, Cook and Holmes 1996) with default double detrending (128 year wavelength and a 50% smoothing spline). In both `dfoliatR` and OUTBREAK we used event thresholds of -1.28 normalized indices, 8 years duration, and allowed for events at the end of series in seven of eight sites. These seven sites were sampled during ongoing outbreak events (Swetnam and Lynch 1993). We found it necessary to be consistent in how we detrended and in what software (e.g., ARSTAN vs `dplR`) we employed because subtle differences in standardized ring-width indices generated between the programs transferred into differences between `dfoliatR` and OUTBREAK. In the end, we chose to only use the standardization output files from ARSTAN, which are easily read into `R` (and then `dfoliatR`) using the `dplR` package.

The `R` code to replicate our comparisons is available from https://github.com/chguiterman/dfoliatR_paper.

3.2. Findings

Across the 43,280 ring-width indices from 222 trees included in our evaluation, we found that `dfoliatR` and OUTBREAK compute identical growth suppression indices at 0.000 precision. We expected this outcome because both programs employ Equation 1. At the tree-level, the programs identified 11,530 years of defoliation. The programs agreed on 97.9% of the years, leaving 927 "difference" years in which only one program identified defoliation on an individual tree. The differences included 102 events on 85 trees. We carefully

inspected each of these events in the full context of each tree's ring-series, and categorized the differences as follows

- *Series-end events* (40% of the total) in which OUTBREAK included "truncated outbreaks" (for seven sites) at the end of each series. In `dfoliatR`, this option is controlled by the "series_end_events" parameter to `defoliate_trees()`. In OUTBREAK, the option appears while changing the duration parameter (option 3). When selected, OUTBREAK will include any sequences of negative indices at the either the beginning or the end of each tree series as a defoliation event, without consideration of either duration or magnitude thresholds. In `dfoliatR`, the duration threshold is omitted and the magnitude threshold is retained in series-end-events. Each of the 13 events included in these differences did not meet the "max_reduction" parameter (-1.28 NGSI) in `dfoliatR` and were excluded. In two cases, OUTBREAK included events at the beginning of the series where `dfoliatR` does not allow truncated events. In four cases, OUTBREAK omitted only the last year of the series because the index was positive. In two cases, `dfoliatR` omitted possible events because it had already included a positive NGSI excursion after the "max_reduction" year, and since it will only allow one excursion on either side of the max year, the events were omitted due to short duration.

- *Sequential events* (36%) in which OUTBREAK omitted events that occur one year prior or one year following an identified event. When this occurs, OUTBREAK selects the event with the greatest negative index year. On two trees, OUTBREAK omitted two of three sequential events. While inspecting these differences, we added an option to `defoliate_trees()` that would "bridge" between sequential events (that surpass magnitude and duration thresholds)into single, long events. We felt that this was ecologically justified, especially for studies of WSBW, because the insect will occasionally defoliate individual trees for decades and reconstructed chronologies show outbreaks as long as **xxx (ANN CHECK)**.

9

- *Undetermined differences* (22%) occurred in cases where OUTBREAK omitted events without clear cause that `dfoliatR` correctly identified as defoliations.

- *Rounding differences* (2%) in the indices either omitted or cut short events on two trees. In both cases the indices were very close to zero, and the difference was less than the precision of the raw data measurement.

(Bridging paragraph?)

At the site level, OUTBREAK and `dfoliatR` produce similar time series of percent trees defoliated (Figure 4), which forms the basis for inferring outbreak occurrence, intensity, and duration. Nearly all of the differences between the programs include either more trees or a longer duration of inferred outbreak by `dfoliatR`. These arise from the inclusion of events omitted by OUTBREAK. These differences translate to greater numbers of trees in defoliation during certain events, or events recorded by single trees during periods of low sample depth.

This evaluation revealed what we believe are short-comings in how OUT-BREAK identifies defoliation events on individual trees. In every one of the 102 cases we inspected, we felt that `dfoliatR` provided a more biologically and statistically appropriate assessment of defoliation. Therefore, `dfoliatR` also did a better job inferring OUTBREAK events and associated statistics at the site level.

4. Availability and installation

The `dfoliatR` library (Guiterman, Lynch, and Axelson 2020) is provided free and open source from the Comprehensive R Archive Network (CRAN; https://cran.r-project.org/). To install `dfoliatR` from CRAN use

```
install.packages("dfoliatR")
```

In each R session, dfoliatR can be loaded via

10

```
library(dfoliatR)
```

Development versions of `dfoliatR` are available on GitHub and installed using the `devtools` library,

```
devtools::install_github("chguiterman/dfoliatR")
```

Issues, bug reports, and ideas for improving `dfoliatR` can be posted to https://github.com/chguiterman/dfoliatR/issues. As an Open Source library, we welcome and encourage community involvement in future development. The best ways to contribute to `dfoliatR` are through standard GitHub procedures or by contacting the first author.

5. Example Usage

Once `dfoliatR` is loaded into an R session (via `library(dfoliatR)`) users can access two sets of tree-ring data to aid in exploring the functions, graphics, and outputs. Each data set consists of individual host-tree series and a local non-host chronology. The host-tree series were standardized using 50-year splines with a 50% frequency response, while the non-host ring-width data were standardized using 150-year splines with a 50% frequency response and then averaged via Tukey's biweight robust mean procedure. Host trees from Demijohn Peak (DMJ; 2902 m asl) in the San Juan Mountains of southern Colorado include Douglas-fir (*Pseudotsuga menziesii*) compared against a local non-host ponderosa pine (*Pinus ponderosa*) chronology (Ryerson, Swetnam, and Lynch 2003). The East Fork site (EFK; 2580 m asl) in the Jemez Mountains of north-central New Mexico includes Douglas-fir and white fir (*Abies concolor*) host trees and a ponderosa pine non-host chronology (Swetnam and Lynch 1993).

With `dfoliatR` loaded, the datasets are accessible using the `data()` function. The data object names are prefixed by their site codes. For instance, the `dmj_*` objects come from the DMJ site and include the host-tree series (`dmj_h`), the non-host chronology (`dmj_nh`), the defoliation series (`dmj_defol`), and the

309 outbreak series (`dmj_obr`). The same suite of data are available for EFK using
310 the `efk_*` prefix.

311   In our example scripts below, `##` and `#` denote user comments, per standard
312 `R` coding, which are colored in brown. Text in blue denote functions; black are
313 loaded objects, and green are quoted variables and links.

314 *5.1. Tree-Level Defoliation Events*

315   The function `defoliate_trees()` performs the GSI indexing procedure on
316 each host-tree series and then identifies defoliation events.

```r
## Load the package
library(dfoliatR)

## Load data
data("dmj_h") # host tree series
data("dmj_nh") # non-host chronology
# View data structure for any data object via
# "View(dmj_h)" or "head(dmj_h)"

## Calculate indices, identify defoliation events
dmj_defol <- defoliate_trees(host_tree = dmj_h,
                             nonhost_chron = dmj_nh,
                             duration_years = 8,
                             max_reduction = -1.28,
                             bridge_events = TRUE,
                             series_end_event = TRUE,
                             list_output = FALSE)
```

317   The result is long-format (stacked) data frame with five variables: "year",
318 "series", "gsi", "ngsi", and "defol_status." The "defol_status" column indicates
319 whether that year has defoliation or not, with a set of factors that include
320 "nd" for non-defoliation year, "defol" for a defoliation year, "max_defol" for

12

the year of maximum suppression (that acts as the basis for individual events), "bridge_defol" to identify years that link subsequent events (only one is present at DMJ), and "series_end_defol" to identify defoliation at the present-end of the series.

Selecting `list_output = TRUE` in `defoliate_trees()` provides a list-object of data frames, each with an `rwl` object that combines the host tree and non-host series and the other columns created by `defoliate_trees()`. This option is not used by subsequent functions in `dfoliatR`, but researchers can examine it to check the results of the GSI calculation (Equation 1), such as the non-host series after scaling to a common variance with a particular host-tree series.

The results of running `defoliate_trees()` can be assessed through graphical and table outputs. The function `get_defol_events()` will provide a list of every defoliation event for every tree, with the corresponding mean "ngsi" value. A summary table of the results for each tree is produced by `defol_stats()` (Table 1).

```
defol_stats(dmj_defol)
View(dmj_defol)
```

The `plot_defol()` function produces a "ggplot" graphics object with line segments showing the measured sequence of each series and a filled segment for each identified defoliation event (Figure 1). The defoliation segments are colored by their relative severity based on their average NGSI value. The default cut-off values between "Severe" and "Moderate" is the overall mean across all events. The cut-off between "Moderate" and "Minor" is the first quartile by default. Users can re-define the breaks to suit their needs via the "breaks" parameter in `plot_defol()`.

```
plot_defol(dmj_defol)
## This creates a ggplot object. Additions can
## be made to adjust plotting themes and
## aesthetics, like color.
```

```
## For more on the features of ggplot see
## https://ggplot2.tidyverse.org/
## Adding 'scale_color_manual()' will
## change the colors that denote
## the average NGSI value for each defoliation
## event.
library(ggplot2)
plot_defol(dmj_defol) +
  scale_color_manual(values = c("red", "orange", "purple"))
```

These output functions aid in assessing the sensitivity of input parameters to `defoliate_trees()`, including the duration and magnitude thresholds for identifying defoliation events. Using `plot_defol()` also provides a direct assessment of the between-tree variability in defoliation.

5.2. Site-Level Events

To infer outbreak events at the site level, the function `outbreak()` composites tree-level defoliation series into a single chronology, with input parameters that control thresholds in the number and proportions of trees recording a defoliation event.

```
## Use the defol object "dmj_defol" created above
dmj_obr <- outbreak(dmj_defol,
                    filter_min_series = 3,
                    filter_min_defol = 1,
                    filter_perc = 25)
```

Input parameters to `outbreak()` include "filter_min_series" to control the chronology cut-off points with regard to sample depth, "filter_min_defol" and "filter_perc" to control the minimum number and percent of trees recording a defoliation event in a given year. `Outbreak()` produces a new data frame with eight variables: "year", "num_defol", "percent_defol", "num_max_defol", "mean_gsi",

14

"mean_ngsi", and "outbreak_status." All of these variables are populated regardless of an inferred outbreak event, providing a continuous disturbance chronology. The "num_max_defol" variable counts the number of trees recording their maximum defoliation in a given year. The "mean_gsi" and "mean_ngsi" variables provide averages of these indices across all available trees. Finally, the "outbreak_status" column shows whether an outbreak event is inferred ("outbreak") or not ("not_obr").

The default plotting function to visualize results from `outbreak()` creates a three-panel graph showing the mean site-level chronology, the percent of trees recording a defoliation, and the sample depth over time (Figure 2).

```
plot_outbreak(dmj_obr, disp_index = "mean_gsi")
```

Inferred outbreak events are shown in the top panel of Figure 2 as the filled-in spaces. Users can change the time series in this panel with the "disp_index" parameter, choosing between the mean NGSI (the default) or GSI.

A summary table of the inferred outbreak events is generated by the `outbreak_stats()` function (Table 2).

```
dmj_obr_stats <- outbreak_stats(dmj_obr)
View(dmj_obr_stats)
```

The output table from `outbreak_stats()` shows the start and end years of each event, along with their corresponding duration, the number and percent of trees in defoliation at the start of the event, the the number of trees recording the outbreak event (in part or full), the minimums of the mean GSI and NGSI indices, and two points of "peak defoliation." The first, "peak_outbreak_year" is the year in which the greatest number of trees are recording the outbreak, and the second, "peak_defol_year" is the year with the lowest average NGSI. These output variables provide a array of options for assessing metrics of the insect defoliation regime, including interval analyses:

15

```
## calculate mean return interval for outbreak start years
intervals <- diff(dmj_obr_stats$start)
mean(intervals) ## 51.2 year mean return interval
```

6. Conclusions

The `dfoliatR` package provides dendroecologists with tools to infer, quantify, analyze, and visualize tree-ring growth suppression events and to reconstruct forest insect defoliator outbreak regimes. It is built on the long-accepted host to nonhost comparison methodology used in the 1980s FORTRAN program OUTBREAK (Swetnam, Thompson, and Sutherland 1985; Swetnam and Lynch 1989, 1993). Key benefits are that `dfoliatR` provides easier control of suppression thresholds, provides additional output tables and high-quality graphics, defines defoliation events more consistently, and operates in the open-source `R` environment.

Using `dfoliatR` requires standardized ring-width measurements from insect host trees and either an indexed tree-ring chronology from local non-host trees or suitable climate chronology. It performs an indexing procedure to remove the climatic signal represented in the non-host chronology from the host-tree series. It then infers defoliation events in individual trees. Site-level analyses identify outbreak events that synchronously affect a user-defined number or proportion of the host trees. Functions are provided for summary statistics and graphics of tree- and site-level series. The package produces publication-quality plots, and tabulates growth suppression indices and tree- and site-level outbreak event statistics for user-defined post-processing needs, including those suitable for charting and tabulating landscape- and regional-level results.

`dfoliatR` adds a new option for dendroentimology to combine, or "bridge," sequential tree-level defoliation events into single events. In practice, we suggest that researchers carefully evaluate if bridging is ecologically applicable to study situations and insect ecology, and to carefully explore the data before deciding whether or not to use this option. It is probably not appropriate

16

for insects with high-frequency, high–severity outbreaks, such as processionary caterpillars. Outbreak reconstructions of insects for which the interval is notably longer than typical outbreak duration, such as Douglas-fir tussock moth (*Orgyia pseudotsugata* (Lepidoptera: Erebidae)), are unlikely to be significantly affected. Species for which outbreak duration varies considerably, for which individual tree resilience to defoliation varies considerably, or which may chronically infest trees or sites, or alternate between chronic, outbreak, and minimal activity states, such as conifer-feeding *Choristoneura*, present more complicated challenges, and the researcher should use discretion with the bridge option.

`dfoliatR` adds to the on-going open-source software development for dendrochronological methods (e.g., Bunn 2008; Brewer 2014; Brewer and Guiterman 2016). The `R` environment enables automation of analyses, allowing input/output processes to become routine, and empowering batch processing of large multi-site projects and sensitivity analyses. It also facilitates additional statistical analyses, such as spectral analyses and superposed epoch analyses (e.g., Malevich, Guiterman, and Margolis 2018), with easy transfer from `dfoliatR` and `dplR` to other libraries in `R`. Source code for `dfoliatR` is available in the Comprehensive R Archive Network (CRAN) and GitHub https://github.com/chguiterman/dfoliatR with updated descriptions and helpful vignettes on the package website https://chguiterman.github.io/dfoliatR/. Researchers wishing to contribute to further development of `dfoliatR` are encouraged to do so via the GitHub repository.

Acknowledgments

17

References

Altman, J., P. Fibich, J. Dolezal, and T. Aakala. 2014. "TRADER: A package for Tree Ring Analysis of Disturbance Events in R." *Dendrochronologia* 32: 107–12. https://doi.org/10.1016/j.dendro.2014.01.004.

Blais, JR. 1954. "The Recurrence of Spruce Budworm Infestations in the Past Century in the Lac Seul Area of Northwestern Ontario." *Ecology* 35 (1): 62–71.

Brewer, Peter W. 2014. "Data Management in Dendroarchaeology Using Tellervo." *Radiocarbon* 56 (4): S79–S83. https://doi.org/10.2458/azu_rc.56.18320.

Brewer, Peter W., and Christopher H. Guiterman. 2016. "A New Digital Field Data Collection System for Dendrochronology." *Dendrochronologia* 38: 131–35. https://doi.org/10.1016/j.dendro.2016.04.005.

Brookes, M. H., Robert W. Campbell, J. J. Colbert, R. G. Mitchell, and R. W. Stark (Technical Coordinators). 1982. "Western Spruce Budworm." USDA Forest Service, Cooperative State Research Service, Technical Bulletin No. 1694. 198 p., Washington D.C.

Bunn, Andrew G. 2008. "A dendrochronology program library in R (dplR)." *Dendrochronologia* 26: 115–24. https://doi.org/10.1016/j.dendro.2008.01.002.

———. 2010. "Statistical and visual crossdating in R using the dplR library." *Dendrochronologia* 28: 251–58. https://doi.org/10.1016/j.dendro.2009.12.001.

Campelo, Filipe, Cristina Nabais, Ana Carvalho, and Joana Vieira. 2016. "tracheideR–An R package to standardize tracheidograms." *Dendrochronologia* 37: 64–68. https://doi.org/10.1016/j.dendro.2015.12.006.

Carus, Serdar. 2004. "Impact of defoliation by the pine processionary moth (*Thaumetopoea pityocampa*) on radial, height and volume growth of Calabrian pine (*Pinus brutia*) trees in Turkey." *Phytoparasitica* 32 (5): 459–69.

———. 2009. "Effects of Defoliation Caused by the Processionary Moth on Growth of Crimean Pines in Western Turkey." *Phytoparasitica* 37 (2): 105–14.

Cook, Edward R, and Richard L Holmes. 1996. "Guide for computer program ARSTAN." *The International Tree-Ring Data Bank Program Library Version 2*, 75–87.

Cook, Edward R., and Paul J. Krusic. 2004. *The North American Drought Atlas*. (Lamont- Doherty Earth Observatory; the National Science Foundation: New York) Available at https://www.ncdc.noaa.gov/data-access/paleoclimatology-data/datasets/tree-ring/drought-variability [Last accessed May 2020].

Guiterman, Christopher, Ann Lynch, and Jodi Axelson. 2020. *DfoliatR: Detection and Analysis of Insect Defoliation Signals in Tree Rings*. R package version 0.1.0. https://doi.org/10.5281/zenodo.3698163.

Holmes, Richard L, and Thomas W Swetnam. 1986. "Dendroecology program library: program OUTBREAK user's manual." *Laboratory of Tree-Ring Research, University of Arizona, Tucson.*

Jevšenak, Jernej, and Tom Levanič. 2018. "dendroTools: R package for studying linear and nonlinear responses between tree-rings and daily environmental data." *Dendrochronologia* 48: 32–39. https://doi.org/10.1016/j.dendro.2018.01.005.

Lara, W., F. Bravo, and C. A. Sierra. 2015. "MeasuRing: An R package to measure tree-ring widths from scanned images." *Dendrochronologia* 34: 43–50. https://doi.org/10.1016/j.dendro.2015.04.002.

Lynch, Ann M. 2012. "What Tree-Ring Reconstruction Tells Us about Conifer Defoliator Outbreaks." In *Insect Outbreaks Revisited*, edited by Pedro Barbosa, Deborah K. Letourneau, and Anurag A. Agrawal, 126–54. Blackwell Publishing Ltd. https://doi.org/10.1002/9781118295205.ch7.

Malevich, Steven B., Christopher H. Guiterman, and Ellis Q. Margolis. 2018. "burnr: Fire history analysis and graphics in R." *Dendrochronologia* 49: 9–15.

19

https://doi.org/10.1016/j.dendro.2018.02.005.

Rathgeber, Cyrille B. K., Fleur Longuetaud, Frédéric Mothe, Henri Cuny, and Gilles Le Moguédec. 2011. "Phenology of wood formation: Data processing, analysis and visualisation using R (package CAVIAR)." *Dendrochronologia* 29 (3): 139–49. https://doi.org/10.1016/j.dendro.2011.01.004.

R Core Team. 2019. *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Ryerson, Daniel E, Thomas W Swetnam, and Ann M Lynch. 2003. "A tree-ring reconstruction of western spruce budworm outbreaks in the San Juan Mountains, Colorado, U.S.A." *Canadian Journal of Forest Research* 33: 1010–28. https://doi.org/10.1139/X03-026.

Sanders, C. J., R. W. Stark, E. J. Mullins, and J. Murphy (eds). 1985. "Recent Advances in Spruce Budworms Research." In *Proceedings of the CANUSA Spruce Budworms Research Symposium, 16-20 September 1984, Bangor, Maine. Canadian Forestry Service, Ottawa, Ontario. 527p.*

Schmitt, D. M., D. G. Grimble, and JL Searcy. 1984. *Managing the spruce budworm in eastern North America.* USDA Forest Service, Cooperative State Research Service, Agriculture Handbook No. 620. 192p., Washington D.C.

Shi, Jingning, Wei Xiang, Qijing Liu, and Sher Shah. 2019. "MtreeRing: An R Package with Graphical User Interface for Automatic Measurement of Tree Ring Widths Using Image Processing Techniques." *Dendrochronologia* 58: 125644. https://doi.org/10.1016/j.dendro.2019.125644.

Speer, James H. 2010. *Fundamentals of Tree-Ring Research.* The University of Arizona Press.

Swetnam, Thomas W., and Ann M. Lynch. 1989. "A tree-ring reconstruction of western spruce budworm history in the southern Rocky Mountains." *Forest Science* 35 (4): 962–86. https://doi.org/10.1093/forestscience/35.4.962.

———. 1993. "Multicentury, Regional-Scale Patterns of Western Spruce Budworm Outbreaks." *Ecological Monographs* 63 (4): 399–424. https://doi.org/10.2307/2937153.

Swetnam, Thomas W., Marna A. Thompson, and Elaine K. Sutherland. 1985. *Using dendrochronology to measure radial growth of defoliated trees.* Vol. 639. United States Department of Agriculture handbook No. 639. 38 p., Washington D.C. www.fs.usda.gov/treesearch/pubs/40898.

Swetnam, TW, CD Allen, and JL Betancourt. 1999. "Applied historical ecology: using the past to manage for the future." *Ecological Applications* 9 (4): 1189–1206. https://doi.org/10.1890/1051-0761(1999)009%5B1189:AHEUTP%5D2.0.CO;2.

van der Maaten-Theunissen, Marieke, Ernst van der Maaten, and Olivier Bouriaud. 2015. "PointRes: An R package to analyze pointer years and components of resilience." *Dendrochronologia* 35: 34–38. https://doi.org/10.1016/j.dendro.2015.05.006.

Wickham, Hadley. 2016. *ggplot2: Elegant Graphics for Data Analysis.* Second edition. Springer-Verlag, New York. https://ggplot2-book.org/.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the Tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.

Wickham, Hadley, Romain François, Lionel Henry, and Kirill Müller. 2020. *Dplyr: A Grammar of Data Manipulation.* R package version 0.8.4. https://CRAN.R-project.org/package=dplyr.

Zang, Christian, and Franco Biondi. 2015. "Treeclim: An R package for the numerical calibration of proxy-climate relationships." *Ecography* 38 (4): 431–36. https://doi.org/10.1111/ecog.01335.

Tables & Figures

Table 1: Summary statistics for individual tree defoliation events.

series	first	last	years	num_events	tot_years	mean_duration
DMJ01	1810	1996	187	4	50	12
DMJ02	1750	1996	247	6	75	12
DMJ03	1830	1996	167	4	42	10
DMJ04	1720	1996	277	8	90	11
DMJ06	1700	1996	297	6	81	14
DMJ07	1710	1996	287	7	96	14
DMJ11	1900	1997	98	2	19	10
DMJ14	1675	1996	322	9	112	12
DMJ15	1730	1996	267	4	64	16
DMJ16	1746	1996	251	7	86	12
DMJ17	1733	1996	264	6	85	14
DMJ22	1720	1996	277	5	56	11
DMJ23	1675	1997	323	8	96	12
DMJ24	1895	1996	102	3	28	9
DMJ25	1680	1996	317	5	70	14
DMJ26	1700	1996	297	6	79	13
DMJ27	1710	1996	287	3	34	11

Table 2: Summary statistics for inferred outbreak events.

| start	end	duration	num_trees_start	perc_trees_start	peak_outbreak_year	num_trees_outbreak	peak_defol_year	min_gsi	mean_ngsi
1680	1699	20	1	33.3	1690	3	1692	0.121	-2.407
1753	1769	17	6	46.2	1754	7	1755	0.343	-1.611
1825	1840	16	11	78.6	1831	12	1826	0.500	-1.304
1849	1865	17	7	46.7	1852	13	1853	0.252	-1.994
1881	1895	15	8	53.3	1886	14	1885	0.262	-1.945
1959	1970	12	7	41.2	1960	15	1965	0.328	-1.830
1987	1996	10	9	52.9	1989	15	1995	0.378	-1.640
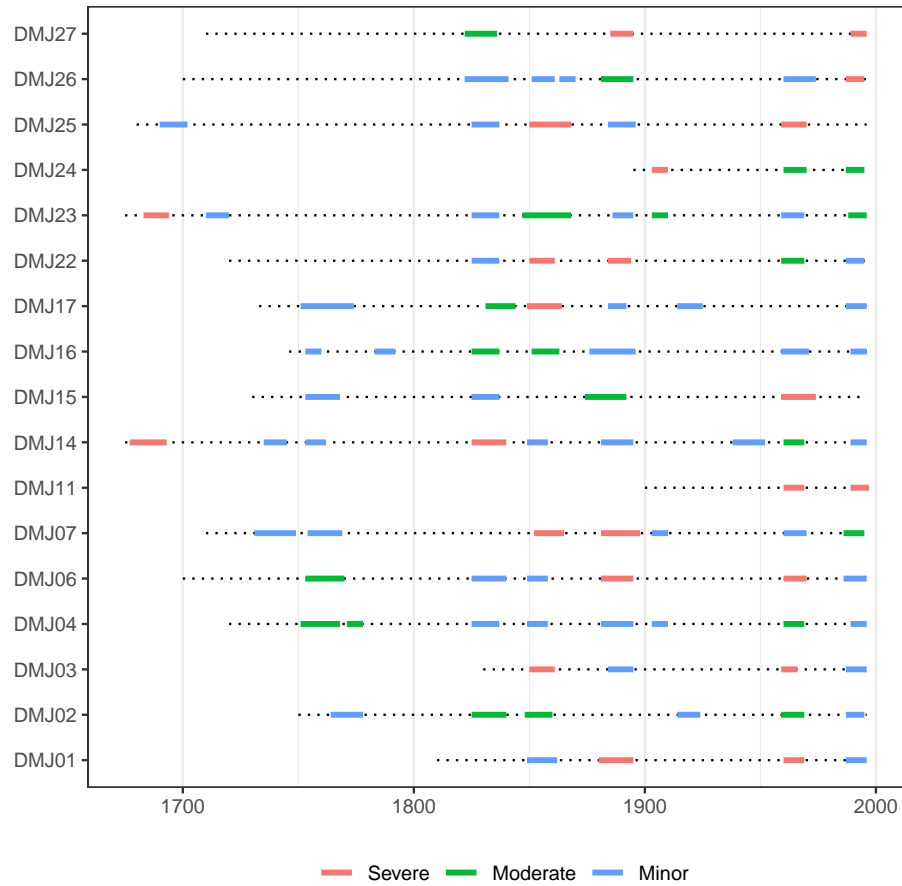
Figure 1: Default dfoliatR graphics for individual trees, produced by the 'plot_defol()' function. Dotted lines represent the series length for each tree and colored segments show periods of defoliation. Users can define the Cut-off values to determine severe-moderate-minor defoliation intensities, and use standard ggplot2 graphics parameters to adjust styles and themes.
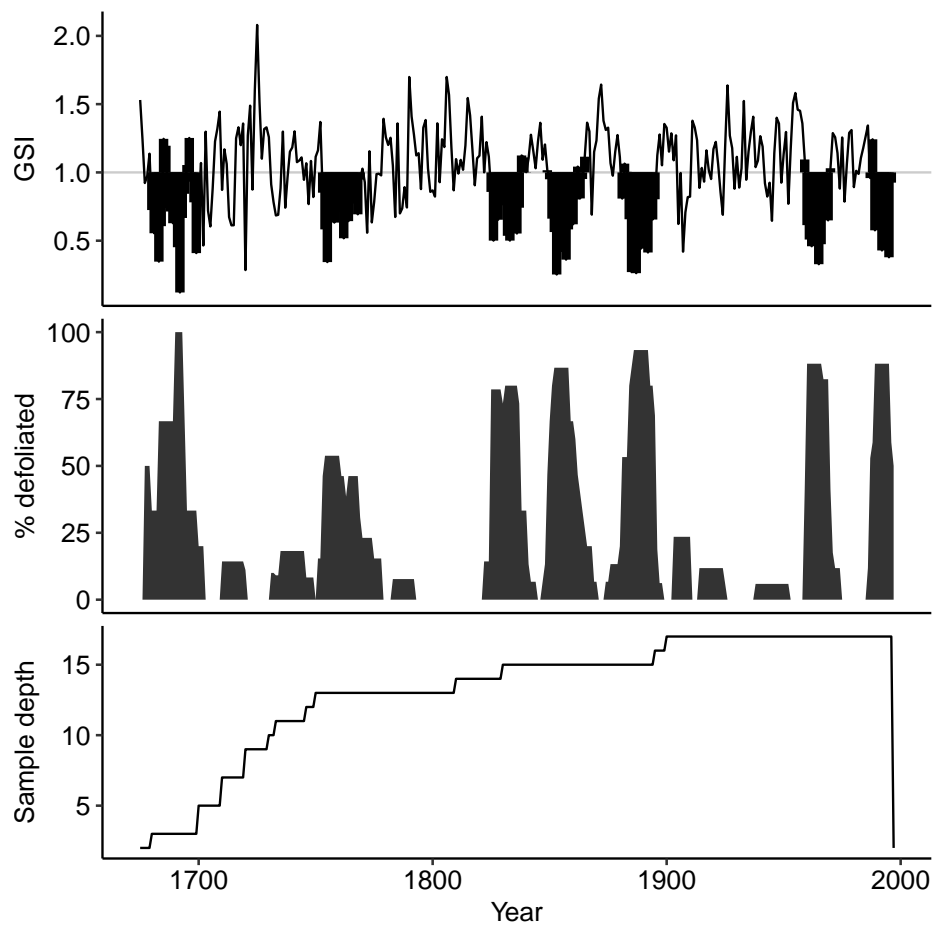
Figure 2: Default graphic to show outbreak events at the DMJ site. Top panel shows the mean GSI through time, with inferred outbreak events filled. The middle panel shows the percent of trees defoliated, and the bottom panel shows the number of trees through time.
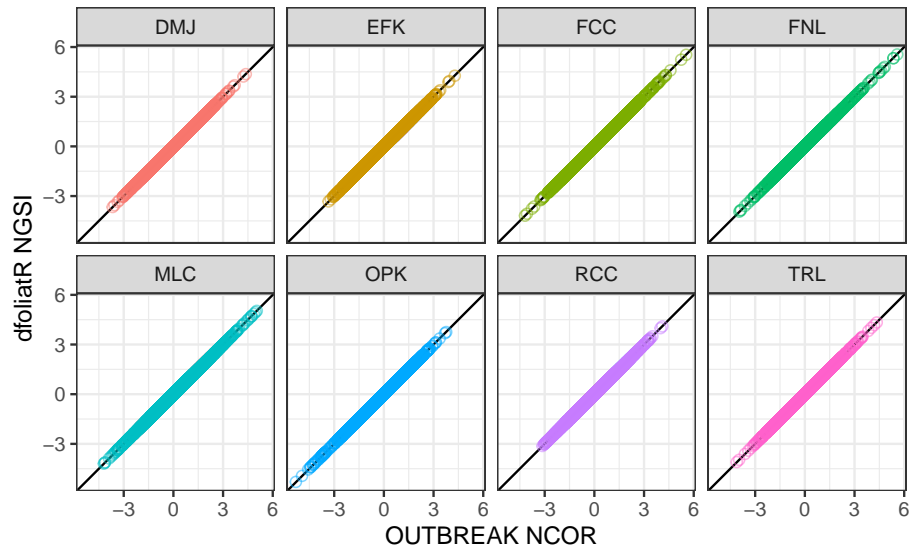
25

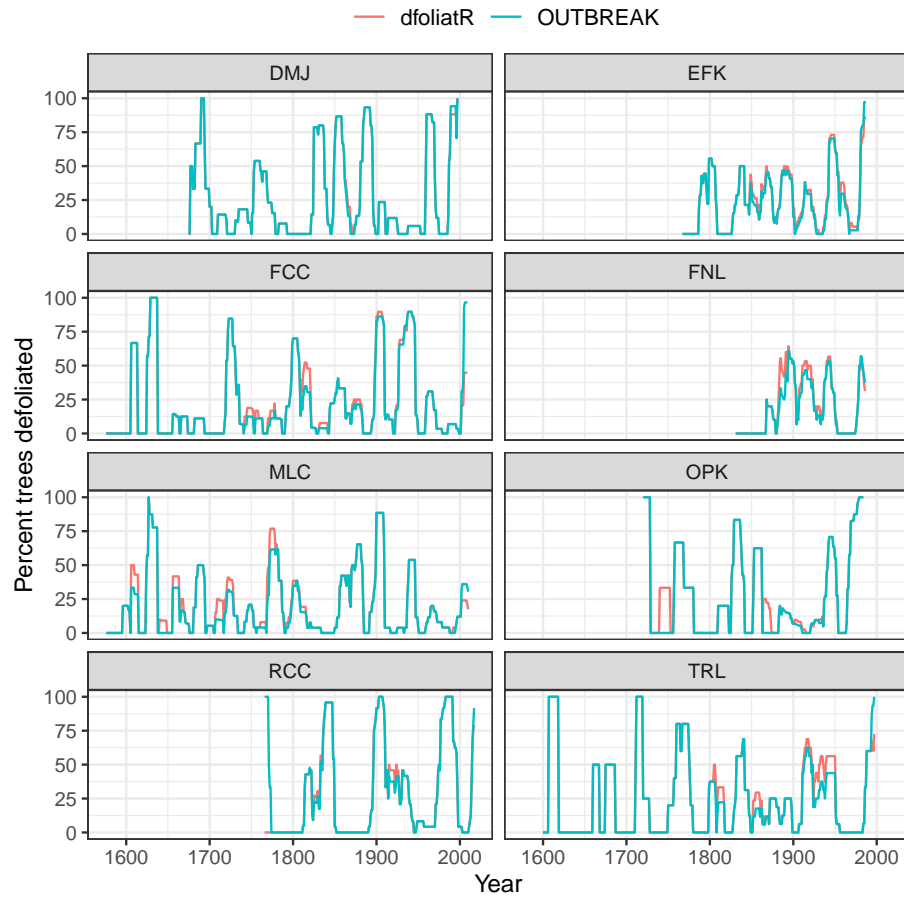Figure 3: Comparison of disturbance indices between dfoliatR and OUTBREAK. Diagonal lines show equal values (1:1).

Figure 4: Comparison of reconstructed western spruce budworm outbreaks computed by dfoliatR and OUTBREAK. Input parameters were identical between programs.