

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^d

^aLaboratory of Tree-Ring Research, University of Arizona, 1215 E Lowell St. Box 210045, Tucson, AZ, 85721

^bThree Pines Forest Research, LLC, PO Box 225, Etna, NH, 03750

^cU.S. Forest Service, Rocky Mountain Research Station, 1215 E Lowell St. Box 210045, Tucson, AZ, 85721

^dDept 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. **dfoliatR** is based on the OUTBREAK program and builds on existing resources in the R computing environment s the well-used **dplR** package. It is designed to aid research in the ecology of insect defoliation events and to reconstruct defoliator outbreak chronologies, but can be applied to other studies where host–non-host comparisons are useful. **dfoliatR** performs an indexing procedure to remove climatic signals in the host-tree series that are represented in the non-host chronology, or other annually-resolved climate 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. We evaluated **dfoliatR** against OUTBREAK, using eight datasets including 222 host-trees, and found that **dfoliatR** improves on OUTBREAK with greater user control, identification of defoliation events, computing capacity, and both the statistical summary and graphical outputs. 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, dendroentomology, OUTBREAK, western spruce

*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)

13 budworm, defoliator regimes, insect outbreak reconstruction

14 1. Introduction

15 Variation in the width and morphology of annual radial growth rings in trees
16 permits dating and quantification of past forest insect defoliator outbreaks. De-
17 foliation can be distinguished from climate- and other disturbance-related influ-
18 ences by comparing ring-width or other annually-resolved features in the wood
19 of host species to that of non-host species or annually-resolved climate records.
20 The effect of defoliation on radial growth of trees has been recognized since the
21 1860s, and used to reconstruct outbreak regimes since the 1950s (Blais, 1954;
22 Alfaro et al., 1982; Lynch, 2012). It was not until the 1980s, however, that pre-
23 cise dendrochronological techniques were applied for inferring defoliation events
24 and reconstructing defoliator outbreak regimes (Swetnam et al., 1985; Speer,
25 2010; Lynch, 2012). The first studies (Swetnam et al., 1985; Swetnam and
26 Lynch, 1989, 1993) focused on developing historical outbreak reconstructions of
27 western spruce budworm (WSBW; *Choristoneura freemani* (Lepidoptera: Tor-
28 tricidae); previously known as *C. occidentalis*). The methodology has since
29 been successfully applied to a wide range of defoliator species, most of which
30 are conifer herbivores, and has evolved in sophistication for a wide range of
31 ecosystem situations (Lynch, 2012).

32 The main dendrochronological tool for inferring, dating, and characteriz-
33 ing defoliator outbreaks from tree-ring records has been the software routine
34 OUTBREAK (Swetnam et al., 1985; Holmes and Swetnam, 1986; Swetnam
35 and Lynch, 1989). OUTBREAK computes indices (described later in detail) of
36 suppressed growth by subtracting a detrended and standardized climate series
37 (a “control” chronology) from individual host-tree detrended and standardized
38 radial growth series after the host and non-host series have been brought to a
39 common variance. The non-host chronology usually consists of a site chronol-
40 ogy developed from non-host tree species growing on a climate-sensitive site,
41 but a gridded climate data point series, like the North American Drought Atlas

42 (Cook and Krusic, 2004) also suffices. If the host and non-host species respond
43 similarly to climate (which can and should be tested), the derived series retains
44 variability that the host and non-host series do not have in common, generally
45 the insect signal and some unexplained variability (noise). The user defines
46 a rule base specifying the magnitude and duration that a period of indexed
47 growth suppression must meet or surpass for a period of suppressed growth to
48 be inferred as a defoliation event at the tree level.

49 Though powerful, OUTBREAK is outdated and increasingly difficult to use
50 in modern computing environments. It was written in FORTRAN V with in-
51 herently severe restrictions, as RAM and disk space were limited at that time
52 (256 kb and 10 MB, respectively) and FORTRAN conventions imposed very
53 strict formatting, file naming, and output conventions. The program lacks a
54 graphical interface or capabilities, forcing users to import generated text files
55 into spreadsheets or other software to assess results and perform analyses. Fur-
56 thermore, OUTBREAK can only handle one test at a time, creating barriers
57 to batch operation and a large burden for researchers with datasets including
58 multiple sites. We developed **dfoliatR** (Guiterman et al., 2020) as an R- and
59 **dp1R**-based library to overcome these issues.

60 **dfoliatR** adds to a growing suite of dendrochronology packages in the R
61 computing environment (R Core Team, 2019). Stemming from the **dp1R** library
62 (Bunn, 2008) that enables R users to read and write an array of tree-ring data
63 formats, standardize ring width series, build and evaluate chronologies, and per-
64 form quality control (to name a few), one can now also measure ring widths from
65 scanned images of prepared samples (Lara et al., 2015; Shi et al., 2019), conduct
66 and check crossdating (Bunn, 2010), analyze sub-annual anatomical features
67 (Rathgeber et al., 2011; Campelo et al., 2016), and perform many analytical
68 tests (Zang and Biondi, 2015; Jevšenak and Levanič, 2018). Tools for assessing
69 stand dynamics and disturbance analyses are under rapid development, with
70 new packages for assessing growth and release events (**TRADER**: Altman et al.,
71 2014), metrics of growth resilience (**pointRes**: van der Maaten-Theunissen et
72 al., 2015), and fire history (**burnr**: Malevich et al., 2018). The key objective

of `dfoliatR` is to provide tools to identify and analyze insect defoliation and outbreak events by building on 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 (`dplyr`: Wickham et al., 2020) and graphics (`ggplot2`: 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 including 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 tabular 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. `dplR` (via the `dplR::treeMean()` function) and `dpl` versions of `ARSTAN` have options for averaging multiple sample series into a tree-level series.

101 At the heart of `dfoliatR` lies two functions: `defoliate_trees()` and `outbreak()`.
 102 These identify defoliation events on individual trees (Figure 1) and then com-
 103 posite across multiple trees to infer stand or site level outbreak events (Figure
 104 2).

105 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

$$\text{GSI}_i = H_i - (\text{NH}_i - \overline{\text{NH}}) \frac{\sigma_H}{\sigma_{\text{NH}}} \quad (1)$$

106 where H and NH are the host-tree series and the non-host chronology, in year i ,
 107 respectively. Only the common period between the host-tree series and the non-
 108 host chronology are used in Equation 1. The host and non-host chronologies are
 109 brought to common variance by scaling the non-host chronology by its mean
 110 $(\overline{\text{NH}})$ and multiplying by the ratio of host and non-host standard deviations
 111 $(\frac{\sigma_H}{\sigma_{\text{NH}}})$, which approximates the variance of the host tree series.

112 Negative departures in the normalized GSI (NGSI, or GSI converted to z-
 113 scores) that surpass user-specified thresholds in duration and magnitude are
 114 defined as *defoliation events*. As in OUTBREAK, the lowest NGSI value in
 115 the particular sequence being assessed must reach the magnitude threshold.
 116 The default setting is -1.28 (NGSI is in units of standard deviation), which
 117 was previously determined to be representative of WSBW effects (Swetnam
 118 and Lynch, 1989) and is commonly used for other species (see Lynch, 2012).
 119 The year with the lowest value is termed the “year of maximum departure”

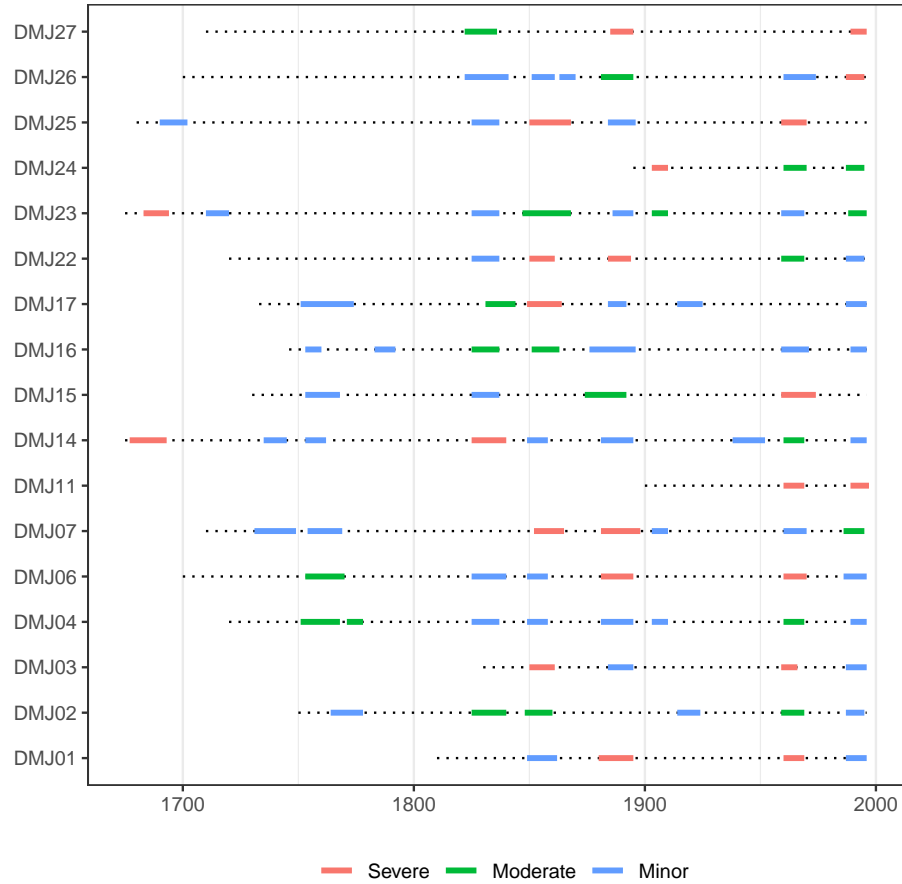


Figure 1: Default **dfoliatR** graphic for individual trees, produced by the `plot_defol()` function. The series names are listed on the Y-axis, dotted lines represent the series length for each tree, and colored segments show periods of defoliation. The colors of defoliation segments represent its severity, for which users can define cut-off values to determine severe–moderate–minor defoliation intensities. The default break points for severity classes are the mean and first quartile for event NGSI values. Colors and other features of the graphic can be adjusted using **ggplot2** parameters, as shown below.

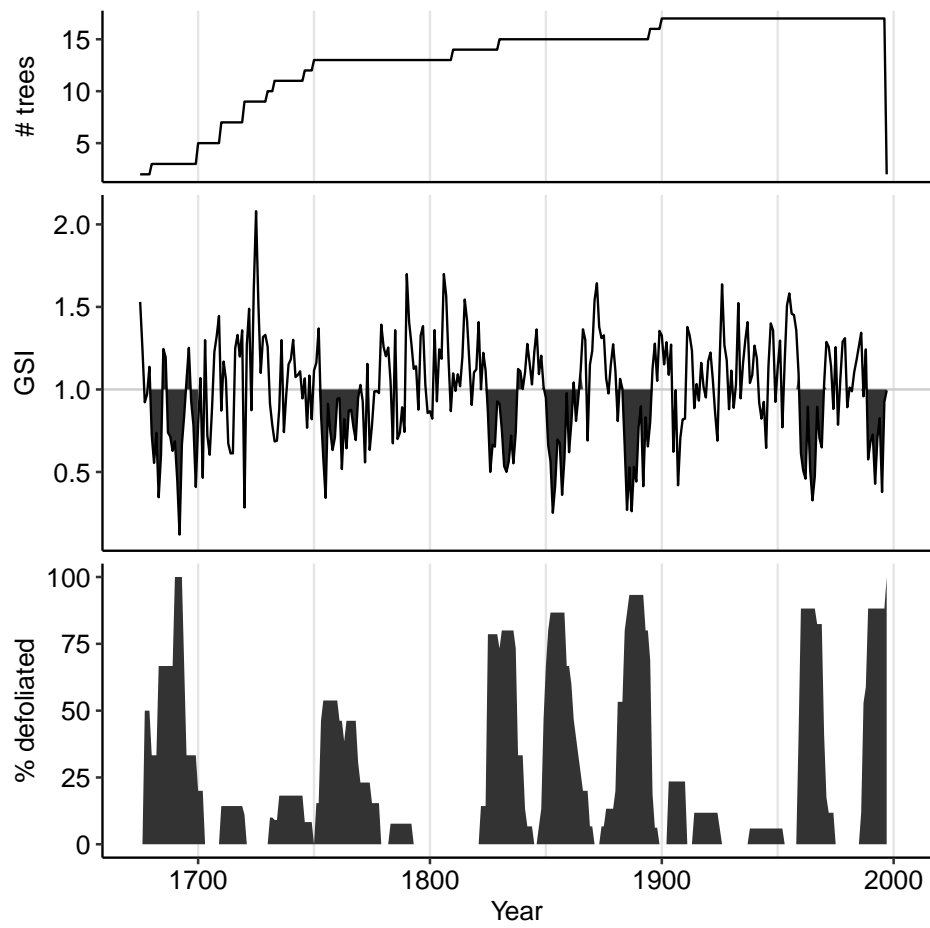


Figure 2: Default graphic showing outbreak events at the DMJ site. Top panel shows the sample depth, the middle panel shows the mean GSI with inferred outbreak events filled, and the bottom panel shows the percent of trees defoliated, used to identify time periods of inferred outbreaks.

120 and becomes a central point in time for assessing other thresholds before being
121 included as a defoliation event. If the year of maximum departure is higher than
122 the threshold (i.e. $\text{NGSI}_{\text{lowest}} > -1.28$), the sequence being assessed is omitted
123 from the event results.

124 Event duration is assessed by examining sequences of negative NGSI (that
125 exceed the magnitude threshold) before and after the year of maximum departure.
126 Each defoliation event is allowed one single-year positive excursion on each
127 side of the year of maximum departure. Duration is computed across the entire
128 sequence that may include these two positive excursions. As in OUTBREAK,
129 the user specifies a duration threshold (minimum number of years) for a departure
130 sequence to be inferred as a defoliation event. The default threshold
131 is eight years, as is commonly used in WSBW studies (Swetnam and Lynch,
132 1989). If the sequence is shorter than the duration threshold, the sequence is
133 omitted from the event results (i.e., both thresholds must be met). Researchers
134 can, and should, adjust the duration and magnitude parameters accordingly
135 and critically evaluate the results, as insect species vary in the length of their
136 outbreaks and the degree to which they can suppress tree growth. OUTBREAK
137 provides two sets of default values, those for WSBW, which typically has lengthy
138 outbreaks, and ones for Douglas-fir tussock moth (*Orgyia pseudotsugata* (Lep-
139 idoptera: Tortricidae)) of three years duration with -1.28 departure threshold
140 that may be suitable for more eruptive species.

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

148 Diverging from OUTBREAK, **dfoliatR** includes an option allowing users
149 to extend defoliation events on individual trees by bridging between sequential
150 events (Figure 3). In cases where two defoliation events are separated by a sin-

151 gle year, bridging will link them into a single event. This option was added to
 152 **dfoliatR** during the testing phase of development, when we realized that OUT-
 153 BREAK deliberately omits sequential, or back-to-back events, even when both
 154 events surpass the magnitude and duration thresholds. Instead, OUTBREAK
 155 will select the one sequential event with the lowest negative departure year. In
 156 every case we assessed (described below) we felt that the OUTBREAK-omitted
 157 defoliation events should have been maintained and recorded. Due in large part
 158 to reconstructions using OUTBREAK (see papers cited by Lynch, 2012), we
 159 now know considerably more about forest defoliator outbreak regimes than we
 160 did in the 1980s when OUTBREAK was under development. We think that two
 161 or more prolonged events separated by a single year should in some situations
 162 be considered a single event. This is particularly relevant to WSBW and spruce
 163 budworm (*C. fumiferana*), for which multiple outbreak regime reconstruction,
 164 as well as other research and forest health observations, show that outbreaks
 165 can be very long (Schmitt et al., 1984; Sanders et al., 1985; Brookes et al., 1987,
 166 and many later publications). Often the greatest growth suppression occurs late
 167 in the outbreak due to lag effects between defoliation and radial growth, and
 168 to cumulative effects accrued on a tree's resources (Brubaker, 1978; Alfaro et
 169 al., 1982; Wickman, 1986; Swetnam and Lynch, 1989, 1993; Mason et al., 1997;
 170 Axelson et al., 2014). We urge caution in using the bridging option, however,
 171 because it may not be appropriate for all studied insects, such as in situations
 172 where impacted stands barely recover from one outbreak before another begins,
 173 as with pine processionary caterpillars (*Thaumetopoea pityocampa* (Lepidoptera:
 174 Thaumetopoeidae)) (Carus, 2004, 2009) or where outbreaks are known to be
 175 very short, such as larch budmoth (*Zeiraphera diniana* Gn.) in the European
 176 Alps (Esper et al., 2007).

177 2.2. Inferring Outbreak Events

178 Defoliation of one or a few trees does not constitute an outbreak. To de-
 179 termine when defoliation becomes an *outbreak event*, **dfoliatR** composites the
 180 individual tree defoliation series into a site-level chronology with the **outbreak()**

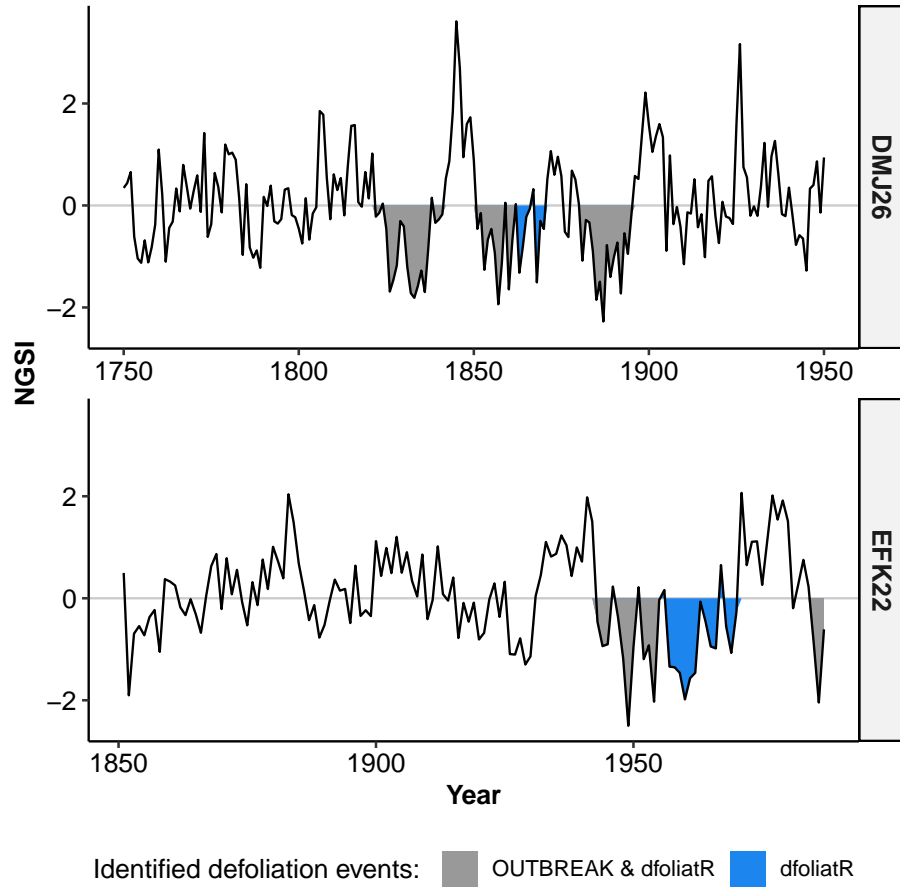


Figure 3: Examples of identified defoliation events on individual trees. For each tree (DMJ26 and EFK22), OUTBREAK and `dfoliatR` identify most of the same events, but there is one added event (in blue) that was omitted by OUTBREAK. These were omitted because they were separated by a single year of positive normalized growth suppression index (NGSI) and OUTBREAK selected the one event with the lowest maximum departure value. `dfoliatR` provides an option to bridge these sequential events into single long events that may better represent the duration of defoliation given the insect and sites under consideration.

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 judgment call as to the confidence ascribed to reduced sample depth toward the ends of their chronologies, thus compensating for the “fading record problem” (Swetnam et al., 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 et al., 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 `OUTBREAK`’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. Our tests used standardized ring-width data from eight host-tree sites spanning the range of `WSBW`. The sites were sampled in British Columbia (Axelson et al., 2015), Wyoming (Axelson et al., 2018), Colorado, and New Mexico (Swetnam and Lynch, 1993; Ryerson et al., 2003). These host data were compared to non-host chronologies from the original studies, but we made no effort here to replicate the reconstructions or analyses of those studies.

We detrended host data for both `dfoliatR` and `OUTBREAK` using `ARSTAN` (version 6.1) with cubic smoothing splines (50% frequency response on 100-150 year wavelengths depending on the site). In both `dfoliatR` and `OUTBREAK` we used event thresholds of -1.28 normalized indices, 8 years duration,

210 and allowed for events at the end of series in seven of eight sites that had known
 211 outbreaks at the time of sampling. We found it necessary to be consistent in
 212 how we detrended and what software we employed (e.g., ARSTAN vs `dp1R`)
 213 because subtle differences in standardized ring-width indices generated between
 214 the programs transferred into differences between `dfoliatR` and OUTBREAK.
 215 In the end, we chose to only use the standardization output files from ARSTAN,
 216 which are easily read into R (and then `dfoliatR`) using the `dp1R` package.

217 The R code to replicate our comparisons is available from https://github.com/chguiterman/dfoliatR_paper.
 218

219 *3.2. Findings*

220 Across the 43,280 ring-width indices from 222 trees included in our eval-
 221 uation, we found that `dfoliatR` and OUTBREAK compute identical growth
 222 suppression indices at 0.00 precision. We expected this outcome because both
 223 programs apply Equation 1 to calculate disturbance indices. At the tree-level,
 224 the programs identified 11,530 total index years with defoliation. The programs
 225 agreed on 97.9% of the years, leaving 927 “difference” years in which only one
 226 program identified defoliation on an individual tree. The differences included
 227 102 events on 85 trees. We carefully inspected each of these events in the full
 228 context of each tree’s ring-series, and categorized the differences as follows

- 229 • *Series-end events* (40% of the total) in which OUTBREAK included “trun-
 230 cated outbreaks” (for seven sites) at the end of each series. In `dfoliatR`,
 231 this option is controlled by the “series_end_events” parameter to `defoliate_trees()`.
 232 In OUTBREAK, the option appears while changing the duration param-
 233 eter (option 3). When selected, OUTBREAK will include any sequences
 234 of negative indices at the beginning and the end of each tree series as a
 235 defoliation event, without consideration of either duration or magnitude
 236 thresholds. In `dfoliatR`, the duration threshold is omitted and the mag-
 237 nitude threshold is retained in series-end-events. Each of the 13 events
 238 included in these differences did not meet the “max_reduction” param-

eter (-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, but `dfoliatR` allowed this single positive excursion. Finally, there were two cases in which `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 back-to-back events that occur one year prior to, or one year following an identified event. When this occurs, OUTBREAK selects the one event sequence with the lowest negative index year (e.g., Figure 3). 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 each surpass the magnitude and duration thresholds) into single, long events. We felt that this was ecologically justified, especially for studies of WSBW, because outbreaks are known to be of long duration and tree-ring reconstructions have shown that outbreaks may persist for as long as 30-50 years at the site level.
- *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.

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. In nearly all site-level comparisons,

268 **dfoliatR** included either more events or it inferred a longer duration outbreak.
269 These differences arise from the inclusion of tree-level events by **dfoliatR** that
270 were omitted by **OUTBREAK** (see note on sequential events above). Thus, in
271 **dfoliatR**, there were a greater number of trees experiencing defoliation during
272 outbreak periods, or outbreaks were represented by a single tree when there was
273 low sample depth.

274 This comparison revealed what we believe are shortcomings in how **OUT-**
275 **BREAK** identifies defoliation events on individual trees. In every one of the
276 102 cases we inspected, we felt that **dfoliatR** provided a more biologically and
277 statistically appropriate assessment of defoliation, translating to more robust
278 inferences of outbreak events and associated statistics at the site level.

279 4. Availability and installation

280 The **dfoliatR** library is provided free and open source from the Compre-
281 hensive R Archive Network (CRAN; <https://cran.r-project.org/>). To install
282 **dfoliatR** from CRAN use

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

283 In each R session, **dfoliatR** can be loaded via

```
library(dfoliatR)
```

284 Development versions of **dfoliatR** are available on GitHub and installed
285 using the **devtools** library,

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

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

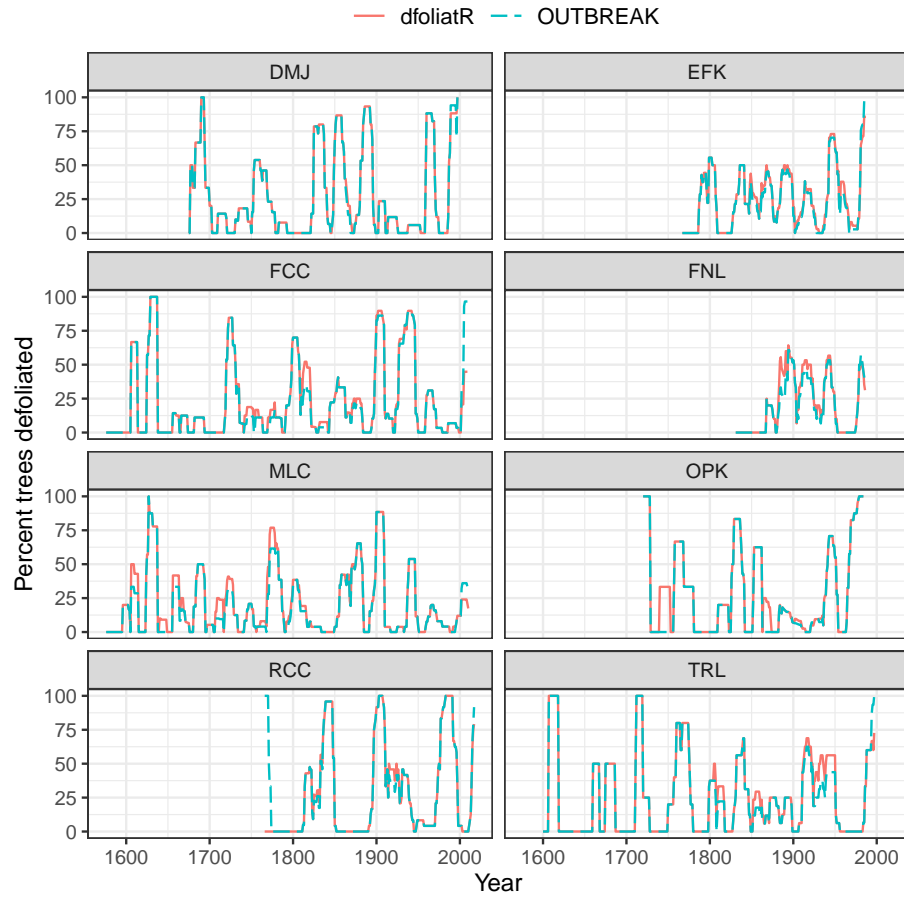


Figure 4: Comparison of reconstructed western spruce budworm outbreaks computed by **dfoliatR** and **OUTBREAK**. Input parameters were identical between programs. Differences arise because **dfoliatR** will identify and record more defoliation events on individual trees.

291 5. Example Usage

292 Once `dfoliatR` is loaded into an R session (via `library(dfoliatR)`) users
293 can access two sets of tree-ring data to aid in exploring the functions, graph-
294 ics, and outputs. Each data set consists of individual host-tree series and a
295 local non-host chronology. The host-tree series were standardized using 128-
296 year splines with a 50% frequency response, while the non-host ring-width data
297 were standardized using 150-year splines with a 50% frequency response and
298 then averaged via Tukey’s biweight robust mean procedure. Host trees from
299 Demijohn Peak (DMJ; 2902 m asl) in the San Juan Mountains of southern
300 Colorado include Douglas-fir (*Pseudotsuga menziesii*) compared against a local
301 non-host ponderosa pine (*Pinus ponderosa*) chronology (Ryerson et al., 2003).
302 The East Fork site (EFK; 2580 m asl) in the Jemez Mountains of north-central
303 New Mexico includes Douglas-fir and white fir (*Abies concolor*) host trees and
304 a ponderosa pine non-host chronology (Swetnam and Lynch, 1993).

305 With `dfoliatR` loaded, the datasets are accessible using the `data()` func-
306 tion. The data object names are prefixed by their site codes. For instance, the
307 `dmj_*` objects come from the DMJ site and include the host-tree series (`dmj_h`),
308 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 denotes functions; black
313 are loaded objects, and green are quoted variables and links. Values or other
314 information provided after equal signs are filenames and parameters provided
315 for this example, and in actual use would be replaced with user-specified infor-
316 mation. In this example “`dmj_h`” and “`dmj_nh`” are the individual-tree host
317 series and non-host site chronology files for the Demijohn site, thresholds are set
318 at 8 years and -1.28 standard deviations, bridging is used, series-end events are
319 included in the interval computations, and comprehensive results information is
320 not included in the output.

321 *5.1. Tree-Level Defoliation Events*

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

```
## 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)
```

324 The result is long-format (stacked) data frame with five variables: “year”,
325 “series”, “gsi”, “ngsi”, and “defol_status.” The “defol_status” column indicates
326 whether that year has defoliation or not, with a set of factors that include
327 “nd” for non-defoliation year, “defol” for a defoliation year, “max_defol” for
328 the year of maximum suppression (that acts as the basis for individual events),
329 “bridge_defol” to identify years that link subsequent events (only one is present
330 at DMJ), and “series_end_defol” to identify defoliation at the present-end of
331 the series.

332 Selecting `list_output = TRUE` in `defoliate_trees()` provides a list-object
333 of data frames, each with an `rw1` object that combines the host tree and non-

334 host series and the other columns created by `defoliate_trees()`. This option
335 is not used by subsequent functions in `dfoliatR`, but researchers can examine
336 it to check the results of the GSI calculation (Equation 1), such as the non-host
337 series after scaling to a common variance with a particular host-tree series.

338 The results of running `defoliate_trees()` can be assessed through graphical
339 and table outputs. The function `get_defol_events()` will provide a list of
340 every defoliation event for every tree, with the corresponding mean “ngsi” value.
341 A summary table of the results for each tree is produced by `defol_stats()` (Table 1).
342

```
dmj_defol_stats <- defol_stats(dmj_defol)
View(dmj_defol_stats)
```

343 The `plot_defol()` function produces a “ggplot” graphics object with line
344 segments showing the measured sequence of each series and a filled segment
345 for each identified defoliation event (Figure 1). The defoliation segments are
346 colored by their relative severity based on their average NGSI value. By default,
347 `plot_defol()` will calculate the average NGSI for all identified events,
348 and assign severity based on the mean and first quartile of the averages. “Se-
349 vere” events have a mean NGSI above the overall average event-period NGSI.
350 “Moderate” events fall between the mean and first quartile. “Minor” events fall
351 below the first quartile. Users can re-define the breaks to suit their needs via
352 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
```

Table 1: Tree-level tabular output provided by the `defol_stats()` function for the DMJ example site.

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

```
## the average NGSI value for each defoliation
## event.
library(ggplot2)
plot_defol(dmj_defol) +
  scale_color_manual(values = c("red", "orange", "purple"))
```

353 These output functions aid in assessing the sensitivity of input parameters
 354 to `defoliate_trees()`, including the duration and magnitude thresholds for

355 identifying defoliation events. Using `plot_defol()` also provides a direct as-
356 sessment of the between-tree variability in defoliation.

357 5.2. Site-Level Events

358 To infer outbreak events at the site level, the function `outbreak()` compos-
359 ites tree-level defoliation series into a single chronology, with input parameters
360 that control thresholds in the number and proportions of trees recording a de-
361 foliation 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)
```

362 Input parameters to `outbreak()` include “`filter_min_series`” to control the
363 chronology cut-off points with regard to sample depth, “`filter_min_defol`” and
364 “`filter_perc`” to control the minimum number and percent of trees recording a
365 defoliation event in a given year. `outbreak()` produces a new data frame with
366 eight variables: “`year`”, “`num_defol`”, “`percent_defol`”, “`num_max_defol`”, “`mean_gsi`”,
367 “`mean_ngsi`”, and “`outbreak_status`.” All of these variables are populated re-
368 gardless of an inferred outbreak event, providing a continuous outbreak recon-
369 struction. The “`num_max_defol`” variable counts the number of trees recording
370 their maximum defoliation in a given year. The “`mean_gsi`” and “`mean_ngsi`”
371 variables provide averages of these indices across all available trees. Finally, the
372 “`outbreak_status`” column shows whether an outbreak event is inferred (“out-
373 break”) or not (“not_obr”).

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

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

377 Inferred outbreak events are shown in the middle panel of Figure 2 as
378 the filled-in spaces. Users can change the time series in this panel with the
379 “disp_index” parameter, choosing between the mean NGSI (the default) or
380 GSI.

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

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

Table 2: Site-level summary statistics for inferred outbreak events provided with the `outbreak_stats()` function for the DMJ example site.

start	end	duration	num_trees_start	perc_trees_start	num_trees_outbreak	peak_outbreak_year	peak_defol_year	min_gsi	min_ngsi
1680	1699	20	1	33.3	3	1690	1692	0.121	-2.407
1753	1769	17	6	46.2	7	1754	1755	0.343	-1.611
1825	1840	16	11	78.6	12	1831	1826	0.500	-1.304
1849	1865	17	7	46.7	13	1852	1853	0.252	-1.994
1881	1895	15	8	53.3	14	1886	1885	0.262	-1.945
1959	1970	12	7	41.2	15	1960	1965	0.328	-1.830
1987	1996	10	9	52.9	15	1989	1995	0.378	-1.640

383 The output table from `outbreak_stats()` shows the start and end years of
384 each event, along with their corresponding duration, the number and percent of
385 trees in defoliation at the start of the event, the the number of trees recording
386 the outbreak event (in part or full), the minimums of the mean GSI and NGSI
387 indices, and two points of “peak defoliation.” The first, “`peak_outbreak_year`”
388 is the year in which the greatest number of trees are recording the outbreak,
389 and the second, “`peak_defol_year`” is the year with the lowest average NGSI.

390 Saving the results of `outbreak_stats()` (the `dmj_obr_stats` object above)
391 provides an array of options for assessing metrics of the insect outbreak regime.
392 Here we use the first year of each outbreak event to calculate the average dura-
393 tion of years between outbreaks, via the `diff()` function in R. The average of
394 those differences, calculated via `mean()` is the mean return interval.

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

395 6. Conclusions

396 The `dfoliatR` package provides dendroecologists with tools to infer, quan-
397 tify, analyze, and visualize tree-ring growth suppression events and to recon-
398 struct forest insect defoliator outbreak regimes. It is built on the long-accepted
399 host to non-host comparison methodology used in the 1980s FORTRAN pro-
400 gram OUTBREAK (Swetnam et al., 1985; Swetnam and Lynch, 1989, 1993).
401 Our evaluation of the two programs revealed that `dfoliatR` excelled in identi-
402 fying defoliation events on single trees, providing researchers with more consis-
403 tent and biologically-justifiable results. Other key benefits are that `dfoliatR`
404 provides easier control of suppression thresholds, additional output tables, high-
405 quality and customizable graphics, and operates in the open-source R environ-
406 ment that is stable across computing platforms and is under active development
407 and maintenance by a large and growing community.

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

419 **dfoliatR** adds a new option for dendroentimology to combine, or “bridge,”
420 sequential tree-level defoliation events into single events. In practice, we suggest
421 that researchers carefully evaluate if bridging is ecologically applicable to study
422 situations and insect ecology, and to carefully explore the data before deciding
423 whether or not to use this option. It is probably not appropriate for insects
424 with high-frequency, high-severity outbreaks, such as processionary caterpillars.
425 Outbreak reconstructions of insects for which the interval is notably longer than
426 typical outbreak duration, such as Douglas-fir tussock moth, are unlikely to be
427 significantly affected. Species for which outbreak duration or individual tree
428 resilience to defoliation varies considerably, or which may chronically infest trees
429 or sites, or alternate between chronic, outbreak, and minimal activity states such
430 as conifer-feeding *Choristoneura*, present more complicated challenges, and the
431 researcher should use discretion with the bridge option.

432 **dfoliatR** adds to the on-going open-source software development for den-
433 drochronological methods (e.g., Bunn, 2008; Brewer, 2014; Brewer and Gui-
434 terman, 2016). The **R** environment enables automation of analyses, allowing
435 input/output processes to become routine, enables efficient sensitivity analy-
436 ses, and empowers batch processing of large multi-site projects. It also fa-
437 cilitates additional statistical analyses, such as spectral analyses and super-
438 posed epoch analyses (e.g., Malevich et al., 2018), with easy transfer from

439 `dfoliatR` and `dplR` to other libraries in R. Source code for `dfoliatR` is avail-
440 able in the Comprehensive R Archive Network (CRAN) and GitHub <https://github.com/chguiterman/dfoliatR> with updated descriptions and helpful vi-
441 gnettes on the package website <https://chguiterman.github.io/dfoliatR/>. Re-
442 searchers wishing to contribute to the further development of `dfoliatR` are
443 encouraged to do so via the GitHub repository.

445 Acknowledgments

446 We are grateful to late Richard Holmes of the Laboratory of Tree-Ring Re-
447 search (LTRR) for his achievements in developing OUTBREAK and many other
448 tree-ring software programs. We also thank Thomas Swetnam for his insights
449 and encouragement, and David Frank for his support. The `dfoliatR` project is
450 funded by the Agnes Haury Visiting Scholars Fellowship (to J. Axelson) at the
451 LTRR, the University of California, Berkeley, and the Western Wildlands Envi-
452 ronmental Threat Center (WWETAC). Additional support was provided by the
453 LTRR and the United States Forest Service Rocky Mountain Research Station.
454 The findings and conclusions in this publication are those of the authors and
455 should not be construed to represent any official USDA or U.S. Government
456 determination or policy.

457 References

- 458 Alfaro, R., Sickle, G. van, Thomson, A., Wegwitz, E., 1982. Tree mortality
459 and radial growth losses caused by the western spruce budworm in a Douglas-fir
460 stand in British Columbia. *Canadian Journal of Forest Research* 12, 780–87.
- 461 Altman, J., Fibich, P., Dolezal, J., Aakala, T., 2014. **TRADER**: A package for
462 Tree Ring Analysis of Disturbance Events in R. *Dendrochronologia* 32, 107–12,
463 doi:10.1016/j.dendro.2014.01.004.
- 464 Axelson, J., Bast, A., Alfaro, R., Smith, D., Gärtner, H., 2014. Variation
465 in wood anatomical structure of Douglas-fir defoliated by the western spruce

466 budworm: a case study in the coastal-transitional zone of British Columbia,
467 Canada. *Trees* 28, 1837–46, doi:10.1007/s00468-014-1091-1.

468 Axelson, J., Csank, A., Lee, R., McGee, J., McNiel, K., 2018. Reconstructing
469 historical *Choristoneura freemani* outbreaks using *Pseudotsuga menziesii* in the
470 greater Yellowstone region, Wyoming, U.S.A. Final Report of the 20th Annual
471 North American Dendroecological Fieldweek (NADEF).

472 Axelson, J.N., Smith, D.J., Daniels, L.D., Alfaro, R.I., 2015. Multicen-
473 tury reconstruction of western spruce budworm outbreaks in central British
474 Columbia, Canada. *Forest Ecology and Management* 335, 235–48, doi:10.1016/j.foreco.2014.10.002.

475 Blais, J., 1954. The recurrence of spruce budworm infestations in the past
476 century in the lac Seul area of northwestern Ontario. *Ecology* 35, 62–71.

477 Brewer, P.W., 2014. Data management in dendroarchaeology using tellervo.
478 *Radiocarbon* Cambridge University Press 56, S79–S83, doi:10.2458/azu_rc.56.18320.

479 Brewer, P.W., Guiterman, C.H., 2016. A new digital field data collection sys-
480 tem for dendrochronology. *Dendrochronologia* 38, 131–35, doi:10.1016/j.dendro.2016.04.005.

481 Brookes, M.H., Campbell, R.W., Colbert, J.J., Mitchell, R.G., Stark, R.W.,
482 1987. Western spruce budworm. USDA Forest Service, Cooperative State Re-
483 search Service, Technical Bulletin No. 1694. 198 p., Washington D.C.

484 Brubaker, L.B., 1978. Effects of defoliation by Douglas-fir tussock moth on
485 ring sequences of Douglas-fir and grand fir. *Tree-Ring Bulletin* 38, 49–60.

486 Bunn, A.G., 2008. A dendrochronology program library in R (**dp1R**). *Den-*
487 *drochronologia* 26, 115–24, doi:10.1016/j.dendro.2008.01.002.

488 Bunn, A.G., 2010. Statistical and visual crossdating in R using the **dp1R**
489 library. *Dendrochronologia* 28, 251–58, doi:10.1016/j.dendro.2009.12.001.

490 Campelo, F., Nabais, C., Carvalho, A., Vieira, J., 2016. **tracheideR**–
491 An R package to standardize tracheidograms. *Dendrochronologia* 37, 64–68,
492 doi:10.1016/j.dendro.2015.12.006.

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

496 Carus, S., 2009. Effects of defoliation caused by the processionary moth on

497 growth of Crimean pines in western Turkey. *Phytoparasitica* 37, 105–14.

498 Cook, E.R., Holmes, R.L., 1996. Guide for computer program ARSTAN.

499 The international tree-ring data bank program library version 2 75–87.

500 Cook, E.R., Krusic, P.J., 2004. The North American Drought Atlas. (Lamont-
501 Doherty Earth Observatory; the National Science Foundation: New York) Avail-
502 able at [https://www.ncdc.noaa.gov/data-access/paleoclimatology-data/datasets/tree-](https://www.ncdc.noaa.gov/data-access/paleoclimatology-data/datasets/tree-ring/drought-variability)
503 [ring/drought-variability](https://www.ncdc.noaa.gov/data-access/paleoclimatology-data/datasets/tree-ring/drought-variability) [Last accessed May 2020].

504 Esper, J., Büntgen, U., Frank, D.C., Nievergelt, D., Liebhold, A., 2007. 1200
505 years of regular outbreaks in alpine insects. *Proceedings of the Royal Society*
506 *B: Biological Sciences* 274, 671–79, doi:10.1098/rspb.2006.0191.

507 Guiterman, C., Lynch, A., Axelson, J., 2020. **dfoliatR**: Detection and Anal-
508 ysis of Insect Defoliation Signals in Tree Rings., [https://chguiterman.github.io/](https://chguiterman.github.io/dfoliatR/)
509 [dfoliatR/](https://chguiterman.github.io/dfoliatR/), doi:10.5281/zenodo.3698163.

510 Holmes, R.L., Swetnam, T.W., 1986. Dendroecology program library: pro-
511 gram OUTBREAK user’s manual. Laboratory of Tree-Ring Research, Univer-
512 sity of Arizona, Tucson.

513 Jevšenak, J., Levanič, T., 2018. **dendroTools**: R package for studying lin-
514 ear and nonlinear responses between tree-rings and daily environmental data.
515 *Dendrochronologia* 48, 32–39, doi:10.1016/j.dendro.2018.01.005.

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

519 Lynch, A.M., 2012. What Tree-Ring Reconstruction Tells Us about Conifer
520 Defoliator Outbreaks, in: Barbosa, P., Letourneau, D.K., Agrawal, A.A. (Eds.),
521 *Insect outbreaks revisited*. Blackwell Publishing Ltd, pp. 126–54, doi:10.1002/9781118295205.ch7.

522 Malevich, S.B., Guiterman, C.H., Margolis, E.Q., 2018. **burnr**: Fire history
523 analysis and graphics in R. *Dendrochronologia* 49, 9–15, doi:10.1016/j.dendro.2018.02.005.

524 Mason, R.R., Wickman, B.E., Paul, H.G., 1997. Radial growth response of
525 douglas-fir and grand fir to larval densities of the douglas-fir tussock moth and
526 the western spruce budworm. *Forest Science* 43, 194–205.

527 Rathgeber, C.B.K., Longuetaud, F., Mothe, F., Cuny, H., Le Moguédec, G.,

2011. Phenology of wood formation: Data processing, analysis and visualisation
 using R (package **CAVIAR**). *Dendrochronologia* 29, 139–49, doi:10.1016/j.dendro.2011.01.004.

R Core Team, 2019. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, <https://www.R-project.org/>.

Ryerson, D.E., Swetnam, T.W., Lynch, A.M., 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, doi:10.1139/X03-026.

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

Schmitt, D.M., Grimbale, D.G., Searcy, J., 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, J., Xiang, W., Liu, Q., Shah, S., 2019. **MtreeRing**: An R package with graphical user interface for automatic measurement of tree ring widths using image processing techniques. *Dendrochronologia* 58, 125644, doi:10.1016/j.dendro.2019.125644.

Speer, J.H., 2010. Fundamentals of Tree-Ring Research. The University of Arizona Press, 333 pp.

Swetnam, T., Allen, C., Betancourt, J., 1999. Applied historical ecology: using the past to manage for the future. *Ecological applications* 9, 1189–1206, doi:10.1890/1051-0761(1999)009[1189:AHEUTP]2.0.CO;2.

Swetnam, T.W., Lynch, A.M., 1989. A tree-ring reconstruction of western spruce budworm history in the southern Rocky Mountains. *Forest Science* 35, 962–86, doi:10.1093/forestscience/35.4.962.

Swetnam, T.W., Lynch, A.M., 1993. Multicentury, Regional-Scale Patterns of Western Spruce Budworm Outbreaks. *Ecological Monographs* 63, 399–424, doi:10.2307/2937153.

Swetnam, T.W., Thompson, M.A., Sutherland, E.K., 1985. Using den-

drochronology to measure radial growth of defoliated trees. United States Department of Agriculture handbook No. 639. 38 p., Washington D.C, 38p. pp, www.fs.usda.gov/treearch/pubs/40898.

van der Maaten-Theunissen, M., van der Maaten, E., Bouriaud, O., 2015. **PointRes**: An R package to analyze pointer years and components of resilience. *Dendrochronologia* 35, 34–38, <http://dx.doi.org/10.1016/j.dendro.2015.05.006>, doi:10.1016/j.dendro.2015.05.006.

Wickham, H., 2016. **ggplot2**: Elegant Graphics for Data Analysis. Springer-Verlag, New York, <https://ggplot2-book.org/>.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Müller, K., Ooms, J., Robinson, D., Seidel, D., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K., Yutani, H., 2019. Welcome to the **Tidyverse**. *Journal of Open Source Software The Open Journal* 4, 1686, <https://doi.org/10.21105/joss.01686>, doi:10.21105/joss.01686.

Wickham, H., François, R., Henry, L., Müller, K., 2020. **dplyr**: A Grammar of Data Manipulation., <https://CRAN.R-project.org/package=dplyr>.

Wickman, B.E., 1986. Radial growth of grand fir and douglas-fir 10 years after defoliation by the douglas-fir tussock moth in the blue mountains outbreak. USDA Forest Service, Pacific Northwest Research Station, Research Paper PNW-367. Portland, Oregon.

Zang, C., Biondi, F., 2015. **treeclim**: An R package for the numerical calibration of proxy-climate relationships. *Ecography* 38, 431–36, doi:10.1111/ecog.01335.