

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

3 Christopher H. Guiterman^{*,a,b}, Ann M. Lynch^{a,c}, Jodi N. Axelson^c

4 ^a*Laboratory of Tree-Ring Research, University of Arizona, 1215 E Lowell St. Box 210045,*
5 *Tucson, AZ, 85721*

6 ^b*Three Pines Forest Research, LLC, PO Box 225, Etna, NH, 03750*

7 ^c*U.S. Forest Service, Rocky Mountain Research Station, 1215 E Lowell St. Box 210045,*
8 *Tucson, AZ, 85721*

9 ^d*Dept of Environmental Science, Policy & Management, University of California, Berkeley,*
10 *Berkeley, CA 94720*

11 **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.

12 *Key words:* Dendroecology, spruce budworm, *Choristoneura*, pandora moth,
13 *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)
Preprint submitted to Dendrochronologia May 6, 2020

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.
17 Defoliation can be distinguished from climate- and other disturbance-related
18 influences by comparing ring-width or other annually-resolved features in the
19 wood of host species to that of non-host species or annually-resolved climate
20 records. The effect of defoliation on radial growth of trees has been recog-
21 nized since the 1860s, and used to reconstruct outbreak regimes since the 1950s
22 (Blais 1954; Lynch 2012). It was not until the 1980s, however, that precise
23 dendrochronological techniques were applied for inferring defoliation events and
24 reconstructing defoliator outbreak regimes (Swetnam, Thompson, and Suther-
25 land 1985; Speer 2010; Lynch 2012). The first studies (Swetnam, Thompson,
26 and Sutherland 1985; Swetnam and Lynch 1989, 1993) focused on developing
27 historical reconstructions of western spruce budworm (WSBW; *Choristoneura*
28 *freemani* (Lepidoptera: Tortricidae); previously known as *C. occidentalis*) out-
29 break regimes. The methodology has since been successfully applied to a wide
30 range of defoliator species, most of which are conifer herbivores, and has evolved
31 in sophistication for a wide range of ecosystem situations (Lynch 2012).

32 The main dendrochronological tool for inferring, dating, and characterizing
33 defoliator outbreaks from tree-ring records has been the software routine OUT-
34 BREAK (Swetnam, Thompson, and Sutherland 1985; Holmes and Swetnam
35 1986; Swetnam and Lynch 1989). OUTBREAK computes indices (described
36 later in detail) of suppressed growth by subtracting a detrended and standard-
37 ized climate series (a “control” chronology) from individual host-tree detrended
38 and standardized radial growth series after the host and non-host series have
39 been brought to a common variance. The non-host chronology usually consists
40 of a site chronology developed from non-host tree species growing on a climate-
41 sensitive site, but a gridded climate data point series, like the North American
42 Drought Atlas (Cook and Krusic 2004) also suffices. If the host and non-host
43 species respond similarly to climate (which can and should be tested), the de-

44 rived series retains variability that the host and non-host series do not have in
45 common, generally the insect signal and some unexplained variability (noise).
46 The user defines a rule base specifying the magnitude and duration that a period
47 of indexed growth suppression must meet or surpass for a period of suppressed
48 growth to 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 performing analyses.
56 Furthermore, OUTBREAK can only handle one test at a time, creating barriers
57 to batch operation and a large burden for researchers with a datasets includ-
58 ing multiple sites. We developed **dfoliatR** as an R- and **dp1R**-based library to
59 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 (Shi et al. 2019; Lara, Bravo, and Sierra
66 2015), conduct and check crossdating (Bunn 2010), analyze sub-annual anotom-
67 ical features (Campelo et al. 2016; Rathgeber et al. 2011), and perform many
68 analytical tests (Jevšenak and Levanič 2018; Zang and Biondi 2015). Tools for
69 assessing stand dynamics and disturbance analyses are under rapid development,
70 with new packages for assessing growth and release events (**TRADER**: Altman et
71 al. 2014), metrics of growth resilience (**pointRes**: van der Maaten-Theunissen,
72 van der Maaten, and Bouriaud 2015), and fire history (**burnr**: Malevich, Gui-
73 terman, and Margolis 2018). The key objective of **dfoliatR** is to provide tools
74 to identify and analyze insect defoliation and outbreak events by building on

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

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

91 2. Overview of the software

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

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

97 Users can develop these data sets in the software of their choosing, such as
98 `dplR` or ARSTAN (Cook and Holmes 1996). It is important that the host-tree
99 data include only one tree-ring series per tree. Both `dplR` and ARSTAN have
100 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 identify outbreak events (Figure 2).

104 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

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

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

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

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

entire negative sequence that allows these two positive excursions. As in OUTBREAK, 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.

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

165 2.2. Inferring Outbreak Events

166 Defoliation of one or a few trees does not constitute an outbreak. To de-
 167 termine when defoliation becomes an *outbreak event*, `dfoliatR` composites the
 168 individual tree defoliation series into a site-level chronology with the `outbreak()`
 169 function. Users have options to define the number and/or the proportion of trees
 170 required for an event to be considered an outbreak. Three parameters control
 171 whether a defoliation event constitutes an outbreak: the minimum number of
 172 trees available, the minimum number of trees recording defoliation, and the
 173 percent of trees recording defoliation. The first allows the researcher to make a
 174 judgement call as to the confidence ascribed to reduced sample depth toward the
 175 ends of their chronologies, thus compensating for the “fading record problem”
 176 (Swetnam, Allen, and Betancourt 1999). The second two parameters adjust the
 177 scale of defoliation considered to be an outbreak. Absolute numbers of trees
 178 and percentages can be applied separately or in conjunction, following filtering
 179 conventions in tree-ring fire history studies (Malevich, Guiterman, and Margolis
 180 2018). We urge users to carefully consider the choice of absolute numbers in
 181 situations where the number of trees represented in the series varies with time,
 182 or the choice of percentages when sample size is small.

183 3. Evaluation

184 3.1. Approach

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

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

203 The R code to replicate our comparisons is available from [https://github.](https://github.com/chguiterman/dfoliatR_paper)
204 [com/chguiterman/dfoliatR_paper](https://github.com/chguiterman/dfoliatR_paper).

205 3.2. Findings

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

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

- 215 • *Series-end events* (40% of the total) in which OUTBREAK included “trun-
216 cated outbreaks” (for seven sites) at the end of each series. In `dfoliatR`,
217 this option is controlled by the “series_end_events” parameter to `defoliate_trees()`.
218 In OUTBREAK, the option appears while changing the duration parame-
219 ter (option 3). When selected, OUTBREAK will include any sequences of
220 negative indices at the either the beginning or the end of each tree series
221 as a defoliation event, without consideration of either duration or magni-
222 tude thresholds. In `dfoliatR`, the duration threshold is omitted and the
223 magnitude threshold is retained in series-end-events. Each of the 13 events
224 included in these differences did not meet the “max_reduction” parameter
225 (-1.28 NGSI) in `dfoliatR` and were excluded. In two cases, OUTBREAK
226 included events at the beginning of the series where `dfoliatR` does not
227 allow truncated events. In four cases, OUTBREAK omitted only the last
228 year of the series because the index was positive. In two cases, `dfoliatR`
229 omitted possible events because it had already included a positive NGSI
230 excursion after the “max_reduction” year, and since it will only allow one
231 excursion on either side of the max year, the events were omitted due to
232 short duration.
- 233 • *Sequential events* (36%) in which OUTBREAK omitted events that occur
234 one year prior or one year following an identified event. When this occurs,
235 OUTBREAK selects the event with the greatest negative index year. On
236 two trees, OUTBREAK omitted two of three sequential events. While
237 inspecting these differences, we added an option to `defoliate_trees()`
238 that would “bridge” between sequential events (that surpass magnitude
239 and duration thresholds) into single, long events. We felt that this was
240 ecologically justified, especially for studies of WSBW, because the insect
241 will occasionally defoliate individual trees for decades and reconstructed
242 chronologies show outbreaks as long as xxx (**ANN CHECK**).

- *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 OUTBREAK 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

```
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 Demi-john 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

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

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

299 5.1. Tree-Level Defoliation Events

300 The function `defoliate_trees()` performs the GSI indexing procedure on
301 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)
```

302 The result is long-format (stacked) data frame with five variables: “year”,
303 “series”, “gsi”, “ngsi”, and “defol_status.” The “defol_status” column indicates
304 whether that year has defoliation or not, with a set of factors that include
305 “nd” for non-defoliation year, “defol” for a defoliation year, “max_defol” for

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

310 Selecting `list_output = TRUE` in `defoliate_trees()` provides a list-object
311 of data frames, each with an `rw1` object that combines the host tree and non-
312 host series and the other columns created by `defoliate_trees()`. This option
313 is not used by subsequent functions in `dfoliatR`, but researchers can examine
314 it to check the results of the GSI calculation (Equation 1), such as the non-host
315 series after scaling to a common variance with a particular host-tree series.

316 The results of running `defoliate_trees()` can be assessed through graphi-
317 cal and table outputs. The function `get_defol_events()` will provide a list of
318 every defoliation event for every tree, with the corresponding mean “ngsi” value.
319 A summary table of the results for each tree is produced by `defol_stats()` (Ta-
320 ble 1).

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

321 The `plot_defol()` function produces a “ggplot” graphics object with line
322 segments showing the measured sequence of each series and a filled segment for
323 each identified defoliation event (Figure 1). The defoliation segments are colored
324 by their relative severity based on their average NGSI value. The default cut-off
325 values between “Severe” and “Moderate” is the overall mean across all events.
326 The cut-off between “Moderate” and “Minor” is the first quartile by default.
327 Users can re-define the breaks to suit their needs via the “breaks” parameter in
328 `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"))

```

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

333 5.2. Site-Level Events

334 To infer outbreak events at the site level, the function `outbreak()` compos-
 335 ites tree-level defoliation series into a single chronology, with input parameters
 336 that control thresholds in the number and proportions of trees recording a de-
 337 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)

```

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

343 “mean_ngsi”, and “outbreak_status.” All of these variables are populated
344 regardless of an inferred outbreak event, providing a continuous disturbance
345 chronology. The “num_max_defol” variable counts the number of trees record-
346 ing their maximum defoliation in a given year. The “mean_gsi” and “mean_ngsi”
347 variables provide averages of these indices across all available trees. Finally, the
348 “outbreak_status” column shows whether an outbreak event is inferred (“out-
349 break”) or not (“not_obr”).

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

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

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

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

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

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

```

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

```

367 6. Conclusions

368 The **dfoliatR** package provides dendroecologists with tools to infer, quan-
 369 tify, analyze, and visualize tree-ring growth suppression events and to recon-
 370 struct forest insect defoliator outbreak regimes. It is built on the long-accepted
 371 host to nonhost comparison methodology used in the 1980s FORTRAN program
 372 OUTBREAK (Swetnam, Thompson, and Sutherland 1985; Swetnam and Lynch
 373 1989, 1993). Key benefits are that **dfoliatR** provides easier control of suppres-
 374 sion thresholds, provides additional output tables and high-quality graphics,
 375 defines defoliation events more consistently, and operates in the open-source R
 376 environment.

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

388 **dfoliatR** adds a new option for dendroentimology to combine, or “bridge,”
 389 sequential tree-level defoliation events into single events. In practice, we sug-
 390 gest that researchers carefully evaluate if bridging is ecologically applicable to
 391 study situations and insect ecology, and to carefully explore the data before
 392 deciding whether or not to use this option. It is probably not appropriate

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

We are grateful to late Richard Holmes of the Laboratory of Tree-Ring Research (LTRR) for his achievements in developing OUTBREAK and many other tree-ring software programs, and we thank Thomas Swetnam for his insights and encouragement. The **dfoliatR** project was supported by the Agnes Haury Visiting Scholars Fellowship at the LTRR, the University of California Berkeley, and the Western Wildlands Environmental Threat Center (WWETAC). Addi-

422 tional support was provided by the LTRR and the United States Forest Service
423 Rocky Mountain Research Station. Any use of trade, firm, or product names
424 is for descriptive purposes only and does not imply endorsement by the U.S.
425 Government.

426 References

427 Altman, J., P. Fibich, J. Dolezal, and T. Aakala. 2014. "TRADER: A
428 package for Tree Ring Analysis of Disturbance Events in R." *Dendrochronologia*
429 32: 107–12. <https://doi.org/10.1016/j.dendro.2014.01.004>.

430 Blais, JR. 1954. "The Recurrence of Spruce Budworm Infestations in the
431 Past Century in the Lac Seul Area of Northwestern Ontario." *Ecology* 35 (1):
432 62–71.

433 Brewer, Peter W. 2014. "Data Management in Dendroarchaeology Using
434 Tellervo." *Radiocarbon* 56 (4): S79–S83. [https://doi.org/10.2458/azu_rc.56.](https://doi.org/10.2458/azu_rc.56.18320)
435 18320.

436 Brewer, Peter W., and Christopher H. Guiterman. 2016. "A New Digital
437 Field Data Collection System for Dendrochronology." *Dendrochronologia* 38:
438 131–35. <https://doi.org/10.1016/j.dendro.2016.04.005>.

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

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

445 ———. 2010. "Statistical and visual crossdating in R using the dplR li-
446 brary." *Dendrochronologia* 28: 251–58. [https://doi.org/10.1016/j.dendro.2009.](https://doi.org/10.1016/j.dendro.2009.12.001)
447 12.001.

448 Campelo, Filipe, Cristina Nabais, Ana Carvalho, and Joana Vieira. 2016.
449 "tracheideR—An R package to standardize tracheidograms." *Dendrochronologia*
450 37: 64–68. <https://doi.org/10.1016/j.dendro.2015.12.006>.

- 451 Carus, Serdar. 2004. “Impact of defoliation by the pine processionary moth
452 (*Thaumetopoea pityocampa*) on radial, height and volume growth of Calabrian
453 pine (*Pinus brutia*) trees in Turkey.” *Phytoparasitica* 32 (5): 459–69.
- 454 ———. 2009. “Effects of defoliation caused by the processionary moth on
455 growth of Crimean pines in western Turkey.” *Phytoparasitica* 37 (2): 105–14.
- 456 Cook, Edward R., and Richard L Holmes. 1996. “Guide for computer pro-
457 gram ARSTAN.” *The International Tree-Ring Data Bank Program Library Ver-*
458 *sion 2*, 75–87.
- 459 Cook, Edward R., and Paul J. Krusic. 2004. *The North American Drought*
460 *Atlas*. (Lamont- Doherty Earth Observatory; the National Science Foundation:
461 New York) Available at [https://www.ncdc.noaa.gov/data-access/paleoclimatology-](https://www.ncdc.noaa.gov/data-access/paleoclimatology-data/datasets/tree-ring/drought-variability)
462 [data/datasets/tree-ring/drought-variability](https://www.ncdc.noaa.gov/data-access/paleoclimatology-data/datasets/tree-ring/drought-variability) [Last accessed May 2020].
- 463 Guiterman, Christopher, Ann Lynch, and Jodi Axelson. 2020. *DfoliatR:*
464 *Detection and Analysis of Insect Defoliation Signals in Tree Rings*. R package
465 version 0.1.0. <https://doi.org/10.5281/zenodo.3698163>.
- 466 Holmes, Richard L., and Thomas W Swetnam. 1986. “Dendroecology pro-
467 gram library: program OUTBREAK user’s manual.” *Laboratory of Tree-Ring*
468 *Research, University of Arizona, Tucson*.
- 469 Jevšenak, Jernej, and Tom Levanič. 2018. “dendroTools: R package for
470 studying linear and nonlinear responses between tree-rings and daily environ-
471 mental data.” *Dendrochronologia* 48: 32–39. [https://doi.org/10.1016/j.dendro.](https://doi.org/10.1016/j.dendro.2018.01.005)
472 2018.01.005.
- 473 Lara, W., F. Bravo, and C. A. Sierra. 2015. “MeasuRing: An R package to
474 measure tree-ring widths from scanned images.” *Dendrochronologia* 34: 43–50.
475 <https://doi.org/10.1016/j.dendro.2015.04.002>.
- 476 Lynch, Ann M. 2012. “What Tree-Ring Reconstruction Tells Us about
477 Conifer Defoliator Outbreaks.” In *Insect Outbreaks Revisited*, edited by Pedro
478 Barbosa, Deborah K. Letourneau, and Anurag A. Agrawal, 126–54. Blackwell
479 Publishing Ltd. <https://doi.org/10.1002/9781118295205.ch7>.
- 480 Malevich, Steven B., Christopher H. Guiterman, and Ellis Q. Margolis. 2018.
481 “burnr: Fire history analysis and graphics in R.” *Dendrochronologia* 49: 9–15.

482 <https://doi.org/10.1016/j.dendro.2018.02.005>.

483 Rathgeber, Cyrille B. K., Fleur Longuetaud, Frédéric Mothe, Henri Cuny,
484 and Gilles Le Moguédec. 2011. “Phenology of wood formation: Data processing,
485 analysis and visualisation using R (package CAVIAR).” *Dendrochronologia* 29
486 (3): 139–49. <https://doi.org/10.1016/j.dendro.2011.01.004>.

487 R Core Team. 2019. *R: A Language and Environment for Statistical Com-*
488 *puting*. Vienna, Austria: R Foundation for Statistical Computing. [https:](https://www.R-project.org/)
489 [//www.R-project.org/](https://www.R-project.org/).

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

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

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

501 Shi, Jingning, Wei Xiang, Qijing Liu, and Sher Shah. 2019. “MtreeRing:
502 An R Package with Graphical User Interface for Automatic Measurement of
503 Tree Ring Widths Using Image Processing Techniques.” *Dendrochronologia* 58:
504 125644. <https://doi.org/10.1016/j.dendro.2019.125644>.

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

507 Swetnam, Thomas W., and Ann M. Lynch. 1989. “A tree-ring reconstruc-
508 tion of western spruce budworm history in the southern Rocky Mountains.”
509 *Forest Science* 35 (4): 962–86. <https://doi.org/10.1093/forestscience/35.4.962>.

510 ———. 1993. “Multicentury, Regional-Scale Patterns of Western Spruce
511 Budworm Outbreaks.” *Ecological Monographs* 63 (4): 399–424. [https://doi.](https://doi.org/10.2307/2937153)
512 [org/10.2307/2937153](https://doi.org/10.2307/2937153).

513 Swetnam, Thomas W., Marna A. Thompson, and Elaine K. Sutherland.
 514 1985. *Using dendrochronology to measure radial growth of defoliated trees*. Vol.
 515 639. United States Department of Agriculture handbook No. 639. 38 p.,
 516 Washington D.C. www.fs.usda.gov/treesearch/pubs/40898.
 517 Swetnam, TW, CD Allen, and JL Betancourt. 1999. "Applied historical
 518 ecology: using the past to manage for the future." *Ecological Applications* 9 (4):
 519 1189–1206. [https://doi.org/10.1890/1051-0761\(1999\)009%5B1189:AHEUTP%](https://doi.org/10.1890/1051-0761(1999)009%5B1189:AHEUTP%5D2.0.CO;2)
 520 [5D2.0.CO;2](https://doi.org/10.1890/1051-0761(1999)009%5B1189:AHEUTP%5D2.0.CO;2).
 521 van der Maaten-Theunissen, Marieke, Ernst van der Maaten, and Olivier
 522 Bouriaud. 2015. "PointRes: An R package to analyze pointer years and com-
 523 ponents of resilience." *Dendrochronologia* 35: 34–38. [https://doi.org/10.1016/](https://doi.org/10.1016/j.dendro.2015.05.006)
 524 [j.dendro.2015.05.006](https://doi.org/10.1016/j.dendro.2015.05.006).
 525 Wickham, Hadley. 2016. *ggplot2: Elegant Graphics for Data Analysis*. Sec-
 526 ond edition. Springer-Verlag, New York. <https://ggplot2-book.org/>.
 527 Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy
 528 McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the
 529 Tidyverse." *Journal of Open Source Software* 4 (43): 1686. [https://doi.org/10.](https://doi.org/10.21105/joss.01686)
 530 [21105/joss.01686](https://doi.org/10.21105/joss.01686).
 531 Wickham, Hadley, Romain François, Lionel Henry, and Kirill Müller. 2020.
 532 *dplyr: A Grammar of Data Manipulation*. R package version 0.8.4. [https:](https://CRAN.R-project.org/package=dplyr)
 533 [//CRAN.R-project.org/package=dplyr](https://CRAN.R-project.org/package=dplyr).
 534 Zang, Christian, and Franco Biondi. 2015. "Treeclim: An R package for the
 535 numerical calibration of proxy-climate relationships." *Ecography* 38 (4): 431–36.
 536 <https://doi.org/10.1111/ecog.01335>.

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

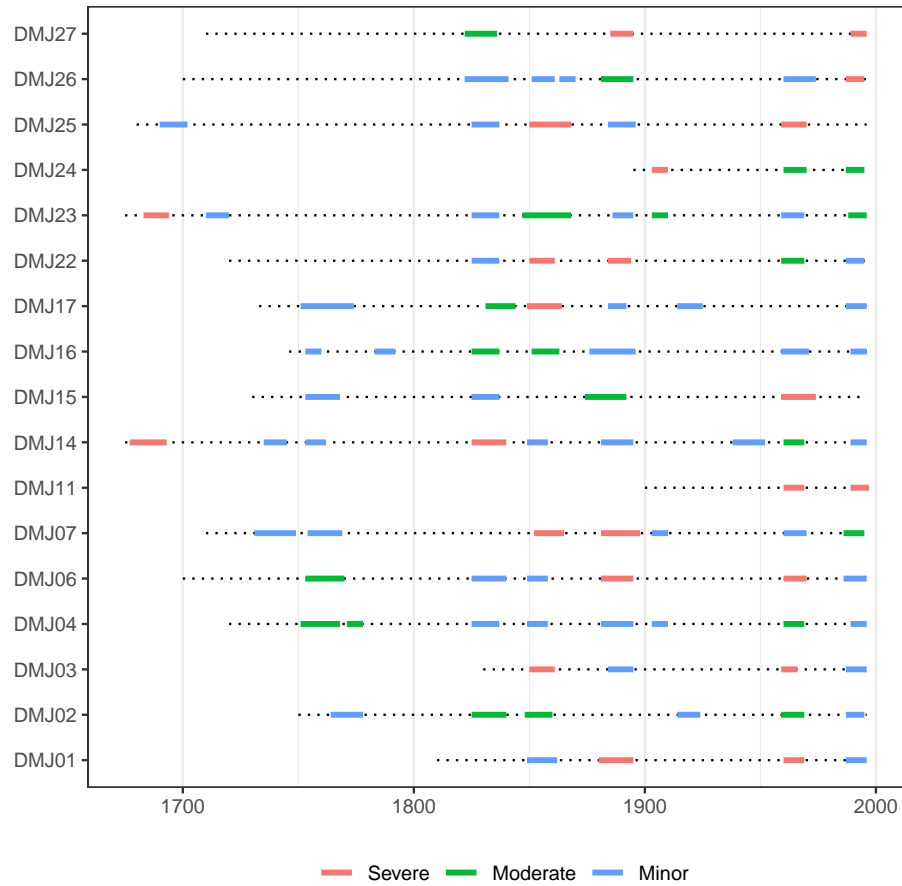


Figure 1: Default dfoliatR graphics for individual trees, produced by the ‘plot_defol()’ function. Dotted lines represent the series length for each tree and colored segments show periods of defoliation. Users can define the Cut-off values to determine severe-moderate-minor defoliation intensities, and use standard ggplot2 graphics parameters to adjust styles and themes.

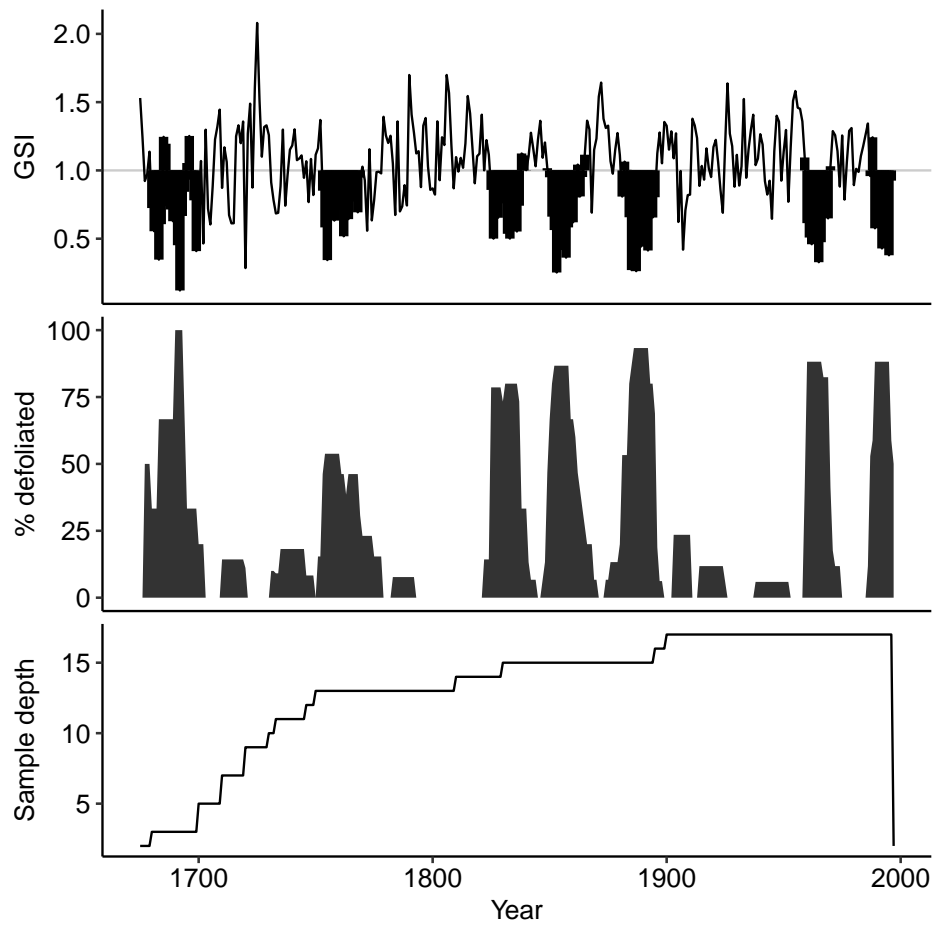


Figure 2: Default graphic to show outbreak events at the DMJ site. Top panel shows the mean GSI through time, with inferred outbreak events filled. The middle panel shows the percent of trees defoliated, and the bottom panel shows the number of trees through time.

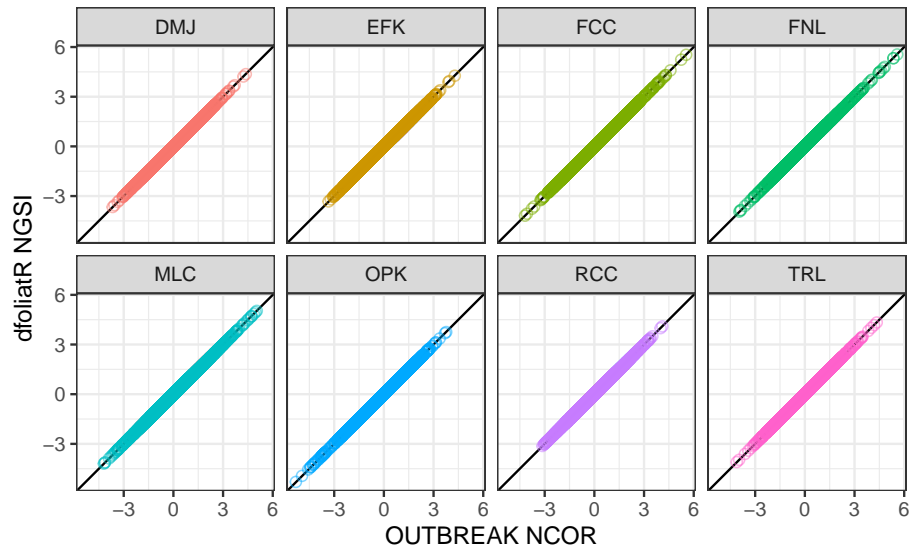


Figure 3: Comparison of disturbance indices between dfoliatR and OUTBREAK. Diagonal lines show equal values (1:1).

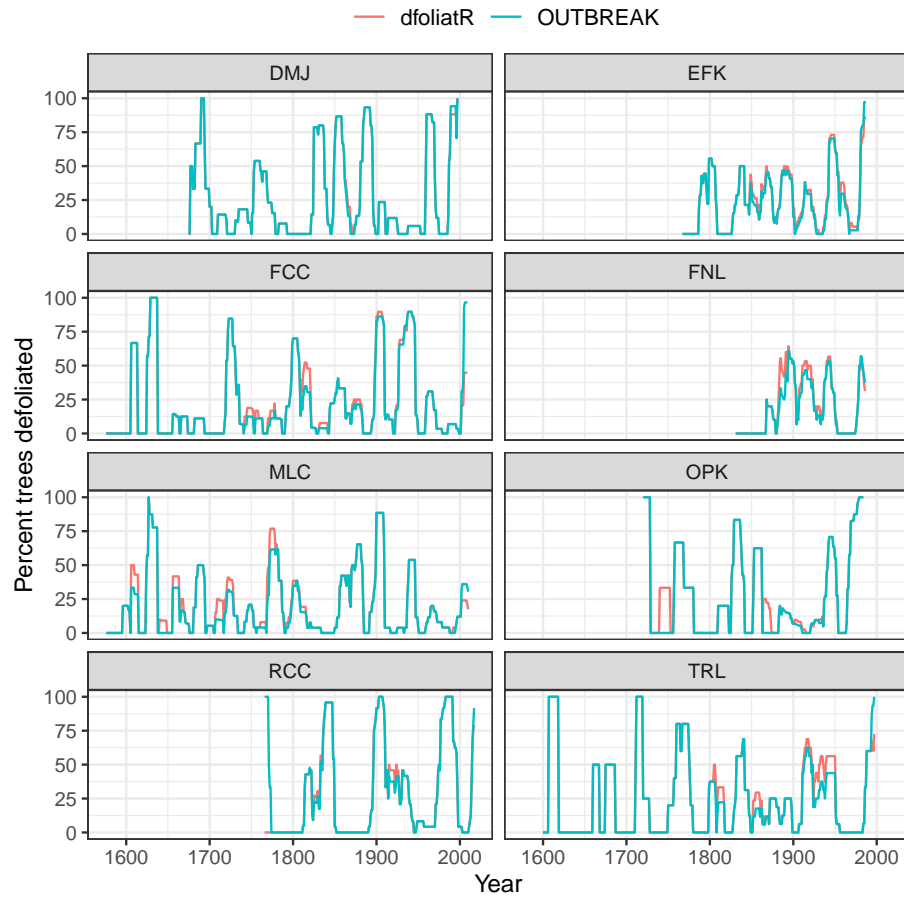


Figure 4: Comparison of reconstructed western spruce budworm outbreaks computed by dfo- liatR and OUTBREAK. Input parameters were identical between programs.