

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

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

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

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

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

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

Abstract

We present a new R package to provide dendroecologists with tools to infer, quantify, analyze, and visualize growth suppression events in tree rings caused by insect defoliation. The ‘dfoliatR’ library is based on the FORTRAN V program OUTBREAK, and builds on existing resources in the R computing environment. ‘dfoliatR’ expands on OUTBREAK to provide greater control of suppression thresholds, additional output tables, and high-quality graphics. To use ‘dfoliatR’ requires standardized ring-width measurements from insect host trees and an indexed tree-ring chronology from local non-host trees. It performs an indexing procedure to remove the climatic signal represented in the non-host chronology from the host-tree series. It then infers defoliation events in individual trees. Site-level analyses identify outbreak events that synchronously affect a user-defined number or proportion of the host trees. Functions are available for summary statistics and graphics of tree- and site-level series.

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

*Corresponding Author

Email addresses: chguiterman@email.arizona.edu (Christopher H. Guiterman), lyncha@email.arizona.edu (Ann M. Lynch), jodi.axelson@berkeley.edu (Jodi N. Axelson)

27 **Introduction**

28 Variation in the width and morphology of annual radial growth rings in wood
29 permits dating and quantification of past forest insect defoliator outbreaks. De-
30 foliation can be distinguished from climate- and other disturbance-related influ-
31 ences by comparing ring-width or other annually-resolved features in the wood
32 of host species to non-host species or to climate records. The effect of defoliation
33 on radial growth of trees has been recognized since the mid-1800s, but work prior
34 to the 1980s was not cross-dated (based on annually-resolved and –dated mea-
35 surements), not standardized (quantitatively controlled for tree geometry and
36 age effects), and suffered from inadequate extraction of the defoliation-related
37 variability from variability associated with climate and other factors (Swet-
38 nam et al. 1985, Speer 2010, Lynch 2012). These and other problems may be
39 circumvented by appropriate use of dendrochronological techniques. Quantita-
40 tive methodology to infer forest defoliator outbreaks from cross-dated tree-ring
41 records was developed in the 1980s by Swetnam et al. (1985) for developing
42 western spruce budworm chronologies (Swetnam and Lynch 1989, 1993). The
43 methodology has since been successfully applied to a wide range of defoliator
44 species, most of which are conifer herbivores, and has evolved in sophistication
45 and application to a wide range of ecosystem situations (Lynch 2012).

46 The main dendrochronological tool for inferring, dating, and quantifying
47 defoliator outbreaks from tree-ring records has been the software routine OUT-
48 BREAK (Swetnam et al. 1985, Holmes and Swetnam 1986, Swetnam and Lynch
49 1989). OUTBREAK computes indices (described later in detail) of suppressed
50 growth by subtracting a detrended and standardized climate series (a “control”
51 chronology, usually a site chronology developed from non-host trees or a gridded
52 climate data point series) from host individual-tree detrended and standardized
53 radial growth series after the host and non-host series have been brought to a
54 common variance. If the host and non-host species respond similarly to climate
55 (which can be tested), the derived series retains variability that the host and
56 non-host series do not have in common, generally the insect signal and noise

57 (unexplained variability). The user defines a rule base specifying the magnitude
58 and duration that a period of indexed growth suppression must meet or surpass
59 for a period of suppressed growth to be inferred as a defoliation event at the
60 tree level.

61 Though powerful, OUTBREAK is outdated and increasingly difficult to use
62 in modern computing environments. It was written in Fortran V with inherently
63 severe restrictions, as RAM and disk space were limited at that time (256 kb and
64 10 MB, respectively) and Fortran conventions imposed very strict formatting,
65 file naming, and output conventions. Windows-based execution operates in a
66 DOS window, Apple-based xxx in xxx, and provides no graphical interface or
67 capabilities. Barriers to batch operation impose burdens for analyses of larger
68 data sets. We developed `dfoliatR` as an R- and `dp1R`-based routine to overcome
69 these issues.

70 `dfoliatR` adds to a growing suite of dendrochronology packages the R com-
71 puting environment (R Core Team 2019). Stemming from the `dp1R` library
72 (Bunn 2008) that enables R users to read and write an array of tree-ring data
73 formats, standardize ring width series, build and evaluate chronologies, and
74 perform quality control (to name a few), one can now also measure ring widths
75 from scanned images of prepared samples (Lara et al. 2015, Shi and Xiang
76 2019), perform and check crossdating (Bunn 2010), and perform many analyt-
77 ical tests (Zang and Biondi 2015, Jevšenak and Levanič 2018). Tools for as-
78 sessing stand dynamics and disturbance analyses are under rapid development,
79 with new packages for assessing growth and release events (`TRADER`: Altman
80 et al. 2014), metrics of growth resilience (`pointRes`: Maaten-Theunissen et
81 al. 2015), and fire history (`burnr`: Malevich et al. 2018). The key objective
82 of `dfoliatR` is to provide tools to identify and analyze insect defoliation and
83 outbreak events by building on the methods employed by OUTBREAK. It capi-
84 tilizes on the robust software already available in R by using `dp1R` data formats
85 for incoming tree-ring series and providing output data formats embodied by
86 the `tidyverse` (Wickham et al. 2019) that include efficient data manipulation
87 (`dplyr`: Wickham et al. 2020) and graphics (`ggplot2`: Wickham 2016).

88 In this paper, we describe the statistical methods employed by `dfoliatR`,
89 its availability, compare results to those produced by OUTBREAK, and present
90 an example analyses. Users need not have much experience in R to replicate the
91 analyses and graphics as presented. The R code presented below is executable
92 in an R session once the required libraries are installed and loaded. Support
93 documentation in addition to this paper is provided within the package via
94 standard help menus (accessed by typing `?` before a function name) and on
95 the package website (<https://chguiterman.github.io/dfoliatR/>), which includes
96 up-to-date vignettes that describe in detail the functionality of the software.
97 Code to create a preprint of this manuscript including the R scripts is available
98 from https://github.com/chguiterman/dfoliatR_paper.

99 Overview of the software

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

- 102 • Standardized ring-width series for individual trees of the host species
- 103 • Standardized tree-ring chronology from a local non-host species

104 Users can develop these data sets in software of their choosing, such as
105 `dplR` (Bunn 2008) or ARSTAN (Cook and Holmes 1996). It is important that
106 the host-tree data include only one tree-ring series per tree. Both `dplR` and
107 ARSTAN have options for averaging multiple sample series into a tree-level
108 series.

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

112 *Identifying Defoliation of Trees*

The `defoliate_trees()` function is usually be the point of entry to the
dfoliatR library. It performs two processes, removing climate-related growth

signals from the host-tree series and then 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 that 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 (following Swetnam et al. 1985, Swetnam and Lynch 1989) for each host tree as

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

where H and NH are the host-tree series and the non-host chronology, in year i , respectively. Only the common period between the host-tree series and the non-host chronology are used in Equation 1. The host and non-host chronologies are brought to common variance by scaling the non-host chronology by its mean ($\overline{\text{NH}} \approx 1.0$) and multiplying by the ratio of host and non-host standard deviations ($\frac{\sigma_H}{\sigma_{\text{NH}}}$), which approximates the variance of the host tree series.

Negative departures in the normalized GSI that surpass user-defined thresholds in duration and magnitude are defined as *defoliation events*. As in OUTBREAK, magnitude is assessed on a single year within the departure sequence. The default setting is -1.28 (in units of standard deviation), which was previously determined to be representative of WSBW effects (**citation?**). Duration is assessed by examining sequences of negative GSI before and after the year of maximum departure. Each defoliation event is allowed one positive excursion on each side of the maximum departure year. Duration is assessed across the entire sequence that includes up to two positive excursions. The default duration is eight years, as is commonly used in WSBW studies (**citation?**). Different species of defoliation insects vary in the length of defoliation and the degree to which they can suppress tree growth. Researchers can, and should, adjust the duration and magnitude parameters accordingly and critically evaluate the results.

Diverging from OUTBREAK, `dfoliatR` allows users to extend defoliation

134 events by bridging successive events and also by allowing potentially short-
 135 duration events that occur at the end of the series. In cases where two defoliation
 136 events are separated by a single year, bridging will link them into a single event
 137 (**Show figure?**). We urge careful use of this option because there is no setting
 138 to limit the number or length of potentially bridged events. The series end
 139 option can be used in cases when the host trees were actively being defoliated
 140 at the time of sampling. This option eliminates the duration parameter for
 141 an event at the recent end of the series, but all other thresholds apply. The
 142 advantage of this parameter is that it can aid in identifying the start-year for
 143 the current defoliation event or outbreak, which is both useful in management
 144 and allows the current event to be included in return-interval estimates.

145 *Inferring Outbreak Events*

146 Defoliation of one or a few trees does not constitute an outbreak. To deter-
 147 mine when defoliation becomes an *outbreak event*, **dfoliatR** composites the in-
 148 dividual tree defoliation series into a site-level chronology with the **outbreak()**
 149 function. Users have options to define the number and/or the proportion of
 150 trees required for an event to be considered an outbreak. Three parameters
 151 control the whether a defoliation event constitutes an outbreak: the minimum
 152 number of trees available, the minimum number of trees recording defoliation,
 153 and the percent of trees recording defoliation. The first allows the researcher to
 154 make a judgement call as to the confidence ascribed to reduced sample depth
 155 toward the ends of their chronologies, thus compensating for the “fading record
 156 problem” (**Swetnam and Fritts?**). The second two parameters adjust the
 157 scale of defoliation considered to be an outbreak. Absolute numbers of trees
 158 and percentages can be applied separately or in conjunction, following filtering
 159 conventions in tree ring fire history studies (Malevich et al. 2018). We urge
 160 users to carefully consider the choice of absolute numbers in situations where
 161 the number of trees represented in the series varies with time, or the choice of
 162 percentages when sample size is small.

163 **Evaluation**

164 *Approach*

165 We tested `dfoliatR` against OUTBREAK by comparing GSI to corrected
166 indices for individual trees and years, outbreak status for individual trees and
167 years, and percentage of trees recording outbreaks at the site level, using raw
168 ring-width data from 8 sites in British Columbia, Colorado, and New Mexico
169 and author-provided non-host site chronologies.

170 We detrended host data for both `dfoliatR` and OUTBREAK using ARSTAN
171 6.1 (**downloaded 21 April 2002 from dpl; xxx Ann check the date, as**
172 **21 April 2002 is 6.05P**) with default double detrending (128 year wavelength
173 and a 50% smoothing spline) and event thresholds of -1.28 normalized index
174 and 8 years. Attempts to detrend with ARSTAN 44 in `dplR` were problematic
175 for two reasons. Some versions of ARSTAN, including ARSTAN 44, do not
176 include an option for producing tree-level averages from multiple cores from the
177 same tree (.tre files). More importantly, ring-width indices from ARSTAN 44
178 truncated the process at either end when sample size fell below 5 xxx (Jodi,
179 what was the minimum?), and produced different values for years near the
180 truncation point than did dpl-ARSTAN 6.05P. The majority of the detrended
181 series values were identical, but not the tails, as the two routines do not use
182 the same data to establish the tails. These differences produce different GSI
183 (`dfoliatR`) and Corrected Indices (OUTBREAK) values at the tails, sometimes
184 considerably, affecting whether or not outbreak events are inferred early or
185 late in the tree series, timing of the first or last event onset, and subsequent
186 computation of return intervals, duration, and periodicity. Note that there
187 is nothing particularly important regarding the choice of ARSTAN 6.1 – it is
188 simply the version that the second author used routinely and had available.

189 *Findings*

190 `dfoliatR` and OUTBREAK compute growth suppresses indices and use
191 them to infer identical tree-level outbreak events, including onset and termina-
192 tion dates and date and magnitude of the maximum growth suppression when

193 `dfoliatR` bridging matches the OUTBREAK protocols. Once the issues with
194 ARSTAN (described earlier) were resolved, `dfoliatR` and OUTBREAK produced
195 identical indices at 0.000 precision.

196 When inferring outbreak events at the tree level, our initial results did not
197 always match OUTBREAK, leading to development of the bridging option.
198 Identifying a minimum within a period while allowing for positive excursions is
199 a tricky computation. OUTBREAK deliberately disallows back-to-back events
200 – after a minimum value within a period of negative values, a second positive
201 index terminates an event. OUTBREAK does not allow a period of highly
202 negative values to be appended to an earlier outbreak if the intervening period is
203 1 year, even if those negative values fall below the minimum in the earlier period.
204 Due in large part to reconstructions using OUTBREAK (see papers cited by
205 Lynch 2012), we now know considerably more about forest defoliator outbreak
206 regimes than we did in the 1980s when OUTBREAK was written. We think that
207 two prolonged events separated by a single year should in some situations be
208 considered a single outbreak at both the tree and site levels. This is particularly
209 relevant to western spruce budworm and spruce budworm (*C. fumiferana*), for
210 which regimes have been reconstructed for several geographic areas and for
211 which researchers and forest health experts have gained considerable knowledge
212 (**Sanders et al. 1985, Brookes et al. 1987 and many later publications**).
213 We now know that the greatest growth suppression often occurs late in the
214 outbreak due to cumulative effects (**rrr**). We developed the bridging option to
215 permit these periods to be linked to preceding events separated by a single year.

216 Bridging operates at the tree level in `dfoliatR`, and these effects generally
217 have minor effects on site-level reconstructions (Fig.Bridging). Occasionally
218 site-level events are inferred differently (Table.Bridging), and this can affect
219 return interval and duration statistics.

220 *Best Practice Recommendations*

221 We encourage authors to report software and library version numbers and
222 download dates where applicable. This is standard practice for statistical pack-

223 ages, but has often been ignored for R- and dplR-based analyses. R-based soft-
224 ware evolves over time, and results cannot always be replicated if the software
225 versions are unknown. Similarly, significant changes (especially fixes) imply
226 that re-analysis of data might differ from earlier published results and inter-
227 pretations. Wherever possible, refereed publications supporting the software
228 should be cited. Our experience with ARSTAN demonstrates the importance of
229 adopting this practice. In the United States, analytic methods not documented
230 thusly and supported by the literature might be questioned if the results are
231 used in a National Environmental Protection Act (NEPA) process.

232 We recommend that researchers compare dfoliatR results with and without
233 bridging, and see which option agrees best with known insect biology and ecol-
234 ogy. Bridging may be a logical choice for insects with prolonged non-outbreak
235 periods, such as western spruce budworm, but not for situations where impacted
236 stands barely recover from one outbreak before another begins, such as occurs
237 with pine processionary caterpillars (*Thaumetopoea pityocampa* (Lepidoptera:
238 Thaumetopoeidae)) (Carus 2004, 2009).

239 Availability and installation

240 The dfoliatR library (Guiterman et al. 2020) is provided free and open
241 source from the Comprehensive R Archive Network (CRAN; [https://cran.r-](https://cran.r-project.org/)
242 [project.org/](https://cran.r-project.org/)). To install dfoliatR from CRAN use

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

243 In each R session, dfoliatR can be loaded via

```
library(dfoliatR)
```

244 Development versions of dfoliatR are available on GitHub and installed
245 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.

Example Usage

In `dfoliatR` we provide two sets of tree-ring data to aid users in exploring the functions, graphics, and outputs. Each set consists of Douglas-fir (*Pseudotsuga menziesii*) host-tree series, standardized with _____, and a local ponderosa pine (*Pinus ponderosa*) non-host chronology. The non-host ring-width data were standardized by _____ and the chronologies averaged following standard procedures (Speer 2010). Data from Demijohn Peak (DMJ; 2902 m asl), in the San Juan Mountains of southern Colorado, come from Ryerson et al. (2003). Data from the East Fork site (EF; 2580 m asl) in the Jemez Mountains of northcentral New Mexico were presented by Swetnam and Lynch (1993).

Tree-Level Defoliation Events

Site-Level Events

Extensions

- Describe how `dfoliatR` can be combined with other R libraries
 - Mapping, what else?

References

Altman, J., P. Fibich, J. Dolezal, and T. Aakala. 2014. TRADER: A package for Tree Ring Analysis of Disturbance Events in R. *Dendrochronologia* 32:107–112.

271 Bunn, A. G. 2008. A dendrochronology program library in R (dplR). *Dendrochronologia* 26:115–124.

272

273 Bunn, A. G. 2010. Statistical and visual crossdating in R using the dplR library. *Dendrochronologia* 28:251–258.

274

275 Cook, E. R., and R. L. Holmes. 1996. Guide for computer program arstan. Pages 75–87 *The international tree-ring data bank program library version*.

276

277 Guiterman, C., A. Lynch, and J. Axelson. 2020. DfoliatR: Detection and analysis of insect defoliation signals in tree rings.

278

279 Holmes, R. L., and T. W. Swetnam. 1986. Dendroecology program library: Program outbreak user’s manual. Laboratory of Tree-Ring Research, University of Arizona, Tucson.

280

281

282 Jevšenak, J., and T. Levanič. 2018. dendroTools: R package for studying linear and nonlinear responses between tree-rings and daily environmental data. *Dendrochronologia* 48:32–39.

283

284

285 Lara, W., F. Bravo, and C. A. Sierra. 2015. MeasuRing: An R package to measure tree-ring widths from scanned images. *Dendrochronologia* 34:43–50.

286

287 Lynch, A. M. 2012. What Tree-Ring Reconstruction Tells Us about Conifer Defoliator Outbreaks. Pages 126–154 *in* P. Barbosa, D. K. Letourneau, and A. A. Agrawal, editors. *Insect outbreaks revisited*. Blackwell Publishing Ltd.

288

289

290 Maaten-Theunissen, M. van der, E. van der Maaten, and O. Bouriaud. 2015. PointRes: An R package to analyze pointer years and components of resilience. *Dendrochronologia* 35:34–38.

291

292

293 Malevich, S. B., C. H. Guiterman, and E. Q. Margolis. 2018. burnr: Fire history analysis and graphics in R. *Dendrochronologia* 49:9–15.

294

295 R Core Team. 2019. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

296

297 Ryerson, D. E., T. W. Swetnam, and A. M. Lynch. 2003. A tree-ring reconstruction of western spruce budworm outbreaks in the San Juan Mountains, Colorado, U.S.A. *Canadian Journal of Forest Research* 33:1010–1028.

298

299

300 Shi, J., and W. Xiang. 2019. MtreeRing: A shiny application for automatic measurements of tree-ring widths on digital images.

301

302 Speer, J. H. 2010. Fundamentals of tree-ring research. Page 333. The
303 University of Arizona Press.

304 Swetnam, T. W., and A. M. Lynch. 1989. A tree-ring reconstruction of
305 western spruce budworm history in the southern Rocky Mountains. *Forest*
306 *Science* 35:962–986.

307 Swetnam, T. W., and A. M. Lynch. 1993. Multicentury, Regional-Scale Pat-
308 terns of Western Spruce Budworm Outbreaks. *Ecological Monographs* 63:399–
309 424.

310 Swetnam, T. W., M. A. Thompson, and E. K. Sutherland. 1985. Using
311 dendrochronology to measure radial growth of defoliated trees. Page 39p.

312 Wickham, H. 2016. *Ggplot2: Elegant graphics for data analysis*. Springer.

313 Wickham, H., M. Averick, J. Bryan, W. Chang, L. McGowan, R. François,
314 G. Grolemund, A. Hayes, L. Henry, J. Hester, M. Kuhn, T. Pedersen, E. Miller,
315 S. Bache, K. Müller, J. Ooms, D. Robinson, D. Seidel, V. Spinu, K. Takahashi,
316 D. Vaughan, C. Wilke, K. Woo, and H. Yutani. 2019. Welcome to the tidyverse.
317 *Journal of Open Source Software* 4:1686.

318 Wickham, H., R. François, L. Henry, and K. Müller. 2020. *Dplyr: A gram-*
319 *mar of data manipulation*.

320 Zang, C., and F. Biondi. 2015. Treeclim: An R package for the numerical
321 calibration of proxy-climate relationships. *Ecography* 38:431–436.