The Wood Image Analysis and Dataset (WIAD): open-access visual analysis tools to advance the ecological data revolution

Running headline

Wood Image Analysis and Dataset

Author details

Tim Rademacher¹,²,³,⁴ Bijan Seyednasrollah²,³, David J. Basler¹,⁴, Jian Cheng⁵, Tessa Mandra¹, Elise Miller¹,⁶, Zuid Lin⁷, David A. Orwig¹, Neil Pederson¹, Hanspeter Pfister⁷, Andrew D. Richardson²,³, Donglai Wei⁷, Li Yao⁵

¹ Harvard Forest, Harvard University, Petersham, Massachusetts, USA.
² School of Informatics, Computing and Cyber Systems, Northern Arizona University, Flagstaff, Arizona, USA.
³ Center for Ecosystem Science and Society, Northern Arizona University, Flagstaff, Arizona, USA.
⁴ Department of Environmental Sciences, University of Basel, Basel, Switzerland.
⁵ Southeast University, Nanjing, Jiangsu, China.
⁶ University of Minnesota Duluth, Duluth, Minnesota, USA.
⁷ School of Engineering and Applied Sciences, Harvard University, Cambridge, Massachusetts, USA.

* Corresponding author, email: trademacher@fas.harvard.edu, 324 North Main Street, Petersham, 01366, Massachusetts, USA.
Abstract

1. Ecological data is collected and shared at an increasingly rapid pace, but it is often shared in inconsistent and untraceable processed forms. Images of wood contain a wealth of information such as colours and textures but are most commonly reduced to ring-width measurements before they can be shared in various common file formats. In fact, no digital libraries of wood samples that have been developed for ecological analysis currently exist.

2. We developed the Wood Image Analysis and Dataset (WIAD) an open-source application including a web-interface to integrate visual analysis of wood samples, such as increment cores, thin sections or x-ray films, basic data processing, and archiving of the images and derived data to facilitate transparency and reproducibility in studies using visual characteristics of wood.

3. WIAD provides user-friendly tools to manipulate images of wood samples, mark and measure wood characteristics such as growth increments, density fluctuations, early- and latewood widths, and fire scars, and to visualise, process, and archive images, metadata, and the derived data.

4. WIAD constitutes a step towards the reproducible automation of tree-ring analysis, while establishing a foundation to create improved community developed repositories and enable novel ecological studies harnessing the wealth of existing visual data.
1. Introduction

Ecological data is collected at an increasingly rapid pace, but to optimise its use data needs to be shared in its raw format with accompanying ancillary and metadata (Michener, 2015). While the community of science using tree rings has been leading data sharing efforts (i.e., International Tree Ring Database and DendroEcological Network), there is still an urgent need for improved tree-ring archives (Babst, Poulter, Bodesheim, Mahecha, & Frank, 2017). In fact, there are no existing digital libraries of images of wood samples that were developed to be used for ecological analysis. Currently, wood samples, such as increment cores, cross sections and microcores, are normally prepared according to standard procedures, extracting visual properties such as ring-width series (Fritts, 1976), blue intensity as a proxy for wood density (J. A. Björklund, Gunnarson, Seftigen, Esper, & Linderholm, 2014), or cellular dimensions (J. Björklund et al., 2017; von Arx & Carrer, 2014). The rich information embedded in the colours, textures and structures of each wood sample, which are used to identify these visual features in the first place, are generally collapsed to a simple one-dimensional geometric measurement (e.g., ring width) before they are shared in one of many common file formats (Brewer, Murphy, & Jansma, 2011).

While multiple specialised tools for digital analysis of wood images exist (i.e., WinDendro, ROXAS, CooRecorder), many require proprietary dependencies excluding users and importantly most are not open source, making it difficult for further development. Further, none of the available tools provide a pipeline that allows users to process, measure, and analyse images as well as archive the raw, derived, and meta-data in one place. Such a pipeline would improve transparency and reproducibility of studies based on wood images by enabling easy
sharing of data and provenance tracing. Shared digital archives of images and derived data would also allow the accumulation of larger collections for more comprehensive studies. By pooling data in such a repository, visual computing algorithms would also have sufficient labels to enable robust automated data processing, such as detection of species, wood morphology, ring boundaries and crossdating. In fact, due to limitations of existing software, many scientists still prefer manual measurements using “linear tables”, although the accuracy of digital analysis was already comparable to manual measurements about a decade ago for some species (Maxwell, Wixom, & Hessl, 2011). Digital technology has advanced rapidly since and offers generally recognised benefits of simplified evaluation, corrections and amendment of previous measurements.

To unlock the wealth of underlying visual data from images of wood, we created an open source pipeline, the Wood Image Analysis and Database (WIAD), that integrates digital analysis and data archiving, while also laying the foundation for open-access sharing of images and derived data. WIAD is an easily-installed, simple R shiny application, whose graphical user interface can currently be accessed at http://wiad.science or run locally. WIAD provides an interactive user-friendly interface for pre- and post-processing of wood images and metadata entry to facilitate a standardized workflow for labeling, measuring and detrending of wood growth. Both the web-interface and locally running copies allow the evaluation and manipulation of both previous measurements (i.e., a review) and new measurements, and associate the data with a minimal set of metadata (Table 1). When running WIAD locally, outputs are saved on the local machine enabling the upload and sharing of local collections at any point thereafter (i.e., upon acceptance for publication). Thus, the user can embargo the data by running a copy.
locally and only uploading the data upon publication as per terms of use. When using the web interface, data is archived straight away to the online repository.

In the following sections, we explain WIAD’s current functionality, provide an illustrative example of how it can be used, describe anticipated developments and the future potential of such a community-developed platform. A community-developed platform to archive, share, and analyse images in a transparent and reproducible way is applicable for a vast array of visual data in ecology such as leaf or root scans. For example, open access, data sharing platforms with analysis capabilities like WIAD already exist for digital images of the canopy in the Phenocam network (Hufkens, Basler, Milliman, Melaas, & Richardson, 2018; Richardson et al., 2018). WIAD aims to provide a blueprint of an integrated repository and analysis framework to advance the ecological data revolution using agile software development and state-of-the-art practices in scientific software development (Ahmed, Zeeshan, & Dandekar, 2014; List, Ebert, &
After the initial installation as an R package, WIAD can be run locally by one simple command. Alternatively, WIAD can be accessed through a web-interface at http://wiad.science. The application currently has four tabs: Toolbox, Plot Board, About and Fair use policy.
2.1 Toolbox

The toolbox provides basic interactive image manipulation tools (Fig. 1), which require the upload of an image in TIFF, PNG or JPEG format. Once uploaded, the image can be rotated, processed with colour transformations (i.e., display true colour, total brightness, or blue channel only), and enlarged using a zoom bar (digital magnification up to a maximum image width of 20000 pixels). The user needs to enter - either manually or via the upload of an XLSX, JSON or CSV file - and confirm metadata, which is then associated with the image via a radio button. To convert the distances in the image from pixels to micrometers, the user needs to provide the true scan resolution (in dots per inch).

Once the metadata is confirmed, the toolbox allows setting, deleting, and inserting five types of labels through a simple graphical user interface. A simple click on the image sets a normal label marking the end of a growth increment (overlaid in yellow), a double click sets a miscellaneous label (overlaid in turquoise), which marks miscellaneous features that can be selected from a pop-up dialog. The user can delete any specific, the last set, or all labels with the “Bin” of the specific data table row, or the “Undo”, and “Erase” button below the image, respectively. The user can switch a normal label to a link label (overlaid in blue) with the “Link” button below the image. Links can bridge gaps in the sample (e.g., two linker labels in a row) or help to relocate the start point for a measurement. Another action button switches a normal label to mark the oldest completely visible ring or poth (overlaid in red). Radial growth (e.g., one-dimensional distance between labels in micrometers) and association of a particular year with each annual growth increment (including a one-year shift when desired in the southern hemisphere sensu Schulman (1956)) are performed automatically and displayed in a data table.
below the image. Checkboxes below the image represent a choice whether the labelling started at the bark side of a core, whether the image contains the pith, and whether years, labels and/or label numbers should be displayed. Additional labels, including missing rings, can be inserted before any existing label by clicking on the "Add" button of the respective row of the data table. Users can also upload images, metadata and labels to evaluate previously analysed images or download a CSV or JSON file, which contains labels, derived measurements and metadata (the later is not included in CSV format) to save their progress or for further analysis offline.

To use the tool for teaching or demonstration, the “Demo” button switches demonstration mode on/off. In demonstration mode, an example image is automatically loaded, but can be replaced, and no data or images are archived.
Figure 1 - Screenshot of WIAD Toolbox with an example image of white pine and uploaded labels from a previous measurement. The metadata automatically populates the fields in the left column, but needs to be confirmed with the radio button below. Normal labels are overlaid onto the image in yellow, miscellaneous labels in turquoise and link labels in blue.

2.2 Plot board

The WIAD plot board tab provides two interactive plots to quickly assess the derived annual growth data and detrended ring-width indices. The upper plot displays the measured annual radial growth increments over time, while the second graph shows ring-width indices based on residuals from an expected growth curve (Cook & Peters, 2016). The user can select a detrending curve from the *dplR* R package (Bunn, 2008) to obtain ring-width indices via controls on the left and examine them in real-time. Finally, a JSON file including the labels, derived measurements, detrended ring-width indices, and metadata can be downloaded from this page.
Figure 2 - Screenshot of the WIAD Plot board showing the controls for detrending on the left and interactive plots of the radial growth increments (blue line and markers; top graph) including a detrending curve (red curve; top graph) and ring-width indices (mauve line and markers; bottom graph).

2.3 About and Fair use policy

The “About” tab provides basic information about WIAD, whereas the “Fair use policy” tab outlines the conditions of use with regard to the tools and data. Images and derived data are publicly and freely available on WIAD for noncommercial reuse under the CC0 Public Domain Dedication (https://creativecommons.org/share-your-work/public-domain/cc0/). However, we
request that the WIAD core development team and data contributors are appropriately acknowledged in work using WIAD by citing this paper and the appropriate software citation. Contributors’ names and email addresses are attached to each datum in the repository through the accompanying metadata and we expect data users to contact data contributors. Data contributors will have a deeper understanding of the data and additional insight; thus we consider them crucial collaborators for any data-user.

3. Example: Measuring ring width and density fluctuations in white pine

To provide an example data set we produced, measured and uploaded images for ten mature white pines (*Pinus strobus* L.) from Harvard Forest, USA. The images and all derived data are available publicly and can be downloaded via the getWiadExampleData function of the WIAD R package. For this data set, we collected two increment cores at breast height using a standard increment corer (5.15mm, Haglöf Company Group, Långsele, Sweden). The cores were glued to wooden mounts and sanded with progressively finer sandpaper (from 80 to 8000 grit) until visual features of the wood were clearly visible by the naked eye.

3.2 Imaging samples

Samples were scanned with an Epson Perfection V500 Photo Scanner (Long Beach, California) at 3200 dpi with an advanced reflective colour-calibration target (10x15cm target printed on Kodak Professional Endura, LaserSoft Imaging, Kiel, Germany). The colour calibration target allows for tracking and potential future correction of drift in colours over time and between scanners. We on a flatbed scanner with a charge-coupled devices (CCD) imaging
array. The CCD imaging arrays have a larger depth of focus compared to the increasingly
popular complementary metal oxide semiconductor arrays, although the later may outperform
CCD arrays in the future (Janesick & Putnam, 2003). Despite higher resolution scans being
generally desirable, as they capture more details, which may lead to additional uses in the
future, larger images also require more digital storage. Therefore, we recommend 3200 dpi,
which balances the tradeoff between visual quality and file size for standard increment cores. In
fact, image rendering is also the computational bottleneck for WIAD, hence larger images slow
down processing. To reduce storage requirements and optimise processing speed images
should always be cropped (using open-source image manipulation programs like GNU Image
Manipulation Program) to only the region of interest prior to upload. For particularly large
samples (e.g., cross-sections), hence files, segmentation into multiple images should be
considered for faster processing and reduced storage costs. However, care should be taken to
avoid changing the resolution or other features of the image when pre-processing scans. Due to
storage limitations we limit the maximum image size to 30 MB when using the web-interface.
When running WIAD offline, the maximum image size can easily be changed via the
maxImageSize variable in the global.R file.

4. Anticipated developments and future potential

4.1 Exploiting the potential of already existing collections by digitisation

Wood is arguably the most abundant biological tissue on earth (Bar-On, Phillips, & Milo,
2018) and provides an imprint of past climates (Mann et al., 2008), disturbances (Bergeron,
Denneler, Charron, & Girardin, 2002; Lorimer, 1984) and biotic interactions (Cook, 1987).
Scientists of multiple disciplines (ecology, physiology, archeology, etc.) have collected millions
of wood samples in the form of increment cores, micro-cores (Rossi, Anfodillo, & Menardi, 2006), and stem cross-sections, which are archived in laboratories across the world. Publicly available digital images of collected samples would move these fields to a big data domain and provide the foundation for major advances in our understanding of the process of wood formation and its abiotic and biotic constraints. Improved tree ring archives have been argued to be essential for earth system science (Babst et al., 2017). More generally, improved archives will be instrumental for teaching the next generation of scientists, reconstructing past environments and will improve forecasts of future carbon sequestration that are based in ecological realities. Digitisation and data sharing have multiple proven benefits such as accelerating science and improving trust in science (Michener, 2015; and references therein). WIAD paves the way to digitise large collections of wood samples and build such improved archives that overcome limitations of current archives (Zhao et al., 2018). Additionally, WIAD could become a simple yet powerful teaching tool in demonstration mode, especially if teaching libraries are shared publicly.

While WIAD provides the foundational tools, several aspects will be refined in future releases including an interface to access and query the database, up- and download collections that were measured offline, association with ancillary data, batch processing of images and integration of data provenance. WIAD requires already a minimal set of metadata. However, to make the database compatible with other ecological data the upload and generation of Ecological Metadata Language (Fegraus, Andelman, Jones, & Schildhauer, 2005; Jones, Schildhauer, Reichman, & Bowers, 2006) is anticipated in future releases. Similarly, WIAD intends to comply with Darwin Core Standard (Wieczorek et al., 2012) to ensure compatibility with other data collections of biological diversity. Finally, the next version should allow the
upload of ancillary data such as data of non-visual chemical or physical properties (isotopic composition, nonstructural carbon concentrations, lignin content, etc.).

4.2 Streamline data processing through automation and enable citizen scientists

To further automate the processing of digital analysis of wood samples low hanging fruit would be the automatic detection of visual properties. Initial algorithms correctly identified almost 100% of rings in four conifer species, roughly 85% in six diffuse-porous species, but only 40-50% in two ring-porous species (Fabijańska, Danek, Barniak, & Piórkowski, 2017; Subah, Derminder, & Sanjeev, 2017), despite being based on comparatively small and homogenous training data. Recently, convolutional neural networks have also succeeded at segmenting anatomical sections of wood (Garcia-Pedrero et al., 2018). The integration of visual computing algorithms into WIAD to automate the analysis and reduce labour time, by suggesting ring boundaries to the user after image upload, is under development. For this purpose, we are currently creating the first fully labeled image library of sufficient size (more than 100 000 labels in 1000 images) (Rademacher et al., in prep.) to train a robust ring detection algorithm that works for ten species including ring-, diffuse- and non-porous wood. Specifically, we adopt the U-Net model (Ronneberger, Fischer, & Brox, 2015) to predict the pixel-wise probability of ring boundaries. Since models trained with the default pixel-wise cross-entropy loss usually yield boundaries with blurry artifacts, we incorporate a perceptual discriminator energy function to improve the sharpness of the boundaries while preserving the connectivity, following the high-level idea of Mosinska et al. (2018). Beyond identifying rings, the identification of other visual features such as wood morphology (e.g., ring-porous, diffuse-porous, and non-porous) density fluctuations or frost and fire scars, could equally be automated using similar techniques.
Other low hanging fruit would be the integration of cross-dating tools that help to visually and statistically crossdate multiple measurement series such as xDateR (Bunn, 2008, 2010) or ringdater (Reynolds, Edge, & Black, 2020). Similar to suggestions of ring boundaries, fast fuzzy subsequence matching (Gong, Fong, & Si, 2019) and unsupervised anomaly detection algorithms for either time-series (Ren et al., 2019) or high dimensional data like raw images (Wang et al., 2019) could be used to assist in the crossdating and identification of missing rings, which would further increase processing speed and improve reproducibility of results.

Beyond streamlining sample processing, these features could enable scientists and potentially even citizen scientists to contribute to the repository in the long-term. However, sufficient skills need to be demonstrated rigorously before expert opinion can be meaningfully complemented. Nonetheless, citizens could already help in the labelling of datasets and thereby provide training data for the refinement of algorithms. CAPTCHA-like interfaces - brief tests used widely online to differentiate humans from bots - could be used to identify ring boundaries and thus crowd-sourcing the labelling of training data and evaluation of existing collections. Citizen-science projects come with tradeoffs in data quality, privacy protection, resource security, transparency, and trust by citizens (Anhalt-Depies, Stenglein, Zuckerberg, & Townsend, 2019), which we will consider during further development.

4.3 Beyond increment cores, tree rings and wood

Although WIAD initially focussed on images of wood samples and scans of increment cores in particular, other ecological data could equally be digitised for the community. Tools for the analysis of images of fine roots could be added to date fine roots in habitats where seasonal
growth rings form (Solly et al., 2018) or to estimate fine root area (Himmelbauer, Loiskandl, & Kastanek, 2004). Another interesting and simple addition would be tools to analyse leaf scans to extract traits such as green chromatic coordinate, leaf area, herbivory damage and other visual characteristics.

In conclusion, WIAD provides powerful tools and exemplifies a new way of dealing with the complexity of visual data in eco-physiology. By offering integrated data processing and archival services, platforms such as WIAD can improve analysis by making data accessible, tracking provenance, assuring metadata standards and, therefore, creating data products that can be used, amalgamated, scaled, built upon, and modified seamlessly.

References


**Author contributions**

TR conceived the idea and developed it with BS and DB. BS developed the code with input and additions from TR and DB. TR cored the trees, mounted and sanded the cores. EM, TM, and TR scanned images and tested the tools. JC, ZL, and DW developed the algorithm for automatic boundary detection. All authors provided ideas on how to improve the tools. TR led
the writing of the manuscript. All authors contributed critically to drafts and approved the final manuscript.

Code and data availability

All data is publicly available through getWIADExampleData function built into the package. The package and underlying code are also publicly accessible under a CC0 Public Domain Dedication on github at https://github.com/bnasr/wiad and is easily installed and loaded in the R statistical software using the following two commands:

```r
if (!require (devtools)) {install.packages (devtools)}
devtools::install_github ("bnasr/wiad")
```

Acknowledgements

TR and TM acknowledge the support from Microsoft “AI for Earth” grant program (271089). AR, EM and TR acknowledge support from the National Science Foundation (DEB- 1741585, HF LTER DEB-1237491 and DEB-1832210). DB acknowledges support through the Swiss National Science Foundation (PSBSP3-168701) and the Harvard Forest Bullard Fellowship. EM was also benefited from NSF DBI-1459519 and the Harvard Forest Summer Research Program in Ecology. We also thank Patrick Baker, Greg Caporaso, Allyson Carroll, Patrick Fonti, George Koch, Rubén D. Manzanedo, Steve Sillett, Laura G. Smith, Yiping Zhang for their valuable feedback.