


APPLICATION PAPER  

Low-cost tree crown dieback estimation using deep learning-based segmentation

Matthew J. Allen¹ , Daniel Moreno-Fernández^{2,3}, Paloma Ruiz-Benito^{2,4}, Stuart W.D. Grieve^{5,6} and Emily R. Lines¹

¹Department of Geography, University of Cambridge, Cambridge, UK

²Forest Ecology and Restoration Group, Departamento de Ciencias de la Vida, Universidad de Alcalá, Madrid, Spain

³Institute of Forest Sciences (INIA-CSIC), Madrid, Spain

⁴Environmental Remote Sensing Research Group, Departamento de Geología, Universidad de Alcalá, Alcalá de Henares, Spain

⁵School of Geography, Queen Mary University of London, London, UK

⁶Digital Environment Research Institute, Queen Mary University of London, London, UK

Corresponding author: Matthew J. Allen; Email: mja78@cam.ac.uk

Received: 16 October 2023; **Revised:** 12 April 2024; **Accepted:** 18 April 2024



Keywords: climate change; deep learning; dieback; drones; forest monitoring

Abstract

The global increase in observed forest dieback, characterized by the death of tree foliage, heralds widespread decline in forest ecosystems. This degradation causes significant changes to ecosystem services and functions, including habitat provision and carbon sequestration, which can be difficult to detect using traditional monitoring techniques, highlighting the need for large-scale and high-frequency monitoring. Contemporary developments in the instruments and methods to gather and process data at large scales mean this monitoring is now possible. In particular, the advancement of low-cost drone technology and deep learning on consumer-level hardware provide new opportunities. Here, we use an approach based on deep learning and vegetation indices to assess crown dieback from RGB aerial data without the need for expensive instrumentation such as LiDAR. We use an iterative approach to match crown footprints predicted by deep learning with field-based inventory data from a Mediterranean ecosystem exhibiting drought-induced dieback, and compare expert field-based crown dieback estimation with vegetation index-based estimates. We obtain high overall segmentation accuracy (mAP: 0.519) without the need for additional technical development of the underlying Mask R-CNN model, underscoring the potential of these approaches for non-expert use and proving their applicability to real-world conservation. We also find that color-coordinate based estimates of dieback correlate well with expert field-based estimation. Substituting ground truth for Mask R-CNN model predictions showed negligible impact on dieback estimates, indicating robustness. Our findings demonstrate the potential of automated data collection and processing, including the application of deep learning, to improve the coverage, speed, and cost of forest dieback monitoring.

Impact Statement

The global increase in forest dieback threatens vital ecosystem services, including habitat provision and carbon sequestration. Traditional field and drone-based monitoring techniques are not cost-effective at large scales. We leverage advancements in low-cost drones and deep learning to assess crown dieback from drone-captured RGB

  This research article was awarded Open Data and Open Materials badges for transparent practices. See the Data Availability Statement for details.

© The Author(s), 2024. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

aerial data, eliminating the need for costly tools like LiDAR and laborious fieldwork. We achieve high accuracy in matching predicted crown footprints with actual forest inventory, demonstrating the effectiveness of these methods for non-expert use in conservation. Our vegetation index-based dieback estimates aligned well with expert field-based estimates, validating this approach. This work enhances forest monitoring by improving coverage, accuracy and speed, and reducing costs, thereby providing a promising tool to address large-scale conservation challenges.

1. Introduction

Forest dieback, the unseasonal loss of crown foliage or mortality of many trees (Mueller-Dombois, 1988), is an early indicator of declining health in forest ecosystems. This degradation directly affects ecosystem services and functions such as carbon sequestration (Baccini et al., 2017) and habitat provision (Watson et al., 2018). Various factors, including the spread of pests, pathogens, and the intensification of drought conditions due to climate change, contribute to the increasing prevalence of crown dieback around the world (Carnicer et al., 2011; Liebholt et al., 2013; Senf et al., 2020). These changes in forest health and structure can result in ecologically significant shifts in composition and function over relatively short time periods (Allen et al., 2015), which can be difficult to detect using traditional sampling strategies (McMahon et al., 2019).

The global increase in dieback and its potential impacts on ecosystem services and functions highlight the urgent need for large-scale, high-frequency monitoring. Large-scale, high-frequency monitoring of forest health and structure would enhance our ability to detect issues such as pest outbreaks, disease, or stress caused by climate change (McMahon et al., 2019) at early stages—which would, in turn, support better-informed decision-making and policy development related to forest management, climate change mitigation, and biodiversity conservation (Asner and Martin, 2016). Moreover, large-scale monitoring techniques enable practitioners to better understand the factors driving dieback, informing more effective and targeted management strategies, and increasing resilience to these threats before they emerge. The instruments and methods to gather and process data for large-scale, rapid monitoring at the level of individual trees have, however, only been recently developed (Diez et al., 2021; Kattenborn et al., 2021; Ecke et al., 2022).

Remote sensing can provide high-resolution spatiotemporal data coverage to meet these needs. Leaf reflectance, including within the visible spectrum, is sensitive to changes in plant physiology, chemistry, and structure due to insect and pathogen infestation as well as water stress (Huang et al., 2019). Such changes in leaf reflectance are observable in the canopy from remotely sensed data (Huang et al., 2019) so high-resolution satellite imagery can likely detect defoliation missed by field-based surveys—for example, in low-severity areas (Bright et al., 2020). Many existing approaches, however, rely on large spatial averages of stress indicators—at the approximately 10 m spatial scale of common non-commercial satellite data products such as Sentinel 1/2 (Lastovicka et al., 2020) or Landsat (Zhang et al., 2021)—and therefore cannot capture information on individual tree health. Tree level information is crucial to understand future forest dynamics—particularly where individual trees may exhibit significantly different responses, such as for disease (Fensham and Radford-Smith, 2021; Hurel et al., 2021; Kännaste et al., 2023) or drought (Teskey et al., 2014; Chen et al., 2022; Fernández-de-Uña et al., 2023). New approaches, leveraging the increasing availability of high-resolution data ($\leq 0.3\text{m}$), are required to identify individual trees within imagery for monitoring.

Traditional segmentation methods—based on handcrafted algorithms and manual feature engineering (Batz and Schäpe, 2000)—have been used to extract Individual Tree Crown (ITC) polygons from aerial imagery for use in some downstream tasks useful for monitoring. For example, Onishi and Ise (2021) applied unsupervised multiresolution segmentation (Batz and Schäpe, 2000) to RGB (Red, Green, and Blue) data of a dense forest canopy in urban Japan, combined with Digital Surface Models (DSMs) to segment ITCs, and classified the results by species or functional type. Although the automatically extracted ITCs contained mostly canopy of the same species, the segmentation algorithm fragmented or merged 76% of tree crowns with neighboring individuals within that species. This method did not

require manual labeling, but the low intra-specific segmentation accuracy means it is not comparable to traditional inventory-based monitoring, as the ITC polygons cannot be used to accurately extract measurements relating to individual trees—such as measurement of stress responses including dieback. Since these responses vary substantially, even within species (Hurel et al., 2021; Fernández-de-Uña et al., 2023), accurate segmentation is crucial. We suggest that a deep learning-based approach is likely to yield better results on data with structurally complex canopies from natural forests, where crowns intersect substantially both within and across species.

Many approaches applying deep learning to segment ITCs exist (Diez et al., 2021). Deep learning was applied to perform segmentation through bounding box delineation (typically referred to as object detection in machine learning literature—for example, in Ren et al., 2016) of trees in both open and closed forest canopies from RGB images in Weinstein et al. (2019). Object detection was performed with a single-stage deep learning detector (He et al., 2015; Lin et al., 2018), achieving a precision of 0.69 and recall of 0.61 in forests from California, at a National Ecological Observation Network (NEON) site. The network was pre-trained using a very large number of noisy labels extracted from LiDAR data and fine-tuned on a smaller number of manual annotations. Precise delineation of individual crowns (typically referred to as instance segmentation in machine learning literature—for example, in He et al., 2017) through deep learning has also been applied produce the polygons required to quantify stress accurately at the individual level (Chiang et al., 2020; Hao et al., 2021; Šandric et al., 2022; Sani-Mohammed et al., 2022; Yang et al., 2022; Ball et al., 2023). Throughout this work, we refer to this type of segmentation, when applied to tree crowns, as “ITC delineation” to provide clarity to practitioners unfamiliar with machine learning literature, although “instance segmentation” could be used interchangeably. Hao et al. (2021), for example, used Mask R-CNN to perform ITC delineation of tree crowns from multispectral 2D imagery coupled with a photogrammetry-derived Canopy Height Model (CHM) in a Chinese fir (*Cunninghamia lanceolata*) plantation, and achieved an F1 score of 0.85. It was demonstrated that these delineations could be used successfully for a simple downstream task—the extraction of individual tree heights by superposing segmentation on the CHM data. The use of data from plantations, however, is unlikely to yield results that are representative of those from natural woodland; stems are typically evenly aged and spaced and do not have structurally complex canopy. Šandric et al. (2022) similarly segmented trees with Mask R-CNN in partly artificial contexts—across five different species in temperate and Mediterranean orchards. Yang et al. (2022) demonstrated that ITC delineation is possible in the canopy of greater structural complexity, showing some crown intersection and overlap, in the heavily managed environment of Central Park in New York City, USA—although delineation was not verified using ground data—and further demonstrated that a simple structural measurement (crown area) could be replicated compared to manual measurement from aerial data by using these delineations downstream. Ball et al. (2023) further demonstrated accurate ITC delineation in tropical forests with a structurally complex canopy, using Mask R-CNN, although differences in leaf reflectance and canopy structure corresponding to species differences may facilitate segmentation more easily than in monospecific canopy, where spectral variation may be relatively lower. Sani-Mohammed et al. (2022) were similarly able to perform ITC delineation using Mask R-CNN in natural temperate forest in Bavaria although only segmented dead trees, which may be easier to identify due to lack of foliage. The use of automatically extracted crown footprints for complex applications such as forest health monitoring, however, is less comprehensively explored.

Several studies exploiting deep learning for automated forest health measurement have emerged in recent years. Some work explores the use of classical computer vision, based on manually engineered algorithms, to perform ITC delineation and then apply deep learning to classify damage levels. Safonova et al. (2019), for example, applied manual filtering and thresholding to extract rectangular patches of treetops and applied a number of CNN-based classifiers downstream to classify damage into one of four levels. Nguyen et al. (2021) similarly applied a manually engineered algorithm for patch extraction, instead based on normalized Digital Surface Models (nDSMs). The extracted patches were classified by damage categorically. Although the classification of tree health into multiple damage levels is a promising approach, as dieback is a symptom with several continuous stages (Ciesla and Donaubauer, 1994), such approaches may be limited by their segmentation accuracy, particularly where individual trees differ

significantly in their response, within and across species. Other works use deep learning-based segmentation for forest health assessment. Schiefer et al. (2023) used a U-net to perform semantic segmentation (per-pixel classification of all pixels within a target image) of standing deadwood in UAV imagery and extrapolated this to the landscape level predictions on time series satellite imagery using Long Short Term Memory Networks (LSTMs). Chiang et al. (2020) and Sani-Mohammed et al. (2022) both used Mask R-CNN to perform ITC delineation of dead tree crowns from aerial imagery. Where dieback is ongoing, however, it requires measurement on a continuous scale or by multiple categories, rather than binary classification (Ciesla and Donaubaue, 1994). Šandric et al. (2022) used deep learning-based ITC delineation via Mask R-CNN to segment crown footprints and performed post hoc analysis using vegetation indices derived from color space transformations to measure tree health on a continuous scale. The approach of using vegetation indices—which do not require additional human labeling—on footprints extracted via deep learning, shows great promise for dieback measurement, but measurements of crown health were not verified versus human measurement. Additionally, the use of plantation data may more easily facilitate this type of forest health measurement—although often monospecific, the canopy formed by artificially spaced trees is easily delineated, simplifying health monitoring. This structural simplicity may not reflect performance in many natural ecosystems.

Here, we develop a new approach for early dieback detection from aerial RGB imagery, using deep-learning based ITC delineation and vegetation indices, and test it in an ecosystem experiencing dieback. Crucially, unlike in previous work (Šandric et al., 2022), we verify our results by comparison to field-based dieback measurement by experts. We use drone data collected in a Mediterranean stone pine (*Pinus pinea*) forest with a structurally complex monospecific canopy, where some individuals are showing signs of drought-induced crown dieback (Moreno-Fernández et al., 2022). The severity of drought is projected to increase across the Mediterranean region (Dubrovský et al., 2014; Hertig and Trambly, 2017), and poses a large-scale threat to the functioning of this biodiversity hotspot. Protection through active monitoring and management of these ecosystems is required (Fernández-Manjarrés et al., 2018; Astigarraga et al., 2020)—but is currently limited by reliance on time-consuming expert manual inventories taken on the ground. In addition to offering a solution to conducting this monitoring at large scale and low cost, drone-based remote sensing is ideal for the monitoring of drought-induced crown dieback, as effects at crown extremities are more easily observable from above than from traditional ground-based visual assessment. We develop a scalable approach based on deep learning and RGB vegetation indices to monitor this dieback at the individual level, and answer the following questions:

1. Is ITC delineation possible in a structurally complex, monospecific canopy?
2. Does individual health assessment based on vegetation indices (Šandric et al., 2022) correlate with field-based estimation, and to what degree is this assessment affected by the accuracy of deep learning-based ITC delineation?

2. Material and methodology

2.1. Data

2.1.1. Study area and inventory

Our dataset covers 1500 ha total of *Pinus pinea* forest, across nine areas, showing signs of climate-induced crown dieback. Drone and ground data were collected in Pinar de Almorox, Spain (40.27 °N, 4.36 °W) in May/June 2021. A map showing the location of the study site can be seen in Figure 1a. The area is under a continental Mediterranean climate with a mean annual rainfall of 568 mm, a mean annual temperature of 14°C, and an altitude ranging from 500 to 850 m above sea level. In addition to the dominant canopy species, *P. pinea*, a smaller number of *Juniperus oxycedrus* L. and *Quercus ilex* can be found in the midstory. The understory further contains *Salvia rosmarinus* L., *Lavandula stoechas* Lam. and *Lamiaceae* shrubs, as well as fallen deadwood. Nine distinct, non-overlapping areas were selected to show a gradient of defoliation, and within these inventory data was taken at 31 sample plots (three to four plots per area).

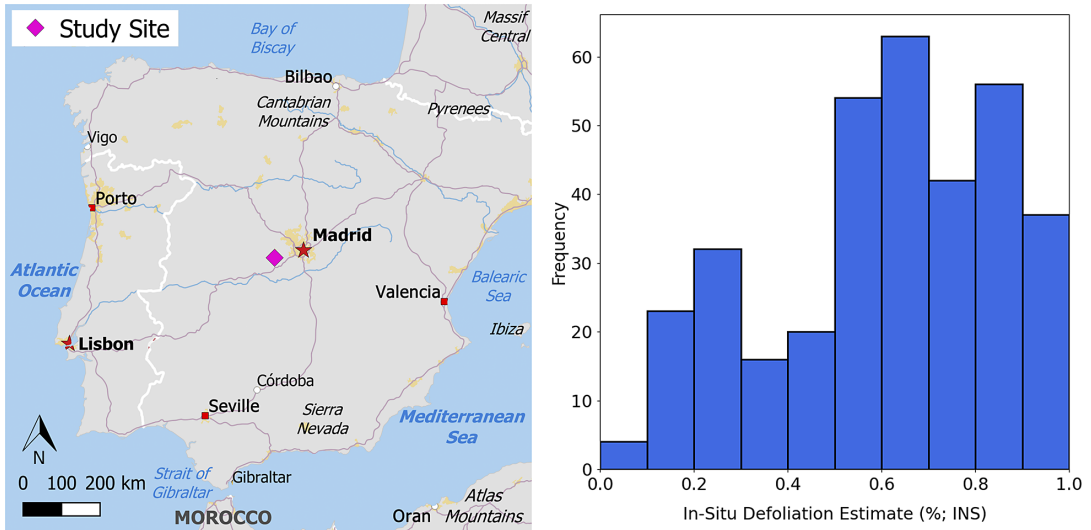


Figure 1. (a) Map showing the location of the site described by our data in the Iberian peninsula. Map data from OpenStreetMap Contributors (2017). (b) Histogram of visually estimated defoliation (percentage needles missing) of adult trees in our field data. An estimate of 1.0 corresponds to a tree for which all foliage is dead, and 0.0 to a tree with completely healthy foliage. A total of 453 adult trees (DBH > 7.5 cm) were surveyed.

Sample plots were circular and of 17 m radius, with a total of 453 adult (diameter at breast height, DBH, > 7.5 cm) trees surveyed in total. Each adult tree was assessed by experts in the field for dieback estimation (percentage defoliation). Crown dieback percentage was estimated visually (with 100% corresponding to wholly defoliated), taking the average score from two experts for each tree within each plot. A histogram of dieback/defoliation percentage for all surveyed trees can be seen in Figure 1b. Trunk locations within each plot were georeferenced using an Emlid Reach RS2 multi-band RTK Global Navigation Satellite System (GNSS) receiver (Ng et al., 2018). See Moreno-Fernández et al. (2022) for further analysis of dieback patterns at the site and additional data collected.

Note that each area surveyed by the drone was not covered entirely by the plot data. Consequently, trees at the edges of each orthomosaic did not have in situ defoliation estimates or geolocated trunk coordinates. These trees were still delineated manually from orthomosaic images at a later stage to provide training data for deep-learning based ITC delineation, but they were not considered in our analysis of dieback estimation.

2.1.2. Drone imagery

Drone flights were used to obtain images over the nine non-overlapping areas (minimum area 16663 m², maximum area 39736 m²), covering all plots, using a DJI Mavic Mini drone. The work was carried out with all relevant permissions, and all national drone regulations and location-specific regulations, including the use of rotor guards, were observed. As this area is of high conservation value, particularly for birds of prey, the areas to survey were discussed and approved with land managers in advance of flights. All flights were performed at a fixed altitude 50 m above take-off location and carried out in May/June 2021 at the same time as ground data collection. Flight start times varied from 09:30 to 17:00 (local time) and conditions varied from overcast to clear, resulting in significant differences in lighting conditions and shadowing between drone flights. Raw images of size 4000 × 2250 px were taken from both nadir and oblique (55° below horizontal; Nesbit and Hugenholtz, 2019) angles, with 95% front and 80% side overlap. Ground sampling distance (GSD) was approximately 3 cm for all areas. Between four and six Ground Control Points (GCPs) were placed in each area and precisely located using the GNSS receiver.

GCPs were matched across images semi-automatically using Agisoft Metashape 1.8.1.13915 (Agisoft LLC, 2022). An orthomosaic was generated for each area, also using Agisoft Metashape. These orthomosaics were cropped to remove edge effects. All data used to train deep learning-based ITC delineation were based on the orthomosaics rather than the original images.

2.1.3. Manual delineation and training data

Orthomosaics for each area were split into 1024×1024 tiles. This resolution was sufficiently large to preserve the visual context around each crown without downsampling, but small enough not to introduce excessive computational overhead. To create training data, all crowns visible in the orthoimagery were delineated manually, including those not covered by field plots, guided by geolocated trunk points where available. Manual delineation was performed in the full-size orthomosaics and the resulting delineations split to match the tiled images. We did not delineate understory growth where it was visible in the drone imagery. A visual example can be seen in Figure 2b. We make this tiled data available in the ubiquitous COCO format (Lin et al., 2015). A summary table describing our data can be seen in Table 1. Note that

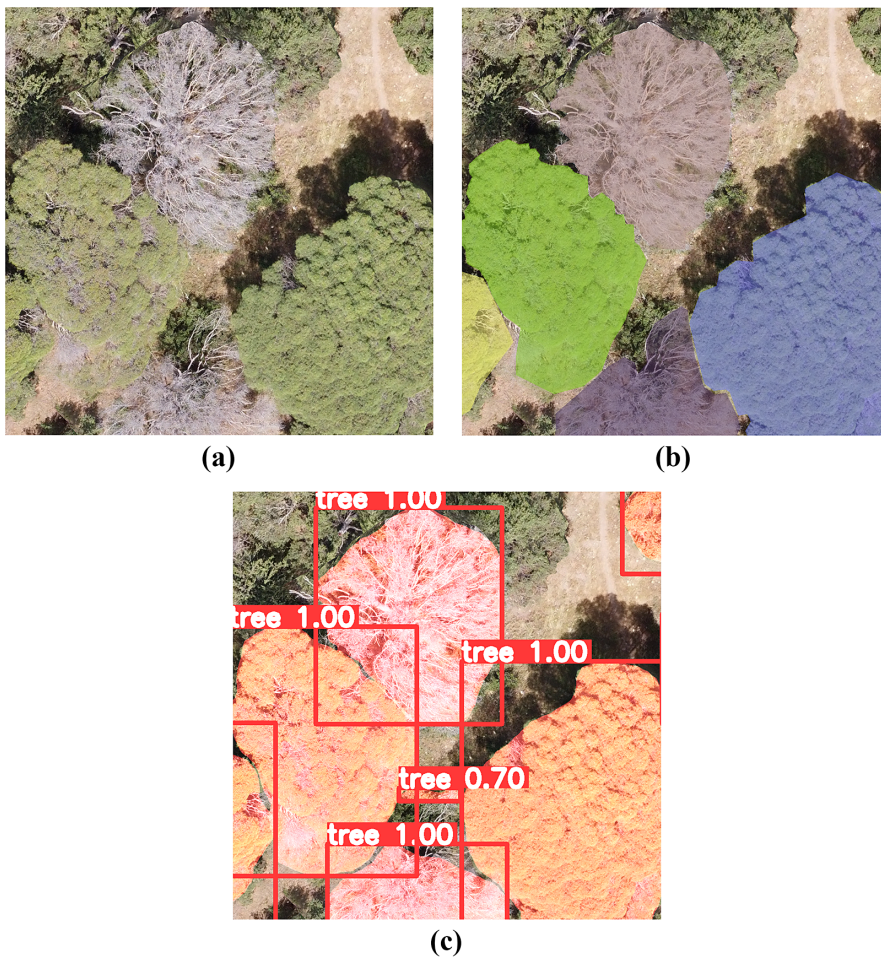


Figure 2. (a) Unlabeled, (b) manual, and (c) automatically predicted crown polygons for both healthy crowns (left, bottom right) and crowns exhibiting dieback (top center, bottom center). Numbers next to the class name “tree” denote confidence score corresponding to each prediction. Images in this figure span approximately 30 m.

Table 1. Summary information for our dataset of tree crowns in the nine areas covered by drone flights in Almorox, Spain

Area no.	Area (m ²)	Dimensions (Px)	No. crowns	Mean crown area (m ²)	No. of tiles
1	17,942	5491 × 8169	87	51.4	49
2	28,516	6837 × 10427	298	28.7	72
3	28,269	8263 × 8553	366	22.1	77
4	19,225	5835 × 8237	326	27.1	56
5	23,288	7489 × 7774	141	50.5	60
6	16,663	6308 × 6604	202	25.1	45
7	39,736	8781 × 11313	391	20.1	108
8	31,001	8599 × 9013	175	50.9	77
9	39,020	9945 × 9809	177	84.1	104
Total/Avg.	243661	N/A	2163	40.0	648

Note. Crown statistics are derived from manual delineations, rather than inventory, as ground sample plots only covered a small area of each orthomosaic.

these statistics are derived from the manually delineated polygons, rather than the inventory data, as the inventory data were not available for all trees, only those in the subplots within each orthomosaic. We did not perform any hyperparameter tuning and left hyperparameters as specified by the original author for each component of our methodology (He et al., 2015, 2017; Akyon et al., 2022). We therefore split our data into training and test sets only using nine-fold cross-validation, geographically, with each of the nine areas corresponding to the orthomosaic generated by one of the nine non-overlapping drone flights (see Section 2.1.2). In this approach, each fold uses one specific area as the test set while the remaining areas were used for training. This methodology means that the individual predicted footprints, which we used to evaluate segmentation performance and estimate dieback, were not generated by a single, universal model. Instead, each prediction was produced by a model trained on eight areas and tested on the remaining area (from which the tile to predict on is drawn). The average performance of our segmentation approach and the dieback estimation comparison was thus calculated by aggregating the results from each of these nine distinct models, on the area that each model was not trained on, providing a performance metric that reflects performance across all areas.

2.2. Methods

2.2.1. ITC delineation via instance segmentation

To delineate individual trees from tiled images, we used the ubiquitous Mask R-CNN framework (He et al., 2017) with a ResNet-101 FPN backbone (He et al., 2015). The backbone was pretrained on ImageNet (Deng et al., 2009), and both random resizing and flipping were used for augmentation. We henceforth refer to crown footprints predicted on individual tiled images as “tiled predictions.”

Tiled predictions were recombined using Slicing Aided Hyper Inference (SAHI; Akyon et al., 2022). We did not change the hyperparameters, used tiles with a relative overlap of 0.2 at inference, and post-processed predictions using greedy Non-Maximum Merging (NMM) with an Intersection-Over-Smaller (IOS) match threshold of 0.5. During ITC delineation (instance segmentation), each predicted crown is given a confidence score, corresponding to how likely it is that it contains a crown. A minimum confidence threshold of 0.3 was imposed on predictions before merging, as in the original SAHI implementation (Akyon et al., 2022). We report segmentation performance using mean Average Precision (mAP; see Beitzel et al., 2009) at two stages—on the tiled dataset, averaged across all tiles within each area, and after the tiled predictions are recombined using SAHI, for each of the nine orthomosaics.

2.2.2. Vegetation index-based dieback estimates at the tree level

We used an iterative approach to match ITCs to GNSS-measured trunk locations for trees with ground-truth GPS field locations (positional error < 1 m) and expert dieback assessment (percentage crown defoliation) in the full-size orthomosaics. To do this, the Euclidean distance was calculated between the centroid of each ITC and all ground truth trunk locations. The pair with the smallest distance was taken to be a match, and both ITC and ground truth (field) location for that pair were not considered as part of further matches. This step was repeated until no ground truth locations remained. We opted to discard ITC-trunk pairs where the distance from the centroid of the predicted crown to the ground-truth trunk location was greater than the square root of the crown area, to avoid including automatic dieback estimates that did not reliably correspond to the supposedly matching in situ estimates. See [Appendix B](#) for pseudocode. We repeated this analysis twice—once using the ITCs predicted using the deep learning-based ITC delineation model, and once using the manually delineated ITCs (which were also used to train the segmentation model).

For the matched crown-trunk pairs, defoliation was estimated via Green Chromatic Coordinate (GCC), a common metric that has been used to track phenology successfully in a range of ecosystems (Richardson et al., 2018), and a similar metric to the vegetation indices used to track tree health in other works (Reid et al., 2016; Šandric et al., 2022). GCC is defined below in [Equation \(1\)](#).

$$\text{GCC} = \frac{G}{R + G + B} \quad (1)$$

Here, R, G, and B refer to the total red, green, and blue pixel values, respectively, summed across the region of interest. ITCs were matched to ground-surveyed trunk locations (see [Appendix B](#)), and drone-derived GCC values were then correlated with field-based percentage of defoliation, using GCC values derived from both the manually labeled and automatically segmented crowns.

3. Results

Our results show that automatic crown delineation was possible in variably packed canopy with low species variation ([Figure 2](#)). We present both per-tile averages and results for full-size orthomosaic inference using SAHI, per-area in [Table 2](#). Visual results can be seen in [Figure 2](#). An average mAP of 0.519 was achieved on the tiled data, with a minimum of 0.473, a maximum of 0.602, and a standard

Table 2. Summary results (crown segmentation mAP) as mean plus/minus standard deviation for ITC delineation of our dataset from Almorox, Spain using Nine-fold cross-validation

Area no.	Tiled (Mask R-CNN)			Orthomosaic (Mask R-CNN + SAHI)		
	mAP	mAP _{0.5}	mAP _{0.75}	mAP	mAP _{0.5}	mAP _{0.75}
1	0.602	0.792	0.681	0.490	0.740	0.518
2	0.507	0.715	0.564	0.424	0.630	0.458
3	0.558	0.777	0.641	0.544	0.779	0.610
4	0.496	0.749	0.547	0.439	0.696	0.483
5	0.515	0.763	0.563	0.315	0.544	0.343
6	0.522	0.756	0.576	0.471	0.695	0.537
7	0.486	0.702	0.545	0.460	0.653	0.501
8	0.473	0.661	0.525	0.377	0.580	0.379
9	0.512	0.763	0.563	0.377	0.661	0.325
Mean ± std.	0.519 ± 0.037	0.742 ± 0.039	0.578 ± 0.047	0.433 ± 0.065	0.664 ± 0.070	0.462 ± 0.090

Note. Mean average precision is reported at IoU thresholds of 0.5 and 0.75 and averaged between 0.5 and 0.95 using intervals of 0.05, as in He et al. (2017).

deviation of 0.037. When recombined for the full orthomosaic using SAHI, mean mAP was reduced to 0.433, with a minimum of 0.377, maximum of 0.544, and a standard deviation of 0.065. We note that performance varied somewhat from area to area. We also note that recombining the tiled predictions increased variation in performance, and decreased mAP for all nine areas.

The correlation between drone and field-based defoliation estimation was strong (Figure 3). We correlated ground estimates of defoliation with GCC estimates derived from crowns segmented in aerial imagery, using both crowns segmented using deep learning (Figure 3a) and crowns segmented by hand (Figure 3b). For the automatically segmented crowns, we found the correlation between calculated GCC and field-based defoliation estimates was significant ($p < 4 \times 10^{-32}$, $R^2 = 0.35$). When the analysis was repeated using the manually segmented (ground truth) crown polygons (Figure 3b), the correlation was significant with equivalent performance ($p < 2 \times 10^{-33}$, $R^2 = 0.34$).

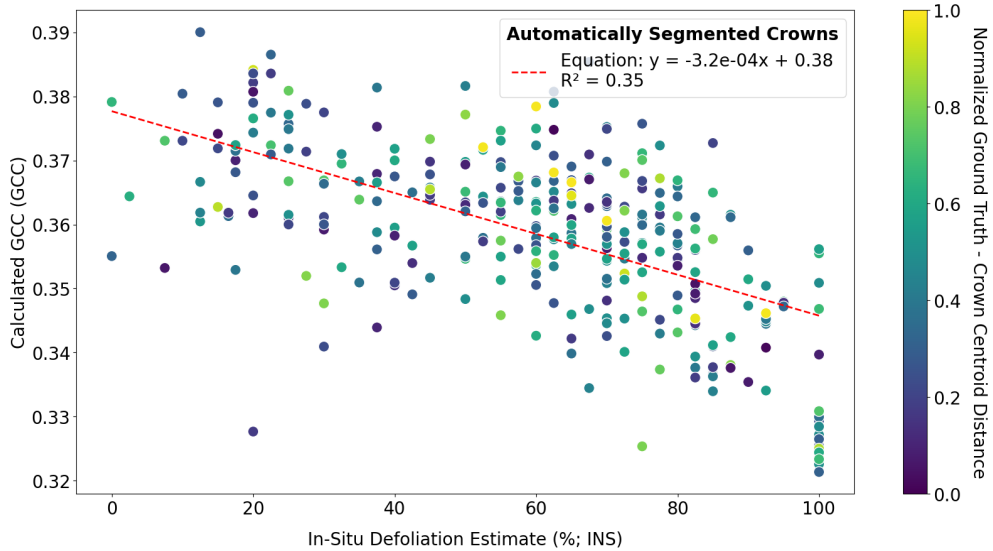
There was little change in the strength of correlation between field-based defoliation estimation and aerial GCC-based estimates, when calculated with the automatically extracted crowns (Figure 3a) versus the manually labeled crowns (Figure 3b). The GCC estimates for each case were matched according to the corresponding ground truth trunk location obtained for each polygon via Algorithm 1, and showed strong correlation against each other in Figure 4 ($R^2 = 0.54$, $\text{RMSE} = 0.01$, $p < 3 \times 10^{-72}$).

4. Discussion

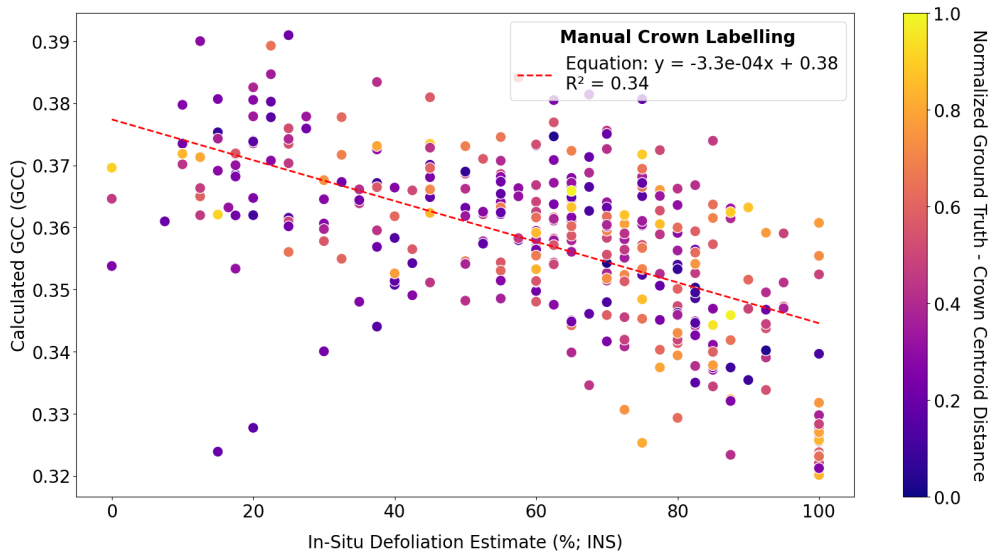
We found that, when evaluated on both the tiled and full-orthomosaic crown segmentation data, Mask R-CNN produced strong results ($\text{mAP} = 0.519 \pm 0.037$), comparable or higher to those on machine learning benchmark data ($\text{mAP} = 0.371$ from Lin et al., 2015). Although results on different data are not directly comparable, these relatively similar mAP scores indicate that ITC delineation in monospecific canopy is no less achievable than segmenting everyday objects from curated imagery—although we stress that the absolute accuracy of the manually annotated training labels is not known. The lack of need for additional technical development underscores the potential of models such as Mask R-CNN for non-expert use, proving their applicability to real-world conservation. Whilst segmentation performance could likely be improved with a more sophisticated or larger model and better postprocessing, these may not be the limiting factors to using such models in a practical context. Performance decreased in almost all cases when recombining tiled to full-size orthomosaic prediction using SAHI.

Direct comparison to previous work performing ITC delineation is difficult due to variation in metrics used to assess performance, differences in canopy structural complexity between ecosystems, and the inclusion or omission of ground-based data to verify labels. Ground-based data serve as a more definitive source of accuracy, and the lack of such data in some studies (Yang et al., 2022; Ball et al., 2023) can potentially lead to misleading performance metrics or less reliable interpretations of model effectiveness. Although mAP and similar derivative metrics are widely accepted for measuring instance segmentation performance on large-scale benchmark datasets such as COCO (Lin et al., 2015), these metrics are not commonly reported in similar works—with authors commonly opting to report performance using metrics such as F1-score (Ball et al., 2023) or precision (Weinstein et al., 2019). The interpretability of these metrics is beneficial, but these are point metrics relying on a single selection of minimum confidence and Intersection-over-Union (IoU) thresholds—which can be adjusted arbitrarily to maximize the target metric (Maxwell et al., 2021b). A higher or lower value of F1-score, precision or recall may not represent a better or worse model for a user, even when trained on the same data. mAP-based metrics are a better reflection of model performance, and higher values indicate that a model is likely to be robust to different manual selections of confidence and IoU thresholds (Maxwell et al., 2021a,b). Given the complex nature of forests, including variation in canopy structural complexity, a model's performance across different ecosystems and conditions relies on its ability to handle observed variability.

We observed a significant negative correlation between field-based defoliation estimates and drone-derived GCC in our site for both automatic ($R^2 = 0.35$; Figure 3a) and manual segmentation ($R^2 = 0.34$; Figure 3b), with a high correlation between estimation using the manual labels versus the automatically



(a) Plot of Estimated Green Chromatic Coordinate (GCC) vs. field-based defoliation (percentage of defoliation), using **automatically segmented crowns**. Linear model derived using Ordinary Least Squares (OLS), $p < 4 \times 10^{-32}$. Since we used nine-fold cross validation, each data point is produced by one of nine models, each of which has not been influenced by the ground-truth trunk location to which it is matched (as it corresponds to a tree in the area comprising the test set).



(b) Plot of Estimated Green Chromatic Coordinate (GCC) vs. field-based defoliation (percentage of defoliation), using **manually segmented crowns**. Linear model derived using Ordinary Least Squares (OLS), $p < 2 \times 10^{-33}$.

Figure 3. Dispersion plots of Estimated Green Chromatic Coordinate (GCC) versus field-based defoliation at the individual tree level.

segmented crowns (Figure 4). The difference between the two GCC estimates for each crown did not grow substantially when the automatically segmented crown footprint was further from the true location (See Appendix A for further details). Our findings suggest that even a basic deep-learning approach performs as effectively as manual annotations in producing ITC footprints for dieback estimation. Investing in more advanced segmentation may not offer significant improvements for estimating dieback or, perhaps, other

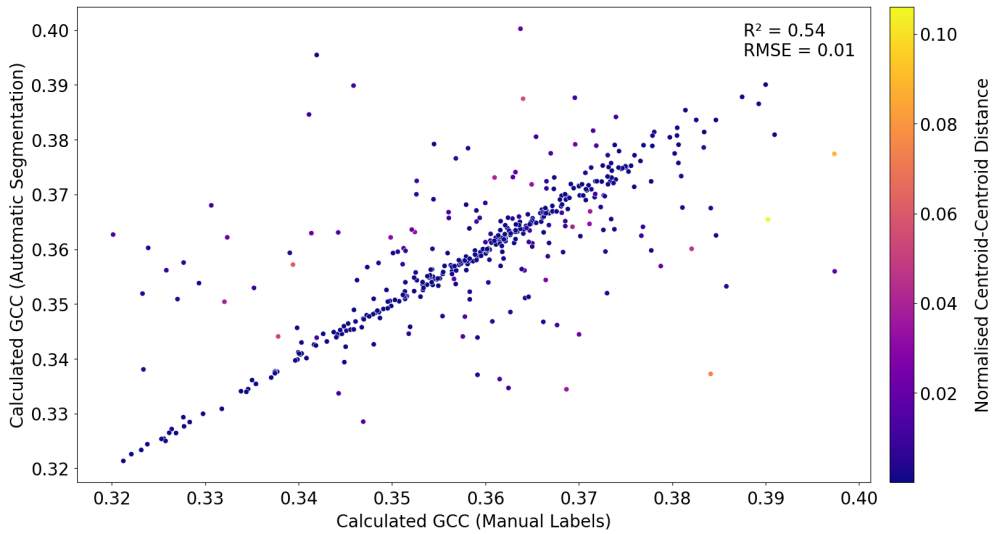


Figure 4. Correlation between Green Chromatic Coordinate (GCC) estimated using manually labeled versus automatically segmented crowns. Crowns are matched according to the corresponding inventory trunk location from the crown matching algorithm. The RMSE between the two GCC estimates was found to be 0.01, with $p < 3 \times 10^{-72}$.

crown metrics such as Leaf Area Index (LAI)—although further verification would be required to confirm this. Manual labels are, however, difficult to verify as being accurate, and it may be the case that training the model based on crowns labeled by humans from above introduces operator bias (Bai et al., 2018; Geva et al., 2019). Verifiable ground-truth labels could be derived, for example, from additional instruments such as Terrestrial Laser Scanners (TLS), but such data was not available for this site. Whether it is possible to match the accuracy of TLS-based segmentation from RGB imagery alone is unknown at this time.

The variance in calculated GCC for a given field-based defoliation percentage may be derived from physical reasons related to field-based estimates. Dieback is not expected to be uniform within each crown (Denman et al., 2022)—so metrics based on per-pixel color averages, without regard to spatial patterns, are unlikely to capture the degree of dieback for each crown with perfect accuracy. This effect is exaggerated by the particular patterns of drought-induced dieback, which typically begins at the extremities of the crown (Denman et al., 2022)—resulting in substantially different observations when viewed from above, as in the aerial images used here, and from below, as would be seen in situ. Although other, similar work, uses different vegetation indices (Šandric et al., 2022), we suggest that the variance in results is unlikely to diminish significantly while still using vegetation indices based on simple color space transformations. Extracting dieback estimates using the deep learning model directly—akin to classification in the context of the original Mask R-CNN model (He et al., 2017), but modified to perform regression—may improve performance as these estimates could leverage spatial patterns in addition to color information. We opted not to use this approach to avoid the need to produce dieback labels for trees visible in orthoimagery that were not already covered by existing field-based estimates.

Our results show that, for data taken with the same instruments, vegetation index-based estimates correlate significantly with expert field-based estimates. However, care should be taken when making direct comparisons between estimates for individual trees. For example—comparing dieback estimates for the same tree from imagery taken at different times, or with different instruments, may generate confusion as measured color intensities will change due to lighting or camera differences. We suggest that color calibrating imagery may improve this, although we did not perform color calibration here as all images were taken with the same instruments and did not make direct comparisons between individual data points obtained under different conditions.

We successfully automated canopy defoliation measurement in a single-species canopy, comprising *P. pinea*, the only species displaying dieback in our ecosystem. Minor modifications are likely to be required for our approach—based on vegetation indices applied to crown footprints extracted using deep learning-based instance segmentation—to be applicable to data of multiple species. Since canopy reflectance may vary due to differences in leaf spectra and crown structure corresponding to species (Ollinger, 2011), we would not recommend directly comparing vegetation index-based estimates between two species. The approach of Šandric et al. (2022) indicates that using the deep learning model to classify crown footprints by species prior to any health monitoring is promising, although this increases the labeling burden on practitioners. It is possible that an increase in open ITC delineation data may help to relieve this problem, by removing the need for site-specific datasets, or by reducing volume requirements on labels via pretraining on such open data. Alternatively, estimating dieback based on spatial patterns rather than simple color-space based estimation may prove to be more comparable, as dieback is not expected to be uniform within each crown (Denman et al., 2022). In the case of our data, the canopy comprises *P. pinea* almost exclusively—which is also the only species displaying dieback in this ecosystem—meaning we do not suffer from such confusion in this work.

Our approach has three important advantages that decrease financial and manual cost to application. Firstly, an increase in processing speed, secondly, the use of a low-cost, widely available drone platform, and thirdly, the reduced reliance on expert analysis in the field. The time savings, for example, of using a large-scale aerial approach such as this rather than a ground method, are significant. Unlike a typical field campaign, including manual inventory and perhaps the use of other instruments such as laser scanners, drone flights for a site can be conducted in hours to days rather than weeks to months. With a trained model, postprocessing these data to extract downstream metrics requires only a few hours (for a dataset of this size on consumer hardware). Short, drone-based campaigns also have smaller requirements for material—with the DJI Mavic Mini used here being available for only a few hundred euros, and being light enough to fly with minimal legal requirements. Combined with the reduced number of person-hours required, the cost of conducting and processing data from a drone-based survey is minimal compared with a traditional field campaign.

Although we used significant manual input at two points during postprocessing, this could be avoided in practice as more open data becomes available (Lines et al., 2022). Firstly, manual GCPs were used to stitch together the orthorectified images. We found orthophotos stitched together without the use of GCPs were of comparable visual quality, although we did not compare the use of these images downstream to avoid the need to repeat the time-consuming labeling process. It is possible that the orthomosaics may be georeferenced less accurately without the use of GCPs, although this could be addressed by mounting a more accurate GPS receiver to the drone, and such RTK systems are increasingly becoming available on consumer drone systems. Secondly, the manual labeling process was time-consuming and required several weeks to complete by a single user. Although we used ground truth labels to verify our results, they may not be required for practitioners to use similar approaches, particularly with community efforts to provide ground-truth training data for similar applications (Puliti et al., 2023). Large foundation models, for example, can produce reasonable segmentation in a wide range of unseen contexts when pretrained on vast quantities of general visual data (Kirillov et al., 2023), although can be difficult to employ using consumer hardware. Some of these approaches rely on context-specific prompting to identify rough areas of interest (Kirillov et al., 2023). As accuracy requirements for prompting are less strict than for full segmentation, methods for generating these prompts developed on open forest data may be sufficiently accurate on unseen ecosystems. Alternatively, smaller models such as the one used here could be pretrained on public ITC delineation data, and used without retraining. Both of these solutions, however, require large-scale data of tree crowns from a diverse set of ecosystems.

In addition to speed and cost savings, the adoption of automated, digital methods also benefits accuracy (Mu et al., 2018) and reduces dataset sampling bias. Typically, an inventory-based approach can gather only simple structural measurements such as trunk diameter or height, which could be used to produce crude estimates of difficult to measure quantities such as crown area. A data-driven approach, using drone-based remote sensing and deep learning, allows for measurement of quantities such as crown area

far more accurately than hand measurement, but without the laborious time requirements of other highly accurate instruments such as TLS scanners. The comprehensive nature of automated surveying also reduces selection bias—for example, vegetation in areas that are difficult to reach by foot, such as on very steep terrain, may differ physiologically from plants found on flat ground. A drone-based survey can include a large number of stems from such areas, whereas an inventory-based survey may only include stems in areas that can be accessed directly.

With these combined improvements in speed, cost, coverage, and accuracy, profound implications for forest management in deteriorating ecosystems are evident. Real-time detection of tree dieback is essential, as declining trees become hotspots for primary pests and pathogens to thrive and multiply. Additionally, such trees are more susceptible to secondary pest attacks (Balla et al., 2021). Regular updates on tree health are vital for guiding management strategies, such as thinning stands to lessen resource competition among surviving trees—crucial under extreme climate conditions—or implementing sanitation cuts to eliminate trees harboring pest or pathogen populations (Roberts et al., 2020).

For this paradigm shift toward automation to be applied in practice, two requirements remain. Firstly, an increase in extensive, diverse open datasets is vital for both training and validation, minimizing the labeling burden on end-users (Lines et al., 2022). Secondly, open and easy-to-use implementations of these tools, applied to common forest management data formats, are needed to minimize barriers to deployment for non-expert users. Some promising initiatives already exist (Ball et al., 2023; Environmental Data Science Book, 2023). With these two requirements fulfilled, the shift to automation in forest management and conservation could be expedited. Focus should therefore be given to fostering a culture of data sharing, and to the development of accessible platforms to deploy these methods. Toward these goals, we make both code and data available for this work, in both common remote sensing and machine learning-readable formats.

5. Conclusions

In this work, we made three key contributions to the state-of-the-art in aerial tree health assessment. Firstly, we demonstrated the feasibility of deep learning-based ITC delineation in structurally complex monospecific natural canopies, where reduced leaf spectra variation between individual trees may result in relatively lower variation in canopy reflectance compared to mixed canopies. Although verifying crown segmentation precisely is challenging, we corroborated our segmentation method using ground truth GPS trunk locations to assess its accuracy. We highlight the need for further verification in future work, potentially through ground-based tools such as Terrestrial Laser Scanning (TLS).

Secondly, our work found that detecting crown dieback using vegetation indices derived from deep learning-based crown footprints is feasible, as suggested by Šandric et al. (2022). For the first time, we validated this type of tree health assessment with field-based expert observations and found that there was strong correlation between our automatic estimates and field-based estimates.

Lastly, we give evidence that our method of tree health assessment is not overly sensitive to segmentation accuracy. Repeating the analysis with ground truth segmentation labels in place of model predictions did not significantly improve the correlation with field-based assessment. This finding suggests developing more precise segmentation methods for tree health assessment may not be necessary, although more verification is required when using more complex or spatially explicit measurements downstream.

Author contribution. All authors contributed to the conceptualization of the work; M.J.A. and E.R.L. designed the methodology; E.R.L., D.M.-F. and R.-B. collected and pre-processed the data; M.J.A. processed the data and wrote the relevant software. M.J.A. analyzed the data and led the writing of the manuscript. All authors contributed critically to the drafts and gave final approval for publication.

Competing interest. The authors declare none.

Data availability statement. Our code is available at github.com/mataln/dlncs-dieback. A permanent record is available at <http://doi.org/10.5281/zenodo.10657079>. Our data is available at <http://doi.org/10.5281/zenodo.10646992>.

Ethics statement. The research meets all ethical guidelines, including adherence to the legal requirements of the study country.

Funding statement. M.J.A. was supported by the UKRI Centre for Doctoral Training in Application of Artificial Intelligence to the Study of Environmental Risks (EP/S022961/1). S.W.D.G. and E.R.L. were funded by a UKRI Future Leaders Fellowship awarded to E.R.L. (MR/T019832/1). P.R.-B. was supported by the Community of Madrid Region under the framework of the multi-year Agreement with the University of Alcalá (Stimulus to Excellence for Permanent University Professors, EPU-INV/2020/010). P.R.-B. and E.R.L. are supported by the Science and Innovation Ministry (subproject LARGE and REMOTE, N PID2021-123675OB-C41 and PID2021-123675OB-C42).

Abbreviations

CHM	Canopy Height Model
COCO	Common Objects in COntext
ExG	Excess Green Index
GCC	Green Chromatic Coordinate
GCP	Ground Control Point
GNSS	Global Navigation Satellite System
GSD	Ground Sampling Distance
IOS	Intersection-Over-Smaller
IoU	Intersection-Over-Union
ITC	Individual Tree Crown
LAI	Leaf Area Index
LiDAR	Light Detection and Ranging
LSTM	Long Short-Term Memory
mAP	Mean Average Precision
nDSM	normalized Digital Surface Model
NEON	National Ecological Observation Network
NMM	Non-Maximum Merging
OLS	Ordinary Least Squares
R-CNN	Region-based Convolutional Neural Network
RGB	Red, Green, and Blue
RMSE	Root-Mean-Square Error
RTK	Real-Time Kinematic
SAHI	Slicing Aided Hyper Inference
TLS	Terrestrial Laser Scanning

References

- Agisoft LLC** (2022) Agisoft Metashape Pro 1.8.3.
- Akyon FC, Altinuc SO and Temizel A** (2022) Slicing aided hyper inference and fine-tuning for small object detection. In *2022 IEEE International Conference on Image Processing (ICIP)*. Bordeaux, France: IEEE, pp. 966–970. <https://doi.org/10.1109/ICIP46576.2022.9897990>.
- Allen CD, Breshears DD and McDowell NG** (2015) On underestimation of global vulnerability to tree mortality and forest die-off from hotter drought in the anthropocene. *Ecosphere* 6(8), art129. <https://doi.org/10.1890/ES15-00203.1>.
- Asner GP and Martin RE** (2016) Spectranomics: Emerging science and conservation opportunities at the interface of biodiversity and remote sensing. *Global Ecology and Conservation* 8, 212–219. <https://doi.org/10.1016/j.gecco.2016.09.010>.
- Astigarraga J, Andivia E, Zavala MA, Gazol A, Cruz-Alonso V, Vicente-Serrano SM and Ruiz-Benito P** (2020) Evidence of non-stationary relationships between Climate and Forest responses: Increased sensitivity to climate change in Iberian forests. *Global Change Biology* 26(9), 5063–5076. <https://doi.org/10.1111/gcb.15198>.
- Baatz M and Schäpe A** (2000) Multiresolution segmentation: An optimization approach for high quality multi-scale image segmentation. In Strobl J, Blaschke T and Griesbner G (eds.), *Angewandte Geographische Informations-Verarbeitung, XII*. Karlsruhe, Germany: Wichmann Verlag, pp. 12–23.
- Baccini A, Walker W, Carvalho L, Farina M, Sulla-Menashe D and Houghton RA** (2017) Tropical forests are a net carbon source based on aboveground measurements of gain and loss. *Science* 358(6360), 230–234. <https://doi.org/10.1126/science.aam5962>.

- Bai Y, Zhang Y, Ding M and Ghanem B** (2018) Finding tiny faces in the wild with generative adversarial network. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Salt Lake City, UT, USA: IEEE, pp. 21–30. <https://doi.org/10.1109/CVPR.2018.00010>
- Ball JGC, Hickman SHM, Jackson TD, Koay XJ, Hirst J, Jay W, Archer M, Aubry-Kientz M, Vincent G and Coomes DA** (2023) Accurate delineation of individual tree crowns in tropical forests from aerial RGB imagery using Mask R-CNN. *Remote Sensing in Ecology and Conservation* 9, 641–655. <https://doi.org/10.1002/rse2.332>.
- Balla A, Silini A, Cherif-Silini H, Chenari Bouket A, Moser WK, Nowakowska JA, Oszako T, Benia F and Belbahri L** (2021) The threat of pests and pathogens and the potential for biological control in forest ecosystems [number: 11 Publisher: Multidisciplinary Digital Publishing Institute]. *Forests* 12(11), 1579. <https://doi.org/10.3390/f12111579>.
- Beitzel SM, Jensen EC and Frieder O** (2009) Map. In Liu L and Özsu MT (eds.), *Encyclopedia of Database Systems*. New York, NY: Springer, pp. 1691–1692. https://doi.org/10.1007/978-0-387-39940-9_492.
- Bright BC, Hudak AT, Egan JM, Jorgensen CL, Rex FE, Hicke JA and Meddens AJH** (2020) Using satellite imagery to evaluate bark beetle-caused tree mortality reported in aerial surveys in a mixed conifer Forest in Northern Idaho, USA. *Forests* 11, 529. <https://doi.org/10.3390/f11050529>.
- Carnicer J, Coll M, Ninyerola M, Pons X, Sánchez G and Peñuelas J** (2011) Widespread crown condition decline, food web disruption, and amplified tree mortality with increased climate change-type drought. *Proceedings of the National Academy of Sciences* 108(4), 1474–1478. <https://doi.org/10.1073/pnas.1010070108>.
- Chen Z, Li S, Wan X and Liu S** (2022) Strategies of tree species to adapt to drought from leaf stomatal regulation and stem embolism resistance to root properties. *Frontiers in Plant Science* 13, 926535. <https://doi.org/10.3389/fpls.2022.926535>.
- Chiang C-Y, Barnes C, Angelov P and Jiang R** (2020) Deep learning-based automated forest health diagnosis from aerial images. *IEEE Access* 8, 144064–144076. <https://doi.org/10.1109/ACCESS.2020.3012417>.
- Ciesla WM and Donaubaue E** (1994) *Decline and Dieback of Trees and Forests: A Global Overview* [Google-Books-ID: FDgYfi1G8LOC]. Rome, Italy: UN FAO.
- Deng J, Dong W, Socher R, Li L-J, Li K and Fei-Fei L** (2009) ImageNet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*. Miami, FL, USA: IEEE, pp. 248–255. <https://doi.org/10.1109/CVPR.2009.5206848>.
- Denman S, Brown N, Vanguelova E and Crampton B** (2022) Chapter 14 - Temperate oak declines: Biotic and abiotic predisposition drivers. In Asiyebu FO and Kovalchuk A (eds.), *Forest Microbiology*. Cambridge, MA, USA: Academic Press, pp. 239–263. <https://doi.org/10.1016/B978-0-323-85042-1.00020-3>.
- Diez Y, Kentsch S, Fukuda M, Caceres MLL, Moritake K and Cabezas M** (2021) Deep learning in forestry using UAV-acquired RGB data: A practical review [number: 14 Publisher: Multidisciplinary digital publishing institute]. *Remote Sensing* 13(14), 2837. <https://doi.org/10.3390/rs13142837>.
- Dubrovský M, Hayes M, Duce P, Trnka M, Svoboda M and Zara P** (2014) Multi-GCM projections of future drought and climate variability indicators for the Mediterranean region. *Regional Environmental Change* 14(5), 1907–1919. <https://doi.org/10.1007/s10113-013-0562-z>.
- Ecke S, Dempewolf J, Frey J, Schwaller A, Endres E, Klemmt H-J, Tiede D and Seifert T** (2022) UAV-based forest health monitoring: A systematic review [number: 13 Publisher: Multidisciplinary digital publishing institute]. *Remote Sensing* 14(13), 3205. <https://doi.org/10.3390/rs14133205>.
- Environmental Data Science Book** (2023) Environmental Data Science Book: A living, open and community-driven online resource to showcase and support the publication of data, research and open-source tools for collaborative, reproducible and transparent Environmental Data Science. <https://doi.org/10.5281/zenodo.7951713>.
- Fensham RJ and Radford-Smith J** (2021) Unprecedented extinction of tree species by fungal disease. *Biological Conservation* 261, 109276. <https://doi.org/10.1016/j.biocon.2021.109276>.
- Fernández-de-Uña L, Martínez-Vilalta J, Poyatos R, Mencuccini M and McDowell NG** (2023) The role of height-driven constraints and compensations on tree vulnerability to drought. *New Phytologist* 239(6), 2083–2098. <https://doi.org/10.1111/nph.19130>.
- Fernández-Manjarrés JF, Ruiz-Benito P, Zavala MA, Camarero JJ, Pulido F, Proença V, Navarro L, Sansilvestri R, Granda E, Marqués L, Temunović M, Bertelsmeier C, Drobinski P, Roturier S, Benito-Garzón M, Cortazar-Atauri IG d, Simon L, Dupas S, Levrel H and Sautier M** (2018) Forest adaptation to climate change along steep ecological gradients: The case of the Mediterranean-temperate transition in South-Western Europe. *Sustainability* 10(9), 3065. <https://doi.org/10.3390/su10093065>.
- Geva M, Goldberg Y and Berant J** (2019) Are we modeling the task or the annotator? An investigation of annotator bias in natural language understanding datasets. *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing - the 9th International Joint Conference on Natural Language Processing*, 1161–1166. <https://doi.org/10.48550/arXiv.1908.07898>
- Hao Z, Lin L, Post CJ, Mikhailova EA, Li M, Chen Y, Yu K and Liu J** (2021) Automated tree-crown and height detection in a young forest plantation using mask region-based convolutional neural network (mask R-CNN). *ISPRS Journal of Photogrammetry and Remote Sensing* 178, 112–123. <https://doi.org/10.1016/j.isprsjprs.2021.06.003>.
- Harris CR, Millman KJ, van der Walt SJ, Gommers R, Virtanen P, Cournapeau D, Wieser E, Taylor J, Berg S, Smith NJ, Kern R, Picus M, Hoyer S, van Kerkwijk MH, Brett M, Haldane A, del Río JF, Wiebe M, Peterson P, Gérard-Marchant P, Sheppard K, Reddy T, Weckesser W, Abbasi H, Gohlke C, Oliphant TE** (2020) Array programming with NumPy. *Nature* 585 (7825), 357–362. <https://doi.org/10.1038/s41586-020-2649-2>.

- He K, Gkioxari G, Dollár P and Girshick R** (2017) Mask R-CNN. *2017 IEEE International Conference on Computer Vision (ICCV)*, 2980–2988.
- He K, Zhang X, Ren S and Sun J** (2015) Deep Residual Learning for Image Recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 770–778.
- Hertig E and Tramblay Y** (2017) Regional downscaling of Mediterranean droughts under past and future climatic conditions. *Global and Planetary Change* 151, 36–48. <https://doi.org/10.1016/j.gloplacha.2016.10.015>.
- Huang C-y, Anderegg WR and Asner GP** (2019) Remote sensing of forest die-off in the Anthropocene: From plant ecophysiology to canopy structure. *Remote Sensing of Environment* 231, 111233. <https://doi.org/10.1016/j.rse.2019.111233>.
- Hurel A, de Miguel M, Dutech C, Desprez-Loustau M-L, Plomion C, Rodríguez-Quilón I, Cyrille A, Guzman T, Alía R, González-Martínez SC and Budde KB** (2021) Genetic basis of growth, spring phenology, and susceptibility to biotic stressors in maritime pine. *Evolutionary Applications* 14(12), 2750–2772. <https://doi.org/10.1111/eva.13309>.
- Kännaste A, Jüriso L, Runno-Paurson E, Kask K, Talts E, Pärlist P, Drenkhan R and Niinemets Ü** (2023) Impacts of Dutch elm disease-causing fungi on foliage photosynthetic characteristics and volatiles in *Ulmus* species with different pathogen resistance. *Tree Physiology* 43(1), 57–74. <https://doi.org/10.1093/treephys/tpac108>.
- Kattenborn T, Leitloff J, Schiefer F and Hinz S** (2021) Review on convolutional neural networks (CNN) in vegetation remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing* 173, 24–49. <https://doi.org/10.1016/j.isprsjprs.2020.12.010>.
- Kirillov A, Mintun E, Ravi N, Mao H, Rolland C, Gustafson L, Xiao T, Whitehead S, Berg AC, Lo W-Y, Dollár P and Girshick R** (2023) Segment Anything. *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, 3992–4003. <https://doi.org/10.48550/arXiv.2304.02643>.
- Lastovicka J, Svec P, Paluba D, Kobliuk N, Svoboda J, Hladky R and Stych P** (2020) Sentinel-2 data in an evaluation of the impact of the disturbances on Forest vegetation. *Remote Sensing* 12(12), 1914. <https://doi.org/10.3390/rs12121914>.
- Liebold AM, McCullough DG, Blackburn LM, Frankel SJ, Von Holle B and Aukema JE** (2013) A highly aggregated geographical distribution of forest pest invasions in the USA. *Diversity and Distributions* 19(9), 1208–1216. <https://doi.org/10.1111/ddi.12112>.
- Lin T-Y, Goyal P, Girshick R, He K and Dollár P** (2018) Focal Loss for Dense Object Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 42(2), 318–327. <https://doi.org/10.48550/arXiv.1708.02002>
- Lin T-Y, Maire M, Belongie S, Bourdev L, Girshick R, Hays J, Perona P, Ramanan D, Zitnick CL and Dollár P** (2015) Microsoft COCO: Common Objects in Context. *ECCV* (5) (2014), 740–755. <https://doi.org/10.48550/arXiv.1405.0312>
- Lines E, Allen M, Cabo C, Calders K, Debus A, Grieve S, Miltiadou M, Noach A, Owen H and Puliti S** (2022) AI applications in forest monitoring need remote sensing benchmark datasets. *2022 IEEE International Conference on Big Data (Big Data)*, pp. 4528–4533. <https://doi.org/10.1109/BigData55660.2022.10020772>.
- Maxwell AE, Warner TA and Guillén LA** (2021a) Accuracy assessment in convolutional neural network-based deep learning remote sensing studies—Part 1: Literature review. *Remote Sensing* 13(13), 2450. <https://doi.org/10.3390/rs13132450>.
- Maxwell AE, Warner TA and Guillén LA** (2021b) Accuracy assessment in convolutional neural network-based deep learning remote sensing studies—Part 2: Recommendations and best practices. *Remote Sensing* 13(13), 2591. <https://doi.org/10.3390/rs13132591>.
- McMahon SM, Arellano G and Davies SJ** (2019) The importance and challenges of detecting changes in Forest mortality rates. *Ecosphere* 10(2). <https://doi.org/10.1002/ecs2.2615>.
- Moreno-Fernández D, Camarero JJ, García M, Lines ER, Sánchez-Dávila J, Tijerín J, Valeriano C, Viana-Soto A, Zavala MÁ and Ruiz-Benito P** (2022) The interplay of the tree and stand-level processes mediate drought-induced Forest dieback: Evidence from complementary remote sensing and tree-ring approaches. *Ecosystems* 25(8), 1738–1753. <https://doi.org/10.1007/s10021-022-00793-2>.
- Mu Y, Fujii Y, Takata D, Zheng B, Noshita K, Honda K, Ninomiya S and Guo W** (2018) Characterization of peach tree crown by using high-resolution images from an unmanned aerial vehicle. *Horticulture Research* 5, 74. <https://doi.org/10.1038/s41438-018-0097-z>.
- Mueller-Dombois D** (1988) Forest decline and dieback — A global ecological problem. *Trends in Ecology & Evolution* 3(11), 310–312. [https://doi.org/10.1016/0169-5347\(88\)90108-5](https://doi.org/10.1016/0169-5347(88)90108-5).
- Nesbit PR and Hugenholtz CH** (2019) Enhancing UAV–SfM 3D model accuracy in high-relief landscapes by incorporating oblique images. *Remote Sensing* 11(3), 239. <https://doi.org/10.3390/rs11030239>.
- Ng KM, Johari, J., Abdullah, S. A. C., Ahmad A and Laja BN** (2018) Performance evaluation of the RTK-GNSS navigating under different landscape. In *2018 18th International Conference on Control, Automation and Systems (ICCAS)*. Pyeongchang, South Korea: IEEE, 1424–1428. Available at <https://ieeexplore.ieee.org/document/8571729> (accessed 18 January 2024).
- Nguyen HT, Lopez Caceres ML, Moritake, K, Kentsch, S., Shu, H., & Diez Y** (2021) Individual sick fir tree (*Abies mariesii*) identification in insect infested forests by means of UAV images and deep learning [Number: 2 Publisher: Multidisciplinary Digital Publishing Institute]. *Remote Sensing* 13(2), 260. <https://doi.org/10.3390/rs13020260>.
- Ollinger SV** (2011) Sources of variability in canopy reflectance and the convergent properties of plants. *New Phytologist* 189(2), 375–394.
- Onishi M and Ise T** (2021) Explainable identification and mapping of trees using UAV RGB image and deep learning. *Scientific Reports* 11(1), 903. <https://doi.org/10.1038/s41598-020-79653-9>.
- OpenStreetMap Contributors** (2017) Planet dump Available at <https://planet.osm.org>.

- Puliti S, Pearse G, Surový P, Wallace L, Hollaus M, Wielgosz M and Astrup R** (2023) For-instance: A UAV laser scanning benchmark dataset for semantic and instance segmentation of individual trees.
- Reid AM, Chapman WK, Prescott CE and Nijland W** (2016) Using excess greenness and green chromatic coordinate colour indices from aerial images to assess lodgepole pine vigour, mortality and disease occurrence. *Forest Ecology and Management* 374, 146–153. <https://doi.org/10.1016/j.foreco.2016.05.006>.
- Ren S, He K, Girshick R and Sun J** (2016) Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 39 (2015), 1137–1149. <https://doi.org/10.48550/arXiv.1506.01497>.
- Richardson AD, Hufkens K, Milliman T, Aubrecht DM, Chen M, Gray JM, Johnston MR, Keenan TF, Klosterman ST, Kosmala M, Melaas EK, Friedl MA and Frohling S** (2018) Tracking vegetation phenology across diverse North American biomes using Pheno Camimagery. *Scientific Data* 5(1), 180028. <https://doi.org/10.1038/sdata.2018.28>.
- Roberts M, Gilligan CA, Kleczkowski A, Hanley N, Whalley AE and Healey JR** (2020) The effect of forest management options on Forest resilience to pathogens. *Frontiers in Forests and Global Change*, 3. <https://doi.org/10.3389/ffgc.2020.00007>.
- Safonova A, Tabik S, Alcaraz-Segura D, Rubtsov A, Maglins Y and Herrera F** (2019) Detection of fir trees (*Abies sibirica*) damaged by the bark beetle in unmanned aerial vehicle images with deep learning [number: 6 Publisher: Multidisciplinary digital publishing institute]. *Remote Sensing* 11(6), 643. <https://doi.org/10.3390/rs11060643>.
- Šandric I, Irimia R, Petropoulos GP, Anand A, Srivastava PK, Plešoianu A, Faraslis I, Stateras D and Kalivas D** (2022) Tree's Detection & Health's assessment from ultra-high resolution UAV imagery and deep learning. *Geocarto International* 37(25), 10459–10479. <https://doi.org/10.1080/10106049.2022.2036824>.
- Sani-Mohammed A, Yao W and Heurich M** (2022) Instance segmentation of standing dead trees in dense Forest from aerial imagery using deep learning. *ISPRS Open Journal of Photogrammetry and Remote Sensing* 6, 100024. <https://doi.org/10.1016/j.ophoto.2022.100024>.
- Schiefer F, Schmidlein S, Frick A, Frey J, Klinke R, Zielewska-Büttner K, Juntila S, Uhl A and Kattenborn T** (2023) UAV-based reference data for the prediction of fractional cover of standing deadwood from sentinel time series. *ISPRS Open Journal of Photogrammetry and Remote Sensing* 8, 100034. <https://doi.org/10.1016/j.ophoto.2023.100034>.
- Senf C, Buras A, Zang CS, Rammig A and Seidl R** (2020) Excess forest mortality is consistently linked to drought across Europe. *Nature Communications* 11(1), 6200. <https://doi.org/10.1038/s41467-020-19924-1>.
- Teskey R, Wertin T, Bauweraerts I, Amez M, McGuire MA and Steppe K** (2014) Responses of tree species to heat waves and extreme heat events. *Plant, Cell & Environment* 38(9), 1699–1712. <https://doi.org/10.1111/pce.12417>.
- Watson JEM, Evans T, Venter O, Williams B, Tulloch A, Stewart C, Thompson I, Ray JC, Murray K, Salazar A, McAlpine C, Potapov P, Walston J, Robinson JG, Painter M, Wilkie D, Filardi C, Laurance WF, Houghton RA, ... Lindenmayer D** (2018) The exceptional value of intact forest ecosystems. *Nature Ecology & Evolution* 2(4), 599–610. <https://doi.org/10.1038/s41559-018-0490-x>.
- Weinstein B, Marconi S, Bohlman S, Zare A and White E** (2019) Individual tree-crown detection in RGB imagery using semi-supervised deep learning neural networks. *Remote Sensing* 11(11), 1309. <https://doi.org/10.3390/rs11111309>.
- Yang M, Mou Y, Liu S, Meng Y, Liu Z, Li P, Xiang W, Zhou X and Peng C** (2022) Detecting and mapping tree crowns based on convolutional neural network and Google earth images. *International Journal of Applied Earth Observation and Geoinformation* 108, 102764. <https://doi.org/10.1016/j.jag.2022.102764>.
- Zhang Y, Ling F, Wang X, Foody GM, Boyd DS, Li X, Du Y and Atkinson PM** (2021) Tracking small-scale tropical forest disturbances: Fusing the Landsat and Sentinel-2 data record. *Remote Sensing of Environment* 261, 112470. <https://doi.org/10.1016/j.rse.2021.112470>.

Appendix

A. Residual Plots

A plot of the residuals corresponding to the OLS of Figure 3, as a function of centroid-trunk match distance, can be seen in Figure 5. The magnitude of the residuals does not seem to increase with match distance for distances less than 1, as per Algorithm 1. We suggest, for this reason, that the matching procedure outlined in Algorithm 1 is not the source of any variation in calculated GCC. This is despite the observation that *P. pinea* trunks often display curvature where the crown center does not coincide with the trunk position. The linear model overestimates GCC for larger in situ defoliation estimates. There is no reason a priori to expect GCC to be a linear function of the in situ visual defoliation estimates.

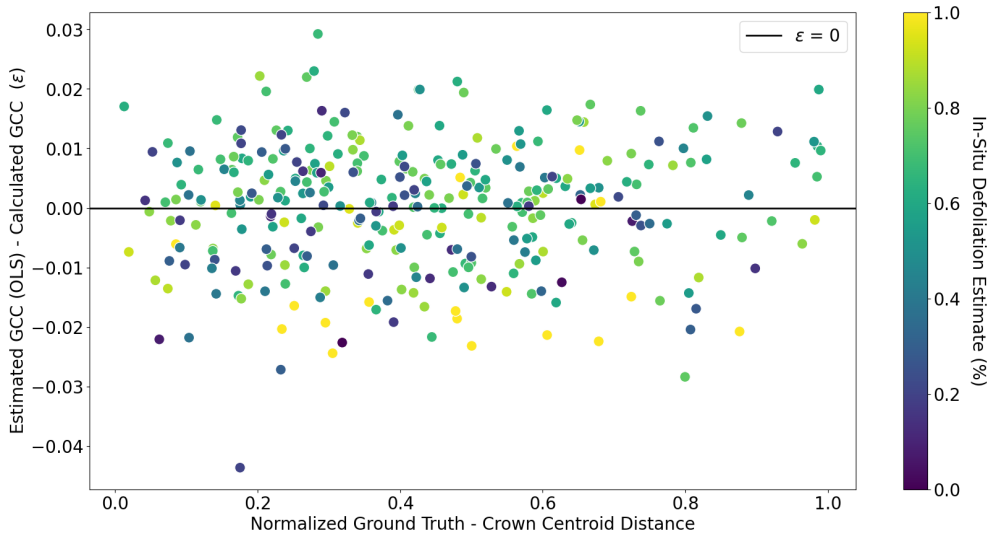


Figure 5. Plot of GCC residuals from OLS versus Distance from inventory trunk location to matched polygon.

B. Crown Matching

We outline the iterative approach used to match ITCs with ground-surveyed trunk locations in Algorithm 1. Refer to <https://numpy.org/doc/> for further details regarding specific functions (Harris et al., 2020).

Algorithm 1 NumPy-like pseudocode for crown matching

```
# pred_polys—list of predicted polygons
# n_trunks—number of ground-truth trunk locations
# n_poly—number of ITCs

dist ← np.empty(n_trunks, n_poly)

while dist.shape[0] > 0 do
  i, j ← np.where(dist == np.min(dist))
  gcc ← calculate_gcc(pred_polys[j])
  dist ← np.delete(dist, i, axis=0)
  dist ← np.delete(dist, j, axis=1)
  pred_polys.pop(j)
end while
```
