

RESEARCH

Open Access



X-ray microtomography and linear discriminant analysis enable detection of embolism-related acoustic emissions

Niels J. F. De Baerdemaeker¹, Michiel Stock², Jan Van den Bulcke^{3,4}, Bernard De Baets², Luc Van Hoorebeke^{4,5} and Kathy Steppe^{1*} 

Abstract

Background: Acoustic emission (AE) sensing is in use since the late 1960s in drought-induced embolism research as a non-invasive and continuous method. It is very well suited to assess a plant's vulnerability to dehydration. Over the last couple of years, AE sensing has further improved due to progress in AE sensors, data acquisition methods and analysis systems. Despite these recent advances, it is still challenging to detect drought-induced embolism events in the AE sources registered by the sensors during dehydration, which sometimes questions the quantitative potential of AE sensing.

Results: In quest of a method to separate embolism-related AE signals from other dehydration-related signals, a 2-year-old potted *Fraxinus excelsior* L. tree was subjected to a drought experiment. Embolism formation was acoustically measured with two broadband point-contact AE sensors while simultaneously being visualized by X-ray computed microtomography (μ CT). A machine learning method was used to link visually detected embolism formation by μ CT with corresponding AE signals. Specifically, applying linear discriminant analysis (LDA) on the six AE waveform parameters amplitude, counts, duration, signal strength, absolute energy and partial power in the range 100–200 kHz resulted in an embolism-related acoustic vulnerability curve (VC_{AE-E}) better resembling the standard μ CT VC (VC_{CT}), both in time and in absolute number of embolized vessels. Interestingly, the unfiltered acoustic vulnerability curve (VC_{AE}) also closely resembled VC_{CT} , indicating that VCs constructed from all registered AE signals did not compromise the quantitative interpretation of the species' vulnerability to drought-induced embolism formation.

Conclusion: Although machine learning could detect similar numbers of embolism-related AE as μ CT, there still is insufficient model-based evidence to conclusively attribute these signals to embolism events. Future research should therefore focus on similar experiments with more in-depth analysis of acoustic waveforms, as well as explore the possibility of Fast Fourier transformation (FFT) to remove non-embolism-related AE signals.

Keywords: Drought-induced embolism formation, Acoustic emissions, Machine learning, Linear discriminant analysis, X-ray computed microtomography, *Fraxinus excelsior* L.

*Correspondence: kathy.steppe@UGent.be

¹ Laboratory of Plant Ecology, Department of Plants and Crops, Faculty of Bioscience Engineering, Ghent University, Coupure links 653, 9000 Ghent, Belgium

Full list of author information is available at the end of the article



© The Author(s) 2019. This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Background

The strategy adopted by vascular plants to absorb and transport water through their conducting xylem tissue during transpiration could be described as brilliant, but lazy. Most other higher species use energy to transport sufficient amounts of water to sustain their metabolism. In contrast, plants exploit the gradient in water potential, from less negative to more negative, to enable water flow through their vascular tissue. Consequently, xylem vessels and tracheids are well adapted to withstand negative water potentials [1, 2], but in drying soil and/or atmospheric conditions, this passive strategy involves the risk of embolism formation, impairing the xylem conducting system [3]. Embolism causes water to be pulled from the vessel or tracheid and being replaced with air, resulting in the formation of emboli [4].

Embolism formation is accompanied by a sudden and rapid release in tension, producing energy waves detectable as acoustic emissions (AE) [5–7]. Milburn and Johnson [8] were the first to register acoustic emissions from a dehydrating leaf petiole, and linked these signals to embolism formation. The first commercially available AE counter to study AEs in wood was developed by Tyree and Sperry [9], and both AE sensors and acquisition systems have been greatly enhanced since then, allowing to record and analyze time, parameter and waveform data of each AE event [10–12]. Frequently used AE waveform parameters to study wood properties are peak amplitude, duration and energy [13], and when adding signal strength and partial power in the range 100–200 kHz [10, 12, 14], they have also been classified as important parameters related to embolism formation.

To quantify drought-induced embolism formation, a xylem vulnerability curve (VC) is typically constructed, relating loss of xylem water transport capacity to xylem water potential [15]. The standard hydraulic method to construct VCs is destructive and discontinuous, creating single VCs from extensive sampling [16]. In contrast, the continuous and non-invasive nature of the AE method allows (i) to develop sample-specific VCs [17], (ii) to characterize anatomical differences in thin drying wood sections [18], and (iii) to be used in outdoor applications [19]. This method can therefore be recommended as a valuable diagnostic tool to assess drought-induced embolism formation. However, aside from the findings reported by Tyree et al. [6] and Lewis [20], the assumption that all recorded AEs represent single embolism events [21] is in most cases invalid [11, 22]. Detecting embolism-related AE is therefore deemed necessary but remains challenging as a wide variety of AE sources is registered during dehydration, including water loss from other xylem elements such as fibers, tracheids and parenchyma [23, 24], mechanical strains [25, 26], dehydration

of bark tissue [18], nanobubble formation [27], Haines jumps [12], and macro- and micro-crack formation [12, 28]. Because AEs do not quantify loss of hydraulic conductivity in the same way as the hydraulic method [15], the AE method has also been cited to be more qualitative than quantitative [16], making detection of embolism-related AE from total measured AE even more important. In order to unravel the link between acoustic emissions and drought-induced embolism formation, additional techniques are required that (i) can visualize the embolization process in order to delimit registered AEs within embolization time intervals, and that (ii) can deduce the acoustic characteristics of the signals in order to link AE to embolism formation.

In recent years, X-ray computed microtomography (μ CT) has evolved from a niche technology [29] into an accessible reference visualization technique in drought-induced embolism research [30–32]. Reconstructed μ CT images visualize the spatial distribution of the linear attenuation coefficient μ of the X-rays within the sample, implying that water-filled vessels appear as grey (high μ) and air-filled vessels as black (low μ) on the image [33]. The μ CT method is therefore capable to non-invasively and continuously visualize embolism formation. It has significantly increased our understanding of drought-induced embolism formation, including (i) resolving issues related to artificial VCs of long-vesselled species constructed with indirect methods [30], (ii) providing evidence on the controversy of embolism refilling under tension [34, 35], and (iii) when vessel dimensions are extracted from which theoretical hydraulic conductance can be calculated, translating qualitative percentage cavitation into quantitative percentage loss of conductance (PLC) [32].

Machine learning has been proven to be an important diagnostic tool for processing, analyzing and deducing biological data [36, 37]. Machine learning is typically divided into unsupervised and supervised learning, where the former is used to model the data structure to identify hidden patterns in unlabeled datasets based on input variables [38], while the latter requires output variables to find a mapping function so that from new input data, output variables can be effectively predicted [38]. Unsupervised learning often starts with an exploratory procedure, like principal component analysis (PCA), to explain the data variance. Correlation matrix plots (i.e., correlograms) are complimentary tools to determine which variables are strongly positively or negatively correlated to one another. Histograms and receiver operating characteristic (ROC) curves can be used on PCA and correlogram results to visualize the underlying data distribution [39]. Supervised learning methods, such as linear discriminant analysis (LDA), incorporate output

variables as labels during data analysis, which is more beneficial when specific patterns need to be extracted. LDA is a generative model [40], with the benefit that the model can be fitted efficiently, because the obtained parameters are directly computed from simple statistics such as average, variance and covariance.

We therefore conducted an experiment consisting of a combination of repeated 30-min rotation μ CT imaging (6×4 -min scans + 6-min break) with supervised LDA during continuous AE registration (with two AE sensors, one 14.0 cm (AE₁) and one 24.3 cm (AE₂) downstream from the scanning position to compare signal output) in a stem of a progressively dehydrating 2-year-old potted *Fraxinus excelsior* L. tree. Histograms of both embolism and non-embolism datasets, classified by μ CT results, and ROC curves of the entire dataset were determined using the AE waveform parameters amplitude, counts, duration, signal strength, absolute energy and partial power in the range of 100–200 kHz to demarcate possible thresholds for these parameters, because these were, according to PCA, most suited for detecting embolism-related AE. The hypothesis is that if only embolism-related AE signals, instead of all registered AE signals, are used to construct acoustic VCs, these will correspond better with the reference VC_{CT} and, hence, improve the characterization of drought-induced vulnerability to embolism formation of species when using acoustic sensors.

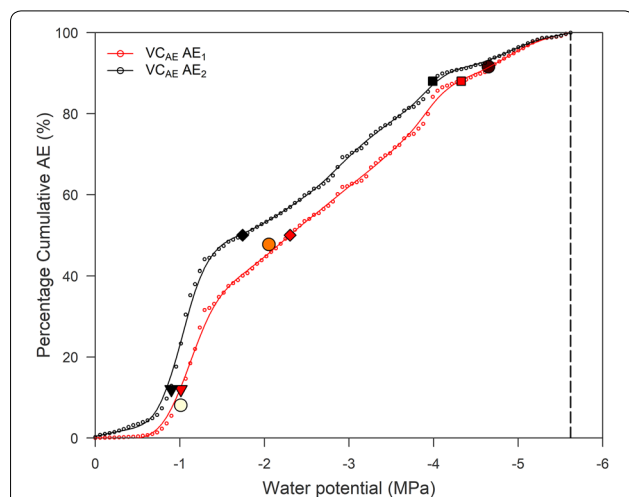


Fig. 1 Acoustic vulnerability curve (VC_{AE}) of *Fraxinus excelsior* L. during dehydration registered by sensor AE₁ (red) and sensor AE₂ (black). Vulnerability values AE₁₂ (filled inverted triangle), AE₅₀ (filled diamond), AE₈₈ (filled square) and AE₁₀₀ (dashed line) are indicated. LDA model outputs for embolism datasets 8, 69 and 120 (open circle) (Table 2), extracted from Fig. 4, are also shown

Table 1 Vulnerability to drought-induced embolism values in *F. excelsior* L. derived from the acoustic vulnerability curve (VC_{AE}) measured with sensor AE₁ and sensor AE₂

Vulnerability value (MPa)	AE ₁	AE ₂
AE ₁₂	-1.02	-0.88
AE ₅₀	-2.30	-1.73
AE ₈₈	-4.32	-3.98
AE ₁₀₀	-5.62	-5.62

Water potential at 12, 50, 88 and 100% cumulative acoustic emissions in *F. excelsior* L. (AE₁₂, AE₅₀, AE₈₈, AE₁₀₀) of sensor AE₁ and sensor AE₂

Results

Unfiltered VC_{AE}

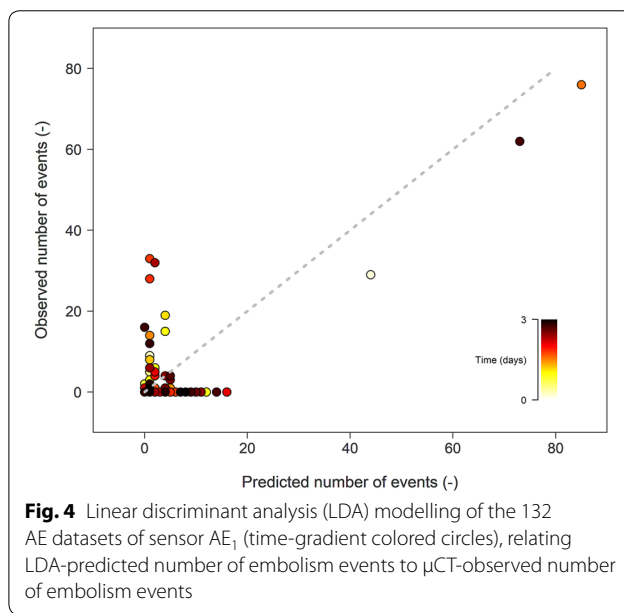
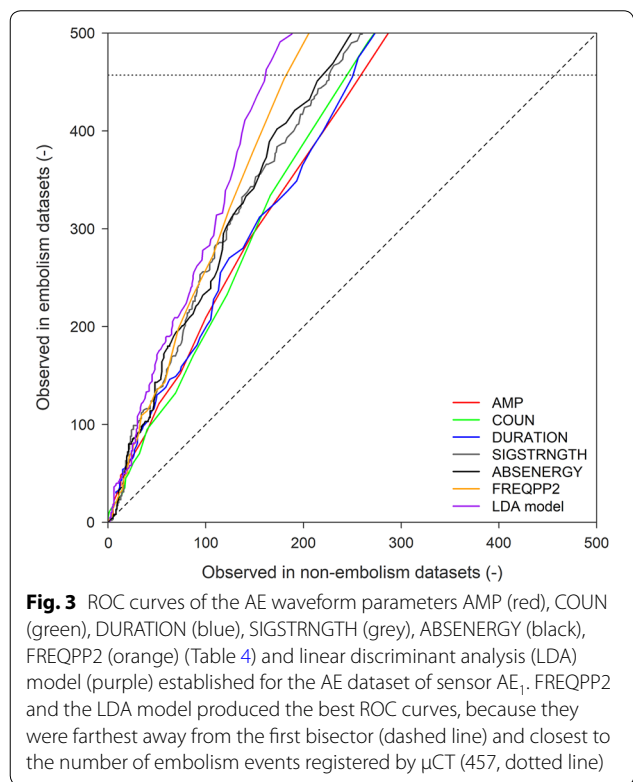
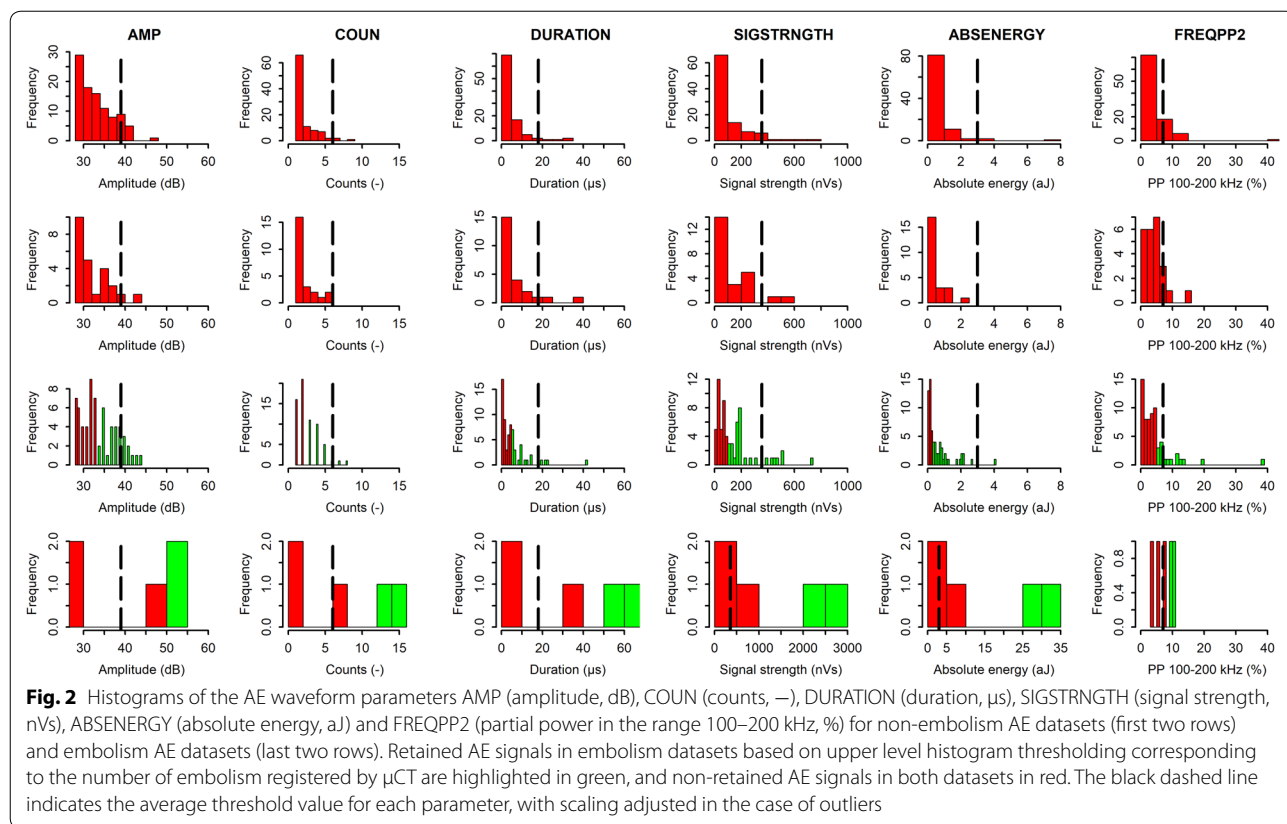
The VC_{AE}s constructed from all AE signals measured by sensors AE₁ and AE₂ were similar in shape, with AE₂ registering almost four times more signals (Fig. 1). Values obtained from VC_{AE2}, especially AE₅₀, were shifted to higher values of vulnerability to drought-induced embolism formation (Table 1).

Histograms and ROC curves

To better link machine learning results of AE signals with μ CT, analysis was conducted on the AE sensor closest to the μ CT scanning point (AE₁). Histogram plots of the AE waveform parameters peak amplitude (AMP), counts from peak amplitude (COUN), duration from peak amplitude (DURATION), signal strength (SIGSTRNGTH), absolute energy (ABSENERGY), and partial power in the frequency range 100–200 kHz (FREQPP2) (Table 4) in both embolism (green) and non-embolism (red) AE datasets showed that their upper level threshold values were most often associated with embolism events recorded by μ CT (Fig. 2).

However, histogram results of AE₁ signals illustrated that thresholds on AE parameters are insufficient to readily distinguish embolism signals (green) from non-embolism signals (red), as the thresholds sometimes included AE signals from non-embolism datasets and sometimes neglected AE signals from embolism datasets (Fig. 2). Thus, a static threshold on these six AE waveform parameters lacked sufficient accuracy.

A static threshold based on the ROC curves of the AE waveform parameters AMP, COUN, DURATION, SIGSTRNGTH, ABSENERGY, and FREQPP2 (Table 4) for the entire AE₁ dataset could not be determined, because all curves had a similar shape, showing no distinct deflection points (Fig. 3). Because the FREQPP2 ROC curve is farthest away from the first bisector, demarcating thresholds on this parameter would be most successful to detect embolism-related AE (Fig. 3), but the lack of a



clear deflection point results in too much uncertainty, suggesting that a FREQPP2 static threshold at the 457 dotted line is insufficient to judge whether retained signals are related to embolism formation or not.

LDA model

To determine which AE₁ signals were related to embolism events, the LDA model was used with the six AE waveform parameters AMP, COUN, DURATION, SIG-STRNGTH, ABSENERGY, and FREQPP2 from the 132 AE₁ datasets and the corresponding μ CT embolism events as labels (Fig. 4). These results are obtained by using X-fold cross-validation. For each AE signal in the different datasets, LDA assigned a probability between 0 and 1. Per dataset, these probabilities were cumulated (predicted number of events) and compared to the number of embolism events detected by μ CT (observed number of events).

Linear discriminant analysis yielded mixed results in detecting embolism-related AE: some AE datasets resulted in a close match between predicted and observed events, while other ones did not. The difference between predicted and observed events increased in function of dehydration time (Table 2).

Linear discriminant analysis shows promising results for larger AE datasets, which included a significant number of embolism events (datasets 8, 69 and 120; Table 2). These datasets were close to the first bisector (dashed line, Fig. 4), and corresponded well with the unfiltered VC_{AE} of sensor AE₁ (Fig. 1). Furthermore, to detect embolism-related AEs, the ROC curve of LDA scored best compared to the ROC curve of the other six AE waveform parameters (Fig. 3).

Despite these positive indications, LDA probabilities attributed to AE signals were generally low (not higher than 0.4). This suggests that LDA might be a first promising step towards detecting embolism-related AE from an acquired AE dataset, yet not accurate enough.

Embolism-related vulnerability

Comparing the embolism-related acoustic VC with the standard μ CT VC (Fig. 5a) illustrates that LDA performs only a little better than the full acoustic VC in detecting embolism-related acoustic emissions mainly because of the over- and underestimation of LDA compared to the visually detected number of embolism events with μ CT (Table 2). The absolute difference in percentage embolism formation to the reference VC_{CT} was calculated for VC_{AE} and VC_{AE-E'} and was over the entire dehydration period slightly lower for the latter, mainly resulting in an overestimation of the number of embolism events compared to μ CT (Fig. 5b).

Linear discriminant analysis detected a total of 518 embolism-related AE signals at the end of dehydration compared to 457 visually detected μ CT embolism events (Table 2). Interestingly, the unfiltered VC_{AE} of

Table 2 Number of embolism events according to X-ray computed microtomography (μ CT) and linear discriminant analysis (LDA) model for the 132 AE datasets of sensor AE₁

AE dataset	μ CT	LDA	AE dataset	μ CT	LDA
1–2	0–2	0–0	67–68	1–0	1–0
3–4	0–2	0–0	69–70	76–0	85–5
5–6	0–1	1–0	71–72	1–0	1–1
7–8	0–29	0–44	73–74	1–0	1–1
9–10	0–6	10–1	75–76	1–0	2–1
11–12	0–9	2–1	77–78	1–0	1–6
13–14	0–3	0–1	79–80	33–0	1–2
15–16	0–1	9–0	81–82	28–0	1–5
17–18	0–1	1–0	83–84	1–0	1–3
19–20	3–0	1–0	85–86	4–0	2–3
21–22	1–0	0–1	87–88	1–0	1–1
23–24	5–0	1–1	89–90	1–0	1–1
25–26	1–0	1–2	91–92	1–0	1–16
27–28	1–0	1–5	93–94	5–0	2–3
29–30	1–0	1–1	95–96	1–0	0–2
31–32	2–0	1–1	97–98	1–0	1–3
33–34	1–0	1–2	99–100	1–0	1–14
35–36	2–0	0–1	101–102	1–0	1–0
37–38	1–0	0–3	103–104	32–1	2–4
39–40	2–0	1–2	105–106	0–6	2–1
41–42	6–0	2–12	107–108	0–4	1–4
43–44	15–0	4–1	109–110	0–1	11–1
45–46	3–0	1–1	111–112	3–0	5–4
47–48	1–0	5–11	113–114	1–0	4–10
49–50	1–19	1–4	115–116	2–0	1–1
51–52	0–2	4–1	117–118	1–4	1–5
53–54	1–0	4–1	119–120	0–62	14–73
55–56	8–0	1–1	121–122	0–16	9–0
57–58	1–0	5–0	123–124	12–0	1–0
59–60	1–0	1–2	125–126	2–0	1–4
61–62	1–0	1–1	127–128	1–0	1–1
63–64	1–0	5–1	129–130	1–0	1–8
65–66	14–0	1–1	131–132	1–0	1–7

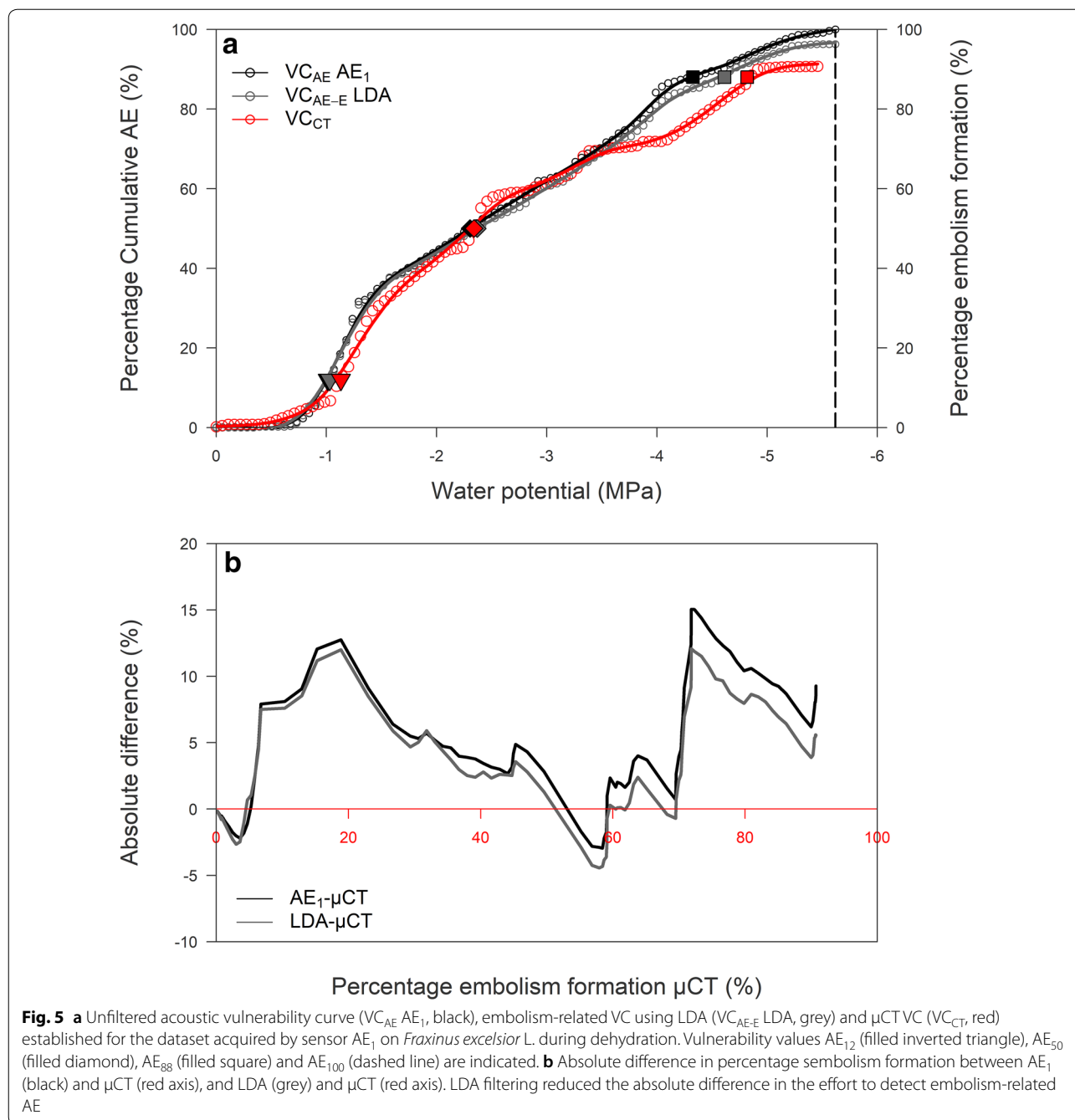
Italic values illustrate sufficient embolism-related AE detection by LDA in accordance with the number of embolism events by μ CT. For every AE dataset, an LDA model was trained on all remaining datasets

sensor AE₁, though registering a total of 25,901 AE signals at the end of dehydration, resulted in an AE₅₀ that only slightly underestimated CT₅₀ with 2% (Table 3; Fig. 5a, b). Divergence in vulnerability was mainly concentrated towards the end of the curve.

Discussion

Detection of embolism-related AE to improve VC_{AE}

Applying LDA on the six AE waveform parameters amplitude, counts, duration, signal strength, absolute



energy and partial power in the range 100–200 kHz and μCT scanning to generate model labels was successful at detecting embolism-related AE events from the dataset both in number and over time (Table 2; Fig. 5). The resulting VC_{AE-E} closely corresponded to VC_{CT} (Table 3; Fig. 5), which is considered the reference to quantify a species' vulnerability to drought-induced embolism formation [30]. Hydraulic P_{50} (i.e., xylem water potential at 50% loss of hydraulic conductivity) derived from the

VC established by Lemoine et al. [41] for 1–3 year-old branches of well-watered 15–20 year-old *Fraxinus excelsior* trees was equal to -3 MPa, which agreed well with the CT_{50} -value of our 2-year-old stem (Table 3). The small difference between the branches and our stem might be attributed to the different techniques used (hydraulic vs μCT). Another reason could be linked to the hydraulic segmentation hypothesis that postulates that angiosperm trunks/stems (30–40 cm diameter) are 0.7–1.8 MPa more

Table 3 Vulnerability-to-cavitation values for *F. excelsior* L. derived from the acoustic vulnerability curve (VC_{AE}) of sensor AE_1 , the embolism-related VC (VC_{AE-E}) based on LDA and the X-ray computed microtomography VC (VC_{CT})

Vulnerability value (MPa)	VC_{AE} AE_1	VC_{AE-E}	VC_{CT}
AE_{12}/CT_{12}	-1.02	-1.03	-1.13
AE_{50}/CT_{50}	-2.30	-2.37	-2.34
AE_{88}/CT_{88}	-4.32	-4.61	-4.82

Water potential at 12, 50 and 88% cumulative acoustic emissions (AE_{12} , AE_{50} and AE_{88}), and 12, 50 and 88% μ CT detected embolism formation (CT_{12} , CT_{50} and CT_{88}) for *F. excelsior* L. of sensor AE_1 , LDA and μ CT

vulnerable than branches (8–14 cm diameter) [42], but is less likely because of the close resemblance in our stem diameter to the investigated branch diameter of Lemoine et al. [41].

The close agreement between LDA VC_{AE-E} and VC_{CT} does however not imply that embolism-related AE signals can be readily distinguished from other AE sources, because the probability of the LDA model attributed to each AE signal was never higher than 0.4. Cumulative probabilities of all AE signals in embolism and non-embolism datasets were indicative for the number of μ CT embolism events (Table 2), but cumulative probabilities of only AE signals detected by LDA (i.e., signals with highest probabilities according to LDA) were not, resulting in only 16 embolism events instead of the detected 457. Uncertainty remains whether AE signals with high values of amplitude, counts, duration, signal strength, absolute energy and partial power in the range 100–200 kHz are indeed typical characteristics of drought-induced embolism events, because these signals also occur in non-embolism AE datasets (Figs. 2, 3, 5b).

Acoustic emissions waveform parameters amplitude, duration, energy, signal strength and partial power in the range 100–200 kHz have previously been associated with embolism formation [10, 12, 14]. We showed that detection based on static thresholds for these parameters did not work, and included non-embolism AE sources (Fig. 2) [11]. This failure in using static thresholds can be attributed to species-specific AE attenuation in wood [13], which is known to decrease with ongoing dehydration [43]. AE sources registered at the start of dehydration will be more attenuated (due to the availability of more water) than at the end, and this attenuation factor is not taken into account when static thresholds on AE waveform parameters are used. Setting the threshold too low might include non-embolism-related AE measured at the end of dehydration, and setting the threshold too high might neglect the embolism events at the start

of dehydration (Figs. 2, 3). Determination of dynamic thresholds that vary with time and incorporate changing attenuation with dehydration might enable separation of embolism from other AE-related sources during dehydration.

The unfiltered VC_{AE} , though constructed from 25,901 cumulated AE signals, was closely related to VC_{CT} (Fig. 5a), both in magnitude and derived vulnerability characteristics (Table 3; Fig. 5b). This offsets the often-misplaced perception of AE being more qualitative, because of the excess of AE signals over embolism events registered during dehydration [16]. Compared to VC_{CT} , this unfiltered VC_{AE} resulted in an AE_{50} which underestimated CT_{50} with only 2%. Incorporation of all AE signals for *F. excelsior* in a VC and using the VC_{AE} endpoint determination of Vergeynst et al. [12] produced a quantitative instead of a qualitative VC. This suggests that all AE signals registered during dehydration can be used to reliably assess drought vulnerability when compared to filtered AE methods [10–12] or reference techniques (hydraulic and μ CT) [6, 7, 14, 44, 45].

Significance of sensor installation to detect embolism-related AE

To reduce the number of AE signals not originating from embolizing conducting elements, sample length (with respect to maximum vessel length) and position of the sensor must be well-conceived [46]. Because *F. excelsior* was dehydrated by exposing the root system, and not by cutting the stem (classic way), maximum vessel length was not an issue in our study, and also cutting artifacts were avoided [16]. However, sensor installation did affect the number of registered AE signals (i.e., 25,901 for AE_1 vs. 90,416 for AE_2) and shifted VC_{AE2} to a slightly higher vulnerability to drought-induced embolism formation (Table 1; Fig. 1). The difference in AE registration was attributed to the installation position of sensor AE_2 , which was just below a leafy non-lignified side branch (Fig. 7), and closer to the tree's foliage than sensor AE_1 (i.e., 22.5 cm for AE_1 vs. 11.7 cm for AE_2). With the frequency of acoustic waves changing on their path through the wood towards the sensor [11, 43], less attenuation of AE sources originating from dehydrating leaves and side branches occurred in AE_2 , resulting in a higher number of detected AE signals above the noise threshold of 28 dB. The higher noise to embolism ratio of sensor AE_2 alongside the closer position of sensor AE_1 to the μ CT scanning point further explains why results of sensor AE_1 were used for machine learning analysis.

The high attenuation factor of wood, especially at the start of dehydration, significantly influences amplitude, frequency, shape-related and energy-related characteristics of registered AE signals [13]. If the effect of distance

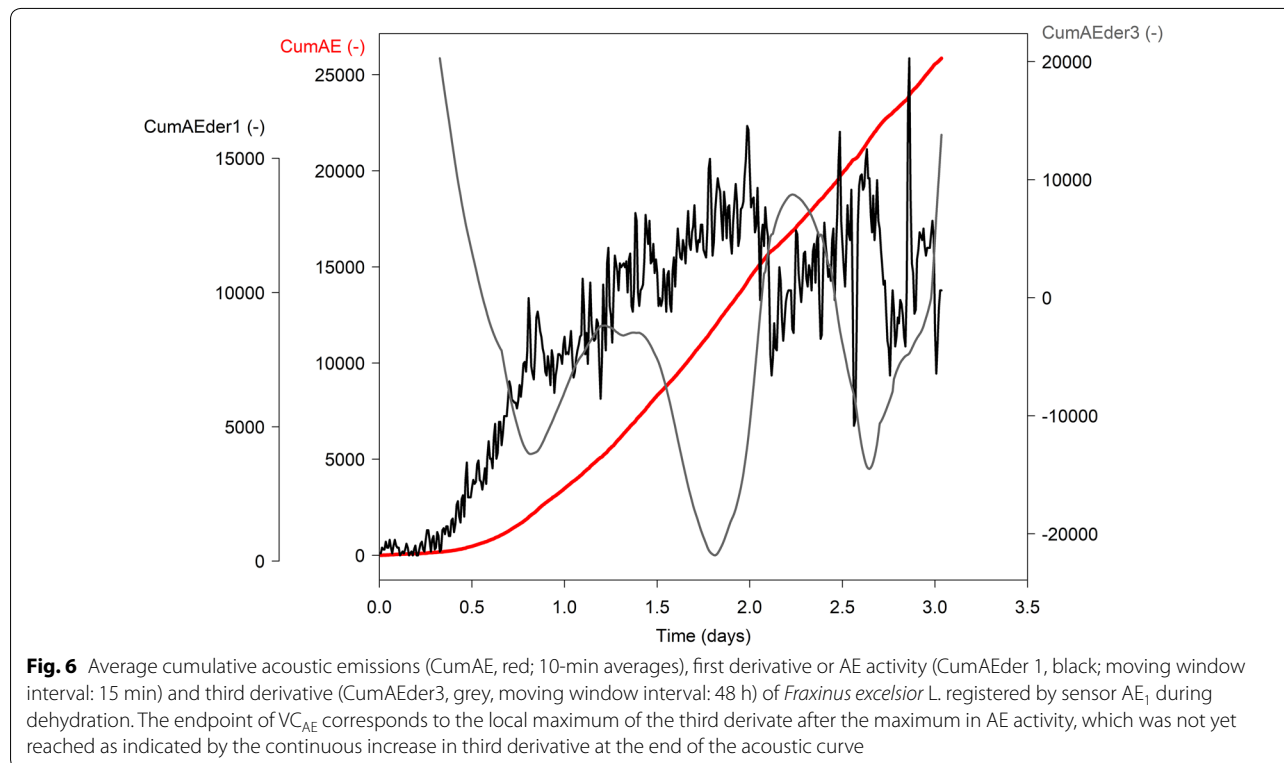
to the AE sensor is not quantified, then any classification approach can incur a large error. It is recommended that the effect of attenuation and its evolution during dehydration is quantified to reduce these errors as much as possible. This requires a mapping of the localization area of AE signals originating from the sample, which can be achieved by installing multiple AE sensors at known distances alongside the sample [13].

Despite the 71% difference in acquired AE signals between sensors AE₁ and AE₂, the VC_{AE}s were similar in shape (Table 1; Fig. 1). Nonetheless, AE₂ was less suited to identify the embolism-related signals, because to successfully detect embolism-related AE via LDA, the embolism to non-embolism signal ratio must be maximized. The embolism to non-embolism signal ratio was already low in sensor AE₁ (25,444 non-embolism signals vs. 457 embolism signals), and increased dramatically in AE₂, explaining why resulting histograms, ROC curves and LDA of sensor AE₁ (Figs. 3, 4) showed difficulties in distinguishing embolism from other AE sources. Sensors should therefore be installed at a sufficiently large distance from the leaves, and in case cut branches or stems are used, sufficiently far from the open cut end of the sample.

Maximum in third derivative to define the VC_{AE} endpoint

The μ CT image taken at the end of the experiment (Fig. 11b) showed that 9% of the vessels were not

embolized. The AE datasets therefore did not include the true VC_{AE} endpoint, and all registered signals were used to construct VC_{AE}, and to detect embolism-related AE signals. In general, the VC_{AE} endpoint is defined by the local maximum of the third derivative of cumulative AE [12]. Because AEs are still recorded after full embolism formation due to a variety of other AE sources related to dehydration [12, 18, 23–28], defining the endpoint of VC_{AE} remains a difficult and challenging task [47], but is crucial to derive physiologically-meaningful vulnerability characteristics. Vergeynst et al. [12] explained in their study that the time of reaching the maximum in the first derivative or AE activity can be used to define the time interval for calculating the third derivative (Fig. 6). As maximum AE₁ activity occurred around 2 days, the third derivative was calculated with a time interval of 48 h. Because the resulting third derivative after the maximum in AE activity kept increasing, the local maximum was never reached for AE₁. This agrees with the μ CT results, which showed that 9% of the vessels were still functional when finishing the last CT-scan, and hence indirectly supports the VC_{AE} endpoint determination described by Vergeynst et al. [12]. Furthermore, xylem water potential registered at maximum AE activity (AE₅₀ = -2.58 MPa) coincided with CT₅₀ (Table 3) [see 44], and further supports the soundness of Vergeynst et al.'s [12] endpoint determination for VC_{AE}.



Future perspectives for embolism-related AE detection

Detection of embolism-related signals from an AE dataset based on LDA modelling using the parameters amplitude, counts, duration, signal strength, absolute energy and partial power in the range 100–200 kHz was promising in the sense that the resulting VC_{AE-E} closely corresponded to the reference VC_{CT} in *F. excelsior* L. (Table 3; Fig. 5), and that the amount of AE signals to construct the VC were efficiently reduced (from 25,901 to 518 signals). However, the low embolism probabilities attributed by LDA to AE signals, as well as the time divergence in embolism formation detection (microseconds for AE vs 4 min for μ CT), hamper the detection of distinct embolism-related AEs from acquired datasets. Multiple combined AE- μ CT experiments on similar samples (age, species, treatment) can provide the necessary training datasets for LDA to better distinguish embolism from non-embolism AE sources, in order, for LDA, to be able to detect embolism-related AE, independent of μ CT measurements, even for species for which no previous training datasets were acquired.

The non-invasive and continuous nature of AE sensors can also be applied to detect other physiologically meaningful AE sources from a dataset. When acoustically measuring and continuous scanning intact and well-watered trees, AE originating from shrinkage and water loss of fibers, tracheids and parenchyma can be captured [12, 23, 24]. Given the theory of water transport dynamics in plants, this shrinkage pattern occurs on a daily basis in well-watered plants, as a result of the time-lag that exists between foliar transpiration and root water uptake [48–50]. Using μ CT scanning to verify that embolism formation does not occur, parameters of the registered AE signals can be analyzed and classified as non-embolism signals. This subset of AE signals could then be removed from acquired AE datasets, hence increasing the embolism to non-embolism signal ratio, and increasing the efficiency of supervised machine learning tools such as LDA to detect embolism-related AE.

Conclusion

Ever since their first use in drought vulnerability research, acoustic emissions have been considered as interesting but unrefined to determine drought-induced embolism formation. The surplus of AE signals registered during dehydration not originating from embolism formation hinders correct quantitative assessment. Utilizing machine learning together with recorded embolism events by μ CT scanning was proposed as a new method to detect embolism-related AE from an AE dataset gathered in a 2-year-old *F. excelsior* L. tree during progressive dehydration. LDA modelling based on the parameters amplitude, counts, duration, signal strength, absolute

energy and partial power in the range 100–200 kHz was found sufficient to detect embolism-related AE probabilities that corresponded well with the μ CT reference ones, but retained signals were still not easy distinguishable from other AE sources. Interestingly, the unfiltered acoustic VC resulted in vulnerability values that were in close agreement to the ones derived from the μ CT VC, hence illustrating for this 2-year-old *F. excelsior* L. tree that unfiltered AE with the third derivative end point determination technique is accurate to determine its vulnerability to drought-induced embolism formation. Future research can still aim at a more in-depth analysis of acoustic waveforms and parameters associated with embolism formation to develop post-processing machine learning tools or state-of-the-art AE sensors that can efficiently filter embolism-related AE signals, without the aid of μ CT. This will further promote the AE method as a reliable and quantitative, powerful diagnostic tool in future drought stress experiments.

Methods

Plant material and experimental setup

Ten 2-year-old *Fraxinus excelsior* L. trees were grown in 3 L pots containing a soil mixture of peat litter, sand and calcium-magnesium based fertilizers in the greenhouse facilities of Ghent University (51° 03' 10.3" N latitude; 3° 42' 32.3" E longitude). Trees were grown under well-watered conditions for 2.5 months (from DOY 68 to DOY 142) during the 2017 growing season. On DOY 139, the tree with the straightest stem was selected for the measurement campaign, and replanted in a custom-built holder designed to keep the tree straight, centered, and tightly fixed during X-ray computed microtomography (μ CT) scanning (Fig. 7; see Additional file 1). The tube enclosing the part of the tree that was scanned was made of carbon fiber (CarbonWinkel.nl, Tilburg, The Netherlands). On DOY 142, the tree was transported to the UGent Centre for X-ray Tomography (UGCT, <https://www.ugct.ugent.be>), Belgium (51° 01' 25.7" N latitude; 3° 44' 26.2" E longitude), where the dehydration experiment took place from DOY 142 till DOY 145.

The tree was first removed from the custom-built holder to wash off the soil mixture, exposing the roots to speed up dehydration during scanning. The tree was re-inserted into the holder, and the loss in root anchoring countered by filling the excess room surrounding the roots with packaging foam. The tree was equipped with two broadband point-contact AE sensors with a flat frequency response between 20 and 1000 kHz (KRNBB-PC, KRN Services, Richland, WA, USA), at a respective distance of 14.0 cm (AE_1) and 24.3 cm (AE_2) downstream from the scanning position. The diameter of the tree, measured with an electronic caliper, was 6.6 mm at the

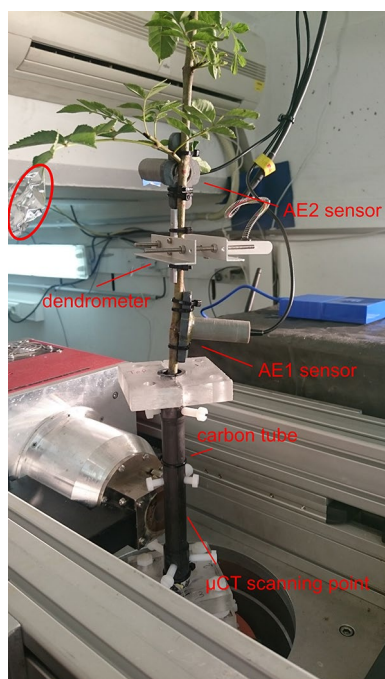


Fig. 7 Experimental setup of the *Fraxinus excelsior* L. tree in the environmental μ CT (EMCT) scanner. The EMCT continuously rotates around the stationary tree without the risk for twisting and winding of sensor cables. The carbon fiber tube was designed to ensure stable and centered positioning of the tree during μ CT scanning. The tree is equipped with two broadband point-contact AE sensors in a pvc spring-containing holder to continuously register AEs from the progressively dehydrating xylem (AE₁, AE₂), and a point dendrometer to continuously register xylem shrinkage. The red circle indicates an aluminum enclosed leaf used for pressure chamber measurements to determine xylem water potential

AE1 sensor, and 6.3 mm at the AE₂ sensor. At the position of the AE sensors, a section of bark (0.5×1.5 cm) was removed with a scalpel to expose the xylem, ensuring a better acoustic coupling with AEs originating from embolizing vessels [10]. To seal the wound and ensure good acoustic coupling, a droplet of vacuum grease (High-Vacuum Grease, Dow Corning, Seneffe, Belgium) was applied between sensor tip and xylem [45]. A compression spring (D22050, Tevema, Amsterdam, The Netherlands) in a small pvc tube was used to press the AE sensors against the xylem. To monitor xylem shrinkage, an additional section of bark (0.5×1.5 cm) was removed between the two AE sensors, the wound was sealed with petroleum jelly to prevent evaporation, the initial diameter (6.4 mm) was measured and the point dendrometer (DD-S, Ecomatik, Dachau, Germany) was installed at a distance of 19.2 cm downstream from the scanning position (Fig. 7; see Additional file 1).

The equipped tree was mounted on the z-stage of the Environmental μ CT scanner (EMCT), a CT scanner

custom-built by the Radiation Physics group (Fig. 7) [51]. This scanner is unique in its operating procedure, because X-ray tube and detector rotate around the stationary sample, opposite to most lab-based μ CT scanners where it is the sample that rotates. As such, the EMCT allows objects to be equipped with peripheral sensors and equipment while still allowing for continuous CT scanning with a maximum rotation speed of one full rotation per 12 s. The scanner is controlled by a LabView interface [52]. See Dierick et al. [51] for more details about the set-up.

Distance between tree and X-ray source was 27 mm (Fig. 7), and the distance between X-ray source and detector 364 mm. The tube voltage was 70 kV, the tube power 8.47 W and no additional filtering was applied. A total of 7200 projections, with an exposure time of 200 ms per projection, were taken over six consecutive full rotations (1200 projections per rotation), with each rotation lasting 4 min, resulting in a total scan run duration of 24 min.

Between each run the scanner was paused for 6 min during the day and 30 min during the night to prevent overheating of the X-ray tube. A total of 15 runs was executed during daytime and 8 runs during evening and nighttime, with the exceptions of DOY 142 with 8 daytime runs (scanner and tree setup preparation), and DOY 145 with 6 daytime runs (end of experiment, including dismantling of the set-up). Reconstructions were automated using a Python wrapper for the Octopus reconstruction [53] software package (currently distributed by TESCAN-XRE, formerly known as XRE, spin-off company of UGCT), and resulted in a 3D reconstruction of a 7.5 mm section of the tree. The reconstructed data consisted of a total stack of 1000 reconstructed 2D slices, and an approximated voxel pitch of $7.5 \mu\text{m}$ was obtained.

During the daytime pauses of the EMCT scanner, measurements of xylem water potential (ψ_x , MPa) were collected with the pressure chamber (Model 1000, PMS Instrument Company, Corvallis, OR, USA). Leaves excised for ψ_x measurements were wrapped in aluminum foil for at least 1 h to ensure equilibration between leaf and stem water potential (Fig. 7). During wrapping and excision, AE detection was put on hold to avoid noise disturbance.

Acoustic emissions sensors and dendrometer were connected to their respective data acquisition systems to enable continuous registration. Dendrometer read-outs were registered every minute via a custom-built acquisition board. The AE signals were amplified by 35.6 decibels (dB) with an amplifier (AMP-1BB-J, KRN Services, Richland, WA, USA) and waveforms of 7168 samples length were acquired at 10 MHz sample rate. The signals were collected using two 2-channel PCI boards and

Table 4 AE waveform parameters of AE signals collected with the software program AEwin, including their respective abbreviation and unit

AE parameter	Abbreviation	Unit
Rise time	RISE	μs
Counts from peak amplitude	COUN	–
Wave energy	ENER	10–14 V ² s
Duration from peak amplitude	DURATION	μs
Peak amplitude	AMP	dB
Absolute frequency	AFRQ	kHz
Root mean square voltage	RMS	μV
Average signal level	ASL	dB
Reverberation frequency	RFRQ	kHz
Initiation frequency	IFRQ	kHz
Signal strength	SIGSTRNGTH	10–9 V s
Absolute energy	ABSENERGY	aJ
Partial power 0–100 kHz	FREQPP1	%
Partial power 100–200 kHz	FREQPP2	%
Partial power 200–400 kHz	FREQPP3	%
Partial power 400–800 kHz	FREQPP4	%
Frequency centroid	FRQC	kHz
Peak frequency	PFRQ	kHz

redirected to the software program AEwin (PCI-2, AEwin E4.70, Mistras Group BV, Schiedam, The Netherlands). A 20–1000 kHz electronic band pass filter was applied and only waveforms above the noise level of 28 dB were retained [12]. AE sensor installation was validated by the pencil lead break test [9, 43, 54]. Each collected AE signal was represented by a total of 18 AE waveform parameters (Table 4), with AE waveform parameters peak amplitude, rise time, counts from peak amplitude, wave energy, and duration from peak amplitude describing the shape of the AE signal (see Additional file 2). Internal clocks of dendrometer and AE acquisition systems were also matched with the EMCT to avoid differences in time between the datasets.

Processing and linking μCT images to AE signals

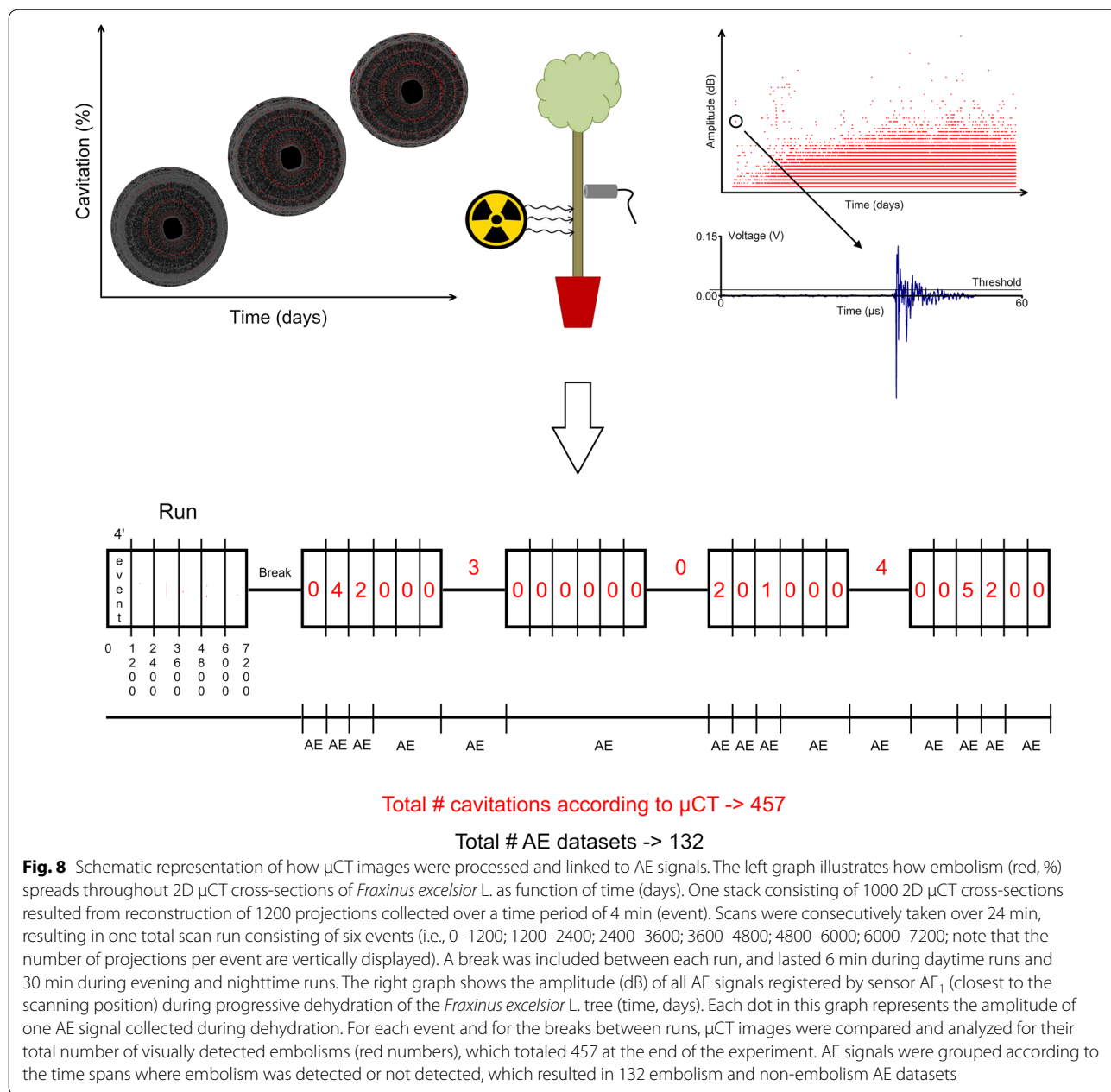
For each event and for the breaks between runs, 50 mid-centered 2D slices were extracted from the total stack (1000 2D slices), combined and reconstructed into single μCT images, which were pairwise compared and the number of visually detected embolisms quantified by the Fiji macro (Fig. 8). Each time an embolism event was detected, the start and end time of the projections used to build the corresponding μCT images were determined, resulting in a timespan for which corresponding AE signals registered by sensor AE₁ (closest to the scanning position) were detected and divided into separate AE embolism datasets (Fig. 8). Also, the start and end

time of the projections where no embolism formation was detected in consecutive events, breaks and runs were determined, resulting in a timespan corresponding to the non-embolism AE datasets (Fig. 8). At the end of the experiment, 457 embolized vessels were detected using the image processing procedure on the μCT data, for which the AE₁ signals were divided into 132 embolism and non-embolism datasets (Fig. 8).

The open source software package Fiji for multidimensional scientific imaging was used to process the reconstructed 2D μCT cross-sections [55]. To automate the processing procedure, we used two custom-written Fiji macros. The first macro was developed to reduce noise, allowing a better comparison between images of consecutive events and runs, by cropping each image as close as possible to the contours of the cross section, and median filtering the cropped images in the z-direction (3D kernel of [51–53]) (Fig. 8). The second macro was used to compare μCT images, by registering images of consecutive time steps using bUnwarpJ to match their contours [56] and taking the difference between two registered consecutive images. The larger size of vessels over other xylem elements results in high absolute differences in corresponding pixels during the transition from water-filled vessels (grey pixel area on μCT image) to embolized vessels (black pixel area on μCT image) between two consecutive μCT images. The function ‘Find Maxima’, manually controlled by a threshold (set at 30), was used to differentiate between true embolism events and noise. Finally, the (x, y) coordinates of the detected embolized vessels were stored.

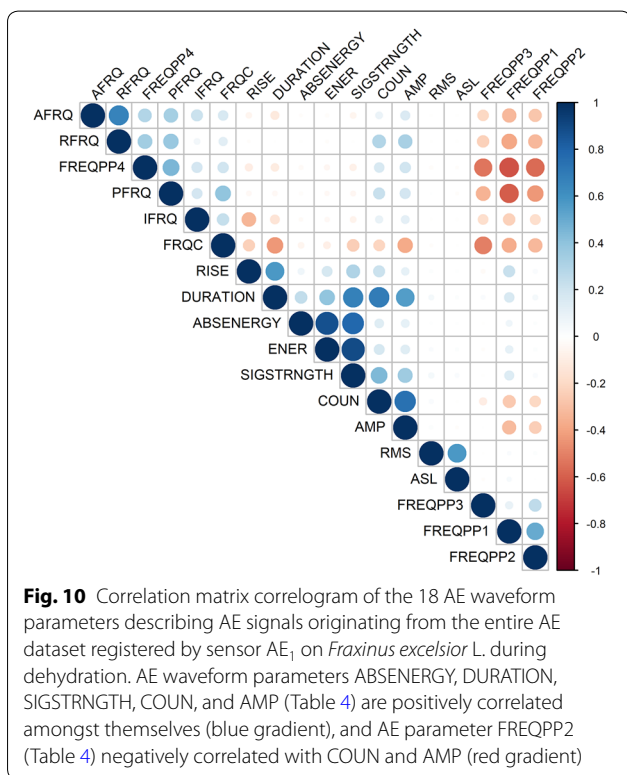
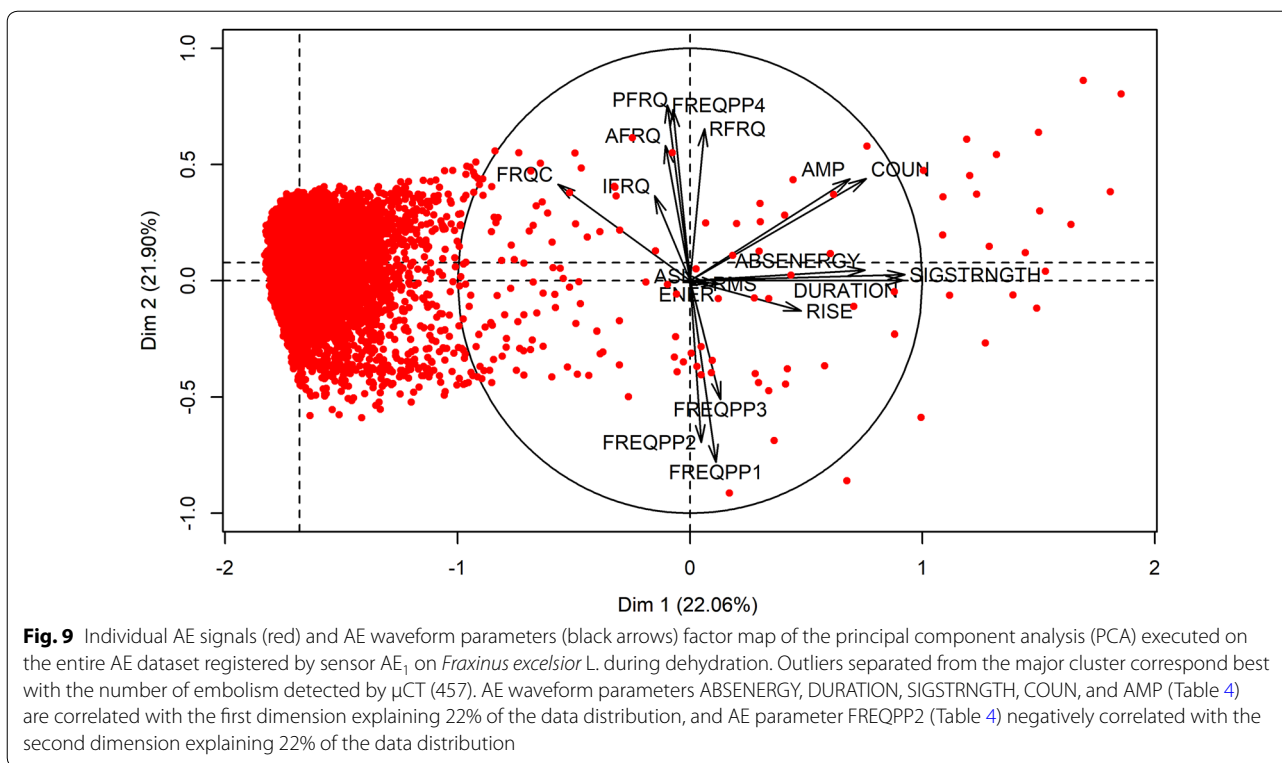
Unsupervised and supervised machine learning

To determine which AE sources are coupled to embolism events, the underlying distribution in recorded AE signals of sensor AE₁ (closest to the μCT scanning point) was determined via principal component analysis (PCA) based on the 18 parameters describing each AE signal (Table 4), and was visualized by combining individual and variable factor map plots with the R package FactoMineR (Fig. 9) [57]. With a total of 25,901 registered AE signals, PCA illustrated that the vast majority of these points were present within a large cluster, while a lower number of signals were separated from the cluster as apparent outliers (Fig. 9). Because μCT detected a total of 457 embolism events, PCA indicated that these outliers were most likely the AE source related to embolism formation. In addition, the correlogram of the correlation matrix between the 18 AE waveform parameters was constructed to visualize the underlying correlations in order to decide which AE waveform parameters were sufficiently related to another to be used as variables in the further detection of embolism-related AE (Fig. 10).



Principal component analysis indicated that outliers in AE signals that were distinctively separated from the major cluster (Fig. 9) mainly followed the direction of the AE waveform parameters ABSENERGY, DURATION, SIGSTRNGTH, COUN, and AMP (Table 4). These parameters were positively and best (length of the arrows) correlated with the first principal component explaining 22.06% of the data distribution. The correlogram also illustrated that the AE waveform parameters ABSENERGY, DURATION, SIGSTRNGTH, COUN, and AMP were positively correlated among themselves (blue

gradient) (Fig. 10). The AE parameter FREQPP2 was negatively and well correlated with the second principal component explaining 21.90% of the data distribution (Fig. 9), and negatively correlated to mainly AE waveform parameters AMP and COUN (red gradient) (Fig. 10). Because Vergeynst et al. [12] indicated FREQPP2 as important in clustering embolism-related AE, and to establish a link between parameters describing the shape of the AE signal and its frequency spectrum, FREQPP2 in addition to ABSENERGY, DURATION, SIGSTRNGTH, COUN, and AMP were selected for the consecutive machine learning



steps to detect embolism-related AE from the total measured signals.

Based on the PCA results, with the six AE waveform parameters as possible sources for embolism formation (outliers Fig. 9), histograms of these parameters were constructed for five randomly selected (from a total of 132) embolism and non-embolism AE datasets, of which two per embolism and non-embolism datasets are shown (Fig. 2), to examine the efficiency of static thresholding to distinguish embolism formation from other AE sources.

Because determining histogram thresholds per AE parameter for the 132 separate AE datasets is too cumbersome, receiver operating curves (ROC) were constructed to determine which AE parameter yielded the most promising threshold on the entire AE dataset. With the interest in distinguishing embolism-related AE from other sources, a two-class prediction problem can be considered in this case, in which the outcomes are labeled either as positive (embolism) or negative (non-embolism). This means that there are four possible outcomes, but for the construction of the ROC curve only the true positive (TP, the actual embolism is predicted correctly, y-axis) and false positive (FP, a non-embolism is predicted as embolism, x-axis) rate are required. For

each AE parameter, the TP versus FP rate enables to determine different static AE parameter thresholds. The first threshold in the ROC curve is the maximum value of each AE parameter over the entire dataset, typically resulting in solely a TP rate, but too strict to detect all the registered embolism events by μ_{CT} . Therefore, each maximum threshold is gradually adjusted and the number of AE signals in the embolism and non-embolism datasets is compared to the total number of embolism events detected by μ_{CT} . The most suited AE parameter to demarcate thresholds on the AE dataset to detect embolism-related AE is determined as the one for which the ROC curve stays as close as possible to the y-axis for the most stringent cut-points (i.e., as far as possible from the first bisector) and then deflects horizontally when the total number of embolism events registered by μ_{CT} (457) is reached.

Linear discriminant analysis (LDA) was used as supervised machine learning method to tackle the AE classification problem. The most straightforward strategy to tackle such a problem is to model the probability of an instance having a certain label given the feature vector x : $P(Y = y | X = x)$, which is called the posterior probability. Here Y is the random variable for the label and X the random variable modelling the features. Labeling an instance is done by assigning it the highest posterior probability, and if the posterior is modeled directly this is known as the discriminative approach (e.g., logistic regression). Using Bayes' rule, the posterior probability can be rewritten as:

$$P(Y = y | X = x) = \frac{P(X = x | Y = y)P(Y = y)}{P(X = x)}, \quad (1)$$

with $P(X = x | Y = y)$ the likelihood of observing a feature vector x in an instance with a label y , $P(Y = y)$ the prior of sampling an instance with a label y , and $P(X = x)$ the evidence or the probability of encountering an instance with this particular feature vector. Note that it is not necessary to compute this last factor explicitly, as it is independent of the label. The label with the highest (arg max ()) posterior probability (y^*) is predicted for a given feature vector x (Eq. 2):

$$y^* = \operatorname{argmax}_y P(Y = y | X = x) \quad (2)$$

In generative models that generate both input and output variables, it is the likelihood and the prior that are modeled using the training data, in contrast to the posterior in the discriminative approach. The posterior probability is then only computed afterwards, using Bayes' rule. As the term generative implies, one can

generate feature vectors associated with a given label. In practice, the model of the likelihood often does a poor job of modeling the conditional feature distribution, but can nevertheless give rise to good predictions. LDA is an example of a simple generative model, where every class is modeled by a normal distribution with the same covariance structure. In case of a binary classification problem, the features of the first class are distributed as $N(\mu_0, \Sigma)$ and of the second class as $N(\mu_1, \Sigma)$, with μ_0 and μ_1 the respective expected value of the feature vector within a class and Σ the covariance matrix. The log-posterior of LDA also gives rise to a linear model.

In this study, μ_0 represents the expected value of a feature vector of a measurement that is not an embolism event and μ_1 the expected feature vector of a feature vector associated with an embolism event. Both are assumed to have the same covariance structure Σ . We have estimated μ_0 and μ_1 by taking a weighted average over the averages feature vector of each dataset. For μ_0 and μ_1 , the weight for each data set is the number of non-embolism and embolism events, respectively, that were detected in a dataset. The global covariance matrix was computed in a similar way, after which a probability was attributed to each registered AE signal based on the weights calculated by the LDA model.

Linear discriminant analysis probabilities were summed for each AE dataset and the resulting sum was indicative for the expected number of AE signals classified as embolism-related (e.g., LDA on AE dataset 8 resulted in 44 embolism-related AE, Table 2). These were then used to construct the acoustic vulnerability curve. To determine whether the probability outcomes of the LDA model were suited to threshold the entire AE dataset, a ROC curve was constructed with the cut points representing a gradual decrease in maximum LDA probability (Fig. 3).

Wood anatomy

A wood sample of ~5 cm in length was taken from the scanned section, and included the marked position of scanning to perfectly match μ_{CT} images with the anatomical cross-section. The sample was preserved in a mixture of 70% ethanol (99%), 15% deionized water and 15% glycerol. A 35 μm thick cross section was cut from the sample at the exact point of scanning with a sliding microtome (Hn-40, Reichert-Jung, Saarland, Germany) at the Department of Biology, Ghent University. The cross section was stained for 15 min with 0.5% w/v astra blue, 0.5% w/v chrysoidine, and 0.5% w/v acridine red

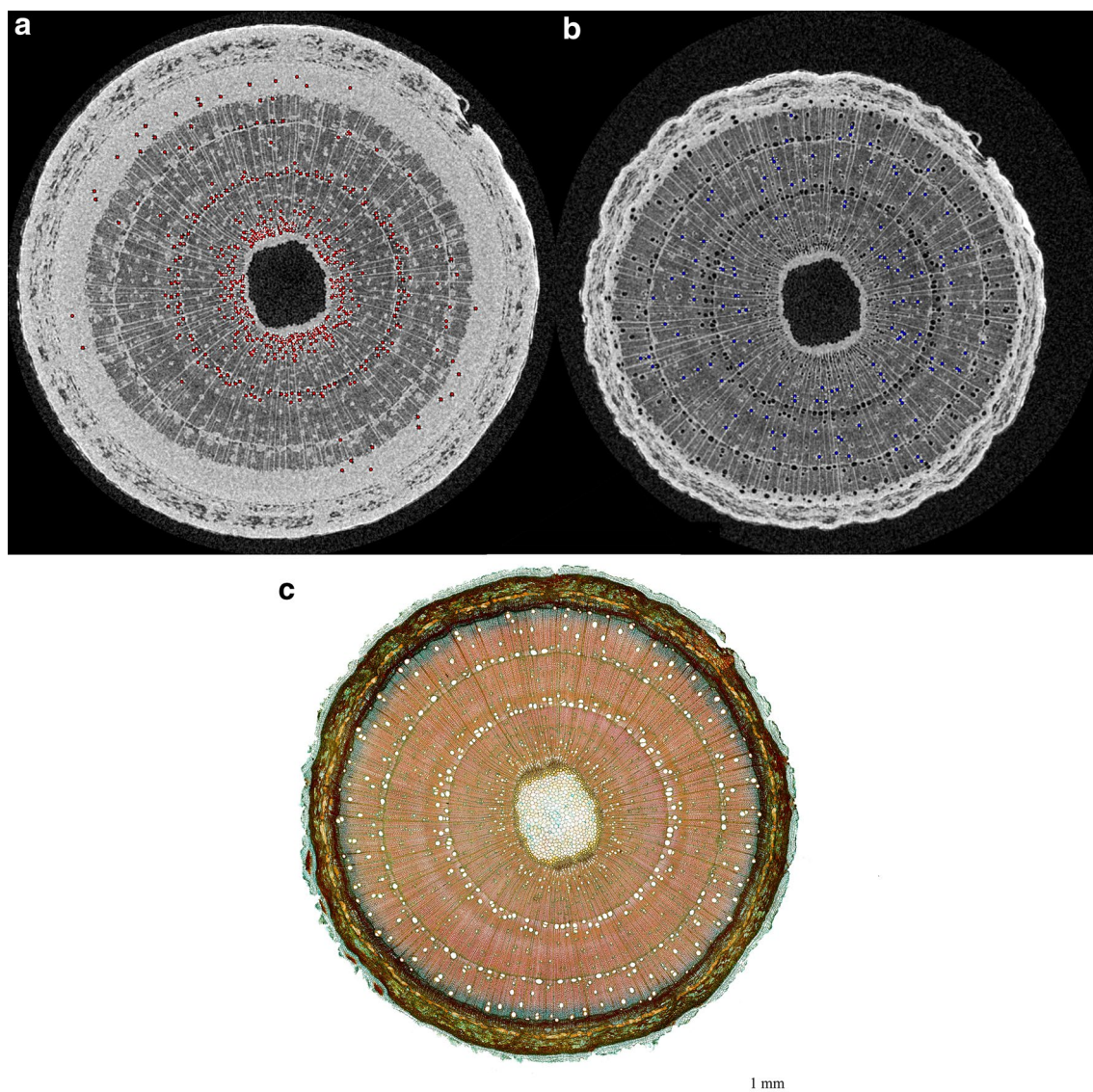


Fig. 11 μ CT image of *Fraxinus excelsior* L. at the start of the dehydration experiment (a), and at the end of the dehydration experiment (b). Native embolized vessels (a), and non-embolized vessels (b) are respectively highlighted in red and blue (b). Resolution = 7.5 μ m. c Stained 35- μ m thick cross-section of a 2-year-old *Fraxinus excelsior* L. stem at the μ CT scanning point showing from inward to outward pit, xylem, cambium and bark. The cross section had a total of 1100 vessels. Scale bar = 1 mm. Credit Dr. Olivier Leroux

and mounted in euparal after dehydration in isopropyl alcohol. Images were captured using a Nikon Ni-U epifluorescence microscope equipped with a Nikon DS-Fi1c camera (Fig. 11c).

Anatomical analysis was restricted to manually counting the number of xylem vessels on the cross-section with the image analysis software Fiji. A total of 1100 vessels was obtained from the cross-section and was used to translate the number of embolized vessels derived from the μ CT images to percentage embolism formation.

Acoustic and μ CT vulnerability curve

A total of 25,901 AE signals was registered by sensor AE₁ and 90,416 by sensor AE₂, which were per sensor cumulated over the measurement period and averaged over 10 min. The endpoint of the acoustic vulnerability curve (VC_{AE}) is normally determined via the local maximum of the third derivative of cumulative AE [12], which was however not yet reached (Fig. 6), because not all vessels were embolized at the end of the dehydration experiment (Fig. 11b). Correct determination of non-embolized

vessels on the μ CT image was facilitated by the anatomical cross-section (Fig. 11c), which perfectly matched with the μ CT scanning point. *Fraxinus excelsior* L. had a total of 1100 vessels, of which 541 were natively embolized (Fig. 11a), and 102 were not embolized at the end of the experiment (Fig. 11b).

Because complete embolism formation was not reached at the end of the dehydration experiment (Figs. 6, 11b), all registered AE signals had to be used in constructing the VC_{AE} . The number of native embolized vessels was taken into account when converting absolute cumulative AE to percentage cumulative AE (%) following the assumption that VCs start from a fully hydrated condition [16]. The unfiltered cumulative AE of sensor AE_1 and AE_2 were translated to percentage cumulative AE by rescaling between 0 and 100% following the technique of Vergeynst et al. [12] (Fig. 6). For the dataset derived from sensor AE_1 , an additional VC_{AE} was constructed based on the LDA model output (VC_{AE-E}). The LDA model detected 518 embolism-related AE signals, which were cumulated over the measurement period, averaged over 10 min, and for LDA rescaled from 0 to 96% as not all vessels were embolized at the end of the experiment (i.e., $(541 + 518/1100) \cdot 100$). The number of embolism formation events derived from the μ CT scans was used to construct a μ CT vulnerability curve (VC_{CT}), which was also averaged over 10 min and rescaled between 0 and 91% (i.e., $(541 + 457/1100) \cdot 100$) to obtain percentage embolism formation (%).

The time axis of the different VCs was replaced with a continuous xylem water potential axis using the stress-strain curve. In this curve, point measurements of xylem water potential or stress (ψ_x , MPa) are plotted against xylem shrinkage or strain ($\Delta d/di$, $\mu\text{m mm}^{-1}$) measured with the dendrometer (see Additional file 3). A segmented-linear regression between ψ_x point measurements and continuous $\Delta d/di$ with two breakpoints was obtained with the segmented R package [58] (see Additional file 3). The three linear regression equations were used to calculate the continuous xylem water potential values.

Statistical analysis

A smooth spline function in the stats library in R software (RStudio version 1.1.419-© 2009–2017 RStudio, Inc.) was fitted to the vulnerability curves. Drought vulnerability values such as the onset of embolism formation (ψ_x at which 12% of embolism-related AE and μ CT occur; AE_{12}/CT_{12}), 50% embolized (ψ_x at which 50% of embolism-related AE and μ CT occur; AE_{50}/CT_{50}), full embolism (ψ_x at which 88% of embolism-related AE and μ CT occur; AE_{88}/CT_{88}) and endpoint of the VCAE (ψ_x at

which 100% of embolism-related AE occur; AE_{100}) were determined [59]. Differences in VCs were quantified using the absolute difference in percentage embolism formation compared to the reference μ CT VC.

Supplementary information

Supplementary information accompanies this paper at <https://doi.org/10.1186/s13007-019-0543-4>.

Additional file 1. Photographic overview of the experimental setup of the *Fraxinus excelsior* L. tree in the custom-built holder with carbon fiber tube designed to ensure stable and centered positioning of the tree during μ CT scanning (A). The tree is equipped with two broadband point-contact AE sensors in a pvc spring-containing holder to continuously register AEs from the progressively dehydrating xylem (AE_1 , AE_2), and a point dendrometer to continuously register xylem shrinkage. This potentiometer-type of dendrometer was selected, as its working principle does not generate a magnetic field, which would otherwise produce acoustic interference.

Additional file 2. Waveform of a registered AE signal from a dehydrating *Fraxinus excelsior* L. stem. AE waveform parameters peak amplitude, rise time, counts from peak amplitude (several threshold crossings indicated with green boxes), wave energy (red) and duration from peak amplitude describe the shape of the AE signal.

Additional file 3. Stress-strain curve (black, open circles) between point measurements of xylem water potential (MPa) and xylem shrinkage ($\Delta d/di$, $\mu\text{m mm}^{-1}$) of *Fraxinus excelsior* L. during dehydration. The segmented-linear regression with two breakpoints (black and red dashed line) divided the dataset in three linear regressions with their own equation and R^2 (black, red, and grey) from which continuous xylem water potential was calculated.

Acknowledgements

The authors wish to thank two colleagues of the Laboratory of Plant Ecology, Ghent University: Erik Moerman for building the custom data acquisition logger, and Geert Favys for the construction of the unique sample holder to hold the tree perfectly in place during the μ CT scanning and for his aid in setting up and breaking down the experimental setup. Sincere gratitude towards our colleague Dr. Olivier Leroux at the Department of Biology, Ghent University for preparing and finalizing the microscopic cross section for further analysis. The Special Research Fund of Ghent University (BOF-UGent) is acknowledged for the financial support of the UGCT Center of Expertise (BOF.EXP.2017.0007).

Authors' contributions

KS and NDB designed the experiment and NDB conducted it. NDB analyzed the data and wrote the manuscript. KS supervised and complemented the writing. MS performed part of the data analysis and helped writing the manuscript. JvDB helped conducting the μ CT measurements, processed the μ CT data and complemented the writing. BDB provided feedback on the statistical and machine learning aspects. LVH provided access to the μ CT facilities and complemented the writing. All authors read and approved the final manuscript.

Funding

This work was supported by the Research Foundation Flanders (FWO), Grant no. G.0941.15N to KS supporting the Ph.D. of NDB. MS is supported by the Research Foundation—Flanders (FWO17/PDO/067).

Availability of data and materials

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Laboratory of Plant Ecology, Department of Plants and Crops, Faculty of Bioscience Engineering, Ghent University, Coupure links 653, 9000 Ghent, Belgium. ²KERMIT, Department of Data Analysis and Mathematical Modelling, Faculty of Bioscience Engineering, Ghent University, Coupure links 653, 9000 Ghent, Belgium. ³UGent-Woodlab-Laboratory of Wood Technology, Department of Environment, Faculty of Bioscience Engineering, Ghent University, Coupure links 653, 9000 Ghent, Belgium. ⁴Ghent University Centre for X-Ray Tomography (UGCT), Proeftuinstraat 86, 9000 Ghent, Belgium. ⁵Radiation Physics Group, Department of Physics and Astronomy, Faculty of Sciences, Ghent University, Proeftuinstraat 86, 9000 Ghent, Belgium.

Received: 18 July 2019 Accepted: 5 December 2019

Published online: 17 December 2019

References

1. Tyree MT, Zimmermann MH. Xylem structure and the ascent of sap. 2nd ed. Berlin: Springer; 2002.
2. Markesteijn L, Poorter L, Paz H, Sack L, Bongers F. Ecological differentiation in xylem cavitation resistance is associated with stem and leaf structural traits. *Plant Cell Environ*. 2011;34:137–48.
3. Sperry JS, Tyree MT. Mechanism of water-stress induced xylem embolism. *Plant Physiol*. 1988;88(3):581–7.
4. Tyree MT, Sperry JS. Vulnerability of xylem to cavitation and embolism. *Annu Rev Plant Biol*. 1989;40(1):19–36.
5. Tyree MT, Dixon MA, Thompson RG. Ultrasonic acoustic emissions from the Sapwood of *Thuja occidentalis* measured inside a pressure bomb. *Plant Physiol*. 1984;74:1046–9.
6. Tyree MT, Dixon MA, Tyree EL, Johnson R. Ultrasonic acoustic emissions from the sapwood of cedar and hemlock: an examination of three hypotheses regarding cavitations. *Plant Physiol*. 1984;75:988–92.
7. Lo Gullo MA, Salleo S. Three different methods for measuring xylem cavitation and embolism: a comparison. *Ann Bot*. 1991;67:417–24.
8. Millburn JA, Johnson RP. The conduction of sap: II. Detection of vibrations produced by sap cavitation in *Ricinus* xylem. *Planta*. 1966;69:43–52.
9. Tyree MT, Sperry JS. Characterization and propagation of acoustic emission signals in woody plants: towards an improved acoustic emission counter. *Plant Cell Environ*. 1989;40:19–38.
10. Mayr S, Rosner S. Cavitation in dehydrating xylem of *Picea abies*: energy properties of ultrasonic emissions reflect tracheid dimensions. *Tree Physiol*. 2011;31:59–67.
11. Wolkerstorfer SV, Rosner S, Hietz P. An improved method and data analysis for ultrasound acoustic emissions and xylem vulnerability in conifer wood. *Physiol Plantarum*. 2012;146:184–91.
12. Vergeynst LL, Sause MGR, De Baerdemaeker NJF, De Roo L, Steppe K. Clustering reveals cavitation-related acoustic emission signals from dehydrating branches. *Tree Physiol*. 2016;36(6):786–96.
13. Beall FC. Overview of the use of ultrasonic technologies in research on wood properties. *Wood Sci Technol*. 2002;36:197–212.
14. Rosner S, Klein A, Wimmer R, Karlsson B. Extraction of features from ultrasound acoustic emissions: a tool to assess the hydraulic vulnerability of Norway spruce trunkwood? *New Phytol*. 2006;171:105–16.
15. Sperry JS, Donnelly JR, Tyree MT. A method for measuring hydraulic conductivity and embolisms in xylem. *Plant Cell Environ*. 1988;11(1):35–40.
16. Cochard H, Badel E, Herbette S, Delzon S, Choat B, Jansen S. Methods for measuring plant vulnerability to cavitation: a critical review. *J Exp Bot*. 2013;64:4779–911.
17. De Baerdemaeker NJF, Salomón RL, De Roo L, Steppe K. Sugars from woody tissue photosynthesis reduce xylem vulnerability to cavitation. *New Phytol*. 2017;216(3):720–7.
18. Kikuta SB. Ultrasound acoustic emissions from bark samples differing in anatomical characteristics. *Phyton*. 2003;43:161–78.
19. De Roo L, Vergeynst LL, De Baerdemaeker NJF, Steppe K. Acoustic emissions to measure drought-induced cavitation in plants. *Appl Sci*. 2016;6:71.
20. Lewis AM. Two mechanisms for the initiation of embolism in tracheary elements and other dead plant cells under water stress. To: Dept. of Organismic and Evolutionary Biology. Advisor: Martin Zimmermann and others. Harvard University.
21. Haacs R, Blank RW. Acoustic emission from drought-stressed red pine (*Pinus resinosa*). *J Acoust Emiss*. 1990;9:181–7.
22. Zweifel R, Zeugin F. Ultrasonic acoustic emissions in drought-stressed trees—more than signals from cavitation? *New Phytol*. 2008;179:1070–9.
23. Ritman KT, Milburn JA. Acoustic emissions from plants: ultrasonic and audible compared. *J Exp Bot*. 1988;39:1237–48.
24. Cochard H, Tyree MT. Xylem dysfunction in *Quercus*: vessel sizes, tyloses, cavitation and seasonal changes in embolism. *Tree Physiol*. 1990;6:393–407.
25. Jackson G, Grace J. Cavitation and water transport in trees. *Endeavour*. 1994;18:50–4.
26. Rosner S, Konnerth J, Plank B, Salaberger D, Hansmann C. Radial shrinkage and ultrasound acoustic emissions of fresh versus pre-dried Norway spruce sapwood. *Trees Struct Funct*. 2010;24:931–40.
27. Schenk JH, Steppe K, Jansen S. Nanobubbles: a new paradigm for air-seeding in xylem. *Trends Plant Sci*. 2015;20(4):199–205.
28. Kowalski SJ, Moliński W, Musielak G. The identification of fracture in dried wood based on theoretical modelling and acoustic emission. *Wood Sci Technol*. 2004;38:35–52.
29. Brodersen CR, McElrone AJ, Choat B, Matthews M, Shackel K. The dynamics of embolism repair in xylem: in vivo visualizations using high-resolution computed tomography. *Plant Physiol*. 2010;154:1088–95.
30. Cochard H, Delzon S, Badel E. X-ray microtomography (micro-CT): a reference technology for high-resolution quantification of xylem embolism trees. *Plant Cell Environ*. 2015;38(1):201–6.
31. Torres-Ruiz JM, Jansen S, Choat B, McElrone AJ, Cochard H, Brodrigg TJ, Badel E, Burrett R, Bouche PS, Brodersen CR, et al. Direct X-ray microtomography observation confirms the induction of embolism upon xylem cutting under tension. *Plant Physiol*. 2015;167:40–3.
32. Choat B, Badel E, Burrett R, Delzon S, Cochard H, Jansen S. Noninvasive measurement of vulnerability to drought-induced embolism by X-ray microtomography. *Plant Physiol*. 2016;170(1):273–82.
33. Suuronen J-P, Peura M, Fagerstedt K, Serimaa R. Visualizing water-filled versus embolized status of xylem conduits by desktop x-ray microtomography. *Plant Methods*. 2013;9:11.
34. Knipfer T, Brodersen CR, Zedan A, Kluepfel DA, McElrone AJ. Patterns of drought-induced embolism formation and spread in living walnut saplings visualized using X-ray microtomography. *Tree Physiol*. 2015;35:744–55.
35. Brodersen CR, Knipfer T, McElrone AJ. In vivo visualization of the final stages of xylem vessel refilling in grapevine (*Vitis vinifera*) stems. *New Phytol*. 2018;217(1):117–26.
36. Swan AL, Mobasher A, Allaway D, Liddell S, Bacardit J. Application of machine learning to proteomics data: classification and biomarker identification in postgenomics biology. *Omic*. 2013;17:595–610.
37. Li Y, Wu F-X, Ngom A. A review on machine learning principles for multi-view biological data integration. *Brief Bioinform*. 2018;19(2):325–40.
38. Sathya R, Abraham A. Comparison of supervised and unsupervised learning algorithms for pattern classification. *IJARAI*. 2013;2:34–8.
39. Hastie T, Tibshirani R, Friedman J, editors. The elements of statistical learning, data mining, inference, and prediction. 2nd ed. New York: Springer; 2009.
40. McLachlan GJ, editor. Discriminant analysis and statistical pattern recognition. New Jersey: Wiley; 2004.
41. Lemoine D, Peltier J-P, Marigo G. Comparative studies of the water relations and the hydraulic characteristics in *Fraxinus excelsior*, *Acer pseudoplatanus* and *A. opalus* trees under soil water contrasted conditions. *Ann For Sci*. 2001;58:723–31.
42. Johnson DM, Wortemann R, McCulloh KA, Jordan-Meille L, Ward E, Warren JM, Palmroth S, Domec JC. A test of the hydraulic vulnerability segmentation hypothesis in angiosperm and conifer tree species. *Tree Physiol*. 2016;36:983–93.

43. Vergeynst LL, Sause MGR, Hamstad MA, Steppe K. Deciphering acoustic emission signals in drought stressed branches: the missing link between source and sensor. *Front Plant Sci.* 2015;6:494.
44. Nolf M, Beikircher B, Rosner S, Nolf A, Mayr S. Xylem cavitation resistance can be estimated based on time-dependent rate of acoustic emissions. *New Phytol.* 2015;208:625–32.
45. Vergeynst LL, Dierick M, Bogaerts J, Cnudde V, Steppe K. Cavitation: a blessing in disguise? New method to establish vulnerability curves and assess hydraulic capacitance of woody tissues. *Tree Physiol.* 2015;35:400–9.
46. Cochard H. Vulnerability of several conifers to air embolism. *Tree Physiol.* 1992;11:73–83.
47. Rosner S. A new type of vulnerability curve: is there truth in vine? *Tree Physiol.* 2015;35:410–4.
48. Schulze ED, Čermák J, Matyssek M, Penka M, Zimmermann R, Vasíček F, Gries W, Kučera J. Canopy transpiration and water fluxes in the xylem of the trunk of *Larix* and *Picea* trees—a comparison of xylem flow, porometer and cuvette measurements. *Oecologia.* 1985;66:475–83.
49. Steppe K, Lemeur R, Samson R. Sap flow dynamics of a beech tree during the solar eclipse of 11 August 1999. *Agric For Meteorol.* 2002;112:139–49.
50. Steppe K, Sterck F, Deslauriers A. Diel growth dynamics in tree stems: linking anatomy and ecophysiology. *Trends Plant Sci.* 2015;20:335–43.
51. Dierick M, Van Loo D, Masschaele B, Van den Bulcke J, Van Acker J, Cnudde V, Van Hoorebeke L. Recent micro-CT scanner developments at UGCT. *Nucl Instrum Methods Phys Res B.* 2014;324:35–40.
52. Dierick M, Van Loo D, Masschaele B, Boone M, Van Hoorebeke L. A LabVIEW® based generic CT scanner control software platform. *J X-ray Sci Technol.* 2010;18(4):451–61.
53. Vlassenbroeck J, Dierick M, Masschaele B, Cnudde V, Van Hoorebeke L, Jacobs P. Software tools for quantification of X-ray microtomography at the UGCT. *Nucl Instrum Methods Phys Res A.* 2007;580:442–5.
54. Sause MGR. Investigation of pencil-lead breaks as acoustic emission sources. *J Acoust Emiss.* 2011;29:184–96.
55. Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, Preibisch S, Rueden C, Saalfeld S, Schmid B, et al. Fiji: an open source platform for biological image analysis. *Nat Methods.* 2012;9:676–82.
56. Arganda-Carreras I, Sorzano COS, Marabini R, Carazo JM, Ortiz-de-Solorzano C, Kybic J. Consistent and elastic registration of histological sections using vector-spline regularization. In: Beichel RR, Sonka M, editors. *Computer vision approaches to medical image analysis. CVAMIA lecture notes in computer science*, vol. 4241. Berlin: Springer; 2006. p. 85–95.
57. Lê S, Josse J, Husson F. FactoMineR: an R Package for multivariate analysis. *J Stat Softw.* 2008;25:1–18.
58. Muggeo VMR. Segmented: an R package to fit regression models broken-line relationships. *R News.* 2008;8:20–5.
59. Domec JC, Gartner BL. Cavitation and water storage capacity in bole xylem segments of mature and young Douglas-fir trees. *Trees Struct Funct.* 2001;15:204–14.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

