Methods in Ecology and Evolution

MISS LAURA ROSE (Orcid ID : 0000-0003-4523-4145)

Article type : Forum Article

Editor: Professor Luca Börger

Accuracy of image analysis tools for functional root traits: A comment on Delory et al., 2017

Comment on:

http://onlinelibrary.wiley.com/doi/10.1111/2041-210X.12771/full

Accuracy and bias of methods used for root length measurements in functional root research

Running title:

Accuracy of root image analysis tools

Laura Rose[1,2]* (ORCID: 0000-0003-4523-4145), Guillaume Lobet[3] (ORCID: 0000-0002-5883-4572)

[1]University of Freiburg, Faculty of Biology, Geobotany, Schaenzlestr. 1, 79104 Freiburg, Germany

[2]Senckenberg Biodiversity and Climate Research Centre (BiK-F), 60325 Frankfurt/Main, Germany

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/2041-210X.13156

[3]Agrosphere, IBG3, Forschungszentrum Jülich, Wilhelm-Johnen-Strasse, 52428 Jülich, Germany

*laura.rose.ecology@gmail.com

Tweetable Abstract:

Root analyses tools WinRhizo™ and IJ_Rhizo estimate traits with comparable accuracy at high but not low resolution

Abstract:

Root traits get increasing attention as functional equivalents of aboveground traits. Image analysis software such as WinRhizo™ and IJ_Rhizo facilitate root trait analyses. Delory et al., 2017 presented a comparison between the accuracy of WinRhizo™ and IJ_Rhizo for measuring root length. We complement their analyses with a comparison of diameter and volume estimates and a comparison of different image resolutions with manual and automatic threshold.

We analysed 100 images of fibrous and tap-root systems, which were obtained by using the root model ArchiSimple. As a result, for each image, diameter, length and volume were known. The images were analysed with WinRhizo™ and IJ_Rhizo and we compared the estimates of diameter, length, and volume to ground-truth values. We further computed relative errors and their magnitude and analysed their dependency on image characteristics and root system properties.

At 1200 and 800 dpi, diameter and length estimates provided by WinRhizo™ and IJ_Rhizo were of comparable accuracy. Diameter errors were balanced. Volume estimates were subjected to a systematic error caused by the assumption of constant diameter. WinRhizo™,

however, provides the opportunity to calculate correctly computed volumes from diameter classes. At 1200 dpi, IJ_Rhizo failed to automatically find an appropriate threshold for pixel classification, which fundamentally decreased accuracy.

The magnitude of diameter errors increased with root overlap for IJ_Rhizo. The length errors increased with increasing root length, overlap and root length density for WinRhizo^{TM}. The magnitude of underestimation of the volume (WinRhizo^{TM}) decreased with volume. It was higher for tap-root than for fibrous root systems. All errors increased with lower resolution.

Our results confirm the results of Delory et al., 2017 regarding the accuracy for length. They further confirm that estimates derived from different software packages or at different resolution should not be compared directly. The characteristics of root systems should be standardized for image analysis. The dependency of errors on the response variable of interest can influence the effect size and increase the probability of errors. Validation of methods should be conducted for each analysed dataset. New image analysis tools should be validated against a real ground-truth.

German Abstract:

Funktionelle Wurzelmerkmale finden verstärkte Beachtung als Äquivalente zu oberirdischen Merkmalen. Bildanalysesoftware, wie WinRhizo^{TM} und IJ_Rhizo, vereinfachen die Analyse. Delory et al., 2017 vergleichen die Genauigkeit dieser beiden Programme bezüglich der Messung von Wurzellängen. Wir ergänzen diese Analyse mit einem Vergleich der Präzision der Durchmesser- und Volumenmessungen bei unterschiedlichen Bildauflösungen sowohl mit manuellem als auch automatisch gesetztem Schwellenwert zur Unterscheidung von Wurzel und Hintergrund.

Wir haben 100 Bilddateien mit WinRhizo^{TM} und IJ_Rhizo analysiert, welche mit dem Wurzelmodel ArchiSimple erstellt wurden. Für jedes Bild waren Durchmesser, Länge und Volumen des Wurzelsystems bekannt und konnten mit den Messwerten der Programme

verglichen werden. In einem zweiten Schritt haben wir die Abhängigkeit der relativen Fehler aller Parameter von Bildeigenschaften und Wurzelsystemeigenschaften überprüft.

Bei einer Bildauflösung von 1200 und 800 dpi war die Genauigkeit der beiden Programme vergleichbar. Die Fehler der Durchmessermessungen waren balanciert. Die Volumenmessungen waren durch einen systematischen Fehler beeinflusst, der durch die Annahme von konstanten Durchmessern verursacht wird. In Falle von WinRhizo™ kann dieser Fehler vermieden werden. Bei einer Auflösung von 1200 dpi war IJ_Rhizo nicht in der Lage automatisch einen angemessenen Schwellenwert zu finden, was zu erheblichen Fehlern führte.

Bei der Analyse mit IJ_Rhizo nahmen die Fehler der Durchmesser mit zunehmender Überlappung der Wurzeln zu. Im Falle von WinRhizo™, stiegen die Fehler der Wurzellänge mit der Gesamtwurzellänge, der Überlappung und der Wurzellängendichte in den Bildern an, während die relativen Fehler der Volumina mit steigendem Gesamtvolumen abnahmen. Alle Fehler nahmen bei verringerter Bildauflösung zu.

Unsere Ergebnisse bestätigen die Ergebnisse von Delory et al., 2017 bezüglich der Genauigkeit der Wurzellängenmessungen. Sie bestärken außerdem die Schlussfolgerung, dass Messungen, die mit verschiedenen Programmen oder bei verschiedenen Auflösungen gemacht wurden, nicht direkt vergleichbar sind. Die Eigenschaften von Bildern sollten möglichst weitgehend standardisiert werden. Die Abhängigkeit der Fehler von der Antwortvariable kann sowohl die Effektgröße als auch die Fehlerwahrscheinlichkeit beeinflussen. Dementsprechend sollte jede Validierung der Methoden datensatzspezifisch erfolgen. Neue Analysemethoden sollten anhand direkter Referenzwerte, nicht nur anderer Programme, getestet werden.

Key-words: fibrous roots, IJ_Rhizo, image resolution, root diameter, root length, root volume, tap-roots, WinRhizo™

Introduction

Functional plant traits have been identified to be useful predictors of plant strategies and ecosystem functions (Diaz et al., 2004). While emphasis was mainly on leaf traits (e.g. Wright et al., 2004), a focus on root traits as equivalently important features of plants has gained momentum only more recently. Diameter, specific root length, and specific root volume (or the more commonly used tissue density) are traits of interest since they are related to resource economic strategies and ecosystem characteristics (Ryser, 2006, Bergmann, Ryo, Prati, Hempel, & Rillig, 2017, Freschet et al., 2017).

Digital image analysis tools facilitate measurements of these traits, but also bear the risk of introducing (systematic) errors if not tested or understood (Ryser, 2006, Lobet et al., 2017, Rose, 2017). In a recent study, Delory and colleagues (2017) compared different root length estimates from digital image analyses using the software packages WinRhizo™ and ImageJ with the macro IJ_Rhizo (IJ_Rhizo hereafter, Pierret, Gonkhamdee, Jourdan, & Maeght, 2013). From comparisons with a ground-truth, they conclude that the length estimates provided by the open source software IJ_Rhizo can be an alternative to WinRhizo™. However, care should be taken, when samples analysed with different software are compared. The ground-truth was derived by tracing roots manually with ImageJ. This very time-consuming procedure can only be used for a limited number of images. Further, and more importantly, it only yields a measure of the total root length, but not of the root diameter or volume.

Previous studies investigating the accuracy of diameter and volume estimates of image analyses software used scans of objects of known diameters (e.g. Bauhus & Messier, 1993, Kimura & Yamasaki, 2003). These objects lack the diameter heterogeneity of real roots, which can introduce severe errors into volume estimates, if it is not accounted for (Rose, 2017). A recently published set of images of ground-truthed, modelled plant root systems (Lobet et al., 2017) provides the unique opportunity to analyse length, volume and diameter of root systems with image analysis software tools.

The purpose of our study was to complement the analysis from Delory et al., (2017) with diameter and volume information. More precisely, the aims of our study were to compare the accuracy of length, volume and diameter estimates provided by WinRhizo^{TM} and IJ_Rhizo using ground-truthed modelled root images, and to identify possible error sources.

Methods

Images and related data

We used a root image library generated as described in Lobet et al. (2017). In short, the root model ArchiSimple (Pagès et al., 2013) was used to simulate 100 different root systems (50 fibrous and 50 tap-rooted). Root systems were grown in a 2D space, to mimic growth in rhizotrons. This was done to avoid any measurement bias that could result from the conversion of a 3D root architecture into a 2D image. Each root system was stored in a Root System Markup Language file (RSML, Lobet et al., 2015) from which we (1) created a black and white JPEG image, resolution 1200 dpi (see Supplementary Fig. 1). For each root system, we extracted two types of ground-truth data. At the root system level, we extracted the total root length (L), the total volume (V) and average diameter (D). An overlapping index was computed as the number of root segments having an overlap with other root segments over the total number of root segments (this was computed directly from each RSML files). In addition, each root system was split in small root segments of constant diameter. From the segments data, for each image, we computed the length-weighted coefficient of diameter variation (CVD, Eq. 1).

$$CVD = \frac{\sqrt{\frac{N\sum_{i=1}^{N} l_i (d_i - D)^2}{(N-1)L}}}{D} \qquad \text{Eq. 1}$$

with N = number of segments, l = segment length, d = segment diameter.

Each image was (2) downsized to resolutions of 800, 600 and 400 dpi, to permit the analysis of the influence of image resolution on accuracy of estimates.

The root length density in the images ranged between 0.38 and 6.44 cm cm^{-2} (mean = 1.31), the average root diameter was 0.22 mm (range 0.16 – 0.33) and the thinnest and thickest root segments in the images had diameters of 0.04 and 0.54 mm, respectively. Image characteristics differed between fibrous and taproot systems (Supplementary Fig. 2).

Image analyses

All images were analysed with WinRhizo™ 2013, Regent method (Régent Instruments Inc 2013) and ImageJ (1.51j8, Schneider, Rasband, & Eliceiri, 2012) with the macro IJ_Rhizo (IJ_Rhizo_v0beta). We conducted batch analyses with the grey level thresholds set to automatic at all four resolutions. We further analysed the images at 800 dpi and 1200 dpi with the threshold value manually set to 144. We considered this value suitable after visual inspection of 10 images at different thresholds using the threshold slider of WinRhizo™. Since no debris was simulated, we used no automatic debris correction. For WinRhizo™, the boundaries of diameter classes were set from 0.1 to 1.9 mm (with an increment of 0.1 mm).

Data analyses

The total image volume (VWR_{SUM}), length (LWR) and length weighted mean diameter (DWR_{wm}) were calculated in excel from values of the different diameter classes of the WinRhizo™ output following Rose (2017). The calculation of a length weighted mean diameter for IJ_Rhizo is not necessary, because it is already provided by the software (DIJc), whereas volume estimates are not provided for diameter classes, which prohibits the calculation of a diameter based volume for IJ_Rhizo. All further analyses were conducted with R (Version 3.2.3, R Core Team, 2015).

We used linear models to analyse the dependency of the diameters estimated by IJ_Rhizo (DIJ, average diameter, and DIJc), the average diameter provided by WinRhizo™ (DWR) and DWR_{wm} on the correct diameter (D). The same procedure was used for the LWR and the Kimura length estimate from IJ_Rhizo (LIJ) and the correct length (L), as well as the VWR_{sum}, the total volume provided by WinRhizo™ (VWR) and the volume from the IJ_Rhizo output (VIJ) and the correct volume (V).

In a next step, we calculated errors for each variable and each image as the relative error (Eq. 2) and the magnitude of the relative error (Eq. 3) as relative differences between the software estimate and the correct (ground-truth) value:

$$Error_{rel} = \frac{Estimate - Correct}{Correct} \times 100 \qquad \text{Eq. 2}$$

$$Error_{mag} = abs\left(\frac{Estimate - Correct}{Correct} \times 100\right) \qquad \text{Eq. 3}$$

Subsequently, we used linear models to test which root system characteristics are potentially influencing the magnitude of errors for the best estimates of D, L and V per software package (Table 1). In this analysis, we included both volume estimates of WinRhizo™, because we hypothesized different structural parameters to influence the different measures (Rose 2017). Tested explanatory variables for errors in diameter were the average diameter, the coefficient of diameter variation and the root overlap, for errors in root length, the total root length, the root length density (RLD) and the overlap. For errors in root volume, we tested the average diameter, the coefficient of diameter variation and the root volume as explanatory variables. We considered linear, exponential and logarithmic relationships and selected the best models based on significance and R^2.

In a last step, we used Wilcoxon tests to analyse whether the relative errors of diameter, length, and volume (VWR_{SUM} only) differed between fibrous and taproot systems.

Results

Relationships to ground-truth and magnitudes of errors at 1200 dpi with automatic threshold

The relationships between the diameter, length and volume estimates of WinRhizo™ and the ground-truth were all highly significant and yielded R^2 values between 0.95 and 1.00 (Table 1). All regression slopes were significant ($P < 0.0001$) and ranged between 0.81 (length WinRhizo™) and 1.01 (length weighted mean diameter). We observed intercepts smaller than 0 for VWR_{SUM} and significantly larger than 0 for LWR (Table 1). The diameter errors ranged between 10 % underestimation (DWR_{wm}) to 14 % overestimation (DWR), with mean magnitudes of 3 %. The length was underestimated on average by 9 % (range 29 – 0 %, Table 1). WinRhizo™ underestimated the true volume up to 40 and 21 %, VWR and VWR_{SUM}, respectively, with error magnitudes of 17 and 10 %. (Table 1).

The relationships between ground truth and IJ_Rhizo estimates at 1200 dpi were all significant, but surprisingly weak, with R^2 values around 0.2 for diameters, and around 0.7 for length and volume (Table 1). None of the intercepts differed from 0. The diameter errors ranged between 72 % underestimation and 11-12 % overestimation, with mean magnitudes of 21 (DIJ) and 16 % (DIJc). The length was overestimated up to 238 %, which did not represent an outlier, and the maximum underestimation was 16 % (mean magnitude 41 %, Table 1). The true volume was constantly underestimated by a minimum of 5 and a maximum of 80 %. (Table 1).

Change of relationships and magnitude of errors at lower image resolution and fixed threshold

Decreasing the resolution from 1200 to 800 dpi increased the magnitude of errors two- to fourfold for all variables estimated by WinRhizo™. It led to flatter slopes for volume and diameter and significant positive intercepts for diameters (Table 1). R^2 values were not affected for length and volume, but the explanatory power of the relationships of diameter estimates to the respective ground-truth decreased from > 95 % explained variance to 55 and 63 % explained variance (DWR, DWR_{wm}, respectively, Table 1). A further decrease of the image resolution to 600 and 400 dpi had no additional effect on relationships or error magnitudes. The application of a fixed threshold alternative to the automatic threshold detection did not influence the errors or relationships in the case of WinRhizo™.

To our great surprise, a decrease in image resolution from 1200 to 800 dpi substantially increased the accuracy of estimates from IJ_Rhizo. The R^2 values for models increased by 0.20-0.35 and the magnitude of errors decreased by 6 (DIJ), 3 (DIJc), 31 (LIJ) and 13 % (VIJ), whereas for diameter estimates the direction of the error changed from mainly under- to overestimation, while the opposite pattern was observed for the length (Table 1). The decrease to 600 dpi increased the magnitude of diameter errors by 8 %, but did not influence the magnitude of length or volume errors. Relative to the 600 dpi resolution, the magnitude of diameter and volume errors doubled at 400 dpi, while the length estimates were not affected (Table 1). In contrast to what we observed for WinRhizo™, applying a fixed threshold of 144 improved the estimates substantially for IJ_Rhizo (Table 1): At 1200 dpi it increased the explanatory power of the linear models fourfold and reduced the magnitude of errors to one third for diameters. R^2 was not affected at 800 dpi, but the magnitude of errors decreased by 6 %. For the length, we observed a 10-fold decrease of error and an increase in explanatory power of 25 % at 1200 dpi, but no noteworthy effect at 800 dpi (Table 1). Appling a fixed threshold reduced the magnitude of errors of the IJ_Rhizo volume estimates from 33 to 12 % at 1200 dpi, but increased it from 20 to 32 % at 800 dpi.

Dependency of errors on root and image properties at 1200 dpi with fixed threshold

At a resolution of 1200 dpi, errors of the diameter and length estimates of WinRhizo™ were smaller for taproot systems than for fibrous root systems, while diameter class based volume estimates revealed more underestimation for taproot systems (Figure 1). For IJ_Rhizo, errors were independent of the root system structure.

The magnitude of relative diameter errors at 1200 dpi decreased with increasing root diameter for IJ_Rhizo (R^2 = 0.08) but not for WinRhizo™ (Figure 2). It increased with root overlap for WinRhizo™ (R^2 = 0.14) and decreased with CVD for WinRhizo™ (R^2 = 0.09) but increased with CVD for IJ_Rhizo (R^2 = 0.08, Figure 2). The magnitude of the relative errors in length increased with increasing root length (R^2 = 0.38), root length densities (R^2 = 0.43) and higher root overlap (R^2 = 0.21) when images were analysed with WinRhizo™, but not when IJ_Rhizo was used (Figure 3). Because of the covariance of the explanatory variables, we tested their effects in a linear model including all three variables in different orders, which revealed a significant effect of density in all cases, whereas the root length and the overlap were significant only when fit first (analysis not shown).

For the root volume estimates VIJ and VWR, the error decreased linearly with the root diameter (R^2 = 0.07 and 0.06, respectively, Figure 4) and increased with the diameter variability within images (CVD, R^2 = 0.57 and 0.52, respectively), but was independent from the root volume. Contrastingly, the errors of the diameter class based root volume estimates of WinRhizo™ decreased with root volume and slightly increased with CVD (R^2 = 0.22 and 0.07) but were independent of the root diameter (Figure 4).

Discussion

The aim of our analyses was to supplement the length estimate comparison between WinRhizo^{TM} and IJ_Rhizo by Delory et al. (2017) with ground-truthed images and with comparisons of accuracy of diameter and volume estimates. Therefore, we first compare our results with the results of Delory et al. (2017), followed by a discussion of error sources and their implications.

Parallels and differences to Delory et al., (2017)

The analysis presented by Delory et al., (2017) represents a comparison of the accuracy of different possible length estimates of WinRhizo^{TM} and IJ_Rhizo, while we only compared the accuracy of their best performing methods of WinRhizo^{TM} (Regent) and IJ_Rhizo (Kimura).

Our analysis confirms the higher root length estimates of IJ_Rhizo compared to WinRhizo^{TM} and the general underestimation of root length by WinRhizo^{TM} reported by Delory et al., (2017) at 400 dpi.

Our results are in agreement with Delory et al., (2017), regarding the increase of length underestimation with increasing root length density for WinRhizo^{TM} but not for IJ_Rhizo. We further show an increase in error magnitude with root overlap and with total root length for WinRhizo^{TM}. The three measures (RLD, overlap and total length) were highly correlated, hence, we ran linear models including all three variables and tested whether changing the order of variables influenced the outcome (analysis not shown). This analysis revealed that the total length and the overlap only had a significant influence when fit before RLD, which in turn was significant in all models irrespective of its position. The result was confirmed by an analysis following Zuur, Ileno, & Elphick (2010), which revealed RLD as the most and total root length as the least important explanatory variable. Consequently, in our study RLD had an effect on error magnitudes additional to its effect on root overlap.

Overall, we can confirm the conclusion of Delory et al., (2017) that root length estimates derived from different software should not be compared directly, because they differ in the magnitude and direction of errors.

Potential error sources and implications for data interpretation

Several recent papers have shown that the automated measurement of root system parameters can be subjected to non-linear errors (Delory et al., 2017, Lobet et al., 2017, Rose, 2017). Here we show that these errors can have different origins.

Firstly, errors clearly increase with lower image resolution. Furthermore, they change not only in magnitude, but also in direction. This is particularly relevant for diameter estimation, which consequently effects volume estimates. As indicated by flat slopes and low R^2 values (< 0.6), the diameter estimates at 400 dpi are not sufficiently representing the true diameters. The direction of the errors of length estimates changed to more underestimation, but the strength of relationships to the ground-truth were unaffected. This allows interpretation of within software package comparisons of different samples, but no comparisons between software packages or between different resolutions. Although the relationships of volume estimates to the ground-truth remained very strong, we do not recommend to measure volume at 400 or 600 dpi or interpret volume estimates from 400 or 600 dpi images, because the magnitude of errors (average 22-39 %) is not acceptable any more. It should also be noted that, for root systems with smaller diameter range than the one used in this study, it might be worth using images with a resolution higher than 800 dpi. In our sample, the minimal diameter was of root segments was 0.04 mm, which, at 800 dpi, is represented by 2 pixels in the images.

Closely related to the problem of image resolution is the importance to choose the correct threshold for pixels to classify as root or background. Both packages allow to set it manually or to use an automatic threshold and to inspect the results visually (skeletonized images IJ_Rhizo, analysed image WinRhizo™). WinRhizo™ further allows to adjust the threshold using the threshold slider. This is a great advantage for heterogeneous image sets, where batch analysis is not an option. Our comparison of the use of different threshold methods revealed a significant difference between the two software packages, especially at 1200 dpi. WinRhizo™, appears to choose the threshold correctly, at least if images have a clear contrast, while IJ_Rhizo introduced large errors by misclassification. Our images contained minor variability in colour intensity (Supplement 3) as a result of image conversion (Minervini, Scharr, & Tsaftaris 2015), which, nevertheless, introduced directional pixel misclassification towards less pixels being classified as root by IJ_Rhizo at 1200 dpi but not 800 dpi (Supplement 3, zoom to see the effect). This explains the great overestimation of root length and underestimation of diameters we observed with automatic threshold at 1200 dpi, since single root segments are partly interpreted as multiple very thin segments by IJ_Rhizo. The effect appears to average off at 800 dpi. Hence, we can not give general advice as to which threshold method to use, but can only emphasize the importance of checking skeletonized images carefully for each image set.

Secondly, errors can result from naïve assumptions about uniformity of root diameters. This is the case for volume estimates based on average diameter and total root length. Both WinRhizo™ and IJ_Rhizo provide volume estimates, which underestimate the true volume systematically. Further, the relative error strongly increased with the diameter heterogeneity. As already discussed by Rose (2017) for WinRhizo™, we can not recommend to use these volume estimates, but instead use volume estimates based on diameter classes (WinRhizo™) or estimate the volume of root fragments manually sorted into different diameter classes.

Finally, and most importantly, errors relate to image characteristics and root system structure, which makes them harder to avoid or account for. In terms of image characteristics, errors have been linked to the level of overlapping roots (Delory et al., 2017), which was the case for diameters estimated by WinRhizo™. It also related to WinRhizo™ length estimates, but showed collinearity with the RLD of the images. Root length density was, with an average of 1.3 cm cm^{-2}, close to the recommended threshold of 1 cm cm^{-2} (Delory et al., 2017), and therefore we did not expect it to be an important error source. It was, however, related to the magnitude of underestimation of root length by WinRhizo™, even when we accounted for overlap and the total root length in the images (analysis not shown). Hence, we recommend keeping RLD as low as possible, when using WinRhizo™ and below 1 cm cm^{-2} when using IJ_Rhizo to avoid the introduction of biased errors. This will also facilitate the reduction of root overlap.

Regarding root system structures, root diameter, the total root length, and root volume proved to be informative predictors of error magnitudes of diameter, length and volume, respectively. This is especially problematic if the error is not balanced, as is the case for WinRhizo™ at our threshold setting. Increasing length underestimation with increasing root length will lead to an underestimation of length differences if complete root systems are scanned. Length, however, is relatively easy to control for during scanning and should be kept as uniform as possible, even if that means splitting some root systems for scanning and others not.

The more challenging variables to estimate correctly are the diameter and the volume of root samples. The magnitude of both errors decreased with increasing values (diameter: WinRhizo™ not significant, volume: IJ_Rhizo not relevant due to wrong assumptions). We assume larger relative errors at small diameters for stochastic reasons, which then in turn influence errors in volume. The same applies to the negative relationship between volume and relative error of volume estimates. Although the diameter did not prove to have a significant effect on the volume error in a single linear regression, multiple linear regression including both, volume and diameter, revealed a negative effect. This was a result of a negative correlation between volume and diameter in our dataset (analyses not shown). Errors in diameter will always have a strong impact on volume estimates, firstly, because of the exponential relationship between both and secondly, because the volume is cumulative and therefore the error is also accumulated. Volume estimates could be improved either by improving diameter estimates via adjusting the pixel classification and the image resolution (see above) or by using non-image based methods like the Archimedes' method (see Birouste, Zamora-Ledezma, Bossard, Pérez-Ramos, & Roumet 2014).

Conclusions

Several image analyses tools have been developed over the last decades (Lobet, Draye, & Périlleux 2013). We compared one open source (IJ_Rhizo) and one commercial (WinRhizo™) software package regarding the accuracy of root diameter, length and volume estimates at different image resolutions with different threshold methods. Regarding diameter and length, we show that both packages provide estimates of comparable accuracy. However, the direction of errors differed between the two and therefore we recommend to not directly compare results gained from the different tools. Further, we can not recommend to use the automatic threshold methods of IJ_Rhizo without carefully checking the skeletonized images or to use IJ_Rhizo to estimate root volume. If WinRhizo™ is used, the volume provided for the different diameter classes should be summed up, because it is not affected by the naïve assumption of diameter homogeneity. We further do not recommend to use a resolution < 1200 dpi for diameter and volume analyses and to be careful how to interpret length estimates derived from low resolution images.

WinRhizo™ is often used as a reference to evaluate the accuracy of new image analyses tools (e.g. Himmelbauer, Loiskandl, & Kastanek 2004, Pierret et al., 2013). We recommend benchmarking at least a subset of analysed images against a true ground-truth to avoid incorporating errors derived from the WinRhizo™ method.

Finally, since errors also changed with changing sample properties (diameter, total length, diameter variability, volume), we strongly recommend to standardize sample size or sample properties as much as possible. Different types of root systems will lead to different errors. Validation of the calculation method should therefore always be performed internally to the dataset.

We are, however, aware that it is not always possible to standardize by the response variable of interest (e.g. volume). Root researchers, when using digital image analyses software, should take into account that even black and white images are subject to errors, which can change with the magnitude of the variable of interest and incorporate this knowledge into the interpretation of their results.

Acknowledgements

We thank Boris Rewald and Christophe Jourdan for encouraging this study. We further thank Benjamin Delory and Grégoire Freschet for comments on an earlier version of the manuscript.

Conflict of interest

The authors declare no conflict of interests.

Author contributions

GL and LR designed the study and analysed the data. GL provided the root images, LR wrote the manuscript with contributions from GL.

Data accessibility

The data used in this research are available here: http://dx.doi.org/10.5281/zenodo.1159846

References

Bauhus, J. & Messier, C. (1999) Evaluation of fine root length and diameter measurements obtained using RHIZO image analysis. *Agronomy Journal* **91**, 142-147.

Bergmann, J., Ryo, M., Prati, D., Hempel, S. & Rillig, M.C. (2017) Root traits are more than analogues of leaf traits: the case for diaspore mass. *New Phytologist* **216**, 1130-1139.

Birouste, M., Zamora-Ledezma, E., Bossard, C., Pérez-Ramos, I.M. & Roumet, C. (2014) Measurement of fine root tissue density: a comparison of three methods reveals the potential of root dry matter content. *Plant and Soil* **374**, 299-313.

Delory, B.M., Weidlich, E.W.A., Meder, L., Lutje, A., Van Duijnen, R., Weidlich, R. & Temperton, V.M. (2017) Accuracy and bias of methods used for root length measurements in functional root research. *Methods in Ecology and Evolution* **8**, 1594-1606.

Diaz, S., Hodgson, J.G., Thompson, K., Cabido, M., Cornelissen, J.H.C., Jalili, A., Montserrat-Marti, G., Grime, J.P., Zarrinkamar, F., Asri, Y., Band, S.R., Basconcelo, S., Castro-Diez, P., Funes, G., Hamzehee, B., Khoshnevi, M., Perez-Harguindeguy, N., Perez-Rontome, M.C., Shirvany, F.A., Vendramini, F., Yazdani, S., Abbas-Azimi, R., Bogaard, A., Boustani, S., Charles, M., Dehghan, M., De Torres-Espuny, L., Falczuk, V., Guerrero-Campo, J., Hynd, A., Jones, G., Kowsary, E., Kazemi-Saeed, F., Maestro-Martinez, M., Romo-Diez, A., Shaw, S., Siavash, B., Villar-Salvador, P. & Zak, M.R. (2004) The plant traits that drive ecosystems: Evidence from three continents. *Journal of Vegetation Science* **15**, 295-304.

Freschet, G.T., Valverde-Barrantes, O.J., Tucker, C.M., Craine, J.M., McCormack, M.L., Violle, C., Fort, F., Blackwood, C.B., Urban-Mead, K.R., Iversen, C.M., Bonis, A., Comas, L.H., Cornelissen, J.H.C., Dong, M., Guo, D.L., Hobbie, S.E., Holdaway, R.J., Kembel, S.W., Makita, N., Onipchenko, V.G., Picon-Cochard, C., Reich, P.B., De La Riva, E.G., Smith, S.W., Soudzilovskaia, N.A., Tjoelker, M.G., Wardle, D.A. & Roumet, C. (2017) Climate, soil and plant functional types as drivers of global fine-root trait variation. *Journal of Ecology* **105**, 1182-1196.

Himmelbauer, M.L., Loiskandl, W. & Kastanek, F. (2004) Estimating length, average diameter and surface area of roots using two different Image analyses systems. *Plant and Soil* **260**, 111-120.

Kimura, K. & Yamasaki, S. (2003) Accurate root length and diameter measurement using NIH Image: use of Pythagorean distance for diameter estimation. *Plant and Soil* **254**, 305-315.

Lobet, G., Draye, X. & Périlleux, C. (2013) An online database for plant image analysis software tools. *Plant Methods* **9**, 38.

Lobet, G., Koevoets, I.T., Noll, M., Meyer, P.E., Tocquin, P., Pages, L. & Perilleux, C. (2017) Using a structural root system model to evaluate and improve the accuracy of root image analysis pipelines. *Frontiers in Plant Science* **8**.

Lobet, G., Pound, M.P., Diener, J., Pradal, C., Draye, X., Godin, C., Javaux, M., Leitner, D., Meunier, F., Nacry, P., Pridmore, T.P. & Schnepf, A. (2015) Root system markup language: Toward a unified root architecture description language. *Plant Physiology*

167, 617-627.

Minervini, M., Scharr, H., & Tsaftaris S.A. (2015) The significance of image compression in plant phenotyping applications. *Functional Plant Biology* **42**, 971-988.

Pagès, L., Bécel, C., Boukcim, H., Moreau, D., Nguyen, C. & Voisin, A.-S. Calibration and evaluation of ArchiSimple, a simple model of root system architecture. *Ecological Modelling,* **290,** 76-84.

Pierret, A., Gonkhamdee, S., Jourdan, C. & Maeght, J.L. (2013) IJ_Rhizo: an open-source software to measure scanned images of root samples. *Plant and Soil* **373**, 531-539.

R Core Team (2015) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Regent, G. (2013) WinRHIZO 2013 reg, User manual. Quebec, Canada.

Rose, L. (2017) Pitfalls in root trait calculations: How ignoring diameter heterogeneity can lead to overestimation of functional traits. *Frontiers in Plant Science* **8**.

Ryser, P. (2006) The mysterious root length. *Plant and Soil* **286**, 1-6.

Schneider, C.A., Rasband, W.S. & Eliceiri, K.W. (2012) NIH Image to ImageJ: 25 years of image analysis. *Nature Methods* **9**, 671-675.

Wright, I.J., Reich, P.B., Westoby, M., Ackerly, D.D., Baruch, Z., Bongers, F., Cavender-Bares, J., Chapin, T., Cornelissen, J.H.C., Diemer, M., Flexas, J., Garnier, E., Groom, P.K., Gulias, J., Hikosaka, K., Lamont, B.B., Lee, T., Lee, W., Lusk, C., Midgley, J.J., Navas, M.L., Niinemets, U., Oleksyn, J., Osada, N., Poorter, H., Poot, P., Prior, L., Pyankov, V.I., Roumet, C., Thomas, S.C., Tjoelker, M.G., Veneklaas, E.J. & Villar, R. (2004) The worldwide leaf economics spectrum. *Nature* **428**, 821-827.

Zuur, A.F., Ieno, E.N. & Elphick, C.S. (2010) A protocol for data exploration to avoid common statistical problems. *Methods in Ecology and Evolution,* **1,** 3-14.

Table 1: Summary of linear models on the relationships between the correct diameter, length and volume (ground truth) and the same variables estimated by WinR and IJ_R and relative errors of the estimated values at four image resolutions. Errors are calculated as (Estimate-Correct)/ Correct.

	Linear Model Resp ~ ground truth						Errors %			
Res	Interc.	SE	$P_{int.}$	Slope	SE	R^2	min	mean	max	mean abs
DIJ										
400	0.140	0.019	***	0.72	0.09	0.41	-13	39	67	39
600	0.109	0.018	***	0.70	0.08	0.42	-25	21	45	23
800	0.086	0.017	***	0.72	0.08	0.46	-29	13	38	15
800_{fix}	0.045	0.018	**	0.77	0.08	0.48	-43	-2	22	9
1200	0.009	0.034	n.s.	0.77	0.15	0.20	-72	-19	11	21
1200_{fix}	0.009	0.007	n.s.	1.02	0.03	0.91	-8	6	22	7
DIJc										
400	0.126	0.015	***	0.79	0.07	0.56	-5	38	74	38
600	0.102	0.014	***	0.72	0.07	0.56	-14	20	47	21
800	0.083	0.013	***	0.72	0.06	0.62	-29	11	29	13
800_{fix}	0.038	0.013	**	0.78	0.06	0.65	-35	-3	16	7
1200	-0.007	0.035	n.s.	0.89	0.16	0.23	-72	-14	12	16
1200_{fix}	0.015	0.005	**	0.97	0.02	0.94	-5	4	13	5
DWR										
400	0.072	0.015	***	0.62	0.07	0.46	-39	-4	24	9
600	0.057	0.014	***	0.66	0.07	0.51	-43	-7	16	9
800	0.044	0.014	**	0.70	0.06	0.55	-44	-9	8	10
800_{fix}	0.048	0.014	**	0.70	0.07	0.53	-43	-8	14	10
1200	0.008	0.005	n.s.	0.97	0.02	0.95	-6	1	14	3
1200_{fix}	0.0109	0.005	**	0.95	0.02	0.95	-6	1	14	3
DWR_{wm}										
400	0.043	0.011	***	0.67	0.05	0.64	-39	-12	4	12
600	0.034	0.012	**	0.70	0.05	0.64	-44	-13	4	13
800	0.028	0.013	**	0.74	0.06	0.63	-44	-13	3	13
800_{fix}	0.031	0.013	**	0.73	0.06	0.61	-44	-12	4	12
1200	-0.007	0.005	n.s.	1.01	0.02	0.96	-10	-2	7	3
1200_{fix}	-0.004	0.005	n.s.	1	0.02	0.96	-10	-2	7	3
LIJ										
400	3.267	0.875	***	0.82	0.01	0.98	-40	-12	11	12
600	2.652	0.812	**	0.85	0.01	0.99	-41	-10	10	11
800	0.897	0.548	n.s.	0.91	0.01	0.99	-41	-9	12	10
800_{fix}	1.589	0.729	**	0.89	0.01	0.99	-41	-11	9	11
1200	-4.035	7.501	n.s.	1.62	0.10	0.74	-16	40	238	41
1200_{fix}	2.436	0.611	***	0.94	0.01	0.99	-18	1	13	4
LWR										

400	2.354	0.746	**	0.76	0.01	0.98	-41	-20	-5	20
600	1.995	0.722	**	0.77	0.01	0.99	-42	-19	-5	19
800	0.450	0.480	n.s.	0.82	0.01	0.99	-41	-19	-5	19
800_{fix}	1.911	0.714	**	0.78	0.01	0.99	-41	-19	-5	19
1200	3.310	0.688	***	0.81	0.01	0.99	-29	-9	0	9
1200_{fix}	3.265	0.679	***	0.81	0.01	0.99	-29	-9	1	9
VIJ										
400	-0.002	0.001	**	1.57	0.03	0.96	-55	34	90	41
600	-0.001	0.001	**	1.22	0.03	0.95	-67	4	56	22
800	-0.002	0.001	**	1.12	0.02	0.96	-70	-9	40	20
800_{fix}	-0.001	0.000	**	0.83	0.01	0.97	-81	-32	9	32
1200	0.001	0.001	n.s.	0.61	0.04	0.69	-80	-33	-5	33
1200_{fix}	0.000	0.000	n.s.	0.86	0.01	0.98	-27	-12	13	12
VWR										
400	-0.002	0.000	***	0.92	0.02	0.97	-77	-32	5	32
600	-0.002	0.000	***	0.84	0.01	0.98	-82	-34	-2	34
800	-0.002	0.000	***	0.80	0.01	0.97	-82	-34	-3	34
800_{fix}	-0.002	0.000	***	0.84	0.01	0.98	-79	-37	-1	37
1200	0.000	0.000	n.s.	0.86	0.01	0.99	-40	-16	4	17
1200_{fix}	0.000	0.000	n.s.	0.86	0.01	0.99	-40	-17	4	17
VWR_{sum}										
400	-0.002	0.000	***	0.92	0.01	0.99	-79	-32	3	32
600	-0.002	0.000	***	0.86	0.01	0.99	-79	-36	-1	36
800	-0.002	0.000	***	0.88	0.01	0.99	-79	-39	-5	39
800_{fix}	-0.002	0.000	***	0.89	0.01	0.99	-82	-33	-1	33
1200	-0.0004	0.0002	**	0.94	0.01	1.00	-21	-10	7	10
1200_{fix}	-0.0004	0.0002	**	0.94	0.01	1.00	-21	-10	7	10

D = diameter, L = length, V = Volume, IJ = IJ_Rhizo, WR = WinRhizoTM, wm = weighted mean, SUM = Sum of diameter classes, fix = fixed threshold, n.s. = not significant, *, **, *** P(Intercept) < 0.05, 0.01, 0.001. All slopes significant P < 0.0001. Highlighted in grey: parameters used for structural analyses

Figures

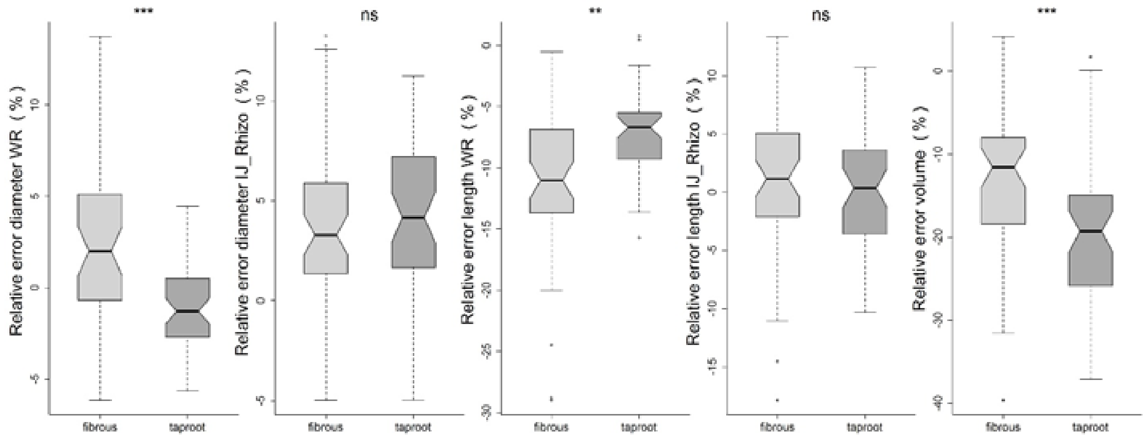

Figure 1: Errors of diameter and length estimates for WinRhizo^{TM} and IJ_Rhizo and diameter class based volume estimates for WinRhizo^{TM} for fibrous and taproot systems at a resolution of 1200 dpi with fixed threshold. Errors are calculated as (Estimate-Correct)/ Correct. **, *** *P* < 0.01, 0.001, respectively, Wilcoxon test.

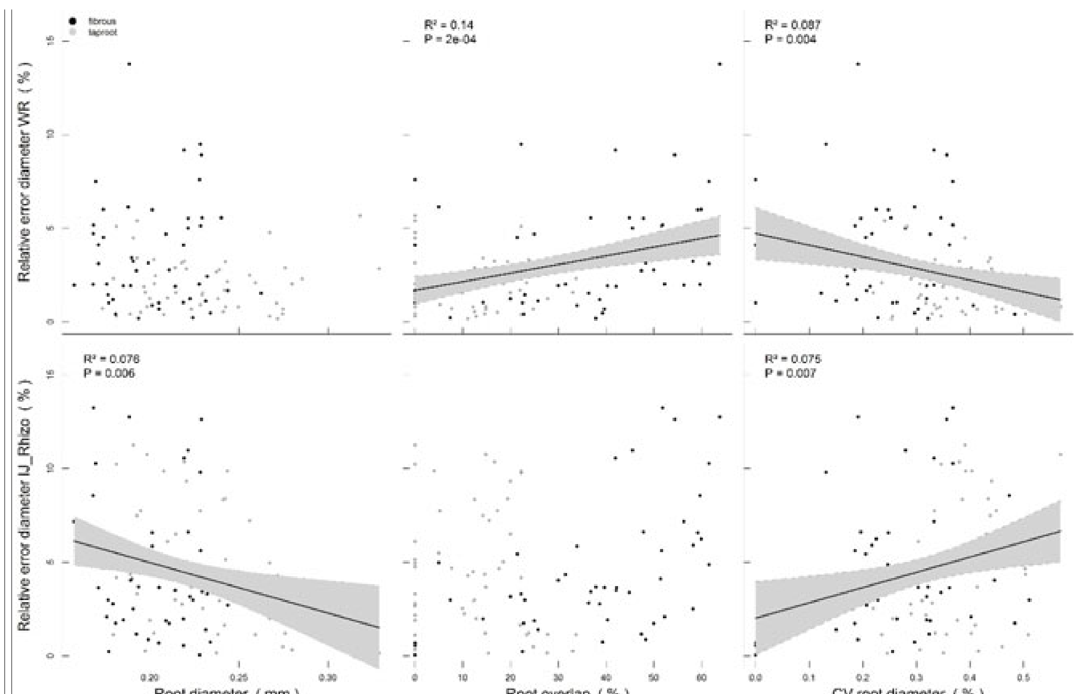

Figure 2: Relationships between the magnitude of the relative error in root diameter and the root diameter, the root overlap and CVD for WinRhizo^{TM} and IJ_Rhizo at 1200 dpi with fixed threshold. Errors are absolutes of (Estimate-Correct)/Correct*100. Root overlap and root diameter were negatively correlated, the other explanatory variables were independent (analysis not shown).

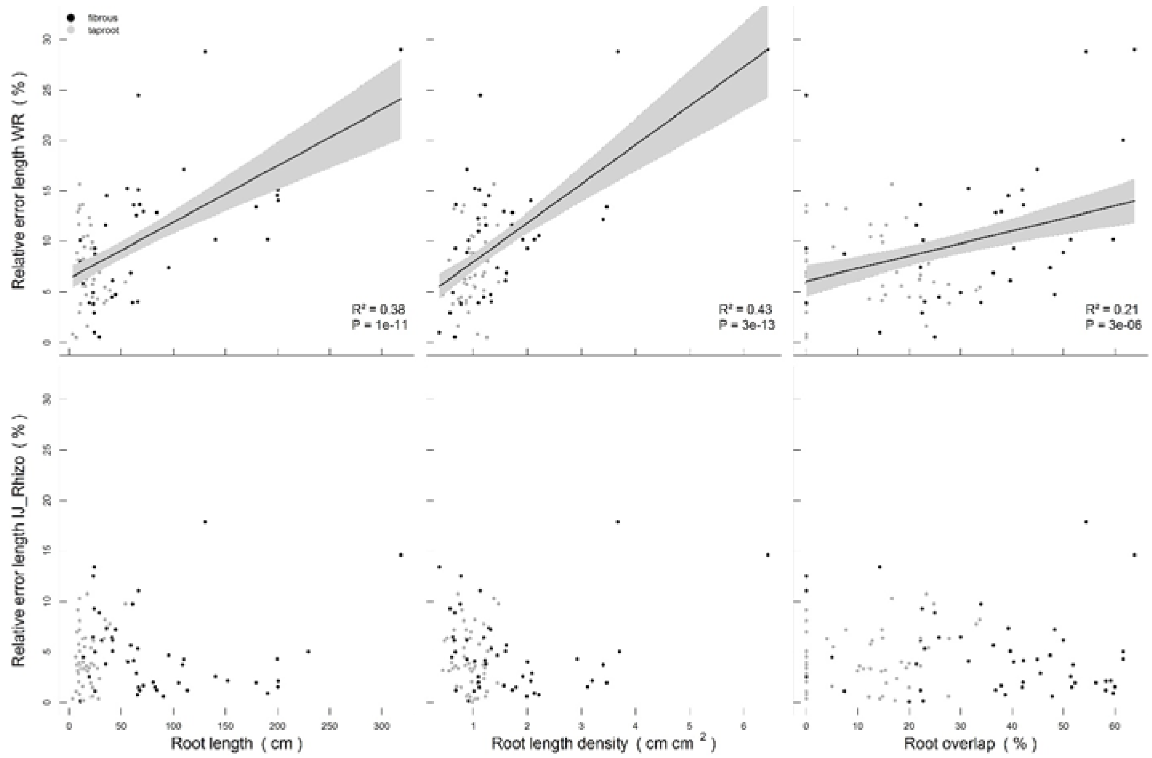

Figure 3: Relationships between the magnitude of the relative error in root length and the root length, root length density and root overlap for WinRhizo^{TM} and IJ_Rhizo at 1200 dpi with fixed threshold. Errors are absolutes of (Estimate-Correct)/Correct*100. All three explanatory variables correlated positively with each other in our dataset.

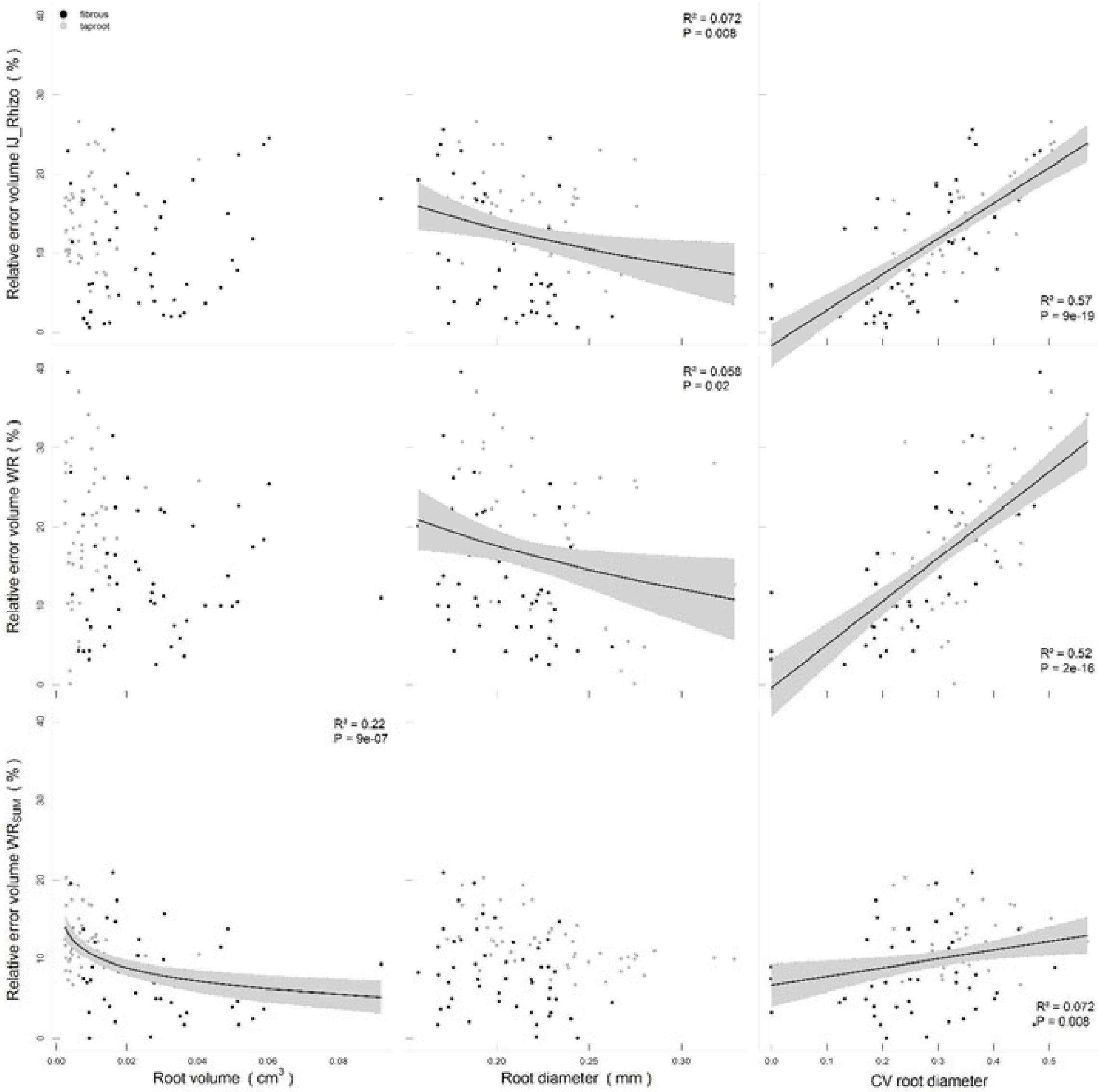

Figure 4: Relationships between the magnitude of the relative error of root volume estimates (VIJ, VWR and VWR_{SUM}) and the root volume, diameter and CVD at 1200 dpi with fixed threshold. Errors are absolutes of (Estimate-Correct)/Correct*100. The root volume correlated negatively with the root diameter in our dataset (analysis not shown).