

A Bayesian model for xylem vessel length accommodates subsampling and reveals skewed distributions in species that dominate seasonal habitats.

Brad Oberle^{1,2,3}, Kiona Ogle⁴, Juan Carlos Penagos Zuluaga⁵, Jonathan Sweeney⁵, Amy E. Zanne^{1,2}

¹Department of Biological Sciences, George Washington University, Washington, DC 20052, USA ²Center for Conservation and Sustainable Development, Missouri Botanical Garden, St. Louis, MO 63166, USA; ³Division of Natural Sciences, New College of Florida, 5800 Bay Shore Rd., Sarasota FL 34243, USA; ⁴Informatics and Computing Program, Northern Arizona University, PO Box 5693, Flagstaff AZ 86011, USA; ⁵Department of Biology, University of Missouri-St. Louis, St. Louis, MO 63121, USA.

Corresponding author: Brad Oberle, boberle@ncf.edu

Date of submission: November 3, 2015

Date of publication: January 27, 2016

Abstract

Vessel length is an important but understudied dimension of variation in angiosperm vascular anatomy. Among other traits, vessel length mediates an important tradeoff between hydraulic efficiency and safety that could influence how plants respond to extreme weather with climate change. However, the functional significance of vessel length variation within individual stems is poorly known, in part because existing data analysis methods handle uncertainty in a way that makes vessel length distributions difficult to compare. We provide a solution to this problem through a hierarchical Bayesian framework for estimating vessel lengths and we demonstrate the flexibility of this method by applying it to data from serial cross sections of dye injected stems. Our approach can accelerate data collection and accommodate associated uncertainties by statistically correcting for bias and error that result from subsampling images. We illustrate our analytical framework by estimating and comparing vessel length distributions for 21 woody species characteristic of a North American forest. The best-fit model corrected for both bias due to secondary growth and sampling error within and among species. Vessel length estimates from this model varied by almost an order of magnitude and parameters of these distributions correlated with point estimates derived from a different, commonly used method. Furthermore, we show how key contrasts can be estimated with the Bayesian framework, and in doing so, we show that the shape of the vessel length distribution differed between ring- and diffuse-porous species, suggesting that within-stem vessel length variation corresponds to water stress seasonality and contributes to landscape-level habitat segregation. Our analysis method revealed the importance of within-stem variation in vessel length, and our results complement work on between-species variation in average vessel length, further illuminating how vascular anatomy can influence woody plants' responses to water stress.

Introduction

Xylem vessels are a key feature of angiosperm vascular anatomy (Jacobsen *et al.* 2012, Tyree and Zimmerman 2002), and their geometry has important implications for plant hydraulic function. Vessel length and width mediate a tradeoff between hydraulic efficiency and safety (Tyree *et al.* 1994, Hacke *et al.* 2006). Wider vessels increase maximum sap flow rates at greater risk of freezing-induced embolism (Cai and Tyree 2010, Davis *et al.* 1999, Hacke *et al.* 2006). Likewise, longer vessels have fewer high resistance end walls (Sperry *et al.* 2005), but may also propagate air embolisms further along the hydraulic pathway (Christman *et al.* 2009, Lens *et al.* 2011). The rare pit hypothesis predicts that long vessels have more pits and a higher probability that one will spread cavitation-induced air bubbles that increase the risk of mortality during drought. Consistent with this prediction, mean vessel length was one of the best predictors of cavitation resistance among seven species in the maple genus *Acer* (Lens *et al.* 2011). Given that trees store tremendous quantities of carbon, predicting which species are most vulnerable to drought has major carbon cycle implications as extreme weather becomes more likely with climate change (Allen *et al.* 2010, McDowell *et al.* 2008).

While species mean vessel length can help predict drought tolerance, this single number belies the complex distribution of vessel length within stems. In many plants, the very longest vessels can be meters long while the vast majority span only a few millimeters (Tyree and Zimmermann 2002). The functional significance of within stem variation in vessel length is largely unknown but may relate to patterns of seasonal variation in vessel diameter. Species with ring-porous wood produce wider vessels early in the growing season and transition to narrower vessels later when moisture becomes scarce (Carlquist, 1975). In contrast, diffuse-porous species produce vessels of similar diameter throughout the growing season. The prevalence of these two wood porosity types varies with seasonality at different scales (Woodcock, 1989). Globally, ring-porous species dominate seasonal temperate forests while diffuse-porous species dominate less seasonal tropical forests (Wheeler 2007, Zanne *et al.* 2006, Pan and Kudo 2012). Within a temperate North American forest, ring-porous species dominate relatively seasonal ridge and slope habitats while diffuse-porous species are more common less seasonal valley bottoms (Zimmerman and Wagner, 1979). To the extent that vessel length and width are correlated (Cai *et al.* 2010, Zimmermann and Jeje, 1981), the decrease in vessel diameter in ring-porous species may either represent avoidance of late season drought stress or tolerance of winter freezing (Ellmore and Ewers, 1986, Sperry *et al.* 1994, Sperry and Sullivan, 1992). Knowing which has important implications for predicting how changes in seasonality will affect tree species' distributions. However, the anatomical basis for apparent differences in performance is unclear because most analyses have focused strictly on within-stem variation in vessel diameter and not vessel length (Lens *et al.* 2011).

Variation in vessel length is relatively understudied in part because the extreme geometry of vessels makes them very difficult to measure (Comstock and Sperry 2000, Ewers and Fisher 1989). Most existing estimates are based on the results of stem infusion experiment which involve injecting a fluid into the cut surface of a stem and tracking its distribution across a series of increasingly distant scoring locations (Cai and Tyree 2014, Scholz *et al.* 2013, Tyree and Zimmerman 2002). Prior to a very recent method based on air injection (Pan *et al.* 2015), injected fluids carried dyes from the injection site down the stem through open vessels. With dye injection, any and all vessels with dye at the scoring location are at least as long as its distance from the injection site provided that two additional assumptions are met (Andre 2005): (1) all vessels terminate within the stem, not side branches and (2) dye moves freely through perforation plates along vessels but not through pit pores between vessels. To meet the first assumption, researchers must inject from stem apex to base (i.e. basipetally) (Lens *et al.* 2011, Scholz *et al.* 2013). Otherwise, the injection medium may flow through vessels out of the main stem into side branches (Zimmermann and Jeje 1981). To meet the second assumption, researchers have developed a silicon based injection medium that moves freely through perforation plates (Hacke *et al.* 2007, Scholz *et al.* 2013). The high resolution necessary to score vessels for dye under UV microscopy can be too fine to image entire cross sections, so researchers sometimes calculate the density of filled vessels within radial subsamples, or sectors, of a cross section (Cai *et al.* 2010, Sperry *et al.* 2005). While this approach can increase accuracy and speed, it may also introduce bias due to secondary growth. Secondary growth tends to add vessels in the basipetal direction that are not continuations of those present at the injection site, cannot receive dye and therefore lower the observed density of vessels due to vessel addition rather than vessel termination.

Following infusion, estimating vessel length requires models that relate the observed changes in the frequency of dyed vessels to the distribution of vessel lengths. The most widely used contemporary approach recognizes that the decreasing frequency of open vessels reflects the increasing cumulative probability of vessel termination with vessel length (Cai and Tyree 2014, Nijssse 2004). Under this interpretation, vessel length depends on the parameters of this cumulative distribution function (CDF). Several studies of vessel length have fit an exponential CDF to observed variation in dyed vessel counts (Cohen *et al.* 2003, Sperry *et al.* 2005, Cai *et al.* 2010, Pan *et al.* 2015). Empirical data may exhibit more or less variance than expected under the exponential CDF (Hacke *et al.* 2007), suggesting that the assumptions of the exponential CDF are often too restrictive to accurately represent vessel length. A generalization of the exponential CDF, the Weibull CDF, incorporates an additional parameter that allows the probability of vessel

termination to vary with vessel length and may be more appropriate for vessel length distributions. Christman, Sperry and Adler (2009) described a method for fitting the Weibull distribution to infusion experiment data (hereafter referred to as the CSA method). While the CSA method is a major improvement over previous approaches, it has some limitations. First, the CSA method assumes a complete and accurate cross section of the stem at every scoring location, precluding radial subsampling and complicating data collection for species with many small vessels. Second, the CSA method only produces point estimates without estimates of uncertainty, which complicates formal comparisons among groups of species. Finally, the CSA method is implemented using proprietary software (i.e. Microsoft (2010) Microsoft Excel, Redmond, Washington) that relies on an optimization algorithm that may not converge on a solution.

The primary objective of this study is to develop a more rigorous analytical method for estimating vessel length distributions that accounts for biases introduced by, for example, secondary growth, and that explicitly quantifies uncertainty via a probability foundation. Second, we aim to apply this analytical method to better understand the variation in angiosperm vessel length by comparing vessel length distributions among 21 co-occurring temperate woody plant species with different wood porosity types. We illustrate how our analytical method can be used to rigorously compare vessel length characteristics among different taxa or functional groups to evaluate the potential hydraulic consequences of variation in vessel lengths. For example, we hypothesized that if different wood porosity types represent different strategies for managing water stress, then species with more variable vessel diameters (i.e., ring-porous wood) would also have more variable vessel length distributions. In particular, we developed a hierarchical Bayesian model for quantifying vessel length distributions, and for conducting relevant contrasts. Importantly, our model accommodates subsampling within stems by statistically correcting for biases that result from secondary growth. Second, our model propagates sampling error within stems and between species for accurate estimates of uncertainty that enable formal comparison of vessel length distributions among groups of species. Third, the Bayesian model is highly flexible and can be modified to accommodate different sampling strategies and measurement approaches. Finally, we implement the Bayesian model in open source software with robust performance (Gelman *et al.* 2004, Ogle and Barber 2012), and make the code available.

Materials and Methods

Sampling methods

Site description

To illustrate our Bayesian approach to estimating vessel length distributions, we analyzed 21 woody species growing at the Tyson Research Center in Missouri, USA (Table 1). Approximately 85% of the site consists of forests on steep limestone ridges that are dominated by oak (*Quercus*) and hickory (*Carya*) species, the remaining forested habitats occur in bottomlands and are dominated by American sycamore (*Platanus occidentalis*) and maple (*Acer*).

Table 1: Study species, wood porosity type (Wheeler 2011), canopy position, and sampling effort information, including visual (v) versus automated (a) image scoring for 21 woody species characteristic of the study site.

Species	Family	Wood porosity	Growth form	Stems	Sectors	Vessels	Scoring
<i>Acer rubrum</i>	Sapindaceae	Diffuse	Canopy Tree	3	63	3097	v
<i>Aesculus glabra</i>	Sapindaceae	Diffuse	Canopy Tree	3	57	4735	a
<i>Amelanchier arborea</i>	Rosaceae	Diffuse	Small Tree	3	54	5117	a
<i>Asimina triloba</i>	Annonaceae	Ring	Shrub	2	39	2433	a
<i>Carya tomentosa</i>	Juglandaceae	Ring	Canopy Tree	3	63	1322	a
<i>Celtis occidentalis</i>	Cannabaceae	Ring	Canopy Tree	3	57	2157	a
<i>Cercis canadensis</i>	Fabaceae	Ring	Small Tree	2	41	1247	a
<i>Cornus florida</i>	Cornaceae	Diffuse	Small Tree	3	59	4149	a
<i>Diospyros virginiana</i>	Ebenaceae	Semi-ring	Canopy Tree	3	62	1814	a
<i>Fraxinus americana</i>	Oleaceae	Ring	Canopy Tree	3	60	853	a
<i>Gleditsia triacanthos</i>	Fabaceae	Ring	Canopy Tree	3	58	1019	a
<i>Juglans nigra</i>	Juglandaceae	Semi-ring	Canopy Tree	2	42	3772	v
<i>Lonicera maackii</i>	Caprifoliaceae	Semi-ring	Shrub	3	63	12795	v
<i>Platanus occidentalis</i>	Platanaceae	Diffuse	Canopy Tree	3	63	4714	a
<i>Prunus serotina</i>	Rosaceae	Diffuse	Canopy Tree	3	62	6897	a
<i>Quercus alba</i>	Fagaceae	Ring	Canopy Tree	3	63	4296	v
<i>Quercus rubra</i>	Fagaceae	Ring	Canopy Tree	3	58	2996	a
<i>Quercus velutina</i>	Fagaceae	Ring	Canopy Tree	3	62	3194	a
<i>Sassafras albidum</i>	Lauraceae	Ring	Small Tree	2	37	1527	v
<i>Ulmus rubra</i>	Ulmaceae	Ring	Canopy Tree	3	52	931	a
<i>Vitis vulpina</i>	Vitaceae	Ring	Liana	3	60	973	a

Sample collection and preparation

During the summer and fall of 2009 and 2011, we identified two to three adult trees from each species growing in typical habitats and collected one stem per tree for a total of two to three replicate stems per species. We stored stems temporarily (a few days to a few weeks) wrapped in plastic bags at 4°C before recutting stems underwater to a total length of 50 cm as measured from the terminal bud. The exact location of the apical cutting site was the closest at which the stem fit into the 0.6 cm diameter compression gland system (PMS Instrument Co. Albany OR, USA), which was typically within 0.1-1 cm of the bud itself. After recutting, we submerged stems in a water bath for 24 hours. Previous studies have suggested little variation in vessel length distributions estimated from fresh versus rehydrated stems (Scholz *et al.* 2013), but some embolisms may have persisted in our stems after submersion.

We prepared silicone by combining a 10:1 ratio of silicone:hardener mix (RTV-141, Rhodia, Cranbury, NJ, USA) with 1% fluorescent optical brightener (Ciba Uvitex OB, Ciba Specialty Chemicals, Tarrytown, NY, USA) (Christman *et al.* 2009). The silicone mixture was stored in a freezer for 24 hours to release air bubbles and avoid hardening (Andre 2005). Upon returning the mixture to room temperature, we placed apical ends of stems into small plastic bags filled with the silicone mixture and sealed them with paraffin film (Scholz *et al.* 2013). Once the bagged apical end was secured in the pressure chamber (Model 1000, PMS Instrument Co. Albany OR, USA), the chamber was pressurized to ~0.5 MPa and left for approximately 4 hours with the opposite end exposed to ambient pressure. This injection protocol may not have been optimal for every specimen. However, considering our goal was to compare variation in vessel length within stems and among species, we chose to inject under standard conditions and let anatomical differences influence silicone movement rather than introducing variation with differing injection protocols. During some injections, silicone traversed the length of the stem indicating that the injection conditions were sufficient to completely fill long vessels in those specimens although we note that complete filling of some of the longest vessels does not guarantee complete filling of especially narrow long vessels. Following silicone injection, we allowed stems to cure at ambient temperature for at least 24 hours.

After the silicone cured, we sectioned stems at 1, 3, 5, 9, 17, 33, and 40 cm from the injection point. We then obtained cross sections using a GSL-1 microtome (Swiss Federal Research Institute WSL, Switzerland) in a range of 15 to 30 μm thick depending on the brittleness of the wood. We semi-permanently mounted cross sections on slides using a glycerol/ethanol mixture. Tearing during sample preparation damaged portions of certain cross sections. Instead of resectioning the stem, which would introduce variation in scoring locations, we identified undamaged radial sectors for further analysis.

Image capture and analysis

Stem cross-sections were imaged using a Zeiss Axioskop UV microscope (Oberkochen, Germany), 5x objective, and a mounted black and white digital camera. Images were taken for up to three complete undamaged sectors per cross section, extending from the edge of the pith to the vascular cambium roughly following radial parenchyma (Table 1). For the larger sectors, multiple images were taken in sequence and later stitched together using the Gnu Image Manipulation Program (GIMP v 2.6).

The images were analyzed with ImageJ software (NIH 2012). Sectors for five species were scored visually (Table 1). Sectors for the remaining species were automatically scored in ImageJ with a minimum vessel area of 78.5 μm^2 and a minimum roundness of 0.8 to minimize scoring errors. All vessels that met these criteria were scored automatically for the presence of fluorescent dye. Preliminary analyses of stems scored both visually and automatically produced similar results (data not shown).

Model description

The data for estimating the parameters of the vessel length distribution consist of counts of dyed and un-dyed vessels in radial sectors at a series of sampling locations (assigned the indexing variable i) within stems. Stems are drawn from different species (j), and species are representative of different groups (k). This nested sampling scheme implies a three-level hierarchy: scored images that are nested within species, and species that are nested within groups. To evaluate the level of complexity necessary to model the data, we compared a series of five increasingly complex models that successively account for the effects of secondary growth and sampling error at different levels. The first and simplest model (A) calculated vessel length based solely on the observed density of dyed vessels in each scored sector, ignoring effects of secondary growth, sampling effort, and group-level variation. The second model (B) built upon model A by explicitly incorporating the effects of secondary growth. The third model (C) expanded on model B by incorporating sampling error. The fourth model (D) expanded on model C by allowing vessel length distributions to differ between canopy and understory species, as defined by the canopy stratum in which plants typically reproduce (Table 1). The rationale for including different canopy positions was simply to improve species-level estimates by partially pooling estimates within these groups (Gelman and Hill 2007), not to test *a priori* hypotheses for variation between groups. The fifth model (E) was similar to the model D, but rather than using canopy position as a grouping

variable, it it used wood porosity types (i.e., ring-, semi-ring and diffuse-porous wood, Table 1), to groups species with *a priori* expected differences. By testing this series of models, we can evaluate the importance of (1) correcting for secondary growth (i.e., model A versus all others), (2) accounting for sampling effects in the context of a model that corrects for secondary growth (i.e., model B versus models C,D and E) and (3) specifying a hierarchical prior for the species-level effects nested in different ecological groupings (i.e., model C versus models D and E). We will present the model starting at the most basic level of this hierarchy (scored sectors within species), which is most similar to existing models.

Weibull distribution of vessel length in the basipetal direction

From the injection site below the apical bud, we assumed that the probability, F , of terminal wall formation—i.e., the probability that an element in the population of vessels changes from dyed (open) to not dyed (closed) with increasing distance, d , from the injection site—is defined by the Weibull CDF for $j = 1 \dots n$ species:

$$F(d) = 1 - e^{-(\lambda_j d)^{c_j}} \tag{1}$$

where λ is the scale parameter and c is the shape parameter. Eq. 1 can be expressed in a simpler form by employing a double log transformation:

$$\ln(-\ln(1 - F(d))) = c_j \ln(\lambda_j) + c_j \ln(d) \tag{2}$$

This transformation linearizes the Weibull CDF, such that the slope equals the shape parameter, and the intercept is the product of the shape and (log) scale parameters. Based on Eq. 2, a goal of our analysis was to estimate λ_j , c_j , and the expected value of d , $E(d)_j$, which is the mean vessel length for each species. Under the Weibull CDF, $E(d)_j$ is given by:

$$E(d)_j = \frac{1}{\lambda_j} \Gamma\left(1 + \frac{1}{c_j}\right) \tag{3}$$

where Γ is the gamma function.

At the level of the scored stem sector, we assumed that the number of open vessels, O , at scoring distance d follows a Binomial distribution. Thus, for $i = 1, 2, \dots, m$ scored sectors:

$$O_i \sim \text{Binomial}(1 - F(d_i), T_i) \tag{4}$$

Where $F(d)$ is given in Eq. 1 and we assume that the parameters (i.e., c and λ) in Eqs. (1) and (2) vary by species, j , thus, $1 - F(d)$ represents the species-specific probability of remaining open at distance d , and T represents the total number of vessels in the population present at the injection site. Equations 1 – 4 represent a simple, non-hierarchical model for vessel length corresponding to model A in our analysis.

Effects of secondary growth

For models B-E, we also estimated the number of vessels in each scored sector that would have been present at the injection site by explicitly modelling and statistically correcting for vessel addition due to secondary growth. We assumed secondary growth adds vessels as an overdispersed Poisson process, which can be represented by the Negative Binomial distribution (Gelman and Hill 2007), such that the increase in total vessel counts, T , with increasing distance from the apical injection site, d , is given by:

$$T_i \sim \text{Negative Binomial}(\mu(d_i), \theta_{j(i)}) \tag{5}$$

where $j(i)$ denotes species j associated with image i , θ is the species-specific overdispersion parameter, and $\mu(d_i)$ is the expected total number of vessels in a scored sector at a given distance. We use a log-link function such that $\ln(\mu(d_i))$ varies linearly with distance, and we allow the parameters in this model to vary by species:

$$\ln(\mu(d_i)) = \alpha_{j(i)} + \beta_{j(i)} d_i \tag{6}$$

where α gives the species-specific (log) expected number of vessels at the injection site ($d = 0$), and β describes the species-specific (log) expected number of vessels added per distance increment (e.g., per m) by secondary growth. Thus, for any scored sector, we estimate the total number of vessels, $\hat{T}(d_i)$, at distance d that are expected to have been present at the injection site as:

$$\hat{T}(d_i) = \frac{T_i}{e^{\beta_{j(i)} d_i}} \tag{7}$$

Substituting $\hat{T}(d_i)$ for T_i in Eq. 4 provides an estimate of the total number of vessels in the sector that is statistically independent of vessel addition. Thus, the model for the observed number of open vessels becomes:

$$O_i \sim \text{Binomial}(1 - F(d_i), \hat{T}(d_i)) \quad (8)$$

The model articulated above is the simplest that estimates species' vessel lengths while correcting for the effects of secondary growth and corresponds to model B in our analysis.

Hierarchical parameter model

For the remaining models (C-E), we build from Eqs. 1-8 by accounting for how differences in sampling effort among species could affect estimates of the vessel length distributions (e.g. Eq. 1). To do this, we treated species as “random effects” such that species-specific parameters of the vessel length distribution (i.e. c_j and λ_j) were treated as random samples from a larger population. This approach borrows strength (Gelman and Hill 2007, Ogle *et al.* 2014) from well-sampled species to inform vessel length parameter estimates of species with fewer samples. In practice, this amounts to estimating the hyperparameters of the population distribution of the slope (i.e., $a = c_j \ln(\lambda_j)$) and intercept (i.e., c_j) in Eq. 2. Because these two terms are expected to be correlated due to the shared factor c_j , we explicitly modeled the covariance structure with a multivariate normal such that for each species j :

$$\begin{pmatrix} a_j \\ c_j \end{pmatrix} \sim \text{Multivariate Normal} \left(\begin{pmatrix} \mu_a \\ \mu_c \end{pmatrix}, \begin{pmatrix} \sigma_a^2 & \rho\sigma_a\sigma_c \\ \rho\sigma_a\sigma_c & \sigma_c^2 \end{pmatrix} \right) \quad (9)$$

where μ_a and σ_a are the mean and standard deviation, respectively, for the population of species-specific intercepts (a 's), μ_c and σ_c are the mean and standard deviation, respectively, for the population of species-specific slopes (c 's), and ρ is the correlation coefficient between the two parameters (a and c). The model above explicitly considers within species sampling error and corresponds to model C in our analysis. A similar approach is possible for modelling species-level parameters of the secondary growth model (e.g. θ , α , and β , in Eqs. 5-6) as random effects, but because our focus is on species-level estimation of vessel length we do not do so in the current analysis.

The overall means (e.g., μ_a and μ_c in Eq. 9) may also differ among groups of species. A grouping with $k = 1, 2, \dots, K$ levels would result in K group-level parameter estimates of $\mu_{a,k}$ and $\mu_{c,k}$. Assigning species to different groups can allow for borrowing of strength within each group, potentially improving parameter estimation at the species level and allowing for direct comparisons among groups with *a priori* expected differences. We evaluated models that used two alternative grouping schemes. First, we grouped species by canopy position as either understory or overstory ($K = 2$ groups). This model corresponded to model D in our analysis. Then we grouped species by wood porosity type ($K = 3$ groups), which corresponded to model E in our analysis.

The above model for vessel lengths from subsampled, dye injected stems belonging to different groups of species is “hierarchical” in the sense that the distribution of the O data ultimately depends on the T data via $\hat{T}(d)$, the distributions of O and T each depend on parameters that are allowed to vary by species (e.g., θ , α , β , λ , c), and the species-level parameters that determine vessel length (e.g. λ , c) depend on group-level hyper-parameters (i.e., μ_a , μ_c , σ_a , and σ_c). We implemented this hierarchical model in a Bayesian framework (Gelman *et al.* 2004), yielding posterior distributions for both species-level quantities of interest (i.e., θ , α , β , λ , c , and $E(d)$) as well as group level quantities of interest. We assigned non-informative, independent priors to the species-level parameters for the secondary growth model, including diffuse normals (mean of zero, variance of 1000) for α and β in Eq. 6 and wide uniform priors, $U(0,100)$, for θ in Eq. 5. We assigned wide uniform priors, $U(0,100)$, for the standard deviation parameters σ_a and σ_c and a uniform prior $U(-1,1)$, for the correlation parameter ρ in Eq. 9. Finally, at the group level, we assigned independent diffuse normal priors (mean of zero, variance of 1000) for μ_a and broad uniform priors $U(0,100)$ for μ_c in Eq. 9.

Model evaluation

We compared models A-E using the posterior predictive loss criterion, D which is defined as the sum of a goodness of fit term, G , and a model complexity penalty term, P , models with lower values of D are considered more adequate (Gelfand and Ghosh 1998). We calculated D for species individually and for the dataset overall by comparing observed data to data simulated under the models. Based on the model with the lowest overall D values, we

further evaluated the overall model fit by calculating the coefficient of determination (r^2) between the predicted (i.e., the mean of the data simulated under the model) and the observed open count values.

In addition to comparing the adequacy of and fit of different models, we tested the accuracy of the simplest model that accounted for both secondary growth and sampling error (Model C) using a set of Monte Carlo simulations (e.g. Tyree and Zimmerman 2002). The simulations spanned the range of vessel length distribution parameter space as represented by four species, *A. glabra*, *A. arborea*, *A. triloba* and *G. triacanthos*, that exhibited extreme values of the shape and scale parameters as estimated from the most adequate model (lowest D , see Results). For each species, we reconstructed twenty infusion experiments typical of our analysis, with 60 vessels scored at each distance in three stems. The number of open vessels at each distance was determined by random draws from the Weibull distribution with each species' posterior mean for c and λ (Eq. 1). The total number of vessels at each distance was adjusted by adding vessels as random draws from the Poisson distribution with a rate parameter based on distance and the species' posterior mean for β , (Eq. 6). We then analyzed the 20 simulated datasets for each species under Model C with the same implementation methods described below. We noted the frequency that the values of the shape and scale parameters used for the simulation were contained within the 95% credible intervals for these parameters.

We also compared species-level estimates of mean vessel length, $E(d)$, and the Weibull parameters (c and λ , Eq. 1) obtained from the most adequate model (lowest D) to those obtained from the CSA method (Christman *et al.* 2009). The CSA method derives vessel length from a single count of all open vessels across the stem cross section and thus does not account for variation among sections. Because our data come from multiple stems, we pooled all intraspecific data at each distance. In many cases, the pooled data consisted of different numbers of sectors (Table 1). While our analysis method accommodates variation in sampling effort, to apply the CSA method, we corrected for sampling effort by dividing the observed pooled open vessel count by the proportion of scored sectors out of the maximum ($n = 9$). Also, the CSA method requires a count of the total number of vessels present at the injection site. To estimate this number for our pooled data, we multiplied the exponentiated posterior mean of the intercept (α , Table 2) by the maximum sampling effort ($n = 9$).

Table 2: Posterior means and 95% credible interval (CI) limits (i.e., 2.5th and 97.5th percentiles) for parameters describing the process of adding vessels by secondary growth (see Eq. 5-6).

Species	β (Eqn (6))			α (Eqn (6))			θ (Eqn (5))		
	Mean	2.5%	97.5%	Mean	2.5%	97.5%	Mean	2.5%	97.5%
<i>Acer rubrum</i>	1.35	0.22	2.49	3.67	3.43	3.92	0.46	0.31	0.65
<i>Aesculus glabra</i>	2.22	1.13	3.29	4.05	3.84	4.27	0.30	0.20	0.43
<i>Amelanchier arborea</i>	2.85	1.84	3.92	4.02	3.80	4.25	0.30	0.21	0.45
<i>Asimina triloba</i>	1.77	0.23	3.35	3.88	3.59	4.19	0.46	0.29	0.73
<i>Carya tomentosa</i>	0.11	-0.73	0.96	3.03	2.86	3.21	0.19	0.11	0.30
<i>Celtis occidentalis</i>	2.16	1.44	2.86	3.22	3.07	3.39	0.13	0.08	0.20
<i>Cercis canadensis</i>	1.78	0.73	2.86	3.10	2.87	3.34	0.21	0.13	0.35
<i>Cornus florida</i>	0.13	-0.71	1.03	4.24	4.06	4.42	0.23	0.15	0.33
<i>Diospyros virginiana</i>	1.69	0.79	2.57	3.09	2.90	3.27	0.20	0.13	0.30
<i>Fraxinus americana</i>	1.31	0.56	2.07	2.43	2.26	2.60	0.11	0.06	0.20
<i>Gleditsia triacanthos</i>	0.60	-0.47	1.71	2.78	2.56	3.00	0.27	0.16	0.41
<i>Juglans nigra</i>	1.45	0.69	2.27	4.25	4.09	4.42	0.12	0.07	0.19
<i>Lonicera maackii</i>	3.60	2.60	4.62	4.62	4.41	4.84	0.33	0.23	0.47
<i>Platanus occidentalis</i>	0.70	-0.07	1.47	4.20	4.04	4.37	0.19	0.13	0.28
<i>Prunus serotina</i>	0.17	-0.53	0.90	4.69	4.54	4.85	0.17	0.11	0.24
<i>Quercus alba</i>	2.33	1.68	2.99	3.81	3.67	3.95	0.13	0.09	0.20
<i>Quercus rubra</i>	-0.41	-1.23	0.44	4.01	3.83	4.20	0.20	0.13	0.29
<i>Quercus velutina</i>	1.86	1.01	2.74	3.62	3.43	3.81	0.24	0.16	0.35
<i>Sassafras albidum</i>	1.52	0.43	2.64	3.45	3.20	3.69	0.22	0.13	0.35
<i>Ulmus rubra</i>	1.44	0.19	2.74	2.67	2.42	2.91	0.31	0.19	0.50
<i>Vitis vulpina</i>	0.64	-0.30	1.59	2.70	2.51	2.89	0.20	0.12	0.30

Finally, the CSA method requires a minimum vessel length value, we used the default of 1 mm for every species. Based on our pooled data, we calculated parameters for the Weibull distribution and mean vessel length using the Microsoft Excel worksheets available at <http://biologylabs.utah.edu/sperry>. We report the point estimates for every species for which the default Generalized Reduced Gradient (GRG) nonlinear solver (Lasdon *et al.* 1978) converged on a solution (Table 3). We then compare our estimates to those on the transformed data by calculating Pearson's correlation coefficients and the 95% confidence interval for the slope of a regression through the origin

Hypothesis testing

We examined how vessel length distributions vary among species of different wood porosity types (e.g. diffuse, semi-ring, and ring-porous) in the context of model E. Diffuse-porous species show little variation in vessel diameter during the growing season, whereas ring-porous species show variation in vessel diameter between earlywood and latewood (Carlquist 1975). Semi-ring-porous species are considered intermediate between the first two. Because the boundaries between semi-ring-porous and other categories can be ambiguous, we calculated a series of different contrasts. First, we contrasted diffuse-porous species against all other species. Then we contrasted ring-porous species against all other species. Finally, we maintained semi-ring-porous species as a third distinct category. We expected that ring-porous species would have vessel length distributions with larger $E(d)$, reflecting longer vessels, and lower values of the shape parameter, c , indicating more highly skewed distributions consisting of both short latewood vessels and long earlywood vessels. We considered a contrast statistically significant if the 95% posterior credible interval for the difference excluded zero.

Model implementation

We implemented models A-E in the open-source, freely available Bayesian statistical software program, OpenBUGS, V3.2.1 rev 781 (Lunn *et al.* 2009) (code provided as Supplementary Material). OpenBUGS employs Markov chain Monte Carlo (MCMC) algorithms to sample from the posterior distributions of model parameters. We ran three parallel chains for 15000 iterations per chain. We assessed convergence of the chains to the posterior using both the built-in Brooks-Gelman-Rubin (BGR) statistic (Brooks and Gelman 1998) and visual examination of plots of the MCMC sequences. We considered chains to have converged on a stationary distribution (the posterior) when the ratio of within chain to between chain variation (\hat{r}) was below 1.1 (Gelman and Hill 2007). Based on these convergence criteria, we discarded the first 5000 samples as burn-in and thinned each chain by a factor of five to reduce within-chain autocorrelation. We also evaluated skewness in the posterior by checking that the posterior median was no more than 5% different from the posterior mean. Posterior summary statistics were based on 6000 roughly independent MCMC samples. For each quantity of interest, we report the posterior mean and 95% credible interval (CI), which is defined by the 2.5th and 97.5th percentiles.

Results

Model evaluation

The most adequate overall model for vessel length included effects for secondary growth, sampling effort, and differences among groups of species. Model D, which classified species by canopy position, had the lowest overall value for the posterior predictive loss criterion, D (Table 3). Based on model D, predicted vessel counts explained 86.6% of the variation in observed vessel counts (Figure 1). The most adequate model varied for different species. For example, the lowest score for D in *Cornus florida*, *Prunus serotina*, *Juglans nigra* and *Quercus rubra*, corresponded to the simplest, non-hierarchical model (A). For the remaining 17 species, models that accounted for the effects of secondary growth outperformed the simplest model. Even though model D was not consistently the best model at the individual species level, it explained a considerable proportion of the observed data for each species. Species-level r^2 values for model D ranged from 48.4% (*Gleditsia triacanthos*) to 93.0% (*C. florida*) (Table 3).

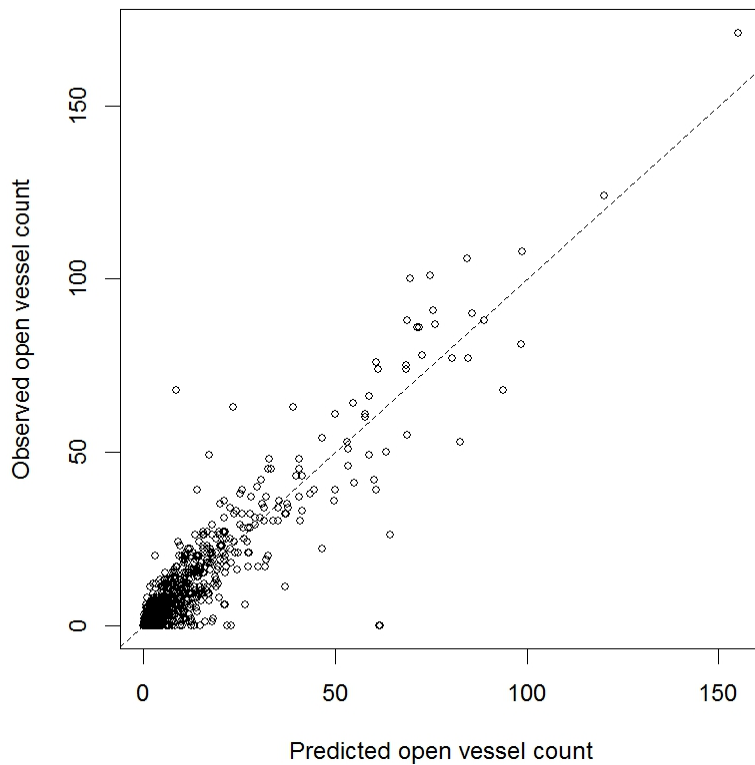


Figure 1: Predicted open vessel counts (posterior mean of simulated data) under the most adequate model (model D) versus observed open vessel counts across all 21 species. Diagonal dashed line represents the 1:1 line.

Table 3: Posterior predictive loss (D) and model goodness-of-fit (r^2) statistics for five alternative models (A-E) of vessel length applied to data from subsamples of dye injected stems, D and r^2 were computed for each species individually and for the overall dataset with the “best” models highlighted in bold (lowest D). Species names that are underlined have a 95% CI for the secondary growth parameter β that overlaps with zero (Table 2). CSA fit indicates whether the CSA method converged on a average vessel length estimate.

Species	D (A)	D (B)	D (C)	D (D)	D (E)	r^2 (D)	CSA fit
<i>Acer rubrum</i>	544.84	546.00	541.76	541.44	543.08	0.896	yes
<i>Aesculus glabra</i>	968.87	965.45	968.33	970.12	971.15	0.666	yes
<i>Amelanchier arborea</i>	3581.64	3562.77	3578.37	3574.86	3555.35	0.885	NO
<i>Asimina triloba</i>	1030.45	1014.46	1022.79	1023.45	1027.69	0.651	yes
<u><i>Carya tomentosa</i></u>	284.47	283.97	290.22	286.01	287.54	0.626	yes
<i>Celtis occidentalis</i>	1235.38	1233.05	1229.57	1230.20	1233.00	0.627	yes
<i>Cercis canadensis</i>	418.25	409.78	410.51	409.07	410.89	0.892	yes
<u><i>Cornus florida</i></u>	3848.09	3863.82	3856.20	3865.09	3860.26	0.930	yes
<i>Diospyros virginiana</i>	778.66	770.94	768.90	767.31	768.43	0.527	yes
<i>Fraxinus americana</i>	502.76	488.42	490.04	490.72	492.55	0.678	yes
<u><i>Gleditsia triacanthos</i></u>	548.34	548.14	551.75	548.29	551.03	0.484	yes
<i>Juglans nigra</i>	804.17	809.87	809.39	809.26	809.60	0.689	yes
<i>Ulmus rubra</i>	11484.09	11481.59	11490.85	11480.79	11495.03	0.777	yes
<u><i>Platanus occidentalis</i></u>	3048.91	3049.19	3045.57	3045.67	3054.48	0.914	yes
<u><i>Prunus serotina</i></u>	6894.03	6905.62	6916.43	6898.56	6935.08	0.866	yes
<i>Quercus alba</i>	1941.03	1936.16	1933.44	1934.99	1938.33	0.721	yes
<u><i>Quercus rubra</i></u>	3047.53	3081.04	3077.21	3065.97	3076.37	0.807	NO
<i>Quercus velutina</i>	1378.58	1365.95	1362.31	1367.46	1363.63	0.872	yes
<i>Sassafras albidium</i>	1005.82	988.98	990.53	990.16	989.41	0.724	yes
<i>Ulmus rubra</i>	520.34	515.07	515.09	512.67	516.00	0.656	yes
<u><i>Vitis vulpina</i></u>	400.69	398.22	398.80	400.72	399.28	0.652	yes
Overall	44266.93	44218.50	44248.06	44212.79	44278.20	0.866	

The analysis of Monte Carlo simulated datasets supported the accuracy of the sampling strategy and model over a range of realistic parameter values. For the 20 simulated datasets representing *A. glabra* which exhibited relatively high values for both the scale and shape parameters (Figure 2, Figure 3) the true shape and scale parameters were always within the estimated 95% CIs. The same was true for all 20 simulated datasets representing *A. triloba* which had a relatively high scale parameter and low shape parameter. For the simulated datasets representing *A. arborea*, which had a low scale parameter and a high shape parameter, the true shape parameter was always within the 95% CI and the true scale parameter was within the 95% CI in 19/20 simulated datasets. Likewise, for *G. triacanthos*, which had relatively low shape and scale parameters, the true shape parameter was always within the 95% CI and the scale parameter was within the 95%CI in 18/20 datasets. Overall, the accuracy rate was over 98%.

Parameter estimates

The most adequate model (model D) consistently recovered effects of secondary growth. The posterior mean for the rate of increase in total vessel counts (β , Eq. 8) was greater than zero for all species except *Q. rubra*, and β was significantly greater than zero for 14 of the 21 species (i.e., the posterior mean for β was greater than zero, and the 95% CI did not contain zero, Table 2).

The posterior estimates of the Weibull parameters, λ and c in Eq. 1, describing the distribution of vessel lengths varied greatly across species. The posterior mean for λ , which under an exponential model corresponds to the expected number of terminations per meter, ranged from 15.8 (*Sassafras albidum*) to 162.9 (*Aesculus glabra*) (Figure 2), across the 21 study species. However, most species (17 of 21) were associated with relatively low values for λ (Figure 2), with posterior means for λ being less than 100 vessel terminations per meter. The posterior mean for c also varied across species, ranging from 0.37 (*Asimina triloba*) to 1.07 (*A. arborea*), indicating that the relationship between the probability of vessel termination and vessel length varied among species (Figure 3). For four species (*A. arborea*, *C. florida*, *Prunus serotina* and *Fraxinus americana*), the 95% CI for c included one, the value at which the Weibull CDF simplifies to the exponential CDF. For the other 17 species, c was significantly less than one, indicating shorter vessels were more likely to terminate. Based on these parameter estimates, the posterior mean for the expected vessel length, $E(d)$ in Eq. 3, varied from 8.8 mm in *A. glabra* to 88 mm in *S. albidum* (Figure 4, Figure S1).

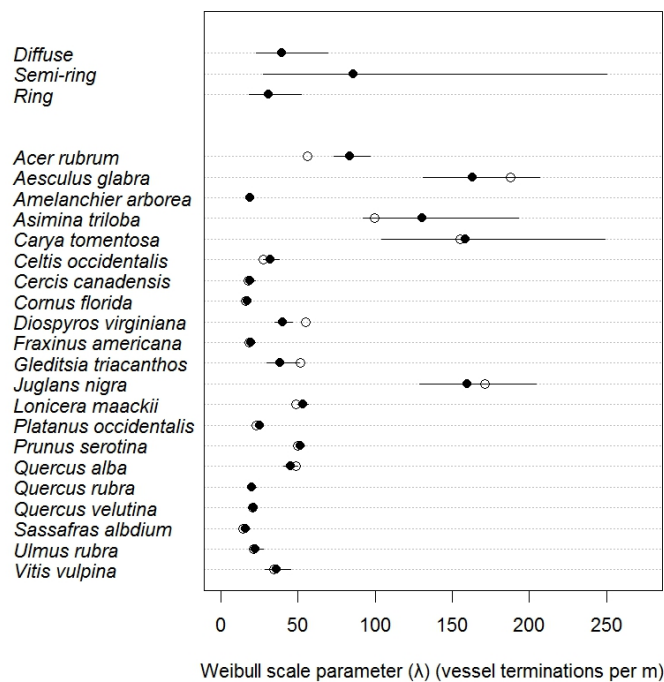


Figure 2: Posterior means (closed circles) and 95% credible intervals (CIs) of the Weibull scale parameter (λ) for 21 species representing three wood porosity types. Open circles represent point estimates for transformed data as generated by the CSA method (i.e., the optimization procedure described by Christman *et al.* (2009)).

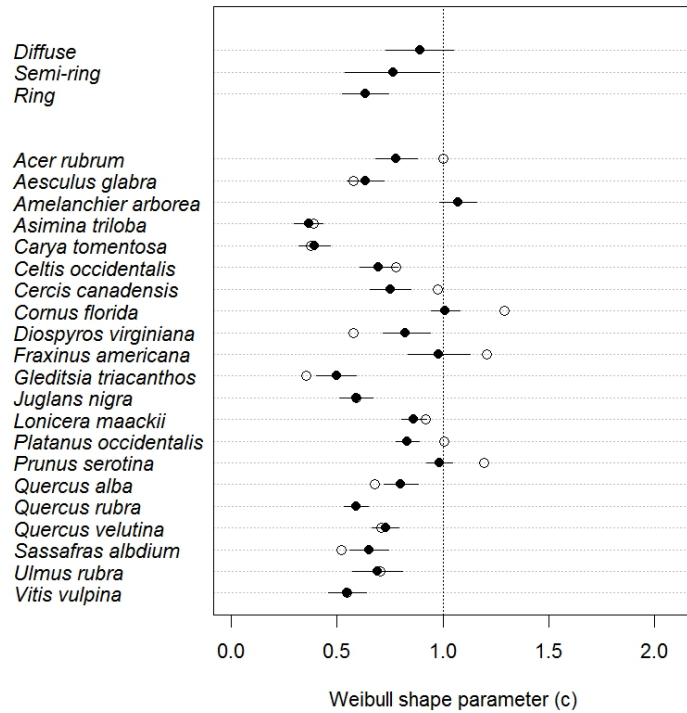


Figure 3: Posterior means (closed circles) and 95% credible intervals (CIs) of the Weibull shape parameter (c) for 21 species representing three wood porosity types. Open circles represent point estimates for transformed data as generated by the CSA method (i.e., the optimization procedure described by Christman *et al.* (2009)). The vertical dashed line denotes $c = 1$, the value at which the Weibull distribution simplifies to the exponential distribution.

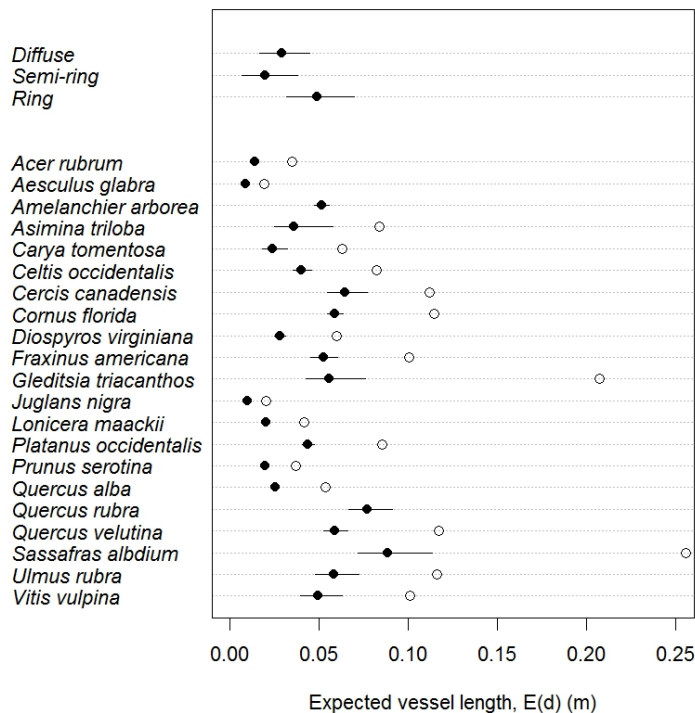


Figure 4: Posterior means (closed circles) and 95% credible intervals (CIs) for the expected vessel length, $E(d)$, for 21 species representing three wood porosity types. These values represent estimates of the mean vessel length of vessels in the basipetal direction from an apical injection point. Open circles represent point estimates for transformed data as generated by the CSA method (i.e., the optimization procedure described by Christman *et al.* (2009)).

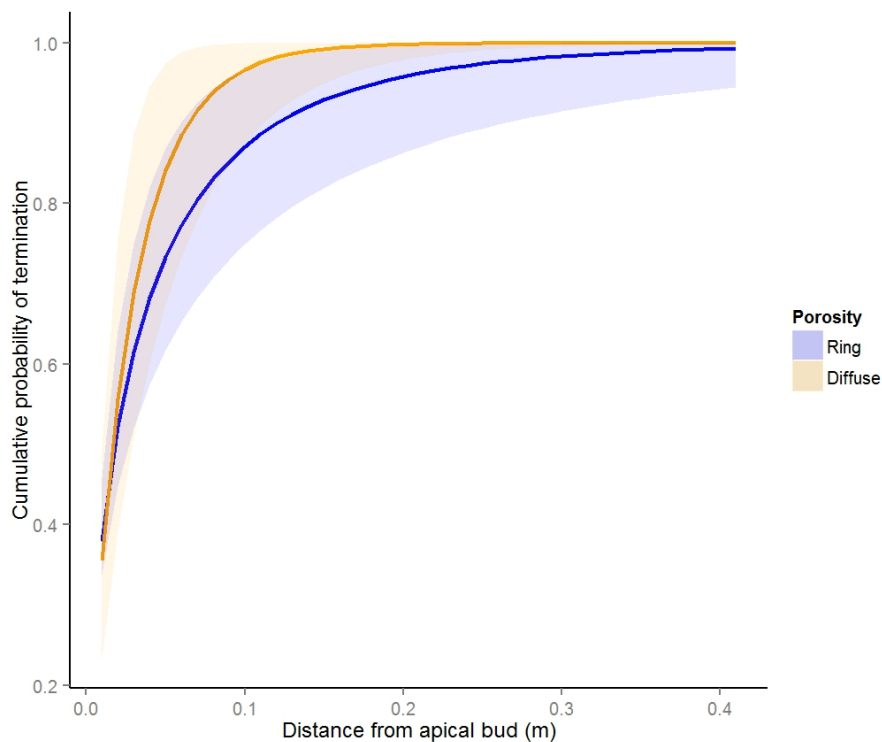


Figure 5: Vessel length distributions for ring-porous and diffuse-porous species based on the posterior means for the shape and scale parameters of each group of species as estimated by model E (see methods) represented as the cumulative distribution function (CDF). The lower cumulative probability of vessel termination for ring porous species is associated with greater skew and longer average length. The shaded areas represent 95% credible intervals.

Differences by wood porosity type

Vessel length distributions differed among species with different wood porosity types (Figure 5). Based on model E, the shape of the vessel length distribution, c , varied such that ring-porous species had lower values of c (posterior mean = 0.632), indicating more strongly right-skewed vessel length distributions and diffuse-porous species had the highest average value of c with a 95% CI that included 1. Semi-ring-porous species had an average value for c that was in between those of the two distinct categories (Figure 3). The difference in the average value of c for ring-porous species versus all other species was significantly less than 0 (contrast posterior mean = -0.195, 95% CI = [-0.371, -0.016]) as was the direct difference between ring versus diffuse-porous species (contrast posterior mean = -0.258, 95% CI = [-0.450, -0.062]). The value of the scale parameter λ did not differ between wood porosity types.

Mean vessel length also varied by wood porosity type. Ring-porous species had the highest expected value on average (posterior mean = 0.049 m), corresponding to the longest average vessel length, while semi-ring-porous species had the lowest average vessel length (posterior mean = 0.020 m), diffuse-porous species were intermediate (posterior mean = 0.029 m) (Figure 4). The average vessel of a ring-porous species was 2.4 cm longer than those from species of other wood porosity types (contrast posterior mean = 0.024 m, 95%CI = [0.004, 0.048]).

Comparison of the Bayesian versus CSA approaches

The parameter estimates obtained from our full model (model D) were comparable to the estimates obtained from the CSA method (Christman *et al.* 2009). The CSA solver algorithm converged for pooled data for 19 species, but failed to produce estimates for *Amelanchier arborea* and *Quercus rubra* (Table 3). Among the 19 species with estimates from both methods, the CSA estimates of λ were highly correlated with the posterior means produced by our Bayesian model and the 95% CI for the slope included 1 (Pearson correlation, $r = 0.97$, $P < 0.001$, slope 95% CI = [0.917 - 1.07]). For 13 of the 19 species, the CSA point estimates of λ fell in the 95% CI derived from the Bayesian model (Figure 2). Similarly, estimates for c were also highly correlated between methods, and the 95% CI for the slope again included 1 (Pearson correlation, $r = 0.90$, $P < 0.001$, slope 95% CI = [0.98 - 1.17]), although only 9 of the 19 CSA point estimates fell within the Bayesian 95% CIs (Figure 3).

While the CSA point estimates generally agreed with the posterior means for λ and c , the estimates of mean vessel length— $E(d)$, which is a deterministic function of λ and c —were quite different between the modeling approaches (Figure 4). Although the CSA point estimates were highly correlated with the posterior means for $E(d)$ (Pearson correlation, $r = 0.90$, $P < 0.001$), the CSA estimates were significantly higher (longer vessels) for all 19 species for which the solver converged. On average, the CSA point estimates of mean vessel length were 2.2 times longer than the posterior means for $E(d)$, (slope 95% CI = [2.02 – 2.59] including a point estimate for *Gleditsia triacanthos* that was almost 4 times longer than the corresponding posterior mean.

Discussion

Advantages of a Bayesian approach to estimating vessel length from subsampled stems

Vessel length is a key trait for understanding angiosperm hydraulic performance. As climate change shifts patterns of water stress, variation in vessel length within and between woody plant species could impact how forests respond to drought (Brandt *et al.* 2014). Despite its importance, vessel length has only been estimated in a few hundred of the hundreds of thousands of angiosperm species (Jacobsen *et al.* 2012). The paucity of vessel length estimates may reflect the difficulty of generating and analyzing the necessary data (Cai and Tyree 2014). As technical advances accelerate data collection, analysis methods must evolve to more accurately capture and compare variability between species and within individual stems (Pan *et al.* 2015).

To facilitate research on this important trait, we developed a flexible hierarchical framework for estimating vessel length from subsamples of stems. Doing so requires accounting for the potential bias and error introduced by subsampling. A potential source of bias in both basipetal and acropetal dye injection samples results from secondary growth which adds vessels in the basipetal direction that may distribute injected dye differently from other vessels. Vessel length models that ignore this additional process may fit the data poorly, or yield biased estimates of vessel lengths. Our results show that the total number of vessels almost always (20/21 species) increased with increasing distance from the stem apex. Correcting for this effect improved model estimates in all but four species. For this reason, we advocate a hierarchical approach to estimating vessel length from stem subsamples that statistically distinguishes changes in dyed vessel density due to vessel termination versus changes attributable to vessel addition.

Our approach to correct for the effects of secondary growth could be modified to improve estimates based on alternative injection methodologies. With acropetal injection, secondary growth is likely to inflate the rate at which dyed vessel counts decline with increasing distance from the injection location. Consequently, a stem that is growing quickly may appear to have shorter vessels than a more slowly growing stem with the same underlying vessel length distribution. For data generated using this method, Eqs. 6-8 could also provide an estimate of the total vessel population that is corrected for secondary growth. However, our model does not accommodate other complications of acropetal injection including branching. It is also possible that changes in the number of vessels along the length of the stem are not solely due to secondary growth. Vessel splitting could also play a role. Murray's law for optimal conductance predicts that vessels should furcate acropetally increasing the total number of vessels from stem base to apex (Denny 2012). While some empirical evidence suggests that a fraction of vessels do furcate in this way (McCulloh *et al.* 2004) our data show the opposite pattern and suggest that the net effect of furcation and secondary growth in our species is balanced towards secondary growth.

In addition to complications from secondary growth, subsampling stems also introduces sampling error. Sampling error can vary from stem to stem or species to species. To mitigate how variation in sampling effort influenced our results, we implemented a random effects approach. This approach weights species-level parameter estimates inversely proportional to their estimated error (Gelman and Hill 2007). Doing so effectively borrows strength from sampling among species to limit how strongly sampling error can influence parameter estimation within a species. It also limits some forms of specification error including pseudoreplication (Schank and Koehnle 2009). This approach easily generalizes to additional levels of pooling. For our dataset, the overall best fit model treated species-level parameters as nested within canopy position levels. By identifying and addressing how secondary growth and sampling error affect vessel length estimates from subsampled stems, our approach can simplify data collection and accelerate research on this critical trait.

With growing interest in the effects of vessel length on plant hydraulics, researchers are generating data in different ways that will require more flexible modelling approaches. Recently, Pan *et al.* (2015) described a dynamic

method based on air injection that can reduce the time needed for data collection from weeks to days. Their results suggested that important variation within and between stems challenged the assumptions of their exponential-gamma model for vessel length distributions. A Bayesian approach to air injection data, similar to the model developed here, could afford the flexibility and power necessary to improve vessel length estimates from these new data and facilitate compilations of vessel length estimates from using different methods (e.g. Ogle *et al.* 2014).

Comparison of the hierarchical Bayesian approach to the CSA method

Not only did our hierarchical Bayesian model provide a better fit to the data than a simpler Bayesian model that ignored secondary growth and sampling effects, it also produced parameter estimates that closely corresponded to those from a very different modeling approach. That is, the posterior mean estimates for the shape (λ) and scale (c) parameters inferred from our model were similar to those generated by the nonlinear optimization approach used by Christman *et al.* (2009) (Figures 2 and 3). While these methods produce correlated results, the implementation of the model in a Bayesian framework has at least two major advantages over the CSA method. First, the Bayesian model generated estimates for every species, whereas the Microsoft Excel implementation of the CSA method failed to converge for two of the 21 species (Table 3). Second, the CSA method only produces point estimates, while the Bayesian approach yields a posterior distribution for each parameter, which incorporates uncertainty from different sources, including variation in sampling effort, secondary growth and species' differences.

The residual differences between estimates may reflect the data transformations necessary to apply the CSA method to our subsampled data or to differences in the modeling assumptions. The biggest difference in the assumptions relates to how the Weibull parameters translate into the underlying vessel length distribution. Our model explicitly estimates observed vessel length basipetal to the injection site near the terminal bud, which is a deterministic function of c and λ (i.e. Eq. 3). Given our methodology, this is equivalent to estimating the injection length distribution rather than the intact vessel length distribution. In contrast, the CSA method aims to include any unobserved and undeveloped vessel length that occurs in the opposite direction above the bud. It is possible that if we had waited for the stem to elongate, the vessels that we measured would have grown approximately twice as long, as implied by the differences in estimated average vessel length. Yet, we are hesitant to model undeveloped vessel length in part because a recently published formula for left-truncated, interval-censored Weibull data (Balakrishnan and Mitra 2012) that is appropriate for estimating vessel lengths in both directions around the injection site differs qualitatively from the formula proposed by Christman, *et al.* (2009). We encourage more theoretical and simulation work to evaluate the properties of these alternative expressions.

Consistent with the difference in the target distribution, vessel length estimates from our model tended to be shorter than existing estimates for certain species. *Acer rubrum*, *Quercus alba* and *Quercus rubra* also occur in the vessel length database compiled by Jacobsen *et al.* (2012); our estimates are 45%, 25%, and 66% of the previously published values respectively. Because we did not vacuum infiltrate our stems prior to injection, some embolisms may have persisted and impeded dye movement, causing vessel lengths to appear relatively shorter. Likewise, differences in vessel anatomy could have influenced the infiltration performance of silicone using our standardized injection procedure. Finally, tylose formation or microbial growth could have occluded vessels, but we note that our storage times were typical of other studies (Hacke *et al.* 2007, Sperry *et al.* 2007) and we did not observe evidence of either in our cross section images. Despite these potential complications, our estimates are consistent with *a priori* expected differences that shed light on the conservation implications of vessel length variation among and within tree species at our site.

Physiological consequences of vessel length variation

Our results also provide new insights into how vessel length varies within stems. Our analysis of differences in the shape of the vessel length distribution among wood porosity types (Figure 2) in a single temperate forest is consistent with the hypothesis that species with more variable vessel diameter distributions also have more variable vessel length distributions. The variation in both vessel diameter and length within ring-porous species is consistent with an ecophysiological strategy that shifts from hydraulic efficiency to safety across the growing season. At this site, ring-porous species in the genera *Quercus* and *Carya* dominate forests on slopes with high seasonal variation in soil moisture (Zimmerman and Wagner 1979) while diffuse-porous species such as *Platanus occidentalis* and *Acer* species

dominate bottomland habitats with more consistent soil moisture availability. While this topographic gradient is associated with seasonal variation in drought stress, it may not be long enough to generate variation in freeze-thaw cycles. In this forest, the hydraulic consequences of developing shorter latewood vessels (i.e., increased resistance to drought-induced cavitation (Christman *et al.* 2009)) may explain the topographic segregation by wood porosity type better than the hydraulic consequences of the characteristic decrease in vessel diameter (i.e., increased resistance against freeze-thaw induced embolism, (Davis *et al.* 1999)). This explanation is consistent with a recent finding that vessel length, not vessel diameter, explains variation in drought tolerance among *Acer* species (Lens *et al.* 2011). It also may explain why ring and diffuse-porous species tend to have similar hydraulic performance under average conditions but different sensitivities to changing water stress (Hacke *et al.* 2006). Additional experiments and observations are necessary to explore whether increased hydraulic efficiency early in the growing season also contributes to habitat segregation and what effects differences in mean vessel length have on demographic responses to drought (Brandt *et al.* 2014)

Similar patterns of variation in vessel length and diameter also have implications for understanding vessel development in different wood porosity types. Nijse (2004) proposed that vessel development links length and width such that diffuse-porous species should have exponential vessel length distributions. Consistent with this hypothesis, the posterior distribution of the shape parameter, c , for diffuse-porous species includes 1, indicating that the exponential CDF is suitable. However, the simple process implied by the exponential CDF did not describe the distribution of vessel lengths in ring and semi-ring-porous species, which have shape parameters significantly less than one. For these species, whether a vessel remains open or closed depends on how long that vessel has been open. This distribution (e.g., the Weibull CDF) arises naturally in systems with fractal branching and could suggest that vessel lengths in ring-porous species reflect a continuum of vessel termination rates (Brown and Wohletz 1995). Linking these theoretical considerations to underlying developmental mechanisms could be a fruitful area for future research.

Conclusions

Vessel length is an important but understudied functional trait. The Bayesian model we present could facilitate further research by informing future data collection. Furthermore, the Bayesian framework could serve as the foundation for more sophisticated models that explicitly evaluate the hydraulic consequences of variation in vessel geometry. Rather than conducting a series of analyses to generate mean trait values and discarding important sources of variability, the Bayesian framework propagates uncertainty (Ogle and Barber 2012), yielding more complete reconstruction of vessel length distribution parameters. Moreover, a Bayesian framework can accommodate more detailed process models (Ogle and Barber 2008, Patrick *et al.* 2009) that allow researchers to accommodate different measurement methods, and to explore the mechanistic basis for differences in performance among plants with different vascular geometries. Models that link vascular anatomy to hydraulic performance are especially important in the context of a changing climate that may superimpose year-to-year variation on the seasonal changes that influence the traits and distributions of temperate trees.

Acknowledgements

The authors thank V. Torrez, M. Evans, B. Cordes, P. Patel, S. Shannafelt and M. Boyher for assistance with the collection and preparation of material. Members of the Zanne and Ogle labs as well as the chief and associate editors provided valuable feedback on the manuscript. We also thank the Tyson Research Center and Washington University in St. Louis for access to the site and logistical support. This work was supported by the National Science Foundation DEB [1302797 to A.E.Z.].

References

- Allen CD, Macalady AK, Chenchouni H, Bachelet D, McDowell N, Venetier M, Kitzberger T, Rigling A, Breshears DD, Hogg EH. 2010. A global overview of drought and heat-induced tree mortality reveals emerging climate change risks for forests. *Forest Ecology and Management* 259: 660–684.
- Andre J-P. 2005. *Vascular organization of angiosperms; a new vision*. Enfield, New Hampshire: Science Publishers, Inc.
- Balakrishnan N, Mitra D. 2012. Left truncated and right censored Weibull data and likelihood inference with an illustration. *Computational Statistics and Data Analysis* 56: 4011–4025.
- Brandt L, He H, Iverson L, Thompson FR, Butler P, Handler S, Janowiak M, Shannon PD, Swanston C, Albrecht M, *et al.* 2014. Central Hardwoods ecosystem vulnerability assessment and synthesis: a report from the Central Hardwoods Climate Change Response Framework project. General Technical Report NRS-124. Newtown Square, PA: U.S. Department of Agriculture, Forest Service, Northern Research Station, 254.
- Brooks SP, Gelman A. 1998. General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics* 7: 434–455.
- Brown WK, Wohletz KH. 1995. Derivation of the Weibull distribution based on physical principles and its connection to the Rosin–Rammler and lognormal distributions. *Journal of Applied Physics* 78: 2758.
- Cai J, Tyree MT. 2010. The impact of vessel size on vulnerability curves: data and models for within-species variability in saplings of aspen, *Populus tremuloides* Michx. *Plant, Cell and Environment* 33: 1059–1069.
- Cai J, Tyree MT. 2014. Measuring vessel length in vascular plants: can we divine the truth? History, theory, methods, and contrasting models. *Trees* 28: 643–655.
- Cai J, Zhang S, Tyree MT. 2010. A computational algorithm addressing how vessel length might depend on vessel diameter. *Plant, Cell and Environment* 33: 1234–1238.
- Carlquist S. 1975. *Ecological strategies of xylem evolution*. University of California Press, xi + 259 pp. + 15 pl.
- Christman MA, Sperry JS, Adler FR. 2009. Testing the ‘rare pit’ hypothesis for xylem cavitation resistance in three species of *Acer*. *The New Phytologist*: 664–674.
- Cohen S, Bennink J, Tyree M. 2003. Air method measurements of apple vessel length distributions with improved apparatus and theory. *Journal of Experimental Botany* 54: 1889–1897.
- Comstock JP, Sperry JS. 2000. Theoretical considerations of optimal conduit length for water transport in vascular plants. *The New Phytologist* 148: 195–218.
- Davis SD, Sperry JS, Hacke UG. 1999. The relationship between xylem conduit diameter and cavitation caused by freezing. *American Journal of Botany* 86: 1367–1372.
- Denny M. 2012. Tree hydraulics: how sap rises. *European Journal of Physics* 33: 43–53.
- Ellmore G, Ewers F. 1986. Fluid flow in the outermost xylem increment of a ring-porous tree, *Ulmus americana*. *American Journal of Botany* 73: 1771–1774.
- Ewers FW, Fisher JB. 1989. Techniques for measuring vessel lengths and diameters in stems of woody plants. *American Journal of Botany* 76: 645–656.
- Gelfand AE, Ghosh SK. 1998. Model choice: a minimum posterior predictive loss approach. *Biometrika* 85: 1–11.
- Gelman A, Carlin JB, Stern HS, Rubin DB. 2004. *Bayesian Data Analysis*. Boca Raton: Chapman and Hall/CRC Press.
- Gelman A, Hill J. 2007. *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press.
- Hacke UG, Sperry JS, Feild TS, Sano Y, Sikkema EH, Pittermann J. 2007. Water transport in vesselless angiosperms: conducting efficiency and cavitation safety. *International Journal of Plant Sciences* 168: 1113–1126.
- Hacke UG, Sperry JS, Wheeler JK, Castro L. 2006. Scaling of angiosperm xylem structure with safety and efficiency. *Tree Physiology* 26: 689–701.
- Jacobsen AL, Pratt RB, Tobin MF, Hacke UG, Ewers FW. 2012. A global analysis of xylem vessel length in woody plants. *American Journal of Botany* 99: 1583–1591.
- Lasdon LS, Waren AD, Jain A, Ratner M. 1978. Design and testing of a generalized reduced gradient code for nonlinear programming. *ACM Transactions on Mathematical Software* 4: 34–49.
- Lens F, Sperry JS, Christman MA, Choat B, Rabaey D, Jansen S. 2011. Testing hypotheses that link wood anatomy to cavitation resistance and hydraulic conductivity in the genus *Acer*. *The New Phytologist* 190: 709–723.
- Lunn D, Spiegelhalter D, Thomas A, Best N. 2009. The BUGS project: Evolution, critique and future directions. *Statistics in Medicine* 28: 3049–3067
- McCulloh KA, Sperry JS, Adler FR. 2004. Murray’s law and the hydraulic vs mechanical functioning of wood. *Functional Ecology* 18: 931–938

- McDowell N, Pockman WT, Allen CD, Breshears DD, Cobb N, Kolb T, Plaut J, Sperry J, West A, Williams DG, *et al.* 2008. Mechanisms of plant survival and mortality during drought: why do some plants survive while others succumb to drought? *The New Phytologist* 178: 719–39.
- Nijse J. 2004. On the mechanism of xylem vessel length regulation. *Plant Physiology* 134: 32–34.
- Ogle K, Barber JJ. 2008. Bayesian Data—Model integration in plant physiological and ecosystem ecology (U Lüttge, W Beyschlag, and J Murata, Eds.). *Progress in Botany* 69: 281–311.
- Ogle K, Barber JJ. 2012. Bayesian Statistics. In: Hastings A, Gross L, eds. *Sourcebook in theoretical ecology*. Berkeley: University of California Press.
- Ogle K, Pathikonda S, Sartor K, Lichstein JW, Osnas JL, Pacala SW. 2014. A model-based meta-analysis for estimating species-specific wood density and identifying potential sources of variation. *Journal of Ecology* 102: 194–208.
- Pan S, Kudo M. 2012. Recognition of wood porosity based on direction insensitive feature sets. *Transactions on Machine Learning and Data Mining* 5: 45–62.
- Pan R, Geng J, Cai J, Tyree MT. 2015. A comparison of two methods for measuring vessel length in woody plants. *Plant, Cell and Environment*. 38: 2519–2526.
- Patrick LD, Ogle K, Tissue DT. 2009. A hierarchical Bayesian approach for estimation of photosynthetic parameters of C(3) plants. *Plant, Cell and Environment* 32: 1695–1709.
- Schank JC, Koehnle TJ. 2009. Pseudoreplication is a pseudoproblem. *Journal of Comparative Psychology* 123: 421–433.
- Scholz A, Klepsch M, Karimi Z, Jansen S, Stephan D, Torres-ruiz JM. 2013. How to quantify conduits in wood? *Frontiers in Plant Science* 4: 1–11.
- Sperry JS, Hacke UG, Feild TS, Sano Y, Sikkema EH. 2007. Hydraulic consequences of vessel evolution in Angiosperms. *International Journal of Plant Sciences* 168: 1127–1139.
- Sperry JS, Hacke UG, Wheeler JK. 2005. Comparative analysis of end wall resistivity in xylem conduits. *Plant, Cell and Environment* 28: 456–465.
- Sperry J, Nichols K, Sullivan J, Eastlack S. 1994. Xylem embolism in ring-porous, diffuse-porous, and coniferous trees of northern Utah and interior Alaska. *Ecology* 75: 1736–1752.
- Sperry JS, Sullivan JEM. 1992. Xylem embolism in response to freeze-thaw cycles and water stress in ring-porous , diffuse-porous , and conifer species. *Plant Physiology* 100: 605–613.
- Tyree MT, Zimmermann MH. 2002. *Xylem structure and the ascent of sap*. Springer Berlin / Heidelberg.
- Tyree MT, Davis S, Cochard H. 1994. Biophysical perspectives of xylem evolution: is there a tradeoff of hydraulic efficiency for vulnerability to dysfunction? *IAWA Journal* 15: 335–360.
- Wheeler EA. 2011. Insidewood - a web resource for hardwood anatomy. *IAWA Journal* 32: 199–211.
- Woodcock D. 1989. Climate sensitivity of wood-anatomical features in a ring-porous oak (*Quercus macrocarpa*). *Canadian Journal of Forest Research* 19: 639–644.
- Zanne AE, Sweeney K, Sharma M, Orians CM. 2006. Patterns and consequences of differential vascular sectoriality in 18 temperate tree and shrub species. *Functional Ecology* 20: 200–206.
- Zimmerman M, Wagner WL. 1979. A description of the woody vegetation of oak-hickory forest in the northern Ozark highlands. *Bulletin of the Torrey Botanical Club* 106: 117–122.
- Zimmermann MH, Jeje AA. 1981. Vessel-length distribution in stems of some American woody plants. *Canadian Journal of Botany* 59: 1882–1892.