**2.2 METHODS**

**Site.** Sepilok Forest Reserve (SFR) is a 4475 ha patch of mostly pristine lowland dipterocarp forest located on the Northeast coast of Borneo in Sabah, Malaysia (5° 10ʹ N, 117° 56ʹ E). The reserve has a highly variable topography, vegetation, flooding frequency and soil characteristics that are partly derived from differential parent materials (Fig. 1a, *for a description of SFR, see* (Baltzer, Thomas, Nilus, & Burslem, 2005; Dent, Bagchi, Robinson, Majalap-Lee, & Burslem, 2006; DeWalt, Ickes, Nilus, Harms, & Burslem, 2006; Margrove et al., 2015)). Sepilok Forest Reserve has a mean annual rainfall of 3136 (±921 SD) mm and a mean annual temperature of 27.3°C (Margrove et al., 2015). Sabah's climate is aseasonal (Born et al., 2014; Margrove et al., 2015; Maycock, Thewlis, Ghazoul, Nilus, & Burslem, 2005); however, there are relatively consistent wetter and drier periods throughout the year (Fig. S2). Climate also varies year on year, with these supra-annual fluctuations attributable to variation in the El Niño Southern Oscillation (ENSO)(Moerman et al., 2013).

**Seed collection.** We selected sixteen species from 3 dipterocarp genera for this study (Table S3) based on their variable adult distributions and seed availability during the 2014 general flowering event. Dipterocarps flower every 2 – 10 years (Peter S Ashton, 1988) often in association with major ENSO events (P. S. Ashton, Givnish, & Appanah, 1988; Curran et al., 1999), and as a result seedling recruitment is intermittent (Peter S Ashton, 1988). Moreover, seeds have a short period of viability and germinate within days of falling from the parent tree leaving no seed bank (O’Brien, Philipson, Tay, & Hector, 2013). As the first year is one of the most sensitive life stages to environmental disturbances (Green, Harms, & Connell, 2014), our study replicated the first year of life as closely as possible with the first period of water inundation occurring ca. six months after germination.

|  |
| --- |
|  |
| **Figure 1 | Macro to micro topographical variation within Sepilok Forest Reserve, and community compositional variation at the macro scale.** A topographical digital terrain model of 3375 ha (ca. 75%) of SFR **(a)**, with the 160 ha forest plot outlined and the distribution of the four 4 ha plots (and one 2 ha plot) for analysis. **(b)** The distribution of two example species, *Shorea leprosula* (white, diamond) and *S. beccariana* (grey, circle), indicating partitioning along topographical gradients within the 160 ha plot. **(c)** Dissimilarity in dipterocarp community compositions reflected by colour variations across the 160 ha plot, extracted from axis 1 of a non-metric multi-dimensional scaling analysis of 12 elevational bands. **(d)** Micro-topographical variation within three example plots, mean water inundation height expressed as the transparent blue (Fig. S1 – pictures of water inundation). |

We collected the seeds for the experiment during the mast flowering in 2014 (July-August) with 13 species harvested directly from SFR and three species (*Shorea waltonii, Dryobalanops lanceolata* and *S. pauciflora*) sourced from nearby forest reserves. These three species are found in SFR and remain in the same floristic region within Borneo (Slik et al., 2003). Seeds were collected directly from the mother tree by shaking the outer fruit-bearing branches using a ‘bigshot' catapult and rope to control for maternal heritage. Species had variable numbers of mothers due to the availability of fruiting individuals, ranging from 1 to 9, with a total of 51 different mother trees used in this experiment. Seeds were germinated in wet jute sacks, transplanted into cylindrical polythene bags 9 x 15 cm filled with a mixture of sawdust and locally sourced alluvial topsoil. After germination, seedlings were grown for three months under 10% ambient light, typical of nursery conditions recreating a medium sized forest gap (Born et al., 2014; Philipson et al., 2014) and watered daily before being transplanted into the forest after discarding their cotyledons and first true leaves had developed.

**Experimental Design.** We planted a total of 2048 seedlings in a stratified randomized block design in October 2014. There were 16 blocks with two plots nested in each. Twelve blocks were within a 160 ha long-term forest dynamics plot (see below for details), and four blocks were outside this area. The furthest distance that a block was from the 160 ha plot boundary was 198 m. Two alluvial floodplains in SFR, separated by a ca. 800 m formation of mudstone hills, were used to distribute the blocks in natural gaps equitably > 100 m apart (Fig 1a & b). Within each block, the two plots were in conditions with an open canopy < 50 m apart. Each plot was in either high or low topographical positions to capture variation in meso-topography and hydrology. We surveyed for prospective blocks fitting these criteria, forming a collection of possible locations. From these, we randomly selected the positions of the final blocks. We planted in each plot four individuals from each of the 16 species (Table S3) while maximizing the possible maternal diversity. We did this by evenly distributing the seedlings derived from different mother trees among plots and blocks within the design framework. To ensure that mortality caused by transplanting seedlings from the nursery to the forest was not a source of error, we replaced any dead or damaged seedlings before the start of the experiment in December. Plots were square with a side length of 2.4m and contained 64 seedling positions spaced at 30 cm apart. Each position within all plots was assigned a random seedling fitting the criteria described above.

**Environmental Variables.** We measured micro-topographic undulations within each plot with a laser level fixed to a pole placed upright ensuring both horizontal tubular bubbles were correct. The level's fan angle laser projected a horizontal beam over the plot. By measuring down from the laser to each seedling using a tape measure, we calculated the relative elevation of each seedling in the plot. The laser level was also used to calculate the difference in elevation between the two plots within each block. We determined the relative seedling elevations within blocks by adding the difference between the plot elevations to the higher plot.

We measured the water depth at the lowest seedling 13 times during the wettest 24 day period. By knowing the relative elevation of all seedlings in every plot we can calculate the mean extent of water inundation for every seedling within a block. Mean flood levels for each block can link plots together. This approach is more desirable than measuring the exact altitude of each plot, as observations indicated that groundwater levels did not play a significant role in determining inundation levels with plots. Furthermore, we found evidence of gleying (mottled oranges and greys, Fig. S3) in many of the plots, indicating that these areas may experience anaerobic conditions although well above river levels.

Flash floods occurred several times during January 2015 in SFR with the most intense rainfall in the morning of January 14th, ­when 178.3 mm fell within a 6 hour period (January’s daily mean precipitation was 12.3 mm 95% CrI = 11.0 – 13.8 mm, Fig. S2). Nineteen plots showed complete and eight plots showing partial submergence (27 out of 32) during this event. However, rivers and surface runoff drained sites quickly, freeing the seedlings from submergence in < 5 hours (mean drainage rate: 26.6 ± 17.5 cm hour-1; n = 4 locations).

Four soil samples were taken from each plot using a soil auger at two corners corresponding to a maximum and minimum elevation within each plot, and two depths (0-15 and 30-50 cm (Sukri, Wahab, Salim, & Burslem, 2012)). The soil stored in a cold dark room for three months to air-dry. The two samples from equal depths within each plot were homogenized, leaving 32 from the surface and 32 from the lower layer of soil. Soil pH was determined by measuring samples after mixing with distilled water (Allen, 1989). Soil texture was used as an edaphic variable to provide a proxy for soil fertility (Hirai et al., 1997; Itoh et al., 2003). We used the pipet method to determine the three soil texture fractions, clay, silt and sand at the plot level (Majalap-Lee, N & Hin, 1992). Organic matter was digested using ammonia solution and hydrogen peroxide with a hot plate. The contents were transferred to a 1L measuring cylinder, filled with distilled water and boiled. Exploiting the differences in settling time of the soil fractions clay, silt, and sand; we pipetted out the solutions and calculated the mean percentage of clay, silt, and sand per plot.

The light environment for each plot was measured using a spherical densiometer. The densiometer was held level in the centre of each plot ca. 30 cm above the ground. We counted the number of quarter grid cells on the concave mirror that were reflecting open canopy as the value of canopy openness(Lemmon, 1956; Philipson et al., 2014).

**Seedling Census.** We took the first seedling census during December 2015, after which, a re-census of the seedlings was taken every three months for one year giving a total of four census intervals. During each census, we measured the diameter and noted the mortality of seedlings. Diameters were measured twice perpendicular at the same marked point 5 cm above the soil with digital callipers. We considered a seedling as dead when there was no green phloem tissue remaining under the stem.

**Digital Terrain Model.** NERC's Airborne Research Facility collected airborne laser scanning (ALS) data for SFR in November of 2014. The sensor used was a Leica ALS50-II LiDAR on a Dornier 228-201 and was flown over the 160 ha permanent forest plot. It emitted pulses at 83.1 Hz, has a field of view of 12° and an impression of about 40 cm. The average point density was 7.3 m-2. The point cloud data were processed to produce a digital elevation model (DEM), a digital terrain model (DTM). Data were processed in LAStools (http:// rapidlasso.com/lastools/). For a full description of the methods used see Jucker *et al.* (2017).

**160 ha Forest plot.** We completed the initial establishment of a 160 ha long-term forest dynamics plot within SFR in 2010. All adult dipterocarps were tagged, identified to species level and mapped (n = 5587). Minimum diameter limits were selected based on the estimated minimum size that a species would flower and the maximum adult size. For sub-canopy dipterocarps species > 30cm DBH (n species = 2) were selected, for large emergents 50 cm DBH (n species = 13) and for those with a strategy somewhere in the middle 40cm DBH (n species = 1) was selected (Table S3). When investigating adult distributions, using different minimum diameter sizes dependent on a species' biology is more appropriate than setting a common minimum limit for all species. This approach explicitly takes into account differences in maximum size, which is important for making meaningful comparisons in a species' distribution. Each adult dipterocarp was assigned an elevation (asl m) transposed from the DTM.

**Four ha plots.** Within SFR, nine 4 ha forest plots were established from 2000 to 2001. Of these, six plots (24 ha) were within lowland dipterocarp forest, and 4 ½ (18 ha) were outside the 160 ha forest plot. All individuals > 5 cm DBH are tagged and have measured DBHs. The most recent census of the plots was from 2013 to 2015. Individual positions are known to occur within a 10 x 10 m grid which divided each 4 ha plot. A differential GPS was used to geo-locate the corners (Geneq SXBlue II), and the layout was used to calculate the approximate position of each tree. We calculate and assigned the mean elevation (asl m) from the DTM for each 100m2 grid tile.

**Seedling mortality data analysis.** To test the hypothesis that species respond differently to water inundation we used a generalized linear mixed effects model assuming a binomial error distribution from the lme4 package (Bates, Maechler Martin, & Walker, 2016) in R v3.3.1(R Core Team, 2016). The flooded explanatory variable had two levels: individuals that typically experienced water inundation during the wet period from November to February (wet), and the control indicating individuals that were on micro-topographical locations that experienced no water inundation (dry). The interaction between flooded and species was key to addressing our hypothesis. In the final model, we controlled for census interval, micro-topography, and the natural logarithm of the diameter of the seedlings as fixed effects. We included two random effects: plot, which was allowed to vary with census interval, and mother controlling for maternal heritage. Other variables considered for inclusion in the model were the site (two alluvial sites), block, pH, soil texture, light, and maximum flash flood levels, and up to all second-order interactions. We removed all variables from the final model that did not reduce the AIC by > 4 points. The model was evaluated using a modified version of the binnedplot function (Gelman et al., 2016) from the arm package in R v3.3.1. to check for normality and independence of binned residuals and ensure random effects followed a Gaussian distribution. The focal effect of this model was the difference between the inundated and non-inundated plants, as this could provide inference on a species' inundation sensitivity. We defined a species seedling's sensitivity to flooding as the absolute risk increase (ARI; also known as the absolute risk reduction). The ARI calculation is the difference between the probability of mortality in the treated group (water inundated individuals) and the probability of mortality in the control group (not inundated)(Schechtman, 2002). As the size of the ARI is the effect size of water inundation on mortality, we henceforth refer to this as the "inundation sensitivity."

**Habitat association within 160 ha plot.** Species relationships with elevation were analysed using the Integrated Nested Laplace Approximation (INLA) Bayesian framework with Stochastic Partial Differential Equation (SPDE) fitted to spatial point-referenced data. INLA executes direct numerical calculation of posterior densities and is a computationally effective substitute to MCMC (Lindgren, Rue, & Lindström, 2011; Rue, Martino, & Chopin, 2009). These tools allowed us to model the probability of a species’ presence or absence at each of the 5587 locations of adult dipterocarp stems in the 160 ha forest plot while accounting for spatial autocorrelation between individuals of the same species. More specifically, the model used to describe the spatial autocorrelation was the Matern correlation function, which depends on a scale parameter and a smoothness parameter. INLA estimated these parameters from flat priors. We constructed a model using species and elevation as the explanatory variables, fitted with the elevation quadratic term, testing five combinations of interactions. For model selection, we used the conditional predictive ordinate (CPO), which computes the density of the posterior predictive distribution at an observation used to select the models which measure the fit through predictive density , essentially a Bayesian leave-one-out cross-validatory predictive check. We checked for possible failures and recomputed these values following Blangiardo and Cameletti (2015). We compared models by calculating . We selected three models based on a hypothesis that approximated the shape of the probability of a species occurrence along an elevation gradient, presenting the results of the model with the lowest CPO. The final model for elevation was:

Credible intervals (95%) were calculated by sampling from the posterior distribution 5000 times.

For each species the derivative of the quadratic equation’s parameters was taken and set to zero . This formula calculates the most probable elevational position at which each species occurs. Simplifying a species' adult distribution to its most probable elevational position gives us a single number to use in the linear model and works as a good proxy for the overall distribution.

**Seedling sensitivity wood density.** We tested the hypothesis that wood density would explain variation in flooding sensitivity, with the expectation that higher wood densities are more resilient than species with lighter wood. A linear model with inundation sensitivity as the response variable and wood density split into high and low portions (than the mean). We found this model to have unequally distributed residuals, with an order of magnitude of more variation in the low wood density group. Hence, we log transforming the response variable.

We constructed a weighted regression model with the response of elevational distribution. The explanatory variables were wood density and water inundation sensitivity. Model weights we set as the frequency of individuals within the 160 ha forest plot area ≥ 50 cm DBH. We ran a global model with the interaction between these two variables. We tested all combinations of this global model and presented the model with the lowest AIC. An ANOVA type II test was used to calculate the variance explained by each variable (Fig. 3a & b). Weighted models sometimes create instability within an analysis, as some high abundant species may have an unduly large effect on the result. To certify that this was not the case, we ran the model with each species removed and found all models were similar to the results presented.

**Verifying the statistical model.** We tested if our statistical model was able to predict the mean elevations of species in the 4 ha plots species distribution outside the 160 ha plot area. We used the mean elevation as low absolute numbers of individuals for some species made modelling this data with a spatial effect impossible. A linear model with elevation (m asl, based on the wood density and flooding sensitivity of a species) as the explanatory variable, responding to the mean elevation a species was found at in the 4 ha plots (calculated by accounting for spatial autocorrelation). We weight this model by abundances. We calculated an additional r2 value for all the elevations of individuals against the species' predicted values.

**Wood density’s elevational distribution.** We tested the hypothesis that the wood density distribution across the elevation gradient was not linearly consistent. High wood density may be distributed across the elevation gradient at low and high elevations, whereas low wood density species were distributed predominantly in lower elevations. We partition the full data set from all plots (178 ha), into quarter hectare units (2500 m2). The mean elevation and mean wood density within each unit was calculated (n = 720; 648 from 160 ha plot, and 72 from the 4 ½ four ha plots). We used a quantile regression, with wood density as the predictor and elevation as the response. Quantile regression allows us to calculate the conditional quantile function for a response variable, thus enabling us to investigate a more complex relationship. The model calculated five quantiles, shown in Fig. 3d (0.025, 0.1, 0.5, 0.9, 0.975). To calculate the confidence around the coefficients, we bootstrap the data (n = 5000).

**Community similarity analysis.** To analyses dipterocarps community composition across elevation gradients, we partitioned the maps into 12 elevation bands so that there were approximately equal densities of dipterocarps within each band (Harrell, 2016) (mean number of trees in each band was 465.25 ± 80.9 SD). Moreover, using the same elevation bands, we conducted a non-metric multi-dimensional scaling analysis (Oksanen *et al.*, 2016) of the communities and extracted the first axes' values to map of Allen, S. E. (1989). *Chemical analysis of ecological materials* (2nd ed.). Oxford: Blackwell Scientific Publications.

Ashton, P. S. (1988). Dipterocarp biology as a window to the understanding of tropical forest structure. *Annual Review of Ecology and Systematics*, *19*(1988), 347–370.

Ashton, P. S., Givnish, T. J., & Appanah, S. (1988). Staggered flowering in the Dipterocarpaceae: new insights into floral induction and the evolution of mast fruiting in the aseasonal tropics. *The American Society of Naturalists*, *132*(1), 44–66.

Baltzer, J. L., Thomas, S. C., Nilus, R., & Burslem, D. F. R. P. (2005). Edaphic specialization in tropical trees: physiological correlates and responses to reciprocal transplantation. *Ecology*, *86*(11), 3063–3077. https://doi.org/10.1890/04-0598

Bates, D., Maechler Martin, & Walker, S. (2016). Package “lme4”: Linear Mixed-Effects Models using “Eigen” and S4. *CRAN Repository*, 1–113. https://doi.org/10.18637/jss.v067.i01

Born, J., Pluess, A. R., Burslem, D. F. R. P., Nilus, R., Maycock, C. R., & Ghazoul, J. (2014). Differing life history characteristics support coexistence of tree soil generalist and specialist species in tropical rain forests. *Biotropica*, *46*(1), 58–68. https://doi.org/10.1111/btp.12083

Curran, L. M., Caniago, I., Paoli, G. D., Astianti, D., Kusneti, M., Leighton, M., … Haeruman, H. (1999). Impact of El Niño and logging on canopy tree recruitment in borneo. *Science*, *286*(5447), 2184–2188. https://doi.org/10.1126/science.286.5447.2184

Dent, D. H., Bagchi, R., Robinson, D., Majalap-Lee, N., & Burslem, D. F. R. P. (2006). Nutrient fluxes via litterfall and leaf litter decomposition vary across a gradient of soil nutrient supply in a lowland tropical rain forest. *Plant and Soil*, *288*(1–2), 197–215. https://doi.org/10.1007/s11104-006-9108-1

DeWalt, S. J., Ickes, K., Nilus, R., Harms, K. E., & Burslem, D. F. R. P. (2006). Liana habitat associations and community structure in a Bornean lowland tropical forest. *Plant Ecology*, *186*(2), 203–216. https://doi.org/10.1007/s11258-006-9123-6

Gelman, A., Su, Y.-S., Yajima, M., Hill, J., Pittau, M., Kerman, J. ouni, … Vincent, D. (2016). Data Analysis Using Regression and Multilevel/Hierarchical Models. *CRAN Repository*, 1–53.

Green, P. T., Harms, K. E., & Connell, J. H. (2014). Nonrandom, diversifying processes are disproportionately strong in the smallest size classes of a tropical forest. *PNAS*, *111*(52), 18649–18654. https://doi.org/10.1073/pnas.1321892112

Harrell, F. E. (2016). Harrell Miscellaneous. *CRAN Repository*, 1–405.

Hirai, H., Matsumura, H., Hirotani, H., Sakurai, K., Ogino, K., & Lee, H. S. (1997). Soils and the distribution of Dryobalanops aromatica and D. lanceolata in mixed dipterocarp forest - a case study at Lambir Hills National Park, Sarawak, Malaysia. *Tropics*, *7*(July), 21–33. https://doi.org/10.3759/tropics.7.21

Itoh, A., Yamakura, T., Ohkubo, T., Kanzaki, M., Palmiotto, P. A., LaFrankie, J. V., … Lee, H. S. (2003). Importance of topography and soil texture in the spatial distribution of two sympatric dipterocarp trees in a Bornean rainforest. *Ecological Research*, *18*(3), 307–320. https://doi.org/10.1046/j.1440-1703.2003.00556.x

Lemmon, P. E. (1956). A spherical densiometer for estimating forest overstory density. *Forest Science*, *2*, 314–320.

Lindgren, F., Rue, H., & Lindström, J. (2011). An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. *Journal of the Royal Statistical Society, Series B*, *73*, 423–498. https://doi.org/10.1111/j.1467-9868.2011.00777.x

Majalap-Lee, N & Hin, C. N. (1992). *Laboratory Manual for Chemcial Analysis*. Sandakan: Forestry Research Centre, Sandakan, Malaysia.

Margrove, J. A., Burslem, D. F. R. P., Ghazoul, J., Khoo, E., Kettle, C. J., & Maycock, C. R. (2015). Impacts of an extreme precipitation event on dipterocarp mortality and habitat filtering in a Bornean tropical rain forest. *Biotropica*, *47*(1), 66–76. https://doi.org/10.1111/btp.12189

Maycock, C. R., Thewlis, R. N., Ghazoul, J., Nilus, R., & Burslem, D. F. R. P. (2005). Reproduction of dipterocarps during low intensity masting events in a Bornean rain forest. *Journal of Vegetation Science*, *16*(6), 635–646. https://doi.org/10.1111/j.1654-1103.2005.tb02406.x

Moerman, J. W., Cobb, K. M., Adkins, J. F., Sodemann, H., Clark, B., & Tuen, A. A. (2013). Diurnal to interannual rainfall δ18O variations in northern Borneo driven by regional hydrology. *Earth and Planetary Science Letters*, *369*–*370*, 108–119. https://doi.org/10.1016/j.epsl.2013.03.014

O’Brien, M. J., Philipson, C. D., Tay, J., & Hector, A. (2013). The influence of variable rainfall frequency on germination and early growth of shade-tolerant dipterocarp seedlings in Borneo. *PLoS ONE*, *8*(7), 1–9. https://doi.org/10.1371/journal.pone.0070287

Oksanen, J., Guillaume Blanchet, F., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., … Wagner, H. (2016). Community ecology package: ordination, diversity and dissimilarities. *CRAN Repository*, 1–292.

Philipson, C. D., Dent, D. H., O’Brien, M. J., Chamagne, J., Dzulkifli, D., Nilus, R., … Hector, A. (2014). A trait-based trade-off between growth and mortality: evidence from 15 tropical tree species using size-specific relative growth rates. *Ecology and Evolution*, *4*(18), 3675–3688. https://doi.org/10.1002/ece3.1186

R Core Team. (2016). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Rue, H., Martino, S., & Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, *71*(2), 319–392. https://doi.org/10.1111/j.1467-9868.2008.00700.x

Schechtman, E. (2002). Odds ratio, relative risk, absolute risk reduction, and the number needed to treat - Which of these should we use? *Value in Health*, *5*(5), 431–436. https://doi.org/10.1046/j.1524-4733.2002.55150.x

Slik, J. W. F., Poulsen, A. D., Ashton, P. S., Cannon, C. H., Eichhorn, K. A. O., Kartawinata, K., … Wilkie, P. (2003). A floristic analysis of the lowland dipterocarp forests of Borneo. *Journal of Biogeography*, *30*(10), 1517–1531. https://doi.org/10.1046/j.1365-2699.2003.00967.x

Sukri, R. S., Wahab, R. A., Salim, K. A., & Burslem, D. F. R. P. (2012). Habitat associations and community structure of dipterocarps in response to environment and soil conditions in Brunei Darussalam, Northwest Borneo. *Biotropica*, *44*(5), 595–605. https://doi.org/10.1111/j.1744-7429.2011.00837.x

how the community similarity changes with elevation (Fig. 1c). Non-metric multi-dimensional scaling collapses information from multiple dimensions into few dimensions similar to a Principal Component Analysis, however, uses rank order rather than Euclidian distances. Although this method does not account for spatial autocorrelation, it does provide some inference as to how communities partition along the elevation gradients.

All linear and ANOVA model residuals had an expectation of zero, equality of variance, independence and followed a Gaussian distribution – unless specified prior. We conducted all analyses in R v3.3.1(R Core Team, 2016). To examine our analysis, pull from GitHub repository: https://github.com/t03jam8/ForestFloodingSensitivityAnalysis.