Online Appendix

Florian Hartig, Claudia Dislich, Thorsten Wiegand and Andreas Huth

July 9, 2013

¹ 1 FORMIND model details

FORMIND is an individual-based, spatially semi-explicit, dynamic forest model. 2 Spatially semi-explicit means that trees are assigned to a spatial 20x20m grid. 3 Trees have no explicit position within a grid cell, and, horizontally, their crown л area is homogeneously distributed across their respective grid cell. Trees mostly interact within grid cells, essentially through the mechanisms that are present 6 in all classical gap models (e.g. Shugart, 1984; Bugmann, 2001): for each tree, growth and establishment depends on the light climate at its crown top. The 8 light climate is determined by the overtopping leaf area. Additional to this central process of competition for light on grid cells, FORMIND implements a 10 number of other processes that act across grid cells such as tree falling and seed 11 dispersal between grid cells. We use the FORMIND model version of Dislich 12 et al. (2009), developed for a tropical rainforest in Ecuardor, with some minor 13 updates that have accumulated since then due to the general development of 14 the model. The FORMIND scheduling within one time step is given in Alg. 1, 15 a visual representation of the model concept in Fig. 1 16

Establishment is modeled as a constant seed rain, meaning that tree regeneration is independent of relative species abundances in the modeled tree community. Provided that species-specific light conditions are met, the number of new recruits appearing on a site is a species specific parameter of the model. There is stochasticity in the recruitment regarding the spatial distribution of recruits.

After establishment, mortality acts on all trees individuals. Mortality originates from four sources: 1) Base mortality: each tree has a species-specific
base mortality rate that is independent of its age and environmental conditions.
2) Small trees have an additional species specific size-dependent mortality (see
Dislich et al., 2009, Appendix A). 3) Self-thinning: when the crown area in a



Figure 1: Formind model principles

28	particular height layer exceeds the plot area, trees are randomly removed until
29	the layer is sufficiently thinned. 4) Gap-formation: When trees larger than a
30	threshold diameter d_f die, they are assumed to fall on a neighboring plot and
31	produce an additional mortality proportional to their crown size on all trees
32	that do not exceed the height of the falling tree by one meter. In that sense,
33	$\operatorname{FORMIND}$ differs from more traditional gap models, where only one "represen-
34	tative" plot is modeled. All four mortality processes are modeled stochastically.

Algo	rithm 1 FORMIND scheduling
1: R	tead initial tree configuration
2: f	$\mathbf{pr} \ t = 1 \ \mathrm{to} \ t_{\mathrm{end}} \ \mathbf{do}$
3:	for all Plots do
4:	${f if}$ Light conditions allow establishment ${f then}$
5:	Establishment of new seedlings from seed bank
6:	end if
7:	for all Trees do
8:	Mortality
9:	end for
10:	Update light climate
11:	for all Trees do
12:	Growth
13:	end for
14:	end for
15: e	nd for

35

2 STATISTICAL ALGORITHM

culated. Productivity and growth are modeled deterministically and depend on 36 tree size and light climate at crown top, corrected by self-shading, according 37 PFT-specific functions for light-response, photosynthesis and respiration. The 38 light climate on the plot is derived by calculating the leaf area contributed by 39 the trees on the plot to different height layers, and from that the light inter-40 ception of the different hight layers. Respiration rates are calculated according 41 to an inverse method which takes maximum growth rates under full light as an input (Dislich et al., 2009, Appendix A). Maximum growth rates can be derived 43 from observations, but in this study, we treat growth rates as parameters that 44 are fit to observed community data. 45

For reasons of computability, seedlings of the same PFT and age within one
plot are grouped into cohorts, which is mathematically identical to calculating
individual trees as there are no stochastic effects on existing tree individuals
except for mortality.

50 2 Statistical Algorithm

The algorithm for Bayesian parameter estimation was implemented in Python 51 2.6, using Scipy, Numpy and parallel python. Parallel python was used to speed 52 up the MCMC algorithm - instead of calculating the posterior value of one new 53 parameter proposal, we always propose n values in parallel (here, n was 6 or 12). 54 If the first value was rejected, the algorithm goes on to check for acceptance of 55 the second value and so on. If one value was accepted, the other values were 56 discarded. The acceptance check was done strictly ordered, so that the order of 57 steps within this algorithm is identical to that of a usual MCMC. The advantage, 58 however, is that time is saved in the case of rejection because practically all 59 our computational costs are for the FORMIND calculations (the estimation of 60 mean and covariance of one parameter combination for a typical situation of 7 61 PFTs on 1 ha over 10.000 yrs with 5yr time steps takes around 20 seconds). 62 Therefore, parallel proposals create a considerable speed-up (maximum a factor 63 n) when there are high rejection rates. Based on the observed rejection rates, 64 we estimate that the average speed-up through parallelization with 6 cores was 65 approximately a factor 3-4. 66

Algorithm 2 MCMC-SPL (parallel version)

- 1: Choose initial condition
- 2: Calculate initial unnormalized posterior value (eqs. 1,2, main text)
- 3: repeat
- 4: Propose n new ϕ according to proposal function $q(\phi \rightarrow \phi')$
- 5: Create *n* proposal ϕ_i and run the model with those in parallel
- 6: repeat
- 7: Estimate $p(\phi'_i | D_{obs})$ according to eq. 1,2, main text
- 8: Accept ϕ'_i with probability $\frac{p(\phi'|D_{obs})q(\phi' \to \phi)}{p(\phi|D_{obs})q(\phi \to \phi')}$, else stay at ϕ
- 9: **until** Acceptance of one ϕ'_i or all n runs tested
- 10: If applicable, adjust $q(\phi \rightarrow \phi')$ according to (Haario et al., 2001)
- 11: **until** Convergence

The algorithm was started with random initial values ϕ^{init} that were gen-67 erated by adding a random parameter vector ϕ^r to the prior best estimate ϕ^* 68 according to $\phi^{\text{init}} = 0.9 \cdot \phi^* + 0.1 \cdot \phi^r$. The best estimate ϕ^* was the "true" 69 value for the virtual tropical forest, and the value from Dislich et al. (2009) 70 for the Ecuadorian parameterization. The proposal function $q(\phi \rightarrow \phi')$ was 71 chosen multivariate normal, with a covariance adaptation according to (Haario 72 et al., 2001) for the parameterizations to the virtual data, and a fixed pro-73 posal function for the parameterization to the Ecuadorian data. The covariance 74 adaptation of Haario et al. (2001) sets the covariance of the proposal function 75 as $\Sigma_i = c * cov_i(p(\Phi|D))$, where i is the i - th step of the algorithm, and the 76 scaling parameter c is a constant whose optimal choice depends on the target 77 function (we used $c = 2.38^2/d$, where d is the number of dimensions of the 78 parameter space). Although the adaptive algorithm leads to more efficient pro-79 posal generation under correlations in the posterior, we noted that there were 80 some remaining inefficiencies in the proposal generation that were probably due 81 to the observed nonlinear and higher-order correlations in the posterior. To 82 minimize the effect of those, we drew did not vary all parameters at once in one 83 step of the MCMC, but first drew two random parameters, and then drew a 84 proposal for those according to Alg. 2. 85

3 Additional analyses of the parameter estimates

In this section, we show more detailed analyses for the results presented in the
main article (in particular Figs. 3,5), and additional analyses that replicates the
setup of Figs. 3,5 with differences in the number of parameters estimated, and

CASE	EXPLANATION	DIMENSIONS					
Parameterization to virtual data, 3 PFTs:							
V1	Data: SDD, GRO, reduced parameters	12,96					
V2	Data: SDD, GRO, full parameters	26,96					
V3	Data: SDD, reduced parameters	12, 48					
V4	Data: total SDD, reduced parameters	12, 16					
V5	Data: BM, reduced parameters	12, 3					
Parameterization to Ecuadorian field data, 7 PFTs:							
E1	Data: SSD	18,112					

Table 1: Overview of parameterizations for different models, parameters and summary statistics. Abbreviations for the data: SSD = Stem size distribution (16 10-cm classes), GRO = size-specific average growth distribution (16 10-cm classes), BM = Biomass. If not stated otherwise, the data type was available for each PFT separately. If we use the mean over all PFTs, we label the data with "total". Full parameters means that the parameters given in tables 1,2 in the online supplementary are under calibration. Reduced parameters means that from table 1, only recruitment, mortality, maximum growth and maximum growth diameter are estimated. Dimensions gives the number of parameters and the number of data points in that order.

in the aggregation type (summary statistics) that are used to compare model 90 results and field data. A summary of the cases considered is provided in Table 1. 91 For all these cases, we show a) a histogram of the marginal posterior density, 92 which allows gaining a better picture of the distribution represented by the violin 93 plots in the main paper, and b) a plot of posterior pair correlation density. The 94 width of the marginal distributions was scaled to the prior width (denoted by 95 the green lines at the sides of the plot). For the virtual case, the red line depicts the "true" parameter value that was used to create the virtual field data. For 97 the dataset from Ecuador, the red line depicts the parameter values chosen by 98 (Dislich et al., 2009). However, as we note in the main text, the model setup 99 was not completely identical, so there are limits in the comparability of the 100 inference with these values. Prior and true parameter values are also provided 101 in Tables 2.3. In the caption of the marginal density plots, we provide some 102 additional information for the runs such as sample size, convergence diagnostics 103 (using Geldman-Rubin, see Gelman and Rubin, 1992; Brooks and Gelman, 1998) 104 and runtime. In all cases, we removed 100.000 samples as burn-in from the 105 chains. 106

¹⁰⁷ 3.1 V1 Parameterization to virtual data, details for results ¹⁰⁸ from the main paper

Figs. 2,3 show detailed plots for Fig. 3 of the main text, which allows a better assessment of the shape of the distributions, and of the parameter correlations.

¹¹¹ 3.2 V2 Parameterization to virtual data, full parameter ¹¹² set

Figs. 4,5 uses the same data as V1, but with a larger number of parameters esti-113 mated. Those additional parameters are parameters for the crown geometry, for 114 specific leaf area, the light extinction coefficient, the leaf area index (LAI) per 115 tree, and tree mortality at gap formation (Table 2). As can be seen, marginal 116 parameter uncertainty considerably increases when fitting a larger number of 117 parameters, which must be due to additional trade-offs between the old and 118 the new parameters with respect to the data. The lower amount of strong pair 119 correlation in Fig. 5 as compared to Fig. 2 suggests that those newly introduced 120 trade-offs are predominantly of higher order and therefore not picked up by the 121 pair correlation plots. Again, we stress that this is neither a fundamental prob-122 lem, nor specific to the simulation-based likelihood approximation, but simply 123 a result from interactions between parameters with respect to the data used 124

for the fit - we would most probably find the same results in a conventional Bayesian analysis. However, the wide marginal distributions that result from these correlations make it difficult to see how parameter uncertainty is affected by the simulation-based approximation and by the choice of model output. The latter is the reason why we estimated only a reduced set of parameters for the main results.

3.3 V3-V5 Parameterization to virtual data with more ag gregated model outputs

The next plots Figs. 6-11 show results from fitting the same parameters as 133 in Fig. 3, main text, but using more aggregated model outputs (i.e. more 134 coarse summary statistics) for the comparison between model and observed 135 data. One can see that width of the posterior distribution generally increases 136 when going to more aggregated representations of the data. Also some new 137 trade-offs between parameters appear while going to more aggregated outputs, 138 while others disappear, potentially indicating higher-order interactions between 139 parameters with respect to this pattern. 140

¹⁴¹ 3.4 E1 Parameterization to Ecuadorian data

Finally, Figs. 12-13 show details of the parameter estimation with field data 142 from Ecuador. Our data consisted of size-abundance distributions only. From 143 our results for the virtual dataset (Fig 7), we know that this data type leads 144 to relatively strong correlations when fitting similar parameters as in V3, which 145 makes the result difficult to interpret. To avoid these correlations, we estimated 146 a lower number of parameters per plant functional type than for the virtual 147 case. However, note that the number of parameters is still larger than for V3, 148 due to the higher number of plant function types. For the model parameters 149 that were not fit to data, we used the values from Dislich et al. (2009). 150



Figure 2: Marginal posterior densities for case V1 in Table 1. Result from 3 chains; sample size per chain: ca. 1.3 million; Gelman-Rubin multivariate potential scale reduction factor: 1.01; runtime: ca. 6 weeks, 6 parallel cores per chain. Model parameters are explained in Table 2



Figure 3: Pair correlation density plot of the posterior distribution for case V1 in Table 1. As in Fig. 3b, main text, the histograms on the diagonal show the marginal distributions for the posterior parameter estimates that were also displayed in the previous figure. The panels in the lower triangle show pairwise correlations between the parameters for which marginals are displayed on the diagonal. The numbers in the upper triangle show Spearman's rank correlation coefficients for the correlations in the lower triangle. Parameter abbreviations are explained in Table 2.

9



Figure 4: Marginal posterior densities for case V2 in Table 1. Result from 3 chains; sample size per chain: ca. 1.4 million; Gelman-Rubin multivariate potential scale reduction factor: 1.17; runtime: ca. 6 weeks, 6 parallel cores per chain. Model parameters are explained in Table 2



Figure 5: Pair correlation density plot of the posterior distribution for case V2 in Table 1. As in Fig. 3b, main text, the histograms on the diagonal show the marginal distributions for the posterior parameter estimates that were also displayed in the previous figure. The panels in the lower triangle show pairwise correlations between the parameters for which marginals are displayed on the diagonal. The numbers in the upper triangle show Spearman's rank correlation coefficients for the correlations in the lower triangle. Parameter abbreviations are explained in Table 2.



Figure 6: Marginal posterior densities for case V3 in Table 1. Result from 2 chains; sample size per chain: ca. 1.4 million; Gelman-Rubin multivariate potential scale reduction factor: 1.02; runtime: ca. 6 weeks, 6 parallel cores per chain. Model parameters are explained in Table 2



Figure 7: Pair correlation density plot of the posterior distribution for case V3 in Table 1. As in Fig. 3b, main text, the histograms on the diagonal show the marginal distributions for the posterior parameter estimates that were also displayed in the previous figure. The panels in the lower triangle show pairwise correlations between the parameters for which marginals are displayed on the diagonal. The numbers in the upper triangle show Spearman's rank correlation coefficients for the correlations in the lower triangle. Parameter abbreviations are explained in Table 2.



Figure 8: Marginal posterior densities for case V4 in Table 1. Result from 2 chains; sample size per chain: ca. 0.7 million; Gelman-Rubin multivariate potential scale reduction factor: 1.04; runtime: runtime: ca. 6 weeks, 6 parallel cores per chain. Model parameters are explained in Table 2



Figure 9: Pair correlation density plot of the posterior distribution for case V4 in Table 1. As in Fig. 3b, main text, the histograms on the diagonal show the marginal distributions for the posterior parameter estimates that were also displayed in the previous figure. The panels in the lower triangle show pairwise correlations between the parameters for which marginals are displayed on the diagonal. The numbers in the upper triangle show Spearman's rank correlations coefficients for the correlations in the lower triangle. Parameter abbreviations are explained in Table 2.

mort3

gro1

gro2

gro3

dia1

dia2

dia3

dia3

recr1

recr2

recr3

mort1

mort2



Figure 10: Marginal posterior densities for case V5 in Table 1. Result from 2 chains; sample size per chain: ca. 1 million; Gelman-Rubin multivariate potential scale reduction factor: 1.08; runtime: ca. 6 weeks, 6 parallel cores per chain. Model parameters are explained in Table 2





Figure 11: Pair correlation density plot of the posterior distribution for case V5 in Table 1. As in Fig. 3b, main text, the histograms on the diagonal show the marginal distributions for the posterior parameter estimates that were also displayed in the previous figure. The panels in the lower triangle show pairwise correlations between the parameters for which marginals are displayed on the diagonal. The numbers in the upper triangle show Spearman's rank correlation coefficients for the correlations in the lower triangle. Parameter abbreviations are explained in Table 2.



Figure 12: Marginal posterior densities for case E1 in Table 1. Red bars show the parameter estimates by Dislich et al. (2009). Result from 3 chains; sample size per chain: 0.9 million; Gelman-Rubin multivariate potential scale reduction factor: 1.19; runtime: ca. 5 weeks, 12 parallel cores per chain. Model parameters are explained in Table 3



Figure 13: Pair correlation density plot of the posterior distribution for case E1 in Table 1. As in Fig. 3b, main text, the histograms on the diagonal show the marginal distributions for the posterior parameter estimates that were also displayed in the previous figure. The panels in the lower triangle show pairwise correlations between the parameters for which marginals are displayed on the diagonal. The numbers in the upper triangle show Spearman's rank correlation coefficients for the correlations in the lower triangle. Parameter abbreviations are explained in Table 3. Note that the 4 growth parameters are assigned to the 7 PFTs as follows: gro1 -> PFT 2, gro2 -> PFT 1, gro3 ->PFTs 3,4, gro4 -> PFTs 5,6,7, which is reflected by the correlation structure.

¹⁵¹ 4 Prior ranges

Tables 2,3 show prior ranges and additional information or the parameters estimated in V1-V5 and E1, respectively.

Parameter	ABBR.	Lower-upper (true)	UNITS
light extinction coefficient	li-ext	0.4-0.8 (0.6)	$[m^2 \cdot m^{-2}]$
m recruitment rate pft1	recr1	50-200 (100)	[ind/ha/yr]
m recruitment rate pft2	recr2	15-50 (30)	[ind/ha/yr]
m recruitment rate pft3	recr3	10-40 (20)	[ind/ha/yr]
min light for establishment pft1	estab1	0-0.3(0.1)	[-]
min light for establishment $pft2$	estab2	0-0.15 (0.05)	[-]
min light for establishment $pft3$	estab3	0-0.1 (0.01)	[-]
mortality pft1	mort1	0-0.25 (0.05)	$[yr^{-1}]$
mortality pft2	mort2	$0-0.1 \ (0.15)$	$[yr^{-1}]$
mortality pft3	mort3	$0-0.05 \ (0.005)$	$[yr^{-1}]$
falling probability of trees	fall	$0.2 - 0.7 \ (0.4)$	[-]
leaf area index per tree	LAI	1.5-2.5 (2)	$[m^2 \cdot m^{-2}]$
$\operatorname{crown} \operatorname{diameter}$	cr-d	$0.12 - 0.2 \ (0.15)$	[-]
$\operatorname{crown}\operatorname{length}$	cr-l	$0.1 \text{-} 0.35 \ (0.25)$	[-]
max dbh growth pf1	gro1	20-80 (41)	[mm/yr]
$\max \ dbh \ growth \ pf2$	$\operatorname{gro2}$	5-15 (9.2)	[mm/yr]
$\max \ dbh \ growth \ pf3$	$\mathrm{gro}3$	2-6 (3.5)	[mm/yr]
start growth pf1 (% of max)	start1	0-100 (40)	[-]
start growth pf2 (% of max)	start2	0-100 (40)	[-]
start growth pf3 ($\%$ of max)	$\operatorname{start3}$	0-100 (40)	[-]
end growth pft1 ($\%$ of max)	end1	0-100 (10)	[-]
end growth pft2 (% of max)	$\mathrm{end}2$	0-100 (10)	[-]
end growth pft3 ($\%$ of max)	end3	0-100 (10)	[-]
$\max \text{ growth diameter pft1}$	dia1	0.0-1.0 $(1/3)$	[-]
$\max \text{ growth diameter pft2}$	dia2	0.0-1.0 $(1/3)$	[-]
$\max \text{ growth diameter pft3}$	dia3	$0.0-1.0 \ (1/3)$	[-]

Table 2: Ranges for the uniform priors used for fitting the model to the virtual data. "Abbr." refers to the parameter abbreviation used in the figures. Lower, upper refers to the lower and upper bound of the uniform prior distributions. True refers to the values used to create the virtual data.

PARAMETER	ABBR.	Lower-upper (Dislich et al.)	UNITS
recruitment rate pft1	recr1	5-100 (50)	[ind/ha/yr]
m recruitment rate pft2	recr2	100-300 (180)	[ind/ha/yr]
$ m recruitment\ rate\ pft3$	recr3	50-250 (130)	[ind/ha/yr]
recruitment rate pft4	recr4	10-100 (50)	[ind/ha/yr]
m recruitment rate pft5	recr5	50-200 (120)	[ind/ha/yr]
m recruitment rate pft6	recr6	100-500 (310)	[ind/ha/yr]
m recruitment rate pft7	recr7	20-100 (50)	[ind/ha/yr]
mortality pft1	$\mathrm{mort1}$	0-0.25 (0.05)	$[yr^{-1}]$
mortality $pft2$	$\mathrm{mort2}$	0-0.1 (0.09)	$[yr^{-1}]$
mortality pft3	$\mathrm{mort}3$	$0-0.1 \ (0.05)$	$[yr^{-1}]$
mortality pft4	$\operatorname{mort} 4$	0-0.25 (0.05)	$[yr^{-1}]$
mortality pft5	$\mathrm{mort}5$	0-0.1 (0.06)	$[yr^{-1}]$
mortality pft6	$\mathrm{mort6}$	$0-0.05 \ (0.018)$	$[yr^{-1}]$
mortality pft7	$\operatorname{mort} 7$	0-0.25 (0.008)	$[yr^{-1}]$
max dbh growth type1	$\operatorname{gro1}$	10-40 (20)	$[\mathrm{mm/yr}]$
$\max \ \mathrm{dbh} \ \mathrm{growth} \ \mathrm{type}2$	$\operatorname{gro}2$	5-30(10)	$[\mathrm{mm/yr}]$
$\max \ \mathrm{dbh} \ \mathrm{growth} \ \mathrm{type3}$	${ m gro}3$	2-30 (6)	$[\mathrm{mm/yr}]$
$\max \ dbh \ growth \ type4$	$\operatorname{gro4}$	1-15(2)	$[\rm mm/yr]$
\max growth diameter pft1	dia1	$0.0-1.0\ (0.33)$	[-]
max growth diameter $pft2$	dia2	$0.0-1.0\ (0.33)$	[-]
\max growth diameter pft3	dia3	$0.0-1.0\ (0.25)$	[-]
max growth diameter $pft4$	dia4	0.0-1.0 (0.33)	[-]
\max growth diameter pft5	dia5	0.0-1.0 (0.2)	[-]
\max growth diameter pft6	dia6	$0.0-1.0\ (0.33)$	[-]
\max growth diameter pft7	dia7	$0.0-1.0\ (0.33)$	[-]

Table 3: Ranges for the uniform priors for the Ecuadorian fit. Note that the grouping is for 7 PFTs, but there are also 4 broader growth types to which the 7 PFTs belong (see Dislich et al., 2009, for details). "Abbr." refers to the parameter abbreviation used in the figures. Lower, upper refers to the lower and upper bound of the uniform prior distributions. Dislicht et al. refers to the parameter values used in Dislich et al. (2009).

154 References

- Brooks, S., Gelman, A., 1998. General methods for monitoring convergence of
 iterative simulations. J. Comput. Graph Stat., 434–455.
- Bugmann, H., 2001. A Review of Forest Gap Models. Clim. Change 51 (3),
 259–305.
- Dislich, C., Günter, S., Homeier, J., Schröder, B., Huth, A., 2009. Simulating
 forest dynamics of a tropical montane forest in South Ecuador. Erdkunde 63,
- 161 <u>347–364</u>.
- doi: 10.1098/rstb.2003.1425
- Gelman, A., Rubin, D., 1992. Inference from iterative simulation using multiple
 sequences. Stat. Sci. 7 (4), 457–472.
- Haario, H., Saksman, E., Tamminen, J., 2001. An adaptive Metropolis algorithm. Bernoulli 7 (2), 223–242.
- Shugart, H. H., 1984. A Theory of Forest Dynamics: The Ecological Implications
 of Forest Succession Models. Springer, New York, USA, New York.