Computational assessment of Amazon forest plots regrowth capacity under strong spatial variability for simulating logging scenarios

Gilles Ardourel, Guillaume Cantin, Benoît Delahaye, Géraldine Derroire, Beatriz M Funatsu, David Julien

Computational assessment of Amazon forest plots regrowth capacity under strong spatial variability for simulating logging scenarios

Gilles Ardourel[*], Guillaume Cantin[†‡], Benoît Delahaye[§], Géraldine Derroire[¶], Beatriz M. Funatsu[‖], David Julien[**]

August 27, 2024

Abstract

In this paper, we assess the regrowth capacity of tropical forest plots by developing an original computational procedure based on statistical model checking methods. We calibrate a new mathematical model of forest dynamics with respect to post-logging data, produced in the Amazon basin. Our mathematical model is determined by a mechanistic system of ordinary differential equations and integrates a new non-linear aging term, which is necessary to reproduce the complex post-logging dynamics of the tropical forest plots. Our method is based on an efficient algorithmic procedure that explores the parameter space of the model and computes a score for each parameter value, depending on the distance between the trajectory of the model and the data. We distinguish a group of four reference plots, for which the logging was the most intense, from another group of five non-reference plots, which are fitted *a posteriori*. Our results provide a set trajectories of the new model, which successfully fit the non-monotone post-logging data on each forest plot, in spite of the high level of biological variability identified in the study site. Rather than a unique set of parameters, we return a small cell of parameters, extracted from the parameter space, which contains in its close vicinity several relevant sets of parameters that are equally able to reproduce the regrowth dynamics of distinct tropical forest plots. The size of the cells that should be extracted from the parameter space increases with the level of biological variability. Finally, we show how to use the calibrated mathematical model for simulating logging scenarios, so as to better understand the temporal dynamics of forest regrowth.

Model Checking – Forest ecosystem – Variability – Parameter synthesis – Land-use data

1 Introduction

Understanding climate-vegetation interactions remains a complex, interdisciplinary challenge. Despite the development of dynamical vegetation models that incorporate several aspects of these interactions [11, 20, 37], modeling of tropical forest dynamics remains particularly challenging, partially due to the large spatial variability of plant species, and to the difficulty in obtaining long-term data on parameters that characterize tree distribution, age, and growth [39, 58]. At the same time, forest management, although it may decrease carbon stocks, is considered as a credible solution for mitigation of climate change effects, and for the

[*]Nantes Université, Laboratoire des Sciences du Numérique de Nantes, CNRS, UMR 6004, École Centrale Nantes, F-44000 Nantes, France.

[†]Nantes Université, Laboratoire des Sciences du Numérique de Nantes, CNRS, UMR 6004, École Centrale Nantes, F-44000 Nantes, France.

[‡]Corresponding author: guillaume.cantin@ls2n.fr

[§]Nantes Université, Laboratoire des Sciences du Numérique de Nantes, CNRS, UMR 6004, École Centrale Nantes, F-44000 Nantes, France.

[¶]CIRAD, UMR EcoFoG (Agroparistech, CNRS, INRAE, Université des Antilles, Université de la Guyane), Kourou, French Guiana, France

[‖]CNRS, UMR 6554 LETG, Bâtiment IGARUN, Campus du Tertre, Nantes Université, France.

[**]Nantes Université, Laboratoire des Sciences du Numérique de Nantes, CNRS, UMR 6004, École Centrale Nantes, F-44000 Nantes, France.

preservation of ecosystem services [18, 24, 25, 44, 46], in a high level of deforestation, degradation and biodiversity loss context [15, 21, 34, 43, 47, 56]. Thus, modeling of forest ecosystems dynamics is essential for the assessment of their effectiveness in terms of forest regeneration, productivity, and estimation of carbon balance [19, 57], and has direct implications for the study of anthropic effects in vulnerable hotspots such as in the Amazon tropical forest [44, 55]. In the past decades, a great number of models have been proposed to study the dynamics of tropical forests. As discussed in the survey [7], these models can be classified into several types. First, statistical models, resulting from a Bayesian approach, show a nice ability to reproduce observation data. However, they often do not provide any explanation on the behavior of the forest ecosystem, in terms of biology or population dynamics. To remedy this lack, numerous phenomenological models have been studied. They can be designed at the individual tree scale (with for instance an agent-based approach, see notably the TROLL model [12] and the survey [59]) or at the populational forest scale (using differential equations). The model of Korzukhin and Antonovsky (presented in [30]), which is at the basis of our work, corresponds to the latter category of models. However, it has never been calibrated to fit with observation data in tropical forests. Since these forests exhibit a high level of complexity and variability, it was a great challenge to adapt and improve it, for performing a parameter estimation of this explanatory model. Our paper has precisely overcome this difficult challenge, by employing an efficient computational method.

Hence, in this paper, we show how a computational approach based on *model checking* methods is successfully applied to model and analyse the regrowth capacity of tropical forest plots. The model checking approach was first introduced in the papers [13] and [51], with the aim to verify, through an efficient algorithmic procedure, if a model of a system satisfies a given property. This algorithmic method was primarily designed for the formal verification of systems arising in engineering or computer programming, and tested on various properties guaranteeing the safety of a system. Its efficiency was recognized by the attribution of the Turing award to their contributors. These methods now admit statistical extensions, which can be used in the context of probabilistic systems, especially for verifying the property of adequacy of a model to observation data (see [33]). We emphasize that these methods, along with their computational results, provide *formal guarantees*, which give a measure of sharpness of the results. Recently, they have been successfully applied in the context of life sciences [29, 32, 52], but their use is unprecedented in the study of tropical forest dynamics. Results obtained from this approach, using a forest dynamic model determined by a system of parametric ordinary differential equations [10, 11], show that the model produces a remarkable fit that successfully explains the observed post-logging trajectories from the Paracou forest station in French Guiana. This in itself is a non-trivial finding, as two specific challenges had to be addressed. First, the post-logging data and the mathematical model are not fully compatible, in the sense that the mathematical model is age-structured, whereas the post-logging data do not distinguish trees of different ages. Consequently, classical mathematical methods of parameter estimation, for the application of which *identifiability* is a necessary condition to be fulfilled, cannot always be applied [22]. This challenge was overcome through an innovative computational approach, relying on *statistical* model checking, for which we perform a thorough exploration of the parameter space of the new forest dynamics model, admitting a nonlinear aging term. While it could be argued that a choice of a simplest mathematical model without any age structure would be more relevant, we will show that the nonlinear age structure is necessary for reproducing a characteristic two-phase regrowth of the tropical forest in a post-logging context.

The second challenge concerns the decision to be made on the output of the computational procedure. Indeed, a great number of candidates that could be selected as best parameters, with respect to arbitrary criteria, are produced by the implementation of our method. Hence, a post-treatment procedure must be designed in order to perform a relevant selection. On this point, we consider three kinds of criteria. First, quantitative criteria, such as the average distance between the trajectory of the model and the data, or the number of points of the trajectory which remain outside an uncertainty tunnel around the data, allow to reduce the number of candidates. Next, qualitative criteria, such as the shape of the trajectory, can underpin the discrimination of candidates. Finally, a careful comparison with ecological or biological indicators allows to reduce the number of candidates to a reasonable level. At the end of the selection procedure, the result we provide is not a unique set of parameters, but rather a small *cell* of parameters, extracted from a fine discretization of the parameter space, which contains several sets of parameters for which the trajectories of the model can fit the data of distinct forest plots. In other words, the high level of spatial variability between several plots of the forest ecosystem is described by a small neighborhood of numerical values contained in the huge parameter space, and not by a single set of parameters.

A second objective we pursue in this paper is to show how the calibrated mathematical model can be used to better anticipate the dynamics of forests which are subject to logging. To that aim, we experiment several plans of stem logging and use our calibrated model to predict various scenarios of forest regrowth, fair with respect to the variability identified between the plots of the forest ecosystem. Our results concord with recent simulations of post-logging regrowth showing that high-intensity logging can have lasting effects on stand regeneration dynamics, with recovery times of biomass and timber volume varying two-fold between moderate and very intense logging conditions [26, 27, 28]. Moreover, by coupling our calibrated model with recent predictions of tree mortality increase in the context of global warming [1, 35, 36, 40] and in particular to the site of Paracou [5], we provide numerical simulations of several potential future climatic contexts. Obviously, our predictions should be considered as *tendencies* rather than a precise reading of a near future. However, these tendencies are sufficiently justified to produce tangible information on the dynamics of forest regrowth.

We emphasize that, in its entirety, our computational procedure is highly innovative and has never been, to the best of our knowledge, applied to the study of tropical forest ecosystems. It has been designed in such a manner that it can be readily applied to a great number of real-world systems where a characteristic variability is observed. Furthermore, it has the ability to validate or to exclude a given model. The ability to exclude an input model, which will be proved in this paper for rejecting a forest model with a linear aging term, represents a very valuable help for the modeler who always searches a better model for describing a system under study. Finally, the implementation and execution of our computational procedure can be automatized and terminate correctly, even on input data which are not fully compatible with the input model. Overall, the flexibility and efficiency of our computational technique are solid guarantees for a genuine advance on calibration methods.

Our paper is organized as follows. In the Material and Methods Section, we first present the Paracou research station, located in the Amazon forest, and its post-logging forest database which we intend to learn the dynamics. We also present the mathematical model of forest growth, whose weakness will be detected by the exclusion mode of our method, and its resulting improvement that integrates a nonlinear aging term, expected to reproduce the forest dynamics. Our computational procedure is then exposed in detail, with a thorough description of its algorithmic progression and a critical analysis of its inputs, outputs and parameters. Next, we give the results of the execution of our calibration method on the Paracou station database. The parameters found for each forest plot are provided, as well as the visualizations of the corresponding trajectories. These results distinguish the fitting of the reference plots, for which the logging was the most intense, from the non-reference plots, which are fitted *a posteriori*. Finally, in the Discussion Section, we compare our results with relevant ecological indicators and we test the robustness of the calibrated forest model by simulating logging scenarios.

2 Material and Methods

2.1 Paracou research station and forest data production protocol

Paracou is a research station located in Sinnamary (French Guiana), dedicated to studying the functioning of the Amazon forest ecosystem (see Figure 1). With 40 years of hindsight on forest dynamics after logging, this facility makes it possible to test the response of tropical forest ecosystems to disturbances exacerbated by global changes. Paracou station is divided into 16 forest plots, which cover a total of 125 hectares on which each of 70 000 trees has been mapped and measured at regular intervals since 1984 (see Figure 2). Indeed, on each forest plot, each tree is referenced by its identifier, its status (0 for dead or 1 for alive), its geographical position (given by local coordinates, converted into global latitude and longitude), its species (completely described by family, genus, species), its size (given by the diameter of the trunk, measured at

breast height), and the year of the measure. For example, the following entry is extracted from plot 1:

$$
\begin{aligned}
\mathrm{T} = \{ & \\
& \mathrm{id} = 75475, \\
& \mathrm{status} = 1, \\
& \mathrm{loc} = (285109.5625, 582983.25), \\
& \mathrm{glob} = (5.27129983901978, -52.9388999938965), \\
& \mathrm{bota} = (\mathit{Euphorbiaceae}, \mathit{Conceveiba}, \mathit{guianensis}), \\
& \mathrm{size} = 56.5, \\
& \mathrm{year} = 1991 \\
\}. &
\end{aligned}
$$

Figure 1: Paracou research station, located in Sinnamary (French Guiana, South America). The station is divided into 16 forest plots, which cover a total of 125 hectares. In this station, 9 of the 16 forest plots have been subject to partial logging in 1988, whereas other plots serve as *control* plots.

This information is crucial and original for understanding the long-term functioning of the forest ecosystem. The station includes several experimental areas that have been set up progressively: in 1984, an experiment of 12 plots of 6.25 ha each, intended to study the impact of different silvicultural interventions; in 1991, 4 new plots dedicated to studies of forest functioning in a little or undisturbed environment; in 2003, a 57 meters high instrumented flux tower overhanging the forest canopy to measure gas exchanges between the ecosystem and the atmosphere. With a time depth of 4 decades, the Paracou forest database has allowed to better understand the dynamics of the Amazon forest [6, 48]. In particular, the coexistence of a great number of trees species has been highlighted in [41].

It is worth emphasizing that these forest plots (see Figure 2) show a high level of variability in time and space (that is, from one plot to another). We also highlight that 9 of the 16 forest plots of Paracou research station have been subject to selective logging in 1988 (see plots 2, 3, 4, 5, 7, 8, 9, 10, 12 in Figure 2), whereas other plots serve as *control* plots. The 9 plots that have been subject to partial logging present a typical dynamical forest regrowth, which can be divided into two characteristic periods: at first, the forest admits a rapid regrowth and the number of trees reaches a maximum which overcomes the maximum number of trees observed when the forest is at equilibrium; then, the number of trees slowly decreases and tends to equilibrium. Hence the maximum which is reached before decreasing towards equilibrium will be qualified as an *over-maximum*. The two-phase regrowth is more visible if the intensity of the logging increases (for instance, it is less intense on plot 7 than on plot 8, as shown in Figure 2). For that reason, plots 3, 4, 8, 10,

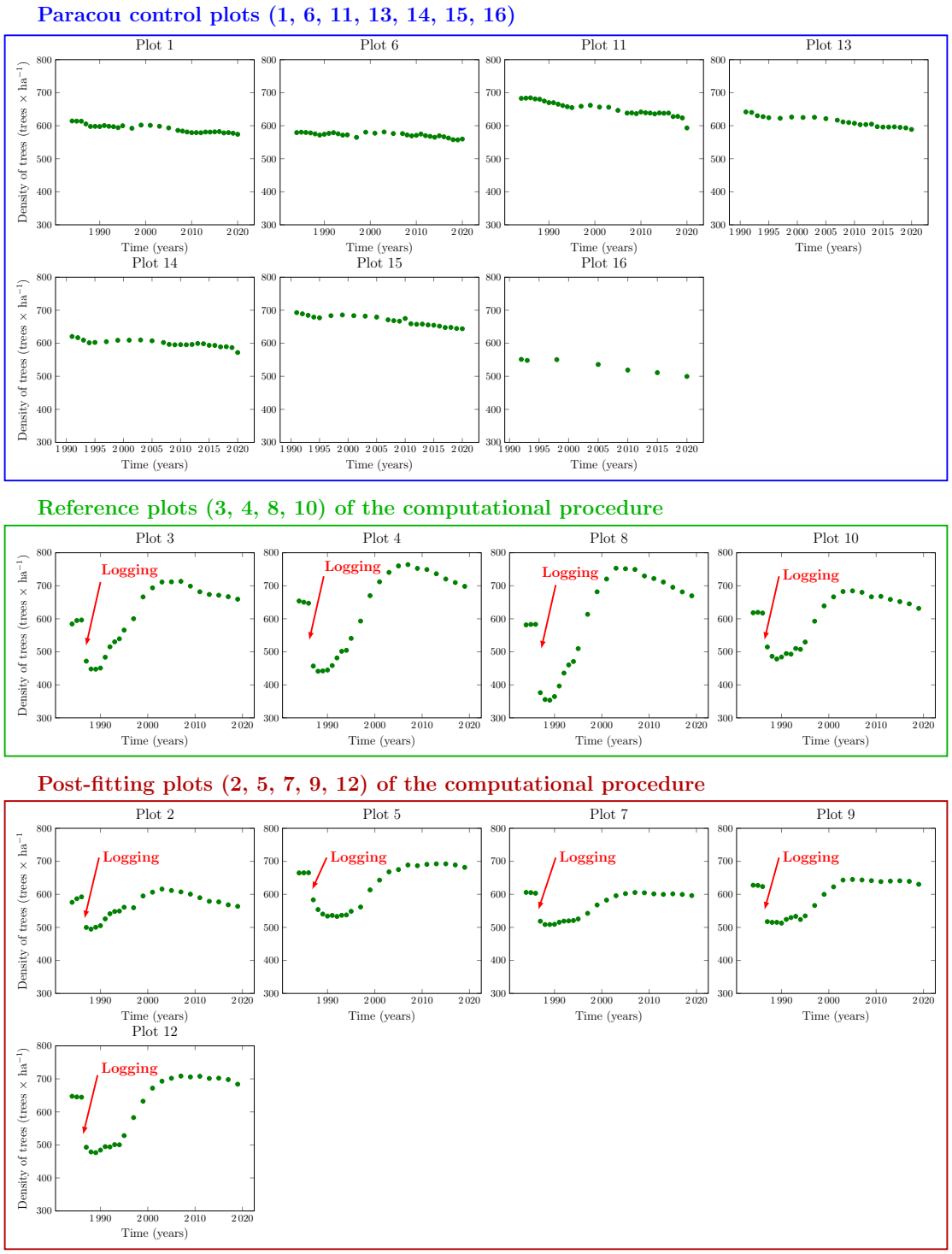


Figure 2: Forest data from Paracou research station. This station is divided into 16 forest plots, on which each of 70 000 trees has been mapped and measured at regular intervals since 1984. 9 of the 16 forest plots (plots 2, 3, 4, 5, 7, 8, 9, 10, 12) have been subject to partial logging in 1988, whereas other plots serve as *control* plots. Plots 3, 4, 8, 10, for which the intensity of logging is the highest, will be considered as *reference* plots for our computational procedure, whereas plots 2, 5, 7, 9, 12 will be treated *a posteriori*.

for which the intensity of logging is the highest, will be considered as *reference* plots for our computational procedure, whereas plots 2, 5, 7, 9, 12 will be treated *a posteriori*. The characteristic two-phase dynamical regrowth can be explained with a biological approach. Roughly speaking, after logging, a great number of young trees invade the free space which is produced by the extraction of timber trees; these young trees then naturally grow together, until a certain maximum number of trees is reached. Then the young trees are subject to competition for space and resources, which provokes an over-mortality which accounts for the slow decrease to equilibrium.

The Paracou forest database therefore represents a genuine opportunity to better understand the biological processes that partially govern the dynamics of tropical forests. In this work, we have extracted from this database the number of trees, for each plot of Paracou station and for each census year since 1984. We have analyzed these data with a new age-structured mathematical model, which integrates a nonlinear aging rate. In this age-structure model, we will distinguish *old* trees from *young* trees by their ability to produce seeds, that is, by their maturity stage. However, we emphasize that although the size of a tree can sometimes be considered as a proxy for its age (see for instance [38]), there is no relevant correlation between the size of a tree and its ability to produce seeds (see notably [45]).

2.2 A new mathematical model with nonlinear aging rate

In [30], a basic age-structured mathematical model was proposed for studying the dynamics of forest ecosystems. This model is given by the following system of two ordinary differential equations:

$$\begin{cases} \dfrac{du}{dt}(t) = \rho v(t) - \gamma\big(v(t)\big)u(t) - fu(t), \\ \dfrac{dv}{dt}(t) = fu(t) - hv(t), \end{cases} \tag{1}$$

with $t > 0$. Here, $u(t)$ and $v(t)$ denote respectively the densities of young trees and old trees in a given forest plot at time t. Following [3], a tree is defined as *old* if and only if it is able to produce seeds. This means that a tree is considered as *old* if it has reached a new maturity stage where it is fertile. Hence, the recruitment term of young trees in the first equation, given by $\rho v(t)$, is linear with respect to the density of old trees $v(t)$. Next, h stands for the death rate of old trees, and f for the aging rate of young trees. The mortality rate of young trees, given by

$$\gamma(v) = a(v-b)^2 + c, \tag{2}$$

with positive coefficients a, b, c, thus depends on the density of old trees, which models the competition between young trees and old trees, notably for water and light. Since $\gamma(v)$ is given by a quadratic expression with $a > 0$, it obviously admits a global minimum which models an optimal situation where the competition between young trees and old trees turns into cooperation. This mathematical model has been widely studied and has served as a basis for several other forest dynamics models [3, 31, 11, 10, 8, 9]. In particular, a spatial extension of the model has been introduced in [31], with an additional equation describing the evolution of seeds, produced by old trees and giving birth to young trees. It is known that the model given by (1) admits three parameters regimes, depending on the values of the recruitment rate of young trees ρ and the death rate of old trees h. The first parameter regime occurs when the death rate of old trees h dominates the recruitment rate of young trees ρ; it is characterized by the attraction of the orbits towards the extinction equilibrium and thus models a dying forest. The second parameter regime occurs when the death rate of old trees h and the birth rate of young trees ρ have the same order; in this case, a persistence equilibrium appears, coexists with the extinction equilibrium, and both equilibria are seen to be locally asymptotically stable. In this parameter regime, the future dynamics of the forest depend on the initial condition: if the initial condition is near the extinction equilibrium, then the forest vanishes; if at the opposite the initial condition is sufficiently far from the extinction equilibrium, then the forest reaches a good health equilibrium. Finally, the third parameter regime occurs when the death rate of old trees h is dominated by the recruitment rate of young trees ρ; it is characterized by a globally stable persistence equilibrium (see Figure 3(a)) and thus models a healthy forest.

Although the forest model (1) was calibrated and shown to fit with real-world boreal forest data [3], we have proved, not surprisingly, that it struggles to reproduce well-known post-exploitation oscillations that are observed in tropical forest plots. This limitation of system (1) has been highlighted using our Statistical

Model Checking Engine. Therefore, we have brought a significant modification to this model, by changing the aging term of young trees. In our new model, we have replaced the linear aging term $\pm fu$ of system (1) by a nonlinear aging term $f(u,v)$ given by

$$f\big(u,v\big) = \alpha uv\left(1 - \frac{u+v}{K_{\max}}\right). \tag{3}$$

The latter expression means that the aging rate of young trees $f(u,v)$ not only depends on the density of old trees v, but also on the total density of trees $u+v$ and on the maximal capacity of the forest plot, which we write $K_{\max}$. Hence, if the density of old trees is small, then the aging rate of young trees is also small. Although it may appear contradictory, it corresponds to the fact that a small density of old trees implies a large amount of available terrain for young trees; at first, a great number of young trees will try to invade this available terrain; but then growing trees will compete for space and a great number of growing trees will die, resulting in a small aging rate. At the same time, if the maximal capacity $K_{\max}$ of the forest plot is small, then, in a process which is similar to the well-known logistic model, the total population will be constrained, resulting a small aging rate. It is worth noting that the density-dependent expression (3) of the aging rate presents similarities with population dynamics models that take into account a strong Allee effect (see for instance [14]). This refinement of the model will allow to better reproduce the characteristic two-phase regrowth of the forest after logging activities.

With the nonlinear aging rate $f\big(u(t),v(t)\big)$ given by (3), our model can now be written as follows:

$$\begin{cases} \dfrac{du}{dt}(t) = \rho v(t) - \gamma\big(v(t)\big)u(t) - \alpha u(t)v(t)\left(1 - \dfrac{u(t)+v(t)}{K_{\max}}\right), \\ \dfrac{dv}{dt}(t) = \alpha u(t)v(t)\left(1 - \dfrac{u(t)+v(t)}{K_{\max}}\right) - hv(t), \end{cases} \tag{4}$$

with $t > 0$ (the parameters and unknowns of the new model are gathered in Table 1).

Table 1: Unknowns and parameters of the mathematical models (1) and (4).

Symbol	**Interpretation**	**Unit**
t	time	year
u	density of young trees	trees $\times$ha^{-1}
v	density of old trees	trees $\times$ha^{-1}
ρ	recruitment rate	year^{-1}
h	death rate of old trees	year^{-1}
a	first coefficient of age competition	trees$^{-2}\times$ ha$^2\times$ year^{-1}
b	second coefficient of age competition	trees$\times$ ha^{-1}
c	third coefficient of age competition	year^{-1}
f	aging rate (model (1))	year^{-1}
α	aging coefficient (model (4))	trees$^{-1}\times$ ha $\times$ year^{-1}
$K_{\max}$	maximal capacity of the forest plot (model (4))	trees $\times$ha^{-1}

As expected, the dynamics of the forest model (4) with nonlinear aging rate are richer than the dynamics of the initial model (1). Notably, the new model (4) admits a large parameter regime with a persistence equilibrium which is reached by the orbits after damped oscillations (see Figure 3 (b)). The dynamics of the new model (4) can be classified into five parameter regimes, depending on the number of zeros of the polynomial equation

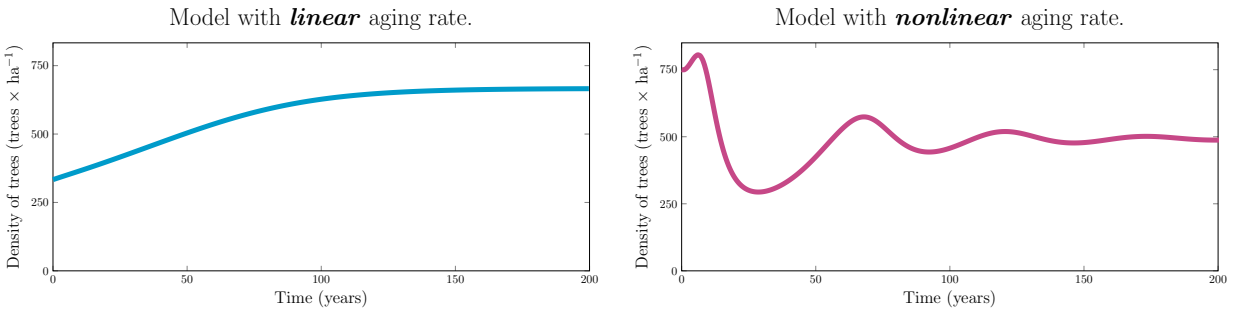
$$P(\bar{v}) = 0, \tag{5}$$

where

$$P(\bar{v}) = \alpha(\rho - h)\bar{v}\big[K_{\max}\gamma(\bar{v}) - (\rho - h)\bar{v} - \bar{v}\gamma(\bar{v})\big] - hK_{\max}\gamma(\bar{v})^2.$$

This polynomial equation is derived from the equations $\frac{du}{dt} = \frac{dv}{dt} = 0$ of the equilibrium points of system (4). Since $P(\bar{v})$ is a polynomial of degree 4, the number of zeros can vary from 0 to 4. As system (4) obviously admits the extinction equilibrium $(0, 0)$ as a uniform equilibrium (that is, for all parameter values), the number of equilibrium points of system (4) varies from 1 to 5.

When the polynomial equation (5) admits 4 solutions, system (4) admits 5 equilibrium points: the extinction equilibrium $(0, 0)$ (which is locally stable) and four persistence equilibrium points (one stable node, one stable focus and two saddles). Next, if (5) admits 3 solutions, then system (4) admits 4 equilibrium points: the extinction equilibrium $(0, 0)$ and three persistence equilibrium points (one stable focus, one saddle-node and one saddle). The saddle-node vanishes if (5) admits only 2 solutions; the stable focus vanishes if (5) admits only 1 solution. Finally, the extinction equilibrium is globally stable if (5) has no solution.



(a) **(b)**

Figure 3: Trajectories of the forest models (1) and (4). **(a)** Model (1) with linear aging: the trajectory is attracted to a persistence equilibrium with a monotone shape. **(b)** Model (4) with nonlinear aging: the trajectory also reaches a persistence equilibrium, after a transitional phase characterized by damped oscillations.

We emphasize that in the parameter regimes where the polynomial equation (5) admits at least 2 solutions, the forest model exhibits damped oscillations which are expected to reproduce the post-exploitation waves observed in deforested or disturbed plots.

For convenience, we have normalized the new model (4) by setting $x = \frac{u}{K_{\max}}$, $y = \frac{v}{K_{\max}}$, where $K_{\max}$ is given by the ratio

$$K_{\max} = \frac{N_{\max}}{S},$$

$N_{\max}$ being the maximal possible number of trees on a plot and S the surface of the corresponding plot. This normalization leads to the system

$$\begin{cases} \dot{x} = \rho y - \tilde{\gamma}(y)x - \tilde{\alpha}xy\big(1 - (x + y)\big), \\ \dot{y} = \tilde{\alpha}xy\big(1 - (x + y)\big) - hy, \end{cases} \tag{6}$$

with dimensionless unknowns x, y, where $\tilde{\gamma}$ is given by

$$\tilde{\gamma}(y) = \tilde{a}(y - \tilde{b})^2 + c. \tag{7}$$

In this normalization, the linear parameters ρ, h, c are unchanged, whereas the nonlinear parameters a, b, α are modified through the relations

$$\tilde{\alpha} = \alpha K_{\max}, \quad \tilde{a} = aK_{\max}^2, \quad \tilde{b} = \frac{b}{K_{\max}},$$

or equivalently

$$\alpha = \frac{\tilde{\alpha}}{K_{\max}}, \quad a = \frac{\tilde{a}}{K_{\max}^2}, \quad b = \tilde{b} \times K_{\max}. \tag{8}$$

In this work, we have validated the dimensionless model (6) with the Statistical Model Checking Engine (SMCE) that we present in the next subsection. In the calibration process of the new mathematical model

(4), we have treated $K_{\max}$ as a varying parameter with respect to each forest plot. Given the outputs of the SMCE, relation (8) allows to express the parameters values in the units of the new model (4). As an output, we have obtained parameters sets for which the trajectories of the model fit with the land use data of Paracou site (see Figures 6 and 8).

2.3 Statistical Model Checking Engine: computational procedure and formal guarantees

Statistical Model Checking (SMC) [33] is a formal technique whose aim is to estimate the probability that a given model satisfies a given property. This technique has been first developed in the context of computer science models, but has recently been used in the context of life sciences, with a special emphasis on models determined by Ordinary Differential Equations [29, 33, 52]. In essence, the model is equipped with a probability measure on its set of traces, and the aim of SMC is to estimate the measure of the subset of traces that satisfy the property of interest. In order to perform this estimation, SMC relies on extensive sampling of the model traces and statistic techniques (in our case the Monte-Carlo method) to compute an estimation of the measure of the subset of traces of interest with formal guarantees regarding the precision and error rate of this estimation.

In the present paper, we have developed a complete computational procedure, which we call the Statistical Model Checking Engine (SMCE), which uses SMC in order to estimate the values of the parameters

$$(\rho, \mu, \alpha, a, b, c) \in (\mathbb{R}^+)^6 \tag{9}$$

of the mathematical model (4), in order to fit the Paracou forest plots data presented in Figure 2. We focus on the forest plots which have been subject to logging and which exhibit a characteristic two-phase regrowth after logging (plots 2, 3, 4, 5, 7, 8, 9, 10, 12). Since the dynamics of the forest presents a strong spatial variability from one plot to another, we search for a small cell $C \subset (\mathbb{R}^+)^6$, which contains distinct parameters sets able to fit the data of each forest plot. In this parameter search, we have to fill a gap between the mathematical model (4) which admits an age structure and the forest data which does not distinguish trees of different ages. To that aim, we introduce an additional parameter ϕ_0, which determines the initial proportion of young trees in each forest plot.

Our computational procedure is divided into several steps, which are illustrated in Figure 4. During the exploration process, the adequacy of a fitting trajectory $\{U(\tau)\}_{t_1 \leq \tau \leq t_2}$ with respect to the corresponding forest plot data $\{D(\tau)\}_{t_1 \leq \tau \leq t_2}$ can be quantified in several ways. In this work, in order to characterize this adequacy, we have chosen to use the relative pseudo-distance Δ between the fitting trajectory and the corresponding forest plot data, defined as follows:

$$\Delta\big(U(\tau), D(\tau)\big) = \sum_{\tau=t_1}^{t_2} \frac{|U(\tau) - D(\tau)|}{|D(\tau)|}, \tag{10}$$

where τ varies over the finite set of temporal coordinates of the data.

We now detail the steps of our procedure, which is schematized in Figure 4.

- *First step.* The SMCE is initialized with three main inputs:
  - The equations of the mathematical model with non linear aging (4) (box ① in Figure 4).
  - A bounded parameter space $\mathscr{B} \subset \mathbb{R}^7$ which stores the domains of the 6 parameters $(\rho, \mu, \alpha, a, b, c)$ of the equations (4) as well as the domain of the additional parameter ϕ_0 which determines the initial proportion of young trees in each forest plot. This parameter space can be written as

    $$\mathscr{B} = [\rho^-, \rho^+] \times [\mu^-, \mu^+] \times [\alpha^-, \alpha^+] \times [a^-, a^+] \times [b^-, b^+] \times [c^-, c^+] \times [\phi_0^-, \phi_0^+], \tag{11}$$

    with lower and upper bounds for each parameter that can be chosen arbitrarily. It is then discretized into a finite number of cells, whose common size can also be chosen arbitrarily. For simplicity, we schematize the seven-dimensional parameter space as a standard three-dimensional space (box ② in Figure 4).

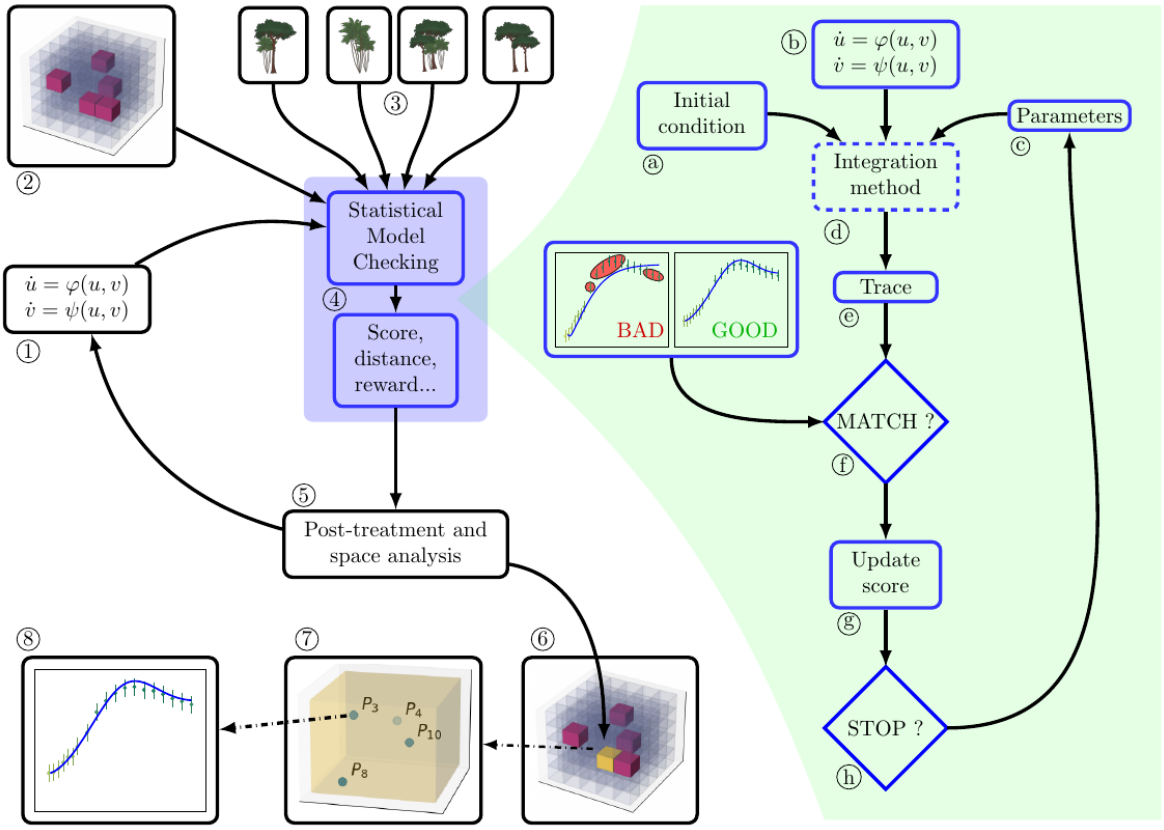


Figure 4: Overview of the SMCE. On the left, the global cartography of the process. On the right, a more precise view on the per-cell process.

  - The post-logging data of 4 forest plots (plots 3, 4, 8, 10 as presented in Figure 2) that we have chosen as reference plots (box ③ in Figure 4). Note that other forest plots will be treated *a posteriori*.

- *Second step.* With its three inputs, the SMCE performs a thorough exploration of the parameter space (see box ④ in Figure 4, whose sub-steps are detailed in the green panel on the right of Figure 4). The following procedure is performed for each reference plot p, regardless of their order.
  - For each cell C in $\mathscr{B} \subset \mathbb{R}^7$, we generate a sample $\mathscr{S}_p(C)$ of N vectors of parameter values:
$$\mathscr{S}_p(C) = \left\{P_i^p(C)\right\}_{1 \leq i \leq N}. \tag{12}$$
    The value of N depends on the chosen precision of the Monte-Carlo procedure (see Theorem 2 in [29]). In this case, we set $N = 100$.
  - Each vector $P_i^p(C) = (\rho_i, \mu_i, \alpha_i, a_i, b_i, c_i, \phi_{0i})$ $(1 \leq i \leq N)$ allows to determine an initial condition (u_0, v_0) as well as the parameter values of the differential equations $(\dot{u}, \dot{v})$ of the input model (boxes ⓐ, ⓑ, ⓒ).
  - The numerical integration of the ordinary differential equations is performed (by using the well-known Runge-Kutta method) for each vector $P_i^p(C)$ of parameter values (box ⓓ), which produces a trace $\{U_i^p(\tau)\}_{t_1 \leq \tau \leq t_2}$ of the input model (box ⓔ).
  - The trace $\{U_i^p(\tau)\}_{t_1 \leq \tau \leq t_2}$ is compared with the input data $\{D^p(\tau)\}_{t_1 \leq \tau \leq t_2}$. In a first step, we compute the number of outliers, i.e. time points where $\frac{|U_i^p(\tau) - D^p(\tau)|}{|D^p(\tau)|} > 0.075$ for the current trace. The trace is selected if there are less than 3 outliers and rejected otherwise (box ⓕ). When the trace is selected, its score is $\Delta(U(\tau), D(\tau))$.

- The score of the cell C from which the sample of vectors $\mathscr{S}_p(C) = \{P_i^p(C)\}_{1\leq i\leq N}$ has been extracted is updated (box ⓖ) as the average of the scores of the traces processed so far.
- Once the N vectors $\{P_i^p(C)\}_{1\leq i\leq N}$ have been treated, the process stops and returns a score S_C^p for each cell C of the parameter space (box ⓗ).

- *Third step.* After the exploration of the parameter space $\mathscr{B}$, for all reference plots, each cell C is given a score S_C^p for each reference plot. Given the finite number of cells which compose the parameter space $\mathscr{B}$, a maximal score $S_{\max}^p$ can easily be determined for each reference plot. At this stage, it is often the case that a high number of cells reach this maximal score or a score which is near this maximum (the correlogram depicted in Figure 5 shows the distribution of the corresponding parameters). Furthermore, the cells reaching the maximal score are not necessarily the same for the four reference plots which we aim to fit. For that reason, a post-treatment is performed (box ⑤) on the set of cells whose score is close to the maximum, in order to select a unique "best" cell, using the following criteria:
  - We start by filtering out the cells that are not close to the maximum for at least one reference plot.
  - We then reject the cells whose values are irrelevant compared to ecological indicators, when such indicators are available.[1]
  - We also reject the cells for which the model trajectories do not present the characteristic two-phase regrowth of the forest[2].

  The application of these criteria allows to reduce the number of candidate "best" cells to a reasonable level. Finally, we arbitrarily choose one of them (box ⑥).

- *Fourth step.* The cell C which is returned by the post-treatment procedure therefore contains four vectors of parameter values P^3, P^4, P^8 and P^{10}, which respectively fit the data of the reference plots 3, 4, 8, 10 (box ⑦). The numerical values of the parameters in these vectors are given in Table 4. Once the vectors P^3, P^4, P^8 and P^{10} are computed, the trajectories of the model are easily reproduced and visualized (box ⑧).

Our computational procedure has been implemented in the `C++` language, and our experiments have been run in a GNU/LINUX environment. The complete code provides the details on the choices of

- the time discretization necessary for the numerical integration of the ordinary differential equations,
- the lower and upper bounds of the parameter space (11),
- the size of the cells which partition the parameter space, and
- the precision of the Monte-Carlo method used in the SMC procedure.

Note that the differential equations of the mathematical model (4) are defined for parameters whose domains can be much larger than their biological counterparts. Assigning bounds to the parameters can be straightforward (e.g. negative birth rate and aging rate can easily be ruled out) but in many cases it is harder to find relevant and precise bounds in the literature. Furthermore, even when such biological bounds can be found, we have to expand them by a margin large enough to better observe the robustness of the model. Observing assertive results outside of biological range is not unexpected, but each case must be assessed by a domain expert:

- if the region found is far away from the biological range, it can be safely dismissed as a mathematical artifact and the parameter space can be reduced accordingly for future experiments;

[1] Although this criterion could be integrated earlier in the process, e.g. when choosing the lower and upper bounds of the parameter domains (11), we prefer to apply it at the post-treatment step in order to better test the ability of the model to reject parameter values of ecological irrelevance.

[2] Given the nature of this last selection criterion, this step is for now done by hand and therefore not included in the code we provide.

- if the region is contiguous to the expected one but contains significantly better results, either they can be validated by the domain expert or a weakness of the model or of the computational procedure is identified.

Note also that we encountered both situations in our preliminary experiments and had to correct the scaling factor and the mortality range, initially based on an overly strict interpretation of studies which were not directly transferable to the specifics of Paracou.

We emphasize that numerous mathematical methods for estimating the parameters values of a given model with respect to real-world data are well-known. However, *identifiability* of the parameters is often a necessary condition to be fulfilled for the application of these methods (see for instance [16] and the references therein). Here, we emphasize that the age structure of the mathematical model (4) is not comptabible with the non-aged forest data. Hence, it is vain to expect that the identifiability property will be satisfied. Therefore, our computational procedure, for the application of which identifiability is *not* a necessary condition, proves to be very efficient. Furthermore, our computational procedure succeeds in fitting simultaneously forest data of several plots which, although being geographically neighbors, exhibit a high level of biological variability.

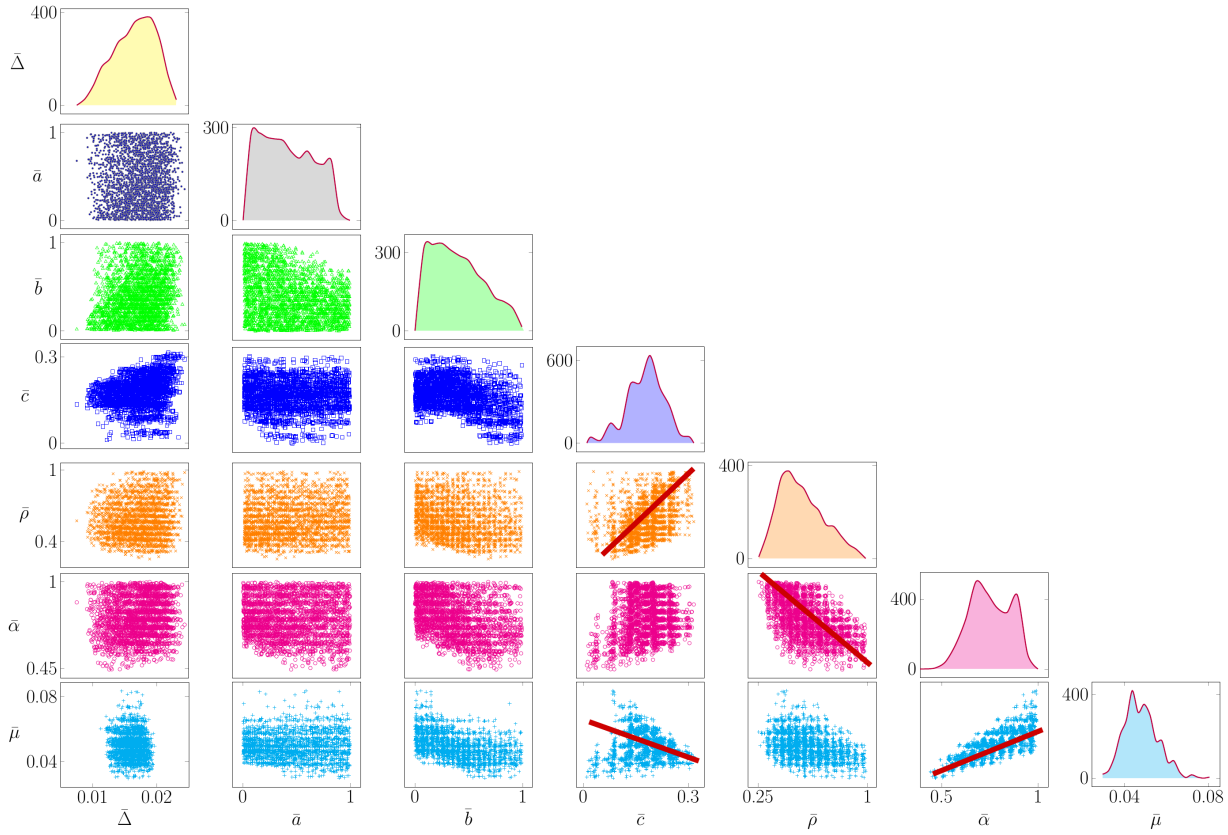


Figure 5: Correlogram illustrating the distributions of the averages of the parameters $(\bar{a}, \bar{b}, \bar{c}, \bar{\rho}, \bar{\mu}, \bar{\alpha})$ of the mathematical model (4) and of the average distance $\bar{\Delta}$ between the trajectory of the model and the forest data, over a filtered sample of simulations of the model. On the diagonal, the distribution of each average allows to identify the regions of the parameter space $\mathscr{B}$ where the computational procedure finds a great number of fitting parameters, and to compare these regions with physical measures or ecological indicators of the forest dynamics. Under the diagonal, the clouds of points show that the parameters of the model are not evidently correlated. Although a couple of clouds exhibit a rough linear shape (e.g. for the pair $(\bar{\alpha}, \bar{\mu})$ on the right bottom), the abscence of correlation proves that the model does not admit too much degrees of freedom (otherwise, the number of parameters could be reduced).

3 Results of the computational procedure

3.1 Validation of the new mathematical model for better understanding the forest regrowth

For better understanding the characteristic two-phase dynamical regrowth of Paracou forest patches, we have chosen a mathematical modeling explanatory approach. Hence, we have considered the new parametric model (4), determined by a system of ordinary differential equations. Recall that this model admits six main parameters: recruitment rate, mortality rate, aging rate, and three competition coefficients (see Equation (4) and Table 1). The new mathematical model (4), which integrates a nonlinear aging term, has been designed by analogy with density-dependent models, which are common in population dynamics. This nonlinear aging term admits rich dynamics which completely differ from the dynamics of model (1). In particular, it admits a large parameter regime with damped oscillations, which are expected to reproduce the two-phase dynamical regrowth of the forest plots which have been subject to selective logging (see Figure 3). Here, our aim is to calibrate the model with respect to the data of Paracou station, that is, to find relevant sets of parameters for which the corresponding trajectories fit the data with an arbitrary precision. This fine calibration of the mathematical model (4) by comparison with Paracou data is the core of our work. It represents the first step in the validation process of the modelling approach and offers two perspectives: the first perspective corresponds to the prediction potential of the calibrated model, which allows to test a high number of relevant scenarios, as will be presented in the Discussion Section; the second perspective is the opportunity for a better understanding of the tropical forest dynamics. If our calibration procedure is mainly numerical, it is also guided by qualitative criteria, the first of these criteria being the characteristic two-phase dynamical regrowth explained before.

Since control plots present an almost constant equilibrium, it is natural to focus on forest plots which have been subject to logging, thus present a dynamical regrowth. For that reason, we have performed the calibration of the model by selecting 4 forest plots with intense logging, namely plots 3, 4, 8, 10, which we call *reference plots*. We will then test *a posteriori* the ability of the model to fit the data of other plots, subject to logging of less intensity.

The results of our computational procedure on the 4 reference plots 3, 4, 8, 10 are presented in Figure 6. We observe that the corresponding trajectories fit the data with a relatively high precision level, which can be measured by the relative distance Δ between the trajectory and the data, given in Table 2 (see Equation (10)).

Table 2: Relative distances between the fitting trajectories and the corresponding data for each reference plot of the computational procedure.

Reference plot	**Relative distance Δ**
3	0.183
4	0.137
8	0.265
10	0.137

Although the level of precision can be measured with several numerical indicators, the fact that only 3 *outliers* (that is, some points which stand outside a tunnel of uncertainty around the data) can be enumerated on a total of 76 points convinces on the quantitative ability of the trajectories to fit the data. In addition, we observe that the two-phase dynamical regrowth and its characteristic over-maximum, which are seen as a qualitative property, are well reproduced by the trajectories. Next, the numerical values of the parameters are given in Table 4 below. We emphasize that the four parameters sets belong to a common discretization cell of the whole parameter space of the model (numbered 47 921 440 in the discretization sequence), hence are close to each other, without however being strictly identical. This common cell, which contains the four parameters sets, thus models both the regrowth dynamics of the whole Paracou tropical forest, and also, through the small distances which separate each parameter set, the intrinsic variability of the forest ecosystem. It can be easily verified that the best parameters presented in Table 4 lead to the dynamics of Figure 3(b), which is characterized by the long time convergence of the trajectories towards a persistence

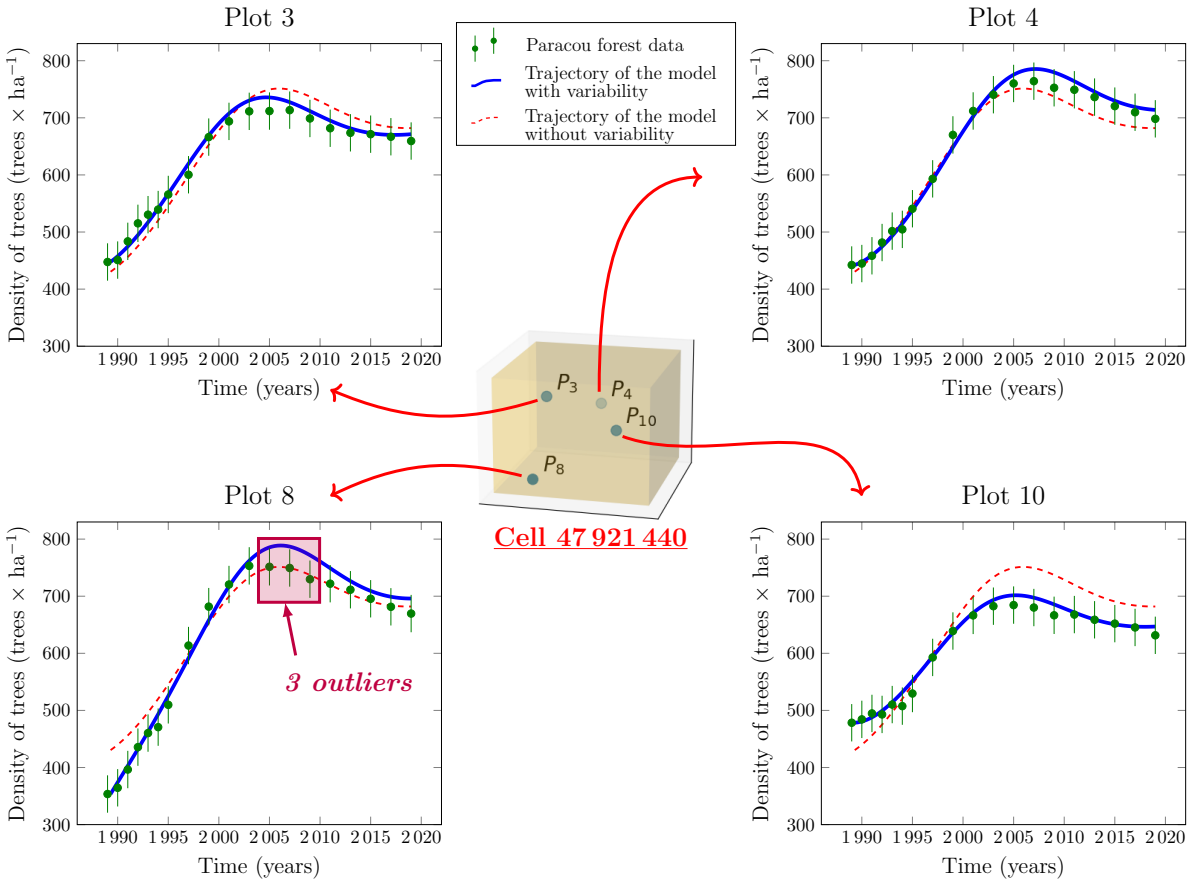


Figure 6: The new mathematical model (4) with nonlinear aging and variability (depicted in solid blue lines) fits the data of reference plots 3, 4, 8, 10 of Paracou station with a high level precision: only 3 outliers can be enumerated on a total of 76 points (on plot 8), and the two-phase dynamical regrowth is well reproduced by the trajectories. Cell 47921440, which contains the four parameters sets, models both the regrowth dynamics of the whole Paracou tropical forest, and also, through the small distances which separate each parameter set, the intrinsic variability of the forest ecosystem. In comparison, the model *without* variability (depicted in dashed red lines) shows a limitation to fit simultaneously the data of the four reference plots, especially on plots 8 and 10.

equilibrium after damped oscillations.

Besides the new mathematical model (4) with *non linear* aging, it is worth noting that the mathematical model with *linear* aging, given by system (1), shows an unsolvable limitation to fit the data of forest plots which have been subject to partial logging. Indeed, when applying our computational procedure to system (1), we obtain trajectories which present two important lacks. The first lack is quantitative: several points of the trajectories remain outside an uncertainty tunnel around the data. The second lack is qualitative: the trajectories admit an increasing shape which fails to reproduce the over-maximum of the number of trees which is characteristic of the forest regrowth after logging. Here, it is interesting to note that the mathematical model (1) was however proved to well reproduce the dynamics of boreal forests (see [3]), which suggests that tropical forests admit characteristic dynamics.

3.2 Post-fitting of non-reference plots

The second step of our computational procedure aims to identify parameters of the mathematical model (4) for fitting the data of non-reference plots 2, 5, 7, 9, 12. Here, rather than exploring again the whole parameter space, we search for parameter values in a neighborhood of the average of the parameters sets obtained for the reference plots 3, 4, 8, 10. In other words, we aim to fit the non-reference plots by exploring only cell 47 921 440, previously identified for fitting the reference plots. The results of this post-fitting procedure are presented in Figure 7, and the corresponding relative distances are gathered in Table 3. We observe again that the resulting trajectories fit the data with a high level of precision, while reproducing the characteristic two-phase regrowth of the forest and its over-maximum.

Table 3: Relative distances between the fitting trajectories and the corresponding data for each non-reference plot of the post-fitting procedure.

Non-reference plot	**Relative distance Δ**
2	0.162
5	0.120
7	0.054
9	0.136
12	0.154

The complete numerical values of the parameters for the non-reference plots 2, 5, 7, 9, 12 are given in Table 5. Overall, cell 47 921 440 covers 9 distinct sets of parameters, which are close to each other while being distinct. These 9 distinct sets of parameters succeed in fitting the data of the forest plots of high biological variability, which have been subject to intense or moderate logging.

Table 4: Best sets of parameters obtained by applying our computational procedure on the new mathematical model (4) for the reference plots 3, 4, 8, 10. The four parameters sets belong to a common discretization cell of the whole parameter space of the model, hence are close to each other, without however being strictly identical.

Parameter	**Plot 3**	**Plot 4**	**Plot 8**	**Plot 10**
ρ (recruitment rate)	0.446	0.436	0.421	0.439
μ (mortality rate)	0.0283	0.0284	0.0296	0.0276
α (aging coefficient)	$9,08 \times 10^{-7}$	$7,92 \times 10^{-7}$	$9,08 \times 10^{-7}$	$9,8 \times 10^{-7}$
a (first competition coefficient)	$0,000630667$	$0,00059125$	$0,000605063$	$0,0006625$
b (second competition coefficient)	498	$524,8$	$522,98$	$480,96$
c (third competition coefficient)	0.0896	0.089	0.0714	0.0854
$K_{\max}$ (maximal capacity)	750	800	790	720
ϕ_0 (initial age proportion)	0.644	0.657	0.547	0.685

3.3 Using the predictive potential of the calibrated model for simulating logging scenarios

Stem logging practices are strictly regulated in the Amazon forest, where a delay of 35 years should be observed between two logging phases [50] (this delay can however be less than 35 years in some countries [49]). In French Guiana, this delay is even longer (65 years), and complemented by a limited timber volume of 20 m^3 per ha, which generates a more important volume of collateral damages of these mechanized practices [48, 17, 49]. In this section, we propose simulations of various logging scenarios in order to test the resilience of the forest ecosystem with respect to a perturbation of its biological dynamics[3]. To that aim, we use the

[3]Along with this paper, we have developed a basic prototype of a user-friendly hub for simulating logging scenarios. This web hub is freely accessible at https://velo.pythonanywhere.com.

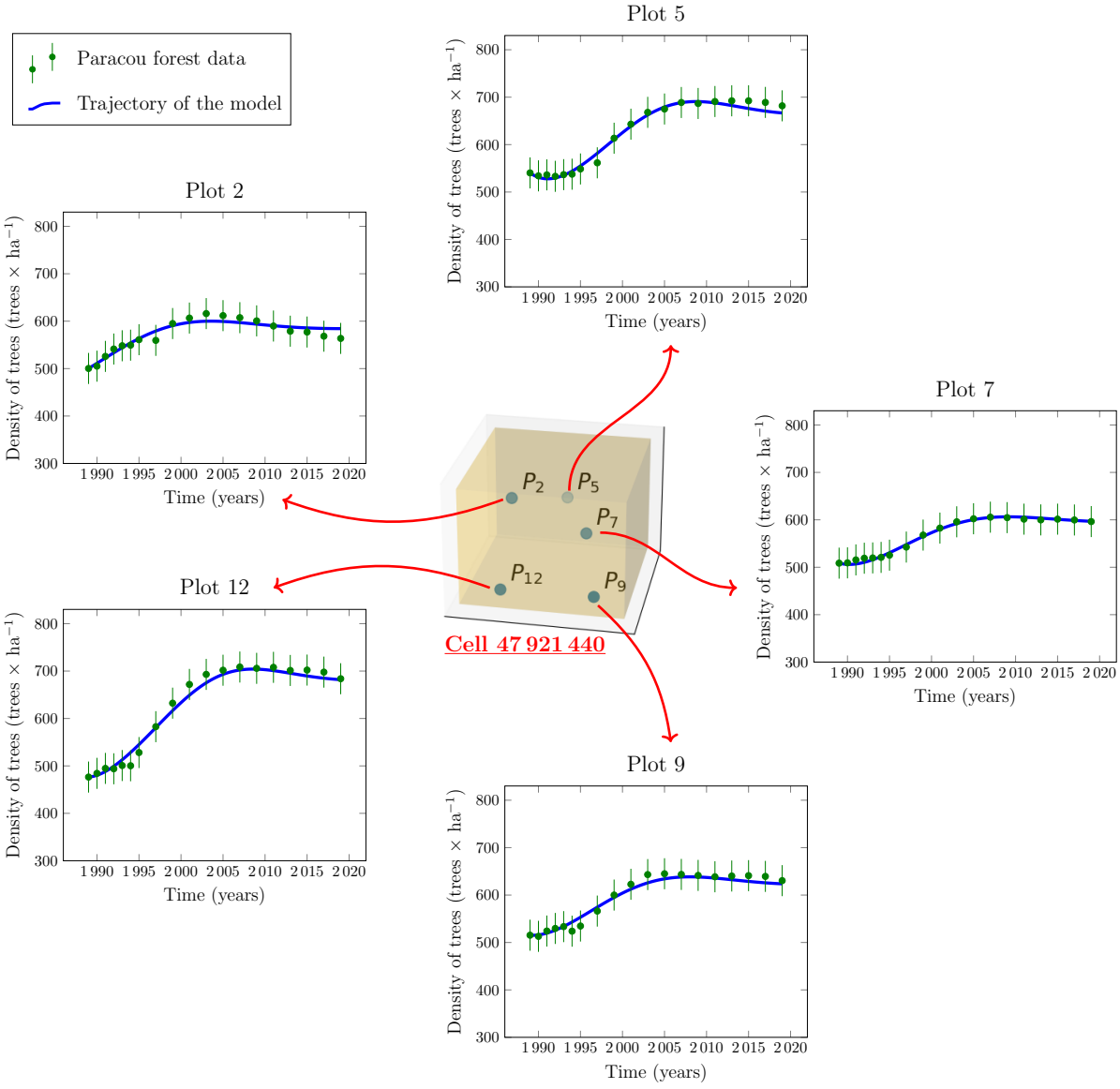


Figure 7: Results of the post-fitting procedure on the non-reference forest plots 2, 5, 7, 9, 12. The parameters are extracted in a neighborhood of cell 47 921 440, which was previously identified for reference plots. The corresponding trajectories fit the data with a high level of precision and reproduce the characteristic two-phase regrowth of the forest.

mathematical model (4) in its calibrated form, with the main parameter values returned by our computational procedure.

So as to experiment the spatial variability of the forest ecosystem, we consider the parameter values of the Paracou forest plots 2, 3, 4, 5, 7, 8, 9, 10, 12 which we have calibrated. Hence, we produce for each scenario 9 distinct temporal trajectories, one for each of those plots. Moreover, we vary the intensity of the perturbation induced by logging (which we denote by I) from 5% to 10%, as well as the time delay between two logging phases (which we denote by ΔT) from 20 years to 80 years. These ranges obviously cover the values recommended by the current legislation. For instance, a 7% intensity of logging represents

Table 5: Numerical values of the parameters of system (4), obtained by applying the post-fitting procedure on plots 2, 3, 5, 7, 12.

Parameter	**Plot 2**	**Plot 5**	**Plot 7**	**Plot 9**	**Plot 12**
ρ	0.36297	0.48267916	0.394853	0.394853	0.5226
μ	0.0313225	0.0313225	0.0313225	0.0313225	0.0313225
α	$7,85 \times 10^{-7}$	$7,35 \times 10^{-7}$	$8,92 \times 10^{-7}$	$8,19 \times 10^{-6}$	$6,98 \times 10^{-6}$
a	$0,000838676$	$0,000750395$	$0,000776012$	$0,000792083$	$0,000731154$
b	$540,6$	$558,0458308$	$548,55$	$572,4$	$620,022$
c	0.076024	0.069875	0.094471	0.0698749	0.07756125
$K_{\max}$	680	760	690	720	780
ϕ_0	0.574147	0.70185208	0.620585	0.620585	0.64380416

approximately 20 m^3 par ha[4]. Furthermore, we experiment both an optimistic future for which the mortality rate of trees remains constant, and a pessimistic future where it increases, according to recent observations and prospective assessments. In particular, since the mortality rate of trees in the Amazon forest was estimated in [40] to have increased by more than 20% in two decades, we have chosen to test a 20% increase of the mortality rate.

Our results are presented in Figure 8. Overall, we observe that the forest ecosystem admits an apparent level of resilience with respect to logging practices. Indeed, in the constant mortality scenario (blue lines in Figure 8 (a)), the densities of trees on each Paracou forest plot resist to the reduction of the number of trees which occurs at each stem logging (visible on each graphic by a discontinuous jump of the trajectories). If the intensity I increases, then the characteristic two-phase regrowth of the forest is accentuated. If the time delay ΔT between two logging phases increases, then the forest can recover its uniform equilibrium before the next logging step. However, this level of resilience of the forest ecosystem is here simulated on the basis of an explanatory modeling approach, which only measures the density of trees, thus does not take into account other physical or biological indicators, such as the level of carbon sequestration or of biodiversity nor the commercial volume [28, 48]. Hence, the resilience is only partial and it must not be concluded that this partial resilience of the forest ecosystem offers an opportunity for an intense stem logging. Furthermore, in the increasing mortality scenario (orange lines), we observe, according to intuition, that a high logging intensity and a short delay between two logging phases can lead to the decline of several forest plots. Indeed, for $I = 35\%$ and $\Delta T = 20$ years, 3 forest plots are shown to wither off, whereas other forest plots, although they admit a neighbor dynamic, resist to the logging practices (see Figure 8 (b)).

To complete these numerical simulations and provide an enlarged vision of the resilience of the forest, we have experimented wider ranges of the logging intensity I (from 0 to 100%) and of the time delay ΔT between two logging phases (from 10 years to 90 years). The results are illustrated in Figure 8 (c), where we depict a 3D histogram with the number of dying plots (vertical axis) with respect to I and ΔT. Here, we call *dying plot* a forest plot whose density of trees decreases to the extinction equilibrium. Between two obvious extreme zones (0 dying plot for small I and large ΔT, 9 dying plots over 9 plots for $I = 100\%$), we observe an abrupt tipping zone which shows that the resilience of the forest ecosystem is limited and can rapidly be destabilized.

Finally, since our computational procedure integrates a post-treatment step, where a parameter cell is arbitrarily chosen in the parameter subspace of candidates likely to be selected as best parameters, it is natural to experiment how the model behaves with respect to this arbitrary choice. Hence, we have simulated several post-treatment choices by selecting 10 distinct parameters cells (depicted in yellow in Figure 8 (d)). Figure 8 (d) shows the distributions of the final densities of trees obtained with parameters taken from these different cells, for plots $2, 5, 7, 9, 12$, using the increasing mortality scenario with $I = 35\%$ and $\Delta = 20$ years. From this figure, one can observe that some plots (here, plots $2, 7$) exhibit a high sensitivity to the choice of the parameter cell, while others (here, plots $5, 9, 12$) are more robust with respect to this choice.

[4]The timber volume can be roughly estimated as a number of trees by using the logging simulator LoggingLab: https://vincyanebadouard.github.io/LoggingLab/.

4 Discussion

4.1 Comparison of the best parameters with ecological indicators

Due to the ecological and biological complexity of tropical forests ecosystems, systematic physical measures of all parameters and of their dynamics are difficult to produce at a satisfactory level. However, some of these parameters can be evaluated, notably the mortality rate. In [54], the mortality rate of trees in tropical forests has been estimated between 0.5% and 2.5%. In [23], the mortality rate is shown to vary up to 3.5%, especially in a post-logging context. Recent studies even reveal an increasing tendency of the mortality rate due to climate change [35, 36, 1, 40]; this tendency has been experimented in Section 3.3, using the simulation ability of our model.

The mortality rates which are returned by our computational procedure (which are between 2.76% and 3.13%, as shown in Tables 4 and 5) seem to slightly overestimate the values obtained by [54], while remaining under the maximal value proposed in [23] and are in agreement with a prospected increase. To the best of our knowledge, sharp physical measures of other parameters of the model (recruitment rate, aging and competition coefficients) are not discussed in the literature. However, rough estimations are known, for instance on the recruitment rate, and can be used to exclude irrelevant values.

Finally, it is interesting to note that the aging and competition parameters values which are returned by our method admit relatively small values. Obviously, these small values should be appreciated in the regard of the degree of nonlinearity which they correspond to, and it should not be concluded that these small terms should be neglected. In particular, the aging coefficient α is estimated of order 10^{-6} or 10^{-7} (see Tables 4 and 5), which is commensurate with the cubical term (3) that it is applied to.

4.2 On the biological variability of the forest plots

The approach developed in the present paper focuses on the possibility, with a single model, to fit observation data of a tropical forest ecosystem which is characterized by a high level of biological variability. This variability is the consequence of several factors, including the heterogeneity of the study site, in terms of ecological and geographical indicators. For instance, the proportion of different topographic positions (bottomland and plateau) varies among the plots, which can partially account for distinct regrowth behaviors of the forest after logging. Hence, the biological variability of the tropical forest ecosystem implies that the data of the forest plots regrowth determine a *range* of trajectories. The width of this range increases with the level of variability. Next, if the data of a given forest plot belong to the center of the range, then the model without variability can be easily calibrated to fit well the data. But if the data deviate away from the center and approach the edge of the range, then the model without variability shows an incapacity to fit the data. This is well illustrated in Figure 6, where the model without variability (depicted in red dashed lines) fits well the data on plots 3 and 4, but is unable to describe the dynamics of plots 8 and 10.

Therefore, in the situation of a high level of biological variability, it is imperative to relax the parameters of the model in such a manner that every area in the wide range of trajectories can be reached by the model. We emphasize that this delicate issue has been successfully overcome by our computational approach, by establishing a correspondence between the range of trajectories and the cells of the parameter space. Overall, the size of the cells that should be extracted from the parameter space increases with the level of biological variability. Finally, we highlight that our approach is easily transposable to other types of ecosystems which exhibit heterogeneity and variability, and not only tropical forests.

4.3 On the age and single-species structure of the model

Although our computational approach leads to a remarkable fit of observation data, on several forest plots exhibiting a strong spatial variability, we emphasize that it provides only partial answers to the general question of better understanding the dynamics of tropical forest ecosystems, especially in a post-logging context.

Indeed, it would be of great interest to enrich the observation data, so far produced by a human recorded protocol of ground measurements, by maturity studies of the tropical trees, so as to better correspond to the age-structure of the model [38], [45]. However, due to practical reasons, such studies are difficult to conduct, since it would ideally require to access the canopy on areas of significant surface. Therefore, the

computational procedure must be ran in such a manner that it can cover a wide range of phenomenological behaviors. This in a way can be viewed as a strength of our statistical model checking framework, in the sense that it overcomes a lack of ecological information, by performing a deep exploration of the parameter set.

Next, we have presented in Section 3.3 a set of numerical simulations of forest dynamics after logging, performed with our calibrated model. As depicted in Figure 8, these numerical simulations can exhibit oscillations in response to a periodic logging activity. However, it is now well described that the forest dynamics after logging exhibit complex behaviors, characterized by a first stage with the invasion of pioneer species, followed by a complex succession of various trees species (see for instance [2], [28] or [53]). Yet, the observation data produced in Paracou station, as presented in Section 2.1, include a botanical description of the trees. Hence, it would be highly beneficial to design a multi-species improvement of the mathematical model (4), in order to reproduce the complex regrowth dynamics of the forest after logging, especially regarding the important number of species registered on the Paracou site (which is approximately equal to 700, see [42]). This improvement could certainly be realized by identifying a small number of functional groups. Otherwise, the dimension of the model (4) could grow over a reasonable level. Another possible approach is to consider the functional traits of the tree species as continuous variables (see for instance [4], where the tree vulnerability to climate change is assessed using continuous traits as indicators of sensitivity). On this point, the explanatory approach of mathematical models determined by differential equations could be coupled with machine learning and feature engineering techniques. This is precisely studied in an on-going research work, that will be published separately in a forthcoming paper.

5 Conclusion and future work

In this paper, we have proposed an original computational technique to calibrate a new mathematical model of tropical forest dynamics. The new mathematical model is determined by an age-structured system of parametric differential equations and integrates a nonlinear aging term which is expected to reproduce the forest regrowth dynamics in a post-logging context. Stemming from the observation that tropical forests exhibit a high level of spatial and biological variability, we have designed our fitting procedure in a sufficiently flexible manner, so that our model can be calibrated on land-use data of several forest plots. Our procedure thus thoroughly explores a huge parameter space and returns a small cell of parameters, whose close vicinity contains parameter values individually fitting the data of distinct forest patches. Furthermore, we manage to overcome a gap between the parameters of our age-structured model and the available land-use data, which do not distinguish trees of different ages. Once calibrated, the forest dynamics model can be used to experiment relevant logging scenarios, in regard of crucial environmental challenges such as carbon sequestration and ecosystem services.

We emphasize that our parametrization technique is generic enough to be readily applied to many other scientific domains where heterogeneity and variability are crucial. Overall, the main features of our computational process are the following:

- the calibration procedure can explore a huge parameter space and searches for a parameter *cell* rather than a parameter *value*;
- it can be ran on any type of parametric model, provided numerical simulations of the input model can be performed;
- it has the ability to validate or to exclude a given model;
- it is based on a statistical verification algorithm that provides sharp guarantees on the output of the method;
- its implementation is automatized and terminates correctly even on data which are not fully compatible with the input model.

It is worth noting that the ability of the procedure to exclude an input model, which has been tested in the present work to reject a model with a linear aging term, helps the modeler in its difficult task to

better understand the dynamics of a given real-world system. Therefore, we believe that our computational procedure represents a genuine advance on the calibration process of ecological or biological models.

In a near future, we aim to explore two research perspectives. The first perspective is to improve our modeling approach on the dynamics of tropical forest ecosystems. Indeed, tropical forest ecosystems are obviously characterized by complex interactions with external constituents, notably the water resource, climate change and anthropic activity. Much effort has been accomplished, at a somewhat abstract level, in order to better understand how to take into account these interactions in an integrative model. For instance, the interaction between the forest and the water resource has been modeled using complex networks of differential equations in [11], or using a reaction-diffusion-advection system in [10]. Therefore, our plan is to make converge an abstract and explanatory process, that allows to integrate the complex mechanisms of the forest ecosystem, with a systematically calibrated approach that guarantees a relevant fitting of each constituent of the forest model with observation data. Depending on the nature of the observation data, different types of sub-models may be considered. This modeling effort should result in a meta-model able to better explain the dynamics of tropical forest ecosystems, and to simulate these dynamics at a fine numerical level. The second perspective concerns the development of a simulation hub for experimenting tropical forest dynamics. Based of a basic prototype, which is already accessible at https://velo.pythonanywhere.com/, the simulation hub will allow a large audience to better understand forest dynamics. It will also allow institutional actors and local stakeholders, who permanently solicit the scientific community, to elaborate adaptation and mitigation strategies to face global warming.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Land use data were extracted from the Paracou Station database, for which access is modulated by the scientific director of the station (https://dataverse.cirad.fr/dataverse/guyaforcirad).

Code availability

The code of our computational procedure is freely accessible at

https://gitlab.univ-nantes.fr/velo_check_forest_model/paracou.

References

[1] I. Aleixo, D. Norris, L. Hemerik, A. Barbosa, E. Prata, F. Costa, and L. Poorter. Amazonian rainforest tree mortality driven by climate and functional traits. *Nature Climate Change*, 9(5):384–388, 2019.

[2] M. R. Amaral, A. J. Lima, F. G. Higuchi, J. Dos Santos, and N. Higuchi. Dynamics of tropical forest twenty-five years after experimental logging in Central Amazon mature forest. *Forests*, 10(2):89, 2019.

[3] M. Y. Antonovsky, R. Fleming, Y. A. Kuznetsov, and W. Clark. Forest-pest interaction dynamics: the simplest mathematical models. *Theoretical Population Biology*, 37(2):343–367, 1990.

[4] I. Aubin, L. Boisvert-Marsh, H. Kebli, D. McKenney, J. Pedlar, K. Lawrence, E. Hogg, Y. Boulanger, S. Gauthier, and C. Ste-Marie. Tree vulnerability to climate change: Improving exposure-based assessments using traits as indicators of sensitivity. *Ecosphere*, 9(2):e02108, 2018.

[5] M. Aubry-Kientz, V. Rossi, G. Cornu, F. Wagner, and B. Hérault. Temperature rising would slow down tropical forest dynamic in the guiana shield. *Sci Rep*, 9:10235, 2019.

[6] B. Brede, L. Terryn, N. Barbier, H. M. Bartholomeus, R. Bartolo, K. Calders, G. Derroire, S. M. K. Moorthy, A. Lau, S. R. Levick, et al. Non-destructive estimation of individual tree biomass: Allometric models, terrestrial and UAV laser scanning. *Remote Sensing of Environment*, 280:113180, 2022.

[7] H. Bugmann and R. Seidl. The evolution, complexity and diversity of models of long-term forest dynamics. *Journal of Ecology*, 110(10):2288–2307, 2022.

[8] G. Cantin. Non-existence of the global attractor for a partly dissipative reaction-diffusion system with hysteresis. *Journal of Differential Equations*, 299:333–361, 2021.

[9] G. Cantin, B. Delahaye, and B. M. Funatsu. On the degradation of forest ecosystems by extreme events: Statistical model checking of a hybrid model. *Ecological Complexity*, 53:101039, 2023.

[10] G. Cantin, A. Ducrot, and B. M. Funatsu. Mathematical modeling of forest ecosystems by a reaction–diffusion–advection system: impacts of climate change and deforestation. *Journal of Mathematical Biology*, 83(6):1–45, 2021.

[11] G. Cantin and N. Verdière. Networks of forest ecosystems: Mathematical modeling of their biotic pump mechanism and resilience to certain patch deforestation. *Ecological Complexity*, 43:100850, 2020.

[12] J. Chave. Study of structural, successional and spatial patterns in tropical rain forests using TROLL, a spatially explicit forest model. *Ecological modelling*, 124(2-3):233–254, 1999.

[13] E. M. Clarke and E. A. Emerson. Design and synthesis of synchronization skeletons using branching time temporal logic. In *Workshop on logic of programs*, pages 52–71. Springer, 1981.

[14] F. Courchamp, L. Berec, and J. Gascoigne. *Allee effects in ecology and conservation*. OUP Oxford, 2008.

[15] D. C. da Cruz, J. M. R. Benayas, G. C. Ferreira, S. R. Santos, and G. Schwartz. An overview of forest loss and restoration in the Brazilian Amazon. *New Forests*, 52(1):1–16, 2021.

[16] L. Denis-Vidal, G. Joly-Blanchard, and N. Verdiere. Identifiability and estimation of nonlinear models: A distribution framework. In *2007 European Control Conference (ECC)*, pages 328–335. IEEE, 2007.

[17] G. Derroire, C. Piponiot, L. Descroix, C. Bedeau, S. Traissac, O. Brunaux, and B. Herault. Prospective carbon balance of the wood sector in a tropical forest territory using a temporally-explicit model. *Forest Ecology and Management*, 497:119532, 2021.

[18] D. Ellison, C. E. Morris, B. Locatelli, D. Sheil, J. Cohen, D. Murdiyarso, V. Gutierrez, M. Van Noordwijk, I. F. Creed, J. Pokorny, et al. Trees, forests and water: Cool insights for a hot world. *Global environmental change*, 43:51–61, 2017.

[19] R. Fischer, F. Bohn, M. D. de Paula, C. Dislich, J. Groeneveld, A. G. Gutiérrez, M. Kazmierczak, N. Knapp, S. Lehmann, S. Paulick, et al. Lessons learned from applying a forest gap model to understand ecosystem and carbon dynamics of complex tropical forests. *Ecological modelling*, 326:124–133, 2016.

[20] R. A. Fisher, C. D. Koven, W. R. Anderegg, B. O. Christoffersen, M. C. Dietze, C. E. Farrior, J. A. Holm, G. C. Hurtt, R. G. Knox, P. J. Lawrence, et al. Vegetation demographics in Earth system models: A review of progress and priorities. *Global change biology*, 24(1):35–54, 2018.

[21] X. Giam. Global biodiversity loss from tropical deforestation. *Proceedings of the National Academy of Sciences*, 114(23):5775–5777, 2017.

[22] K. Godfrey and J. DiStefano III. Identifiability of model parameter. *IFAC Proceedings Volumes*, 18(5):89–114, 1985.

[23] S. Gourlet-Fleury, J.-M. Guehl, and O. Laroussinie. *Ecology and management of a neotropical rainforest: lessons drawn from Paracou, a long-term experimental research site in French Guiana*. Elsevier, 2004.

[24] C. Haga, W. Hotta, T. Inoue, T. Matsui, M. Aiba, T. Owari, S. N. Suzuki, H. Shibata, and J. Morimoto. Modeling tree recovery in wind-disturbed forests with dense understory species under climate change. *Ecological Modelling*, 472:110072, 2022.

[25] V. H. Heinrich, R. Dalagnol, H. L. Cassol, T. M. Rosan, C. T. de Almeida, C. H. Silva Junior, W. A. Campanharo, J. I. House, S. Sitch, T. C. Hales, et al. Large carbon sink potential of secondary forests in the Brazilian Amazon to mitigate climate change. *Nature communications*, 12(1):1–11, 2021.

[26] B. Herault, J. Ouallet, L. Blanc, F. Wagner, and C. Baraloto. Growth responses of neotropical trees to logging gaps. *Journal of applied ecology*, 47(4):821–831, 2010.

[27] U. Hiltner, A. Huth, A. Bräuning, B. Hérault, and R. Fischer. Simulation of succession in a neotropical forest: High selective logging intensities prolong the recovery times of ecosystem functions. *Forest Ecology and Management*, 430:517–525, 2018.

[28] U. Hiltner, A. Huth, B. Hérault, A. Holtmann, A. Bräuning, and R. Fischer. Climate change alters the ability of neotropical forests to provide timber and sequester carbon. *Forest Ecology and Management*, 492:119166, 2021.

[29] D. Julien, G. Cantin, and B. Delahaye. End-to-end statistical model checking for parametric ODE models. In *International Conference on Quantitative Evaluation of Systems*, pages 85–106. Springer, 2022.

[30] M. Korzukhin and M. Antonovsky. *Population level models of forest dynamics*. WP-89-074, 1989.

[31] Y. A. Kuznetsov, M. Y. Antonovsky, V. Biktashev, and E. Aponina. A cross-diffusion model of forest boundary dynamics. *Journal of Mathematical Biology*, 32(3):219–232, 1994.

[32] C. Largouët, M.-O. Cordier, Y.-M. Bozec, Y. Zhao, and G. Fontenelle. Use of timed automata and model-checking to explore scenarios on ecosystem models. *Environmental Modelling & Software*, 30:123–138, 2012.

[33] A. Legay, B. Delahaye, and S. Bensalem. Statistical model checking: An overview. In *International conference on runtime verification*, pages 122–135. Springer, 2010.

[34] R. Lorenz, A. Pitman, and S. A. Sisson. Does Amazonian deforestation cause global effects; can we be sure? *Journal of Geophysical Research: Atmospheres*, 121(10):5567–5584, 2016.

[35] Y. Luo and H. Y. Chen. Observations from old forests underestimate climate change effects on tree mortality. *Nature communications*, 4(1):1–6, 2013.

[36] Y. Luo and H. Y. Chen. Climate change-associated tree mortality increases without decreasing water availability. *Ecology Letters*, 18(11):1207–1215, 2015.

[37] I. Martínez Cano, E. Shevliakova, S. Malyshev, S. J. Wright, M. Detto, S. W. Pacala, and H. C. Muller-Landau. Allometric constraints and competition enable the simulation of size structure and carbon fluxes in a dynamic vegetation model of tropical forests (lm3ppa-tv). *Global Change Biology*, 26(8):4478–4494, 2020.

[38] M. Martínez-Ramos and E. R. Alvarez-Buylla. How old are tropical rain forest trees? *Trends in plant science*, 3(10):400–405, 1998.

[39] R. G. Mateo, G. Arellano, V. Gómez-Rubio, J. S. Tello, A. F. Fuentes, L. Cayola, M. I. Loza, V. Cala, and M. J. Macía. Insights on biodiversity drivers to predict species richness in tropical forests at the local scale. *Ecological Modelling*, 473:110133, 2022.

[40] N. McDowell, C. D. Allen, K. Anderson-Teixeira, P. Brando, R. Brienen, J. Chambers, B. Christoffersen, S. Davies, C. Doughty, A. Duque, et al. Drivers and mechanisms of tree mortality in moist tropical forests. *New Phytologist*, 219(3):851–869, 2018.

[41] A. Mirabel, B. Hérault, and E. Marcon. Diverging taxonomic and functional trajectories following disturbance in a neotropical forest. *Science of the Total Environment*, 720:137397, 2020.

[42] J.-F. Molino and D. Sabatier. Tree diversity in tropical rain forests: a validation of the intermediate disturbance hypothesis. *Science*, 294(5547):1702–1704, 2001.

[43] S. Morand and C. Lajaunie. Outbreaks of vector-borne and zoonotic diseases are associated with changes in forest cover and oil palm expansion at global scale. *Frontiers in veterinary science*, page 230, 2021.

[44] C. A. Nobre, G. Sampaio, L. S. Borma, J. C. Castilla-Rubio, J. S. Silva, and M. Cardoso. Land-use and climate change risks in the Amazon and the need of a novel sustainable development paradigm. *Proceedings of the National Academy of Sciences*, 113(39):10759–10768, 2016.

[45] J. J. O'Brien, S. F. Oberbauer, D. B. Clark, and D. A. Clark. Phenology and stem diameter increment seasonality in a Costa Rican wet tropical forest. *Biotropica*, 40(2):151–159, 2008.

[46] S. Panda, D. Amatya, J. Grace, P. Caldwell, and D. Marion. Extreme precipitation-based vulnerability assessment of road-crossing drainage structures in forested watersheds using an integrated environmental modeling approach. *Environmental Modelling & Software*, 155:105413, 2022.

[47] C. Park, K. Takahashi, J. Takakura, F. Li, S. Fujimori, T. Hasegawa, A. Ito, and D. Lee. How will deforestation and vegetation degradation affect global fire activity? *Earth's Future*, 9(5):e2020EF001786, 2021.

[48] C. Piponiot, G. Derroire, L. Descroix, L. Mazzei, E. Rutishauser, P. Sist, and B. Hérault. Assessing timber volume recovery after disturbance in tropical forests–a new modelling framework. *Ecological Modelling*, 384:353–369, 2018.

[49] C. Piponiot, E. Rutishauser, G. Derroire, F. E. Putz, P. Sist, T. A. West, L. Descroix, M. C. Guedes, E. N. H. Coronado, M. Kanashiro, et al. Optimal strategies for ecosystem services provision in amazonian production forests. *Environmental Research Letters*, 14(12):124090, 2019.

[50] F. E. Putz, P. Sist, T. Fredericksen, and D. Dykstra. Reduced-impact logging: challenges and opportunities. *Forest ecology and management*, 256(7):1427–1433, 2008.

[51] J.-P. Queille and J. Sifakis. Specification and verification of concurrent systems in cesar. In *International Symposium on programming*, pages 337–351. Springer, 1982.

[52] S. Ramondenc, D. Eveillard, L. Guidi, F. Lombard, and B. Delahaye. Probabilistic modeling to estimate jellyfish ecophysiological properties and size distributions. *Scientific reports*, 10(1):1–13, 2020.

[53] S. Schmitt, I. Maréchaux, J. Chave, F. J. Fischer, C. Piponiot, S. Traissac, and B. Hérault. Functional diversity improves tropical forest resilience: Insights from a long-term virtual experiment. *Journal of Ecology*, 108(3):831–843, 2020.

[54] D. Sheil and R. M. May. Mortality and recruitment rate evaluations in heterogeneous tropical forests. *Journal of ecology*, pages 91–100, 1996.

[55] C. A. d. Silva Junior, M. Lima, P. E. Teodoro, J. F. d. Oliveira-Júnior, F. S. Rossi, B. M. Funatsu, W. Butturi, T. Lourençoni, A. Kraeski, T. D. Pelissari, et al. Fires drive long-term environmental degradation in the amazon basin. *Remote Sensing*, 14(2):338, 2022.

[56] C. H. Silva Junior, A. Pessoa, N. S. Carvalho, J. B. Reis, L. O. Anderson, and L. E. Aragão. The Brazilian Amazon deforestation rate in 2020 is the greatest of the decade. *Nature Ecology & Evolution*, 5(2):144–145, 2021.

[57] C. Temperli, C. Blattert, G. Stadelmann, U.-B. Brändli, and E. Thürig. Trade-offs between ecosystem service provision and the predisposition to disturbances: a NFI-based scenario analysis. *Forest Ecosystems*, 7:1–17, 2020.

[58] A. R. Townsend, G. P. Asner, and C. C. Cleveland. The biogeochemical heterogeneity of tropical forests. *Trends in ecology & evolution*, 23(8):424–431, 2008.

[59] B. Zhang and D. L. DeAngelis. An overview of agent-based models in plant biology and ecology. *Annals of Botany*, 126(4):539–557, 2020.

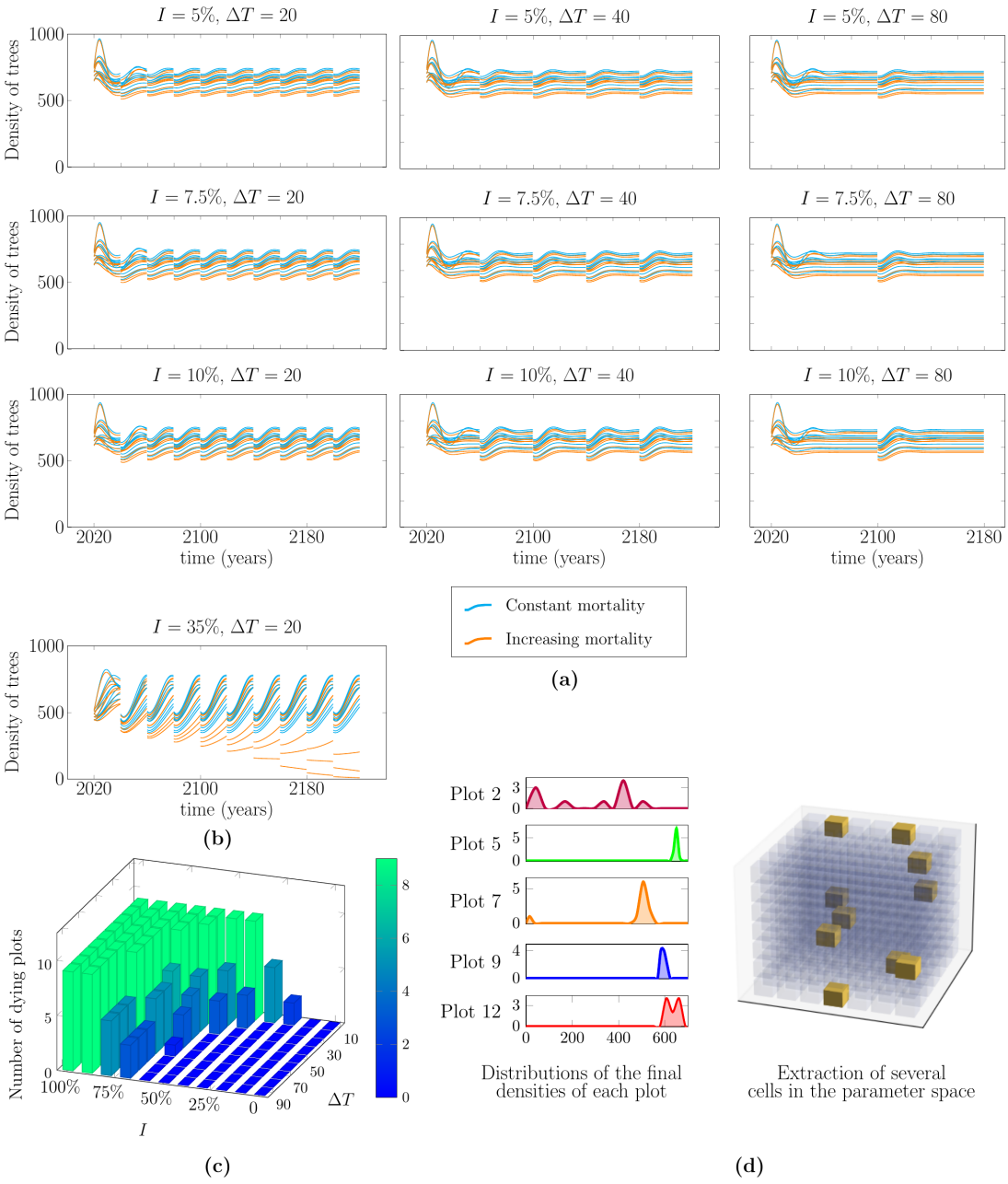


Figure 8: Numerical simulations of the mathematical model (4) for testing relevant logging scenarios. The intensity of the perturbation induced by logging and the time delay between two logging phases are denoted by I and ΔT respectively. **(a)** Trajectories of the model showing the time evolution of the densities of trees on the Paracou forest plots 2, 3, 4, 5, 7, 8, 9, 10, 12 for a constant mortality scenario (blue lines) or for an increasing mortality scenario (orange lines). Each discontinuity jump corresponds to a logging phase. The logging intensity I varies from 5% to 10% and the time delay ΔT varies from 20 to 80 years. **(b)** In the case of an increasing mortality, three plots are shown to decline for a larger value $I = 35\%$ and $\Delta T = 20$ years. **(c)** 3D histogram showing the number of dying plots out of a total of 9 plots, for an enlarged range of values where the logging intensity I varies from 0 to 100% and the time delay ΔT varies from 10 years to 90 years. **(d)** Distributions of the final densities of trees for plots 2, 5, 7, 9, 12, after a selection of several acceptable cells in the parameter space. The graphics show that some plots exhibit a high sensitivity (plots 2, 7), while others are more robust (plots 5, 9, 12).