

**Luiselli, Juliette**

Master Thesis M1 IMA Lis Life Sciences

10/02/2019 — 18/06/2019

2019 — 2020

École Normale Supérieure, Biology Department

PSL Université Paris

*Distinguishing different forms of competition  
in a mechanistic model of eco-evolutionary dynamics*

Rosindell Lab (Silwood Park Campus)

Imperial College (UNITED KINGDOM, London)

Supervisor: ROSINDELL, James, Doctor

**Imperial College  
London**



## **Abstract**

The biodiversity of ecological communities is influenced by ecological and evolutionary processes, yet both are rarely studied together as models often focus on a single spatial, temporal or taxonomic scale. The aim of this work is to predict the community assembly forces shaping biodiversity, as well as other classical evolutionary and ecological parameters from empirical data. To do this, I expanded an existing innovative mechanistic model of community assembly (“massive eco-evolutionary synthesis simulations”, or MESS), which uses information from three biodiversity axes – species richness and abundance; population genetic diversity; and trait variation – to distinguish between competition, environmental filtering and neutrality. Building on the foundations provided by MESS, I added essential components required to properly detect and characterise elements of selection. This includes a new form of competition, explicitly comparing each pair of individuals and distinguishing between inter- and intraspecific competition. The results show that the new form of competition resembles greatly the neutral model, blurring the frontier between selection and neutrality. It also shows that the lack of ability to determine between discrete community assembly models is in fact indicative of a mixed reality in which both filtering and competition act, but in different amounts and sometimes weakly, highlighting the necessity of a more cohesive view of community assembly forces rather than a discrete dichotomous one.

## Glossary

- **MESS** : Massive Eco-Evolutionary Synthesis Simulations
- $m$  : migration rate (at each “birth” event, the probability that the individual comes from the meta community)
- $s$  : speciation rate (at each “birth” event, the probability that it is followed by a speciation event)
- $J$  : number of individuals in the local community
- $s_E$  : ecological strength (represents the strength of selection for the environmental filtering of the competition)
- $z$  : trait value of an individual (no dimension, represents a generic trait)
- **neutral model**: community assembly model where all individuals have the same death probability, regardless of their species identity or trait value.
- **environmental filtering model**: community assembly model where the death probability depends on the distance between the trait value of an individual and the filtering optimum of the local community.
- **mean competition model**: community assembly model where the death probability depends on the distance between the trait value of an individual and the mean trait value of the local community.
- **pairwise competition model**: community assembly model where we compare all pair of individuals. This does not take the species into account, just the trait value.
- **interaction matrix model**: community assembly model where we compare all pair of individuals, their trait difference being weighted by a different factor if they are from the same species or from different species
- **continuous model**: mixed community assembly model with an amount of **neutrality**, **environmental filtering** and **competition** (following the interaction matrix model).

# Summary

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Material &amp; Methods</b>	<b>2</b>
2.1	The core MESS model . . . . .	2
2.2	Pairwise competition . . . . .	4
2.3	Interaction matrix . . . . .	5
2.4	Testing convergence to apparent neutrality . . . . .	5
2.5	From discrete to continuous assembly model . . . . .	6
2.6	Code availability and architecture . . . . .	7
2.7	Inference methods . . . . .	7
2.8	Empirical examples . . . . .	7
<b>3</b>	<b>Results</b>	<b>8</b>
3.1	Different forms of competition show different behaviors . . . . .	8
3.2	Convergence of pairwise competition model to neutrality . . . . .	9
3.3	Continuous model . . . . .	11
3.4	Fitting to empirical data . . . . .	12
<b>4</b>	<b>Discussion</b>	<b>13</b>
4.1	Convergence of competition models to apparent neutrality . . . . .	13
4.2	Empirical results . . . . .	15
<b>5</b>	<b>Perspectives and future developments</b>	<b>16</b>
<b>6</b>	<b>Conclusion</b>	<b>17</b>
	<b>References</b>	<b>18</b>

# 1 Introduction

Unraveling the ecological and evolutionary mechanisms that shape biodiversity is a major challenge in Biology. Many models have been developed attempting to reveal the diverse biological processes behind the evolution of biodiversity (24; 33). However, most of these attempts apply exclusively to ecology or evolution, because models focus on different time scales and different levels of structure (e.g. species or populations). Indeed, ecological models tend to be individual-based (11; 25) and focus on relatively short timescales, while evolutionary frameworks tend to focus on longer timescales and study species dynamics (1; 19). Nevertheless, models emerge attempting to connect both fields (30), since it is now recognised that the ecological and evolutionary aspects of biodiversity are interlaced (29). A classical approach is to model evolutionary dynamics in an individual-based manner (24), in order to possibly integrate different species but also different sub-populations of these species.

Many ecological and evolutionary processes are thought to have an influence on structuring the ecological communities, yet there is little agreement on which processes are most important, and methods to infer processes from data are needed. Two particularly prominent processes are environmental filtering and competition. Yet, a lot can also be done without modeling explicit difference between individuals from different species. According to the environmental filtering paradigm, the environment selects against certain species by favoring certain phenotypes over others (4; 15). Selection can also occur through competition between species and individuals that can lead to competitive exclusion (1; 17; 34). Yet, the diversity and relative abundances of species in ecological communities could also be explained by a random walk, with death probabilities independent of species identity (22; 11). This hypothesis is called the neutral theory of biodiversity. All these processes affect various aspects of empirical data: the species richness and abundances distribution (18), as well as population genetic diversities (3), and trait variation of the species in a phylogenetic context (14). Each of these data axes captures different pieces of information: changes in species abundance take place on ecological timescales while the accumulation of genetic variation occurs over thousands of generations (16) and phylogenetic diversity over millions of generations (32).

The MESS model (“Massive Eco-Evolutionary Synthesis Simulations”), developed by OVERCAST et al. 2020 (21), is built upon classical biogeography and community ecology

theory (11; 21), while also incorporating micro- and macro-evolutionary processes. MESS proposes an inference method to determine which process among neutral, environmental filtering and competition has predominated the community assembly in a local community. In addition to that, MESS infers other underlying parameters of the community assembly, such as the migration and speciation rates. Yet, MESS relies on a simplified view of competition, where the distance between an individual's trait value and the mean trait value of the population determines the individual's survival in the face of competition (21; 28). MESS also has other drawbacks, such as discriminating strictly among the community assembly models at the beginning of the simulation without any possibility to have communities shaped by several models in different proportions, which impedes its ability to diagnose competition and filtering as we expect these processes to be mixed in nature. In the environmental filtering case, individuals whose phenotype is closest to a target (the filtering optimum) are favored. This means some individuals can have a huge advantage if they are close enough to the optimum and so be virtually immortal as long as other species are present.

In this project, I built on the MESS simulation method to develop a more biologically realistic competition model, which enables better understanding of the influence of competition and its various forms on the community assembly and more generally on the three axes of biodiversity data. I also developed a continuous approach for the community assembly model, enabling the different models to shape biodiversity in a single simulation. The leading idea is to approach a more realistic view of community assembly and to achieve better predictions of the underlying parameters.

## 2 Material & Methods

### 2.1 The core MESS model

MESS (21) simulates a metacommunity and a local community being colonised. The metacommunity is simulated first with a constant birth-death process, using fixed speciation and extinction rates, until the specified number of species is reached. Traits are simulated on the resulting phylogeny using a Brownian motion process, and abundances per species are sampled from a log-normal distribution.

The metacommunity is large and static with respect to the timescale of processes in

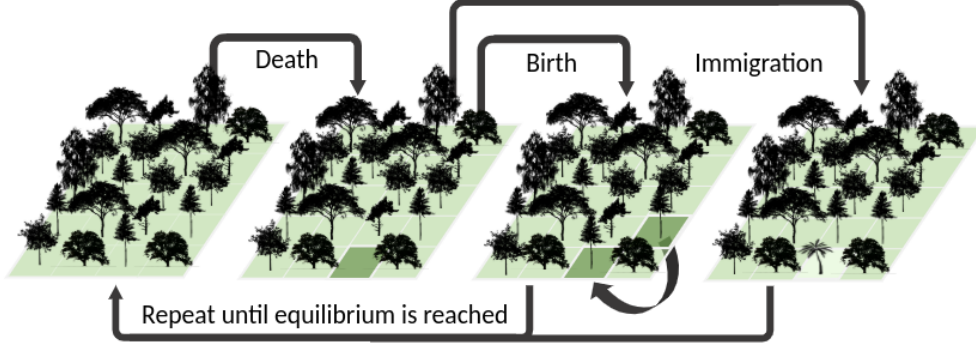


Figure 1: **Functioning of the MESS model**, adapted from ROSINDELL et al. 2011 (26). At each time step, an individual dies and is replaced either by an immigrant or by a descendant from an individual from the local community.

the local community evolution. At each time step for the local community, an individual dies and is replaced either by a descendant from an individual of the local community, or by a migrant from the mainland at rate  $m$  (Fig 1). Speciation events occur at rate  $s$ : the new species inherits its trait value from its parent species, with some normally distributed variation. The size of the local community ( $J$ ) is fixed throughout the simulation. At the end of a simulation, the genetic diversity is simulated for each species' population based on the histories of colonization and abundance changes through time. This enables the calculation of the local genetic diversity and genetic divergence from the mainland (20).

Model	Death probability of the individual $i$ with trait value $z_i$
Neutral	$P(z_i) = \frac{1}{J}$ (1)
Environmental filtering	$P(z_i) = 1 - \exp\left[-\frac{1}{s_E}(z_i - z_E)^2\right]$ (2)
Mean competition	$P(z_i) = \exp\left[-\frac{1}{s_E}(z_i - \bar{z})^2\right]$ (3)

Table 1: **Death probabilities associated to the trait values.** Values are afterwards normalized to sum up to 1.  $z_i$  is the trait value of individual  $i$ .  $J$  is the size of the local community.  $s_E$  is the ecological strength.  $z_E$  is the filtering optimum.  $\bar{z}$  is the mean trait value for the local community.

The death probabilities of the individuals vary depending on the chosen community assembly model. In the neutral case, it is a uniform distribution. In the environmental filtering paradigm, it is based on the distance between the individual's trait value and the filtering optimum. Finally, in a context of competition, it is based on the distance between the individual's trait value and the mean trait value for the local population (21; 28). The ecological strength  $s_E$  refines the effect of competition or environmental filtering by increasing

or decreasing the impact of the distances over the death probabilities (Table 1). See Appendix A.1 for a detailed table of all parameters.

To describe the results of any simulation, summary statistics are computed for each of the three target data axes (species abundances, population genetic variation and trait values). These statistics include generalized Hill numbers (6) – which capture magnitudes and relative evenness of distribution of quantities within the local community for each axis – and other classical data descriptors, such as mean, median, standard deviation, skewness and kurtosis. For more details, see OVERCAST et al. 2020 (21).

## 2.2 Pairwise competition

Competition in MESS originally relies on a comparison between each individual’s trait value and the mean trait value for the local population, following RUFFLEY et al. (28) (eq. 3). Because it is collapsing all trait differences into one value, this simplifying assumption can generate counter-intuitive results: the density of species across the trait axis might be bimodal, leading to an obvious gap around the mean which is wrongly penalised by the algorithm while denser areas are favored (Fig. 2).

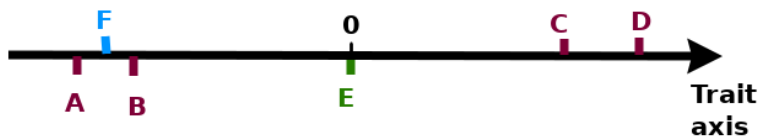


Figure 2: **Schematic example of species distribution in the trait space.** If the red points are existing individuals, classical theory would predict E to outcompete F because he is in an available ecological niche, while the competition model in MESS predicts the opposite.

In order to make the model more biologically realistic, I implemented a pairwise competition model, which compares each individual to all the others.  $z_i$  is the trait value for individual  $i$ ,  $z_j$  for individual  $j$  and  $s_E$  is the ecological strength (eq. 4). The obtained values are normalized throughout the population to sum up to 1. This added complexity had a huge cost on computation time because every pair needs to be checked instead of just each individual against the mean. This issue was tackled by essential code optimisation carried out as part of this project, and a careful test of the optimised code (Appendix A.3).



$$P(z_i) = \sum_{j \neq i} \exp \left[ -\frac{(z_i - z_j)^2}{s_E} \right] \quad (4)$$

### 2.3 Interaction matrix

The pairwise competition model does not allow the refinement of the strength of the intraspecific competition: individuals of the same species have the exact same trait value and thus the exponential is always equal to 1. In order to allow the strength of competition to vary, and also to allow for mutualistic interactions to exist (reducing the death probability instead of increasing it), a square matrix  $\alpha$  of size  $J$  was introduced to specify the strength and sign of interaction for each pair of individual  $ij$ , based on their respective species (eq. 5).

$$P(z_i) = \sum_{j \neq i} -\alpha_{ij} \exp \left[ -\frac{(z_i - z_j)^2}{s_E} \right] \quad (5)$$

I explored different possibilities to parameterize this  $\alpha$  interaction matrix, with the aim to limit the number of additional parameters and to produce realistic results. The most promising was to have separate but fixed for the simulation interspecific and intraspecific terms. The intraspecific term was constrained to be in absolute value greater than or equal to the interspecific term to prevent one species dominating the local community (Appendix A.5). Positive interactions have not yet been investigated.

### 2.4 Testing convergence to apparent neutrality

To understand the apparent convergence of the pairwise competition to neutrality, I studied the trait distribution, which gives a graphical representation of the population, for several sets of parameters to compare our expectations with the outputs. The death mechanism being the only difference between the community assembly models, the distribution of death probabilities was also plotted to understand whether the convergence in summary statistics is due to a convergence in death probabilities. To study whether the death probabilities distribution was significantly different from a uniform distribution, I conducted a Pearson's chi-squared test: 10 000 individuals are drawn with replacement from the local community, and the obtained distribution is compared to a uniform expectation. This binary

descriptor of neutrality is averaged over several generations to give the proportion of nearly neutral step in a simulation, the metric used the figures.

## 2.5 From discrete to continuous assembly model

To tackle the issue of the strict distinction between the different community assembly models, I proposed a continuous approach of the model. In the continuous formulation, the assembly model is a mosaic of filtering, competition, and neutral models, their respective amount being represented by coordinate on a simplex. For each simulation, the death function is drawn according to these proportions at each time step. Biologically, this means that an individual can die from being poorly adapted to its environment (filtering), being out-competed for resources (competition), or just having bad luck (neutral).

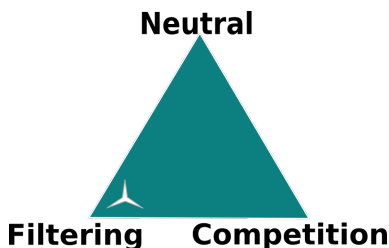


Figure 3: **Graphical representation of the continuous community assembly model:** 5% Pairwise competition, 5% Neutral and 90% Filtering

Instead of predicting a single value, the inference will have to predict two values (amount of competition, amount of filtering, the rest being neutral), which could have a negative impact on the efficiency of the inference. Yet, this could also enable more accurate predictions, a continuous approach of the community assembly processes being more biologically realistic.

Since all death probabilities are used in each simulation, the ecological strength parameter  $s_E$  cannot remain unified between environmental filtering and competition throughout the software. Considering that the interspecific and intraspecific competition terms are sufficient to refine the strength of the competition in the pairwise competition steps, I removed the ecological strength from these simulation when using the continuous model.

To approach the behavior of the new versions of MESS, I reproduced the figures 2 and 3 from OVERCAST et al. 2020 for the [pairwise model](#), the [pairwise model with varying inter- and intraspecific competition](#) (called interaction matrix) and the continuous model. For the continuous model, simulations were sorted according to the community assembly model with

the highest proportion of the three.

## 2.6 Code availability and architecture

The code used for the simulations is available on GitHub in its three versions (pairwise competition, interaction matrix and continuous). I introduced the possibility to specify the random seed for any simulations to make our tests repeatable, and all simulation results are available online with the used parameters for reproducibility (Appendix A.2).

## 2.7 Inference methods

To perform the inference based on MESS simulations, I first ran several tens of thousands simulations with parameters spanning over wide range of possible values (Appendix A.9), and computed the summary statistics for all these runs and for the empirical data. The MESS inference procedure relies on a classifier and a regressor, trained with Random Forest strategies (5) and implemented with the scikit-learn module for python (v0.20.3, (23)). To quantify the accuracy of both the classifier and the regressor, simulated data are split into 5 groups and each is used successively as test set and the others as training set. The real parameters used for each simulation being known, the accuracy can be computed.

The prior value used for the inferences are based on OVERCAST et al. 2020 (21), in order to enable a comparison of the results between both methods. The competition model used here are both new models (pairwise competition and interaction matrix), and not the mean competition model.

## 2.8 Empirical examples

I reanalyzed the data sets from OVERCAST et al. 2020 as a reference for comparing the predicted results: 1) the spider community from Reunion island (8); 2) weevil communities from two Mascarene islands (Reunion and Mauritius) (12); 3) three subtropical rain forest tree communities (27); 4) Galapagos snail communities collected from Santa Cruz (13; 31).

To compare the precision of the different predictions obtained with our different inferences (with and without the continuous model, and with interspecific competition lower or equal to intraspecific competition in the discrete case), I qualitatively compared the prediction errors of the different regressions, and the length of the 95% intervals prediction as well

as the relative error for the prediction (Appendix A.11). To compare the predicted community assembly models, I plotted the confidence in each model in the discrete case, and the estimation of proportions in the continuous case (Fig. 10).

### 3 Results

#### 3.1 Different forms of competition show different behaviors

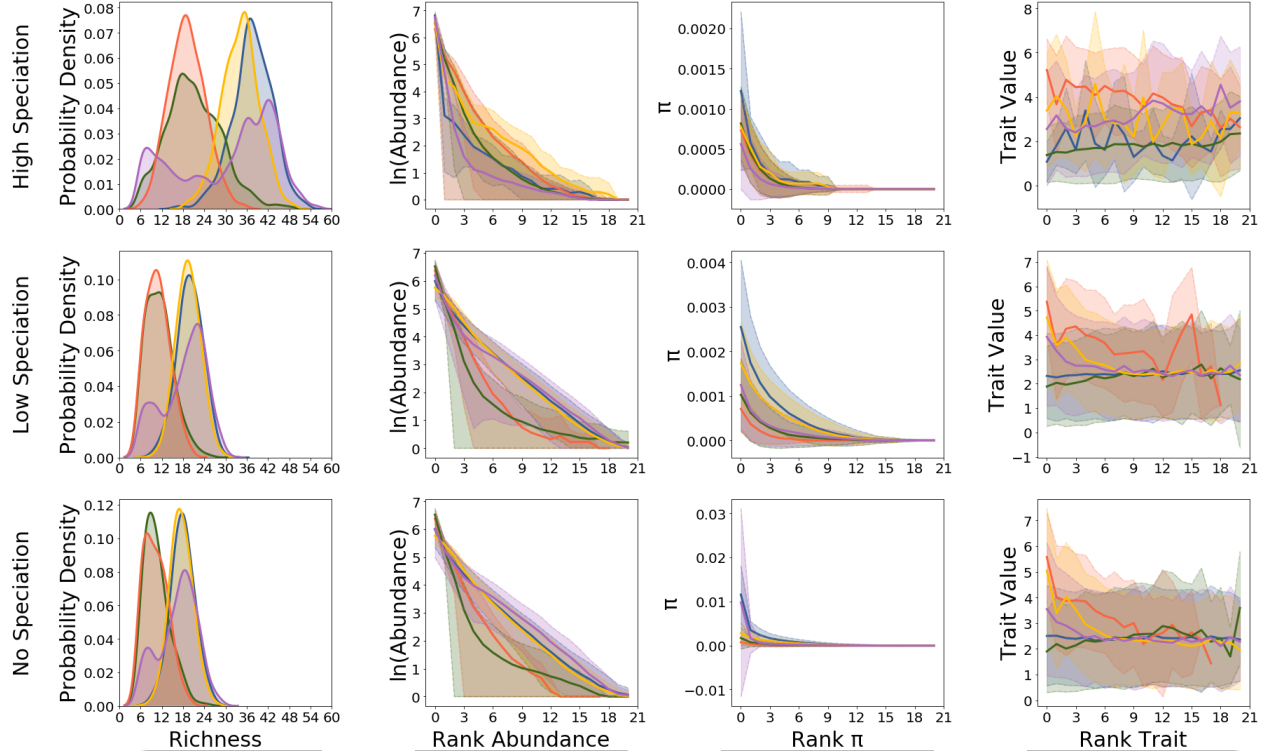


Figure 4: **Effect of varying speciation rate and community assembly model on summary statistics.** Species richness, rank abundance, rank genetic diversity, and rank trait values for over 1000 simulations generated under **neutral**, **mean competition**, **pairwise competition**, **interaction matrix** and **environmental filtering** scenarios with **time fixed at 500 generations**, with  $J=1000$ ,  $s_E=0.1$  and  $m=5e-3$ . Rows of panels correspond to simulations with high ( $s = 0.0001$ ), low ( $s = 0.00005$ ) and no ( $s = 0$ ) speciation. In the left column of panels, plots indicate the distribution of richness across simulations. In the rank plots, thick lines indicate average rank values and shaded areas show plus and minus one standard deviation.

I first verified that the results of the pairwise competition significantly differed from the mean competition model by inspection of figure 4 (reproduction of Fig 2 of OVERCAST et al. 2020 (21)). The **pairwise competition** model differs highly from the **mean competition** model. The exploration of the behavior of the different community assembly models through time also shows that for all the represented variables, the **pairwise competition** model is closer to the **neutral** model than any other (see Fig 5, reproduction of Fig 3 from OVERCAST et al.

2020 (21)). The behavior of [this version of MESS](#) shows intermediate results between the [pairwise model](#) and the [filtering model](#) (Fig 4, 5) in most cases, and behaves quite differently at the end of the simulations. Although  $\Lambda=1$ , the hill numbers have not reached a plateau.

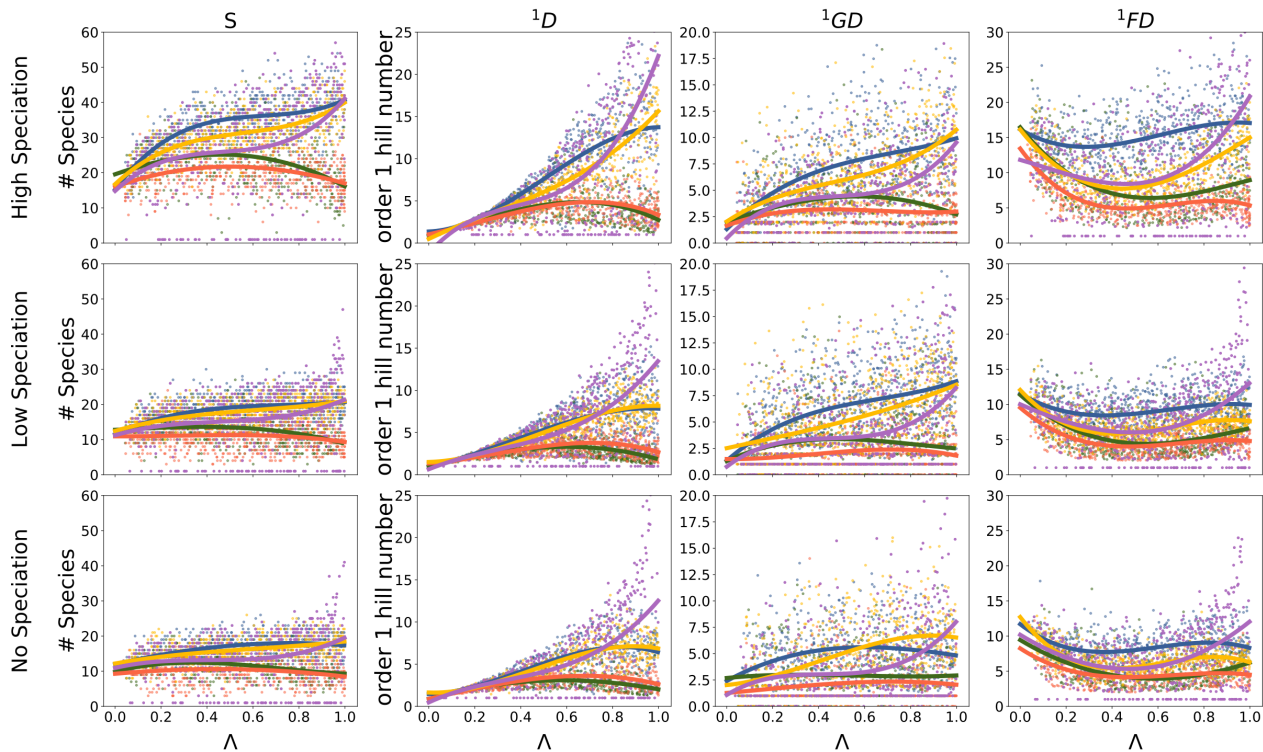


Figure 5: **Community summary statistics through time for neutral and non-neutral models.** This plot depicts the temporal change in selected summary statistics for the four focal community assembly models at three different speciation rates: No, Low, and High corresponding to  $s = 0, 0.0005, 0.005$ , respectively. Community assembly models depicted are [neutral](#), [mean competition](#), [pairwise competition](#), [interaction matrix](#) and [environmental filtering](#). Each subpanel shows the resultant summary statistic for over 1500 simulations **equally spaced through time** for each model class, with  $J=1000, s_E=0.1$  and  $m=5e-3$ . Simulated values are depicted as points, and a least squares polynomial is fit to better illustrate the trajectory. The far left column of panels illustrate species richness on the y-axes (S). The y-axes of the remaining columns illustrate the Hill number of order 1 for abundance, genetic diversity, and trait values, respectively.

### 3.2 Convergence of pairwise competition model to neutrality

For the 4 models the trait distribution corresponds to our expectations (Appendix A.8). The even distribution across the trait axis in the pairwise competition model case could lead to equal death probabilities for each individual. After 160 generations, all death probabilities tend to resemble a uniform distribution (Appendix A.4), however this convergence happens earlier in the simulation in the case of the pairwise competition model (Fig 7): at 64 generations the death probabilities distribution is already flat, while this is not the case for the other 2 models. This could explain the apparent convergence of the pairwise competition

simulations to neutrality.

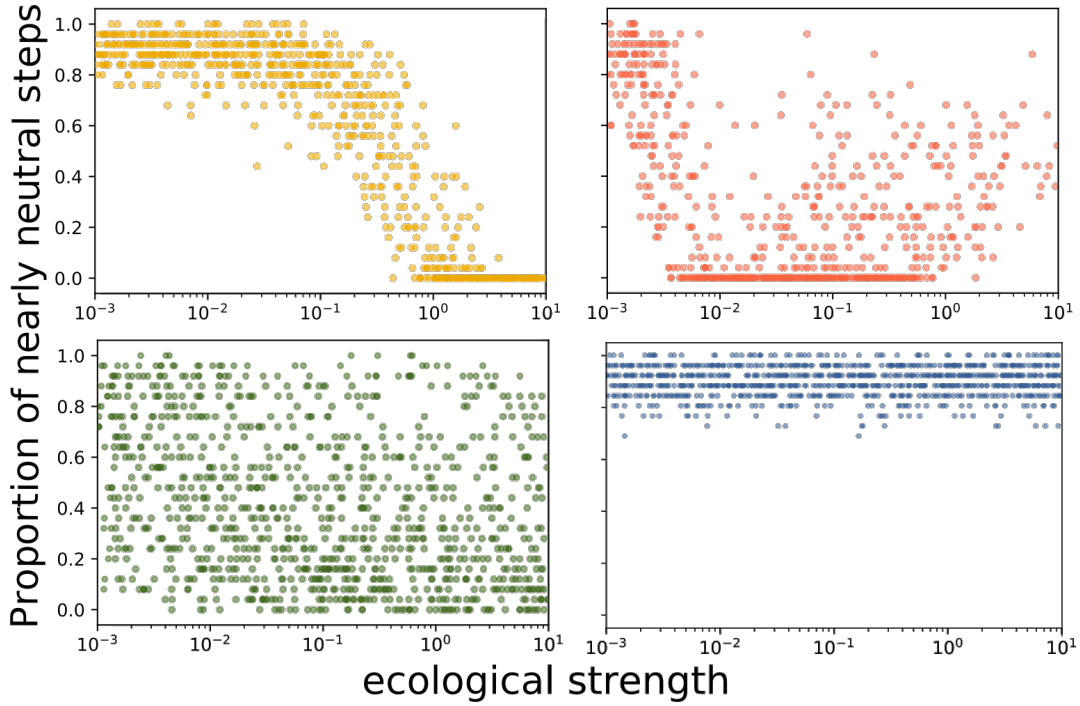


Figure 6: Percentage of almost neutral time-steps with different ecological strength parameters for 3612 simulations, over the 100 last generations (400 to 500).

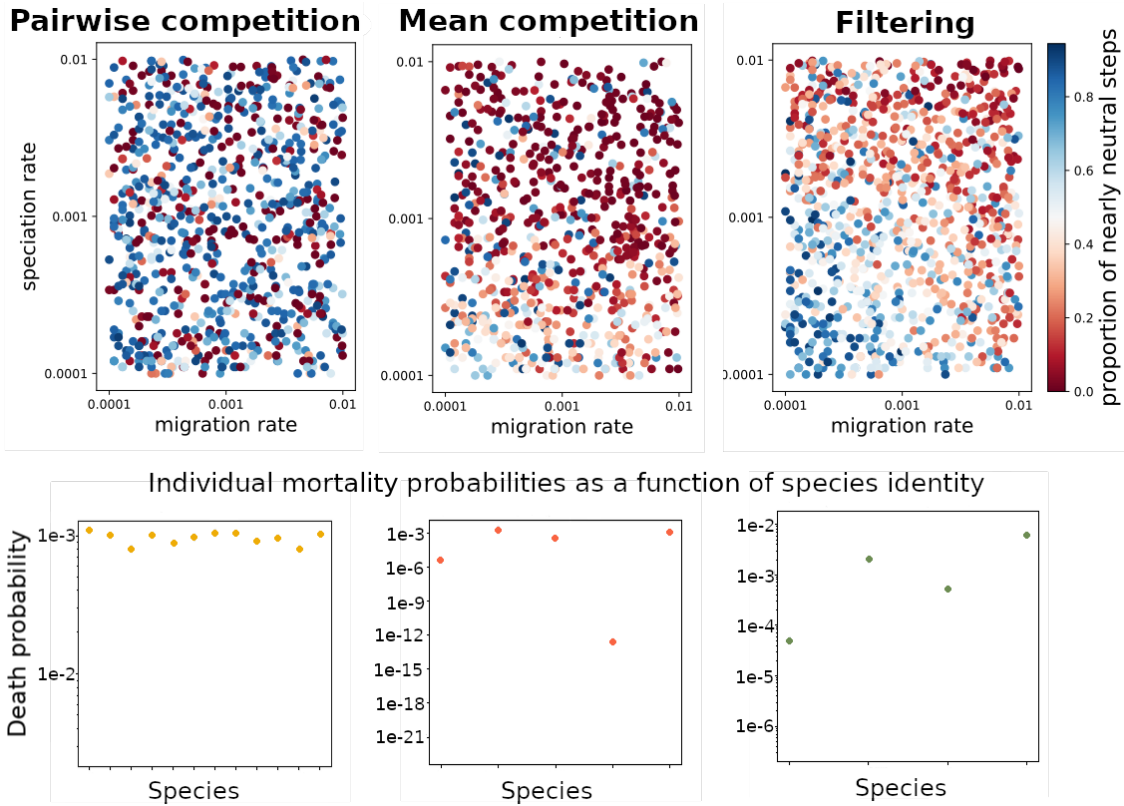


Figure 7: Percentage of almost neutral time-steps for different migration and speciation rates over the 100 last generations and representative death probabilities per species after 64 generations. See Appendix A.6 for the neutral simulations.

For the **pairwise competition** model, the lower the ecological strength is, the more uniform the death probabilities distribution is (Fig. 6). The transition period is between  $5 \cdot 10^{-2}$  and 3, which is within the tested range for the inferences. This effect is also detectable in the **filtering** model, even if it is less clear: natural selection acts differently in both models. However, no clear pattern emerge for the **mean competition** model. The dynamic through time confirms that simulations become more neutral as time passes in the pairwise competition case, but not for the other models (Appendix A.7).

The impact of speciation and migration rates over the proportion of nearly neutral death steps is clear for the environmental filtering model and also distinguishable for the mean competition model (Fig 7): the lower the rates, the more neutral the results are. There was no discernible impact of migration or speciation on proportion of apparently neutral steps in the pairwise competition model

### 3.3 Continuous model

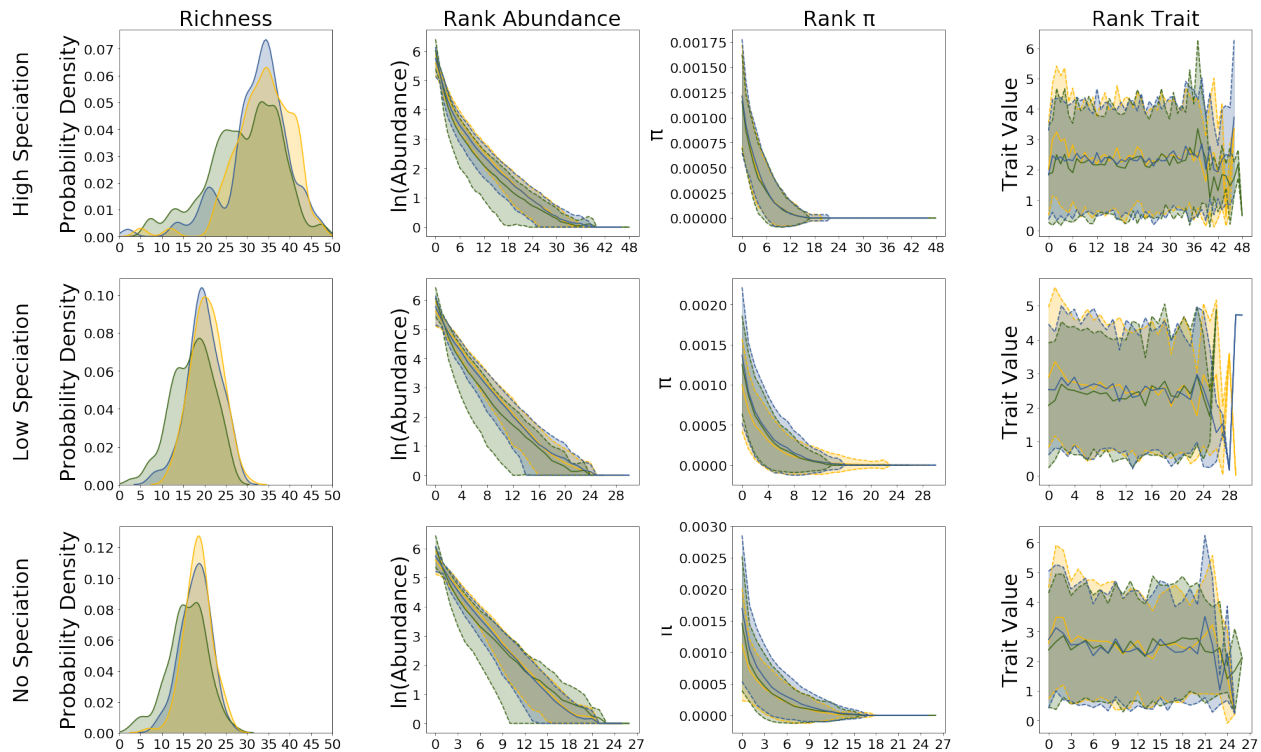


Figure 8: **Effect of varying speciation rate and community assembly model on summary statistics.** Species richness, rank abundance, rank genetic diversity, and rank trait values for 855 simulations generated with time fixed at 500 generations, with various proportions of the community assembly models (more **neutral**, **competition** or **environmental filtering**). See Fig 4 for description.

The competition model used for the continuous implementation is the interaction matrix.

The results shown in Fig 8 are quite different from those produced by the discrete community assembly model. For example, the distributions of the number of species are highly variable and no longer normal and the three curves are overlapping. When looking at the changes across time however (Fig. 9), the distinction between the different predominating models is clearer, especially at the later stages of the simulations.

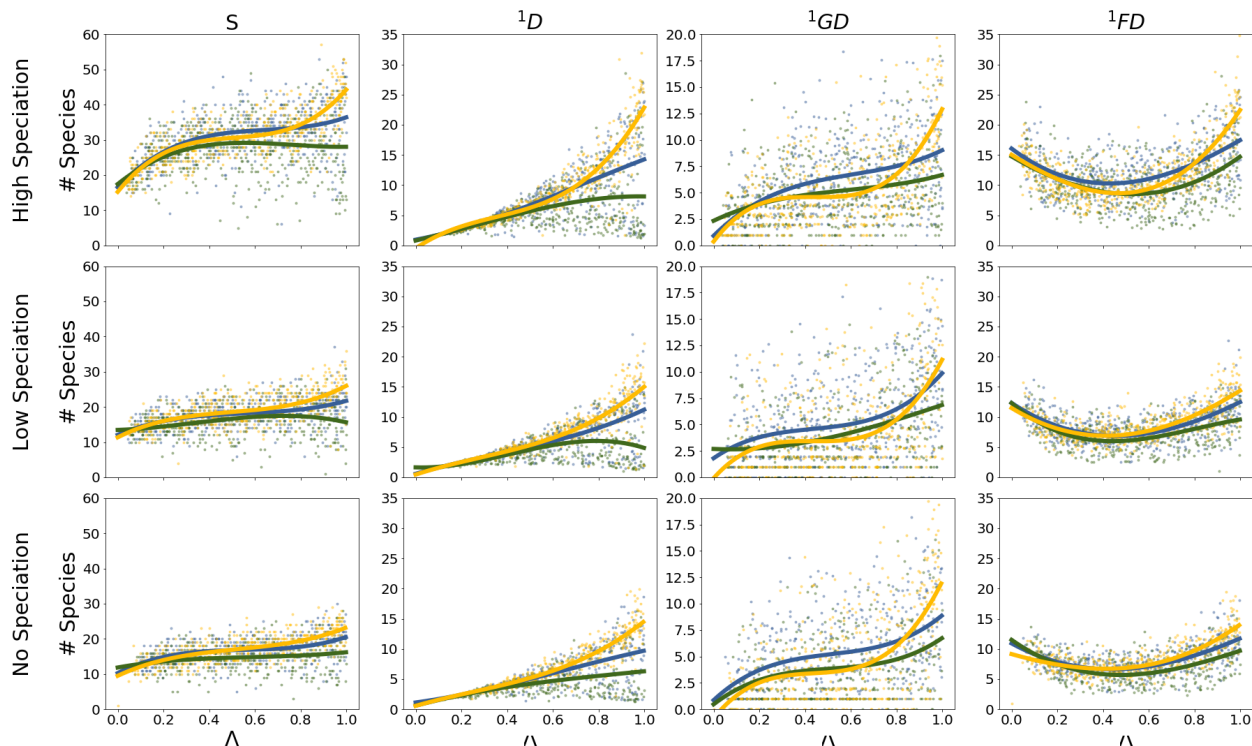


Figure 9: **Effect of varying speciation rate and community assembly model proportions on summary statistics.** Majority community assembly models depicted are **neutral**, **competition** and **environmental filtering**. Each subpanel shows the resultant summary statistic for over 1200 simulations equally spaced through time for each model class. See Fig 5 for description.

### 3.4 Fitting to empirical data

The performance of the classifier in the discrete community assembly model has been assessed with confusion matrices (Appendix A.10). These show that the community assembly model used in the simulation is easily recovered by our inference method. This proves that the community assembly model could be recovered, if what happens in nature is close to what is modeled by MESS. The performance of the regressor was assessed through cross validation plots (Appendix A.10). For both discrete community assembly inferences, the regression provides encouraging results, with a high predictability reached for  $\alpha$ ,  $J$  and  $\nu$ . However, some parameters (such as  $m$  or  $s_E$ ) are harder to predict. In contrast our inference



method does not perform well enough to enable to do predictions for empirical data with the continuous community assembly model (Appendix A.10.4).

The predicted community assembly model is consistent throughout the different inference methods for the same data set, and differs from one empirical data set to the other (Fig 10). However, results for the other parameters are harder to interpret because of high variability in the predicted values and low confidence in those (Appendix A.11). The performances of the different version has been compared quantitatively through the width of the 95% confidence intervals and the relative error (Appendix A.11), and qualitatively with the posterior predictive check plots (Appendix A.12). The inference method performing best is the interaction matrix.

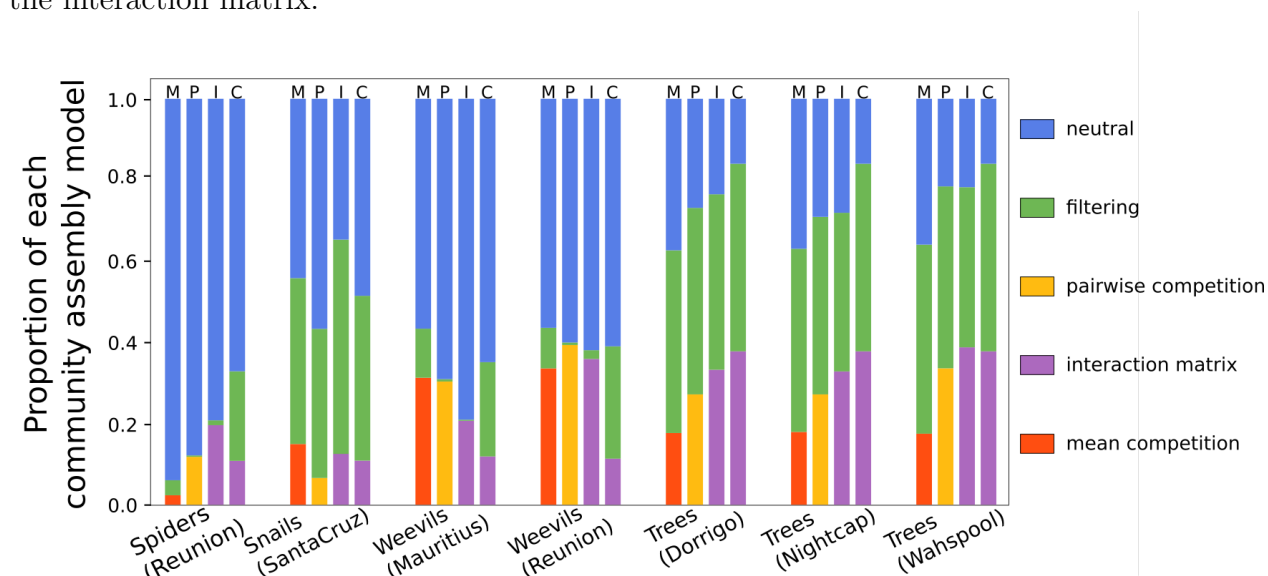


Figure 10: **Predicted model for each set of empirical data.** Successively for each data set : the original prediction from OVERCAST et al. 2020 (M), the prediction for the pairwise competition model (P), for the interaction matrix (I) and the continuous model (C).

## 4 Discussion

### 4.1 Convergence of competition models to apparent neutrality

Strongly accurate descriptions of biodiversity are needed to discern competition from neutrality in the context of more accurate models of competition than previously studied in related work (21) producing results resembling neutral results. The theory of emergent neutrality argues that the mechanisms behind niche theory and those behind the neutral theory could lead to the same observable results in the community structure (10). One way

to explain this apparent convergence is that mechanisms which appear to be neutral are regulated by more complex underlying competition mechanisms (hidden niches (2)), often not directly modeled but ruling the community assembly. Another explanation is that species either end up very similar in traits (so they don't compete strongly) or different (so they don't compete at all) but not intermediate, where the niches have a more obvious role in dynamics. In this perspective, the neutral theory is a simplification of what is observed and a useful tool to reduce the number of parameters, since it produces results similar to selection with independent niches (7). Our results support this view, and contrast with the previous finding that neutral and non-neutral processes (both environmental filtering and competition) were easily distinguished (21).

The death probabilities distribution shows that the apparent convergence between the neutral and pairwise models is linked to a convergence toward an equilibrium: individuals that are less fit than the others are quickly counter-selected, until all remaining individuals are almost equivalent and thus have almost equal probabilities of dying (Fig. 7). Two factors are essential in this convergence: its speed and its degree. In this perspective, a stronger selection – induced by higher migration and speciation rates, making new individuals available – accelerates the reach of the equilibrium, and so the convergence toward neutrality. However, lower rates are associated to a higher “neutrality” in Fig 7 for the mean and filtering models, which could be explained by the fact that less new individuals arriving through migration or speciation implies that the selection is less efficient and so less detectable. So convergence is faster with higher rates, but also less accurate. The convergence toward an equilibrium in the pairwise competition case is both quicker and more accurate than for the other models. Yet, even if the resulting mechanisms after some generations appear to be the same, communities are structured in a different way, and the trait distribution (Appendix A.8) clearly shows some differences: traits are spread out evenly across trait space in the pairwise competition case and randomly distributed in the neutral case. This proves that all models can be distinguished when considering enough variables or data axes. Here, the species abundance distribution does not enable to distinguish between neutral and non-neutral, but summary statistics associated with trait values could, proving the necessity of mechanistic models spanning over multiple data axes such as MESS.

Our simulations for Fig. 4 and 5 all start with the similar initial conditions (for  $J$ ,  $m$ ,

s, etc.), so it takes time for the process and its effect on the community to be noticeable in summary statistics. Therefore, if processes are variable across time in the wild they will be harder to detect and we must assume that processes remain similar for a long period of time. The comparison with the interaction matrix model must be cautious since the number of varying parameters is not the same (varying inter- and intraspecific competition strengths instead of all parameters being fixed).

Knowing that the inferences from the original publication (21) stated that the communities were mostly neutral, the convergence observed could have a great impact over the community assembly model predicted and could shift the prediction toward competition. Nevertheless, the fine resolution of data produced with the MESS model could enable to discriminate between neutral and non-neutral community assembly processes in empirical data. The neutral model requires less parameters and should therefore be favoured by default, but MESS analyses different data axes and that could enable the detection of an underlying competition.

## 4.2 Empirical results

The processes we want to detect are assumed to be relatively constant through time. This is not the case for the continuous model, which is consistent with the high prediction errors (Appendix A.10.4). This could be inherent to the data generated by the continuous model: since everything is highly random at each time step, no pattern can be distinguished. This is demonstrated in Fig 8: all curves overlap, which is the expected outcome since all simulations are a mix of all models. The necessary consistence of processes across time is also probably contradicted in the filtering case. It might be that after the filtering has happened, it is no longer the dominant force as all present species are already suitable – if the environment remains constant, and our inference method does not account for varying forces across time.

The confusion matrices shows that the prediction of the community assembly model is quite accurate when done for simulated data (Appendix A.10.1). These these data being designed by these models, empirical data may however be less easy to sort, especially if the forces vary across time. Other untested forms of community assembly could be closer to the empirical data than what we have in our simulated benchmark.

The predicted confidence intervals for fitted model parameters were sometimes very broad for the other variables (up to 95% of the *prior*, Appendix A.11). Future works could improve the resolution by running more simulations within these intervals (*a posteriori* distribution of the parameters). Nevertheless, our results show that the new and biologically more realistic form of competition developed here is now detected, presumably because of trait data, despite the birth death process of individuals mimicking neutrality. The fact that the species abundance distribution does not allow to distinguish competition from neutrality had already been stated in literature (9), but taking into account other data axis such as trait values by MESS enables this distinction.

A two dimensional trait space could also stabilize the continuous model by allowing more complex trade-offs in trait selection between competition and environmental filtering. Another hypothesis to explain the high regression error is that it is due to an insufficient number of simulated data sets with respect to the number of parameters to infer: we added a new parameter for the community assembly model and the number of needed simulations grows exponentially with the number of parameters to infer. Increasing the number of summary statistics, if they provide new information, could be another lead. Nevertheless, the continuous model predicts proportions close to the confidences for the choice of the discrete model. This potentially validates our continuous approach, and proportions of the different community assembly models have more ecological meaning than the confidence in an extreme model dominated by only a single force. These encouraging results should therefore be further studied.

## 5 Perspectives and future developments

MESS is constrained by numerous simplifications that are essential to develop any model and to hold to reasonable computation times. However, our developments on the original MESS model suggest possible enhancements and also make the technical improvements necessary to tractably use these enhancements.

The initiation in MESS relies on the simplifying assumption that the local community is totally invaded by a single species at time 0. MESS includes a feature allowing to begin the simulations with a single individual (not explored in this report), but even in this perspective

the maximal size of the populations is fixed and once reached the number of individuals does not vary. A sense of density-dependence, with explicit indirect competition for resources and possibly a varying amount of available resources instead of the explicit direct competition between individuals, could be introduced to counter this simplification. However, these mechanisms are very complex and would affect tractability.

The model could be further developed to incorporate other features. For instance, the phenotypic representation of each individual - its trait value - is only one dimensional in the current version of the model, however interesting dynamics of trade-off (especially with the continuous community assembly model and the eventual antagonistic influences of competition and environmental filtering) could be explored in two dimensions. Expanding the possible phenotypes could also lead to make the corresponding fitness landscape more complex, it could comprise several peaks of fitness, as with a sum of Gaussian functions for instance. The possibility of including other types of interactions (Table 2) has already been discussed here and included in the model, but not thoroughly tested nor used for the inferences yet and could be further explored.

	-	0	+
-	competition	amensalism	predation/parasitism
0	amensalism	neutralism	commensalism
+	predation/parasitism	commensalism	mutualism

Table 2: **The different types of interactions between individuals**

Evolution is depicted in MESS through the possibility of point-wise speciation in the local community, which includes a variation in the trait value for the species. However, many evolutionary models includes more complex dynamics, including non-homogeneous species (each individual has a unique trait value) and a protracted speciation model (24). Having this possibility of character displacement would probably greatly change the dynamic of the simulation and interesting patterns could emerge. However, this would question the notion of species in itself, and increase the computation costs in time and space.

## 6 Conclusion

The MESS model is an individual-based mechanistic model of community assembly taking into account evolutionary and ecological processes such as dispersal, drift, competition

and speciation. It generates data over three axes used to describe the biodiversity: the genetic diversity, the species abundances and its distribution and the trait diversity. Using a high number of simulations and machine learning tools, MESS provides a framework to test our model against real world data, and to predict the strength of the different community assembly processes – competition, environmental filtering and neutrality – from empirical data. In this project, I changed the form of competition used in MESS to make it more biologically realistic with pairwise comparison, which resulted in simulations resembling substantially the neutral simulations. Competition can still be detected in empirical biodiversity data and distinguished from neutrality thanks to the wide range of data used by MESS, especially the traits data which are very different in both cases. The introduction of a continuous community assembly model showed encouraging results since it reproduced the confidence in each model given by the discrete models, and gives a better interpretation of a mix of factors. This tends to prove that empirical data are indeed produced by a mix of the studied processes, and this mix is increasingly easy to diagnose with more data and more advanced inferences with mechanistic models as we do here.

## Acknowledgment

This Master 1 thesis was financed by the École Normale Supérieure of Paris and took place at the Department of Life Sciences (Imperial College, Silwood Park) under the supervision of Dr. James ROSINDELL, who I would like to thank warmly for his reception and support. I would like to wholeheartedly thank Dr. Isaac OVERCAST (IBENS) for his participation in the project, as well as Pr. Hélène MORLON (IBENS) and her team for their support. Finally, I would also like to thank Flore VANDIER and Paul BANSE for useful comments on the manuscript and Pr. Guillaume BESLON and his team (INRIA) for enriching discussions on the subject.

## References

- [1] Aristide L. and Morlon H., Understanding the effect of competition during evolutionary radiations: an integrated model of phenotypic and species diversification, *Ecology Letters*, 22(12):2006–2017, 2019.
- [2] Barabás G., D’Andrea R. et al., Emergent neutrality or hidden niches?, *Oikos*, 122(11):1565–1572, 2013.
- [3] Baselga A., Gómez-Rodríguez C. et al., Multi-hierarchical macroecology at species and

- genetic levels to discern neutral and non-neutral processes, *Global Ecology and Biogeography*, 24(8):873–882, 2015.
- [4] Bazzaz F.A., Habitat selection in plants, *The American Naturalist*, 137:S116–S130, 1991.
- [5] Breiman L., Random forests, *Machine Learning*, 45(1):5–32, 2001.
- [6] Chao A., Chiu C.H. et al., Unifying species diversity, phylogenetic diversity, functional diversity, and related similarity and differentiation measures through hill numbers, *Annual Review of Ecology, Evolution, and Systematics*, 45(1):297–324, 2014.
- [7] Chisholm R.A. and Pacala S.W., Niche and neutral models predict asymptotically equivalent species abundance distributions in high-diversity ecological communities, *Proceedings of the National Academy of Sciences*, 107(36):15821–15825, 2010.
- [8] Emerson B.C., Casquet J. et al., A combined field survey and molecular identification protocol for comparing forest arthropod biodiversity across spatial scales, *Molecular Ecology Resources*, 17(4):694–707, 2017.
- [9] Etienne R.S. and Rosindell J., Comment on “global correlations in tropical tree species richness and abundance reject neutrality”, *Science*, 336(6089):1639–1639, 2012.
- [10] Holt R., Emergent neutrality, *Trends in Ecology & Evolution*, 21(10):531–533, 2006.
- [11] Hubbell S.P., *The Unified Neutral Theory of Biodiversity and Biogeography (MPB-32)*, 2001.
- [12] Kitson J.J.N., Warren B.H. et al., Community assembly and diversification in a species-rich radiation of island weevils (coleoptera: Cratopini), *Journal of Biogeography*, 45(9):2016–2026, 2018.
- [13] Kraemer A.C., Philip C.W. et al., Trade-offs direct the evolution of coloration in galapagos land snails, *Proceedings of the Royal Society B: Biological Sciences*, 286(1894):20182278, 2019.
- [14] Kunstler G., Lavergne S. et al., Competitive interactions between forest trees are driven by species’ trait hierarchy, not phylogenetic or functional similarity: implications for forest community assembly, *Ecology Letters*, 15(8):831–840, 2012.
- [15] Lebrija-Trejos E., Pérez-García E.A. et al., Functional traits and environmental filtering drive community assembly in a species-rich tropical system, *Ecology*, 91(2):386–398, 2010.
- [16] Leffler E.M., Bullaughey K. et al., Revisiting an old riddle: What determines genetic diversity levels within species?, *PLoS Biology*, 10(9):e1001388, 2012.
- [17] Macarthur R. and Levins R., The limiting similarity, convergence, and divergence of coexisting species, *The American Naturalist*, 101(921):377–385, 1967.
- [18] McGill B.J., Etienne R.S. et al., Species abundance distributions: moving beyond single prediction theories to integration within an ecological framework, *Ecology Letters*, 10(10):995–1015, 2007.
- [19] Morlon H., Phylogenetic approaches for studying diversification, *Ecology Letters*, 17(4):508–525, 2014.

- [20] Overcast I., Emerson B.C. et al., An integrated model of population genetics and community ecology, *Journal of Biogeography*, 46(4):816–829, 2019.
- [21] Overcast I., Ruffley M. et al., A unified model of species abundance, genetic diversity, and functional diversity reveals the mechanisms structuring ecological communities, in review.
- [22] Patel M. and London I.C., Neutral theory: Species in space and time, page 33, 2015.
- [23] Pedregosa F., Varoquaux G. et al., Scikit-learn: Machine learning in python, *Machine Learning in Python*, page 6, 2011.
- [24] Rosindell J., Harmon L.J. et al., Unifying ecology and macroevolution with individual-based theory, *Ecology Letters*, 18(5):472–482, 2015.
- [25] Rosindell J., Hubbell S.P. et al., The unified neutral theory of biodiversity and biogeography at age ten, *Trends in Ecology & Evolution*, 26(7):340 – 348, 2011.
- [26] Rosindell J. and Phillimore A.B., A unified model of island biogeography sheds light on the zone of radiation: A unified model of island biogeography, *Ecology Letters*, 14(6):552–560, 2011.
- [27] Rossetto M., McPherson H. et al., Where did all the trees come from? a novel multispecies approach reveals the impacts of biogeographical history and functional diversity on rain forest assembly, *Journal of Biogeography*, 42(11):2172–2186, 2015.
- [28] Ruffley M., Peterson K. et al., Identifying models of trait-mediated community assembly using random forests and approximate bayesian computation, *Ecology and Evolution*, 9(23):13218–13230, 2019.
- [29] Schoener T.W., The newest synthesis: Understanding the interplay of evolutionary and ecological dynamics, *Science*, 331(6016):426–429, 2011.
- [30] Thuiller W., Münkemüller T. et al., A road map for integrating eco-evolutionary processes into biodiversity models, *Ecology letters*, 16(0):94–105, 2013.
- [31] Triantis K.A., Rigal F. et al., Discordance between morphological and taxonomic diversity: land snails of oceanic archipelagos, *Journal of Biogeography*, 43(10):2050–2061, 2016.
- [32] Uyeda J.C., Hansen T.F. et al., The million-year wait for macroevolutionary bursts, *Proceedings of the National Academy of Sciences*, 108(38):15908–15913, 2011.
- [33] Vellend M., Conceptual synthesis in community ecology, *The Quarterly Review of Biology*, 85(2):183–206, 2010.
- [34] Weiher E., Freund D. et al., Advances, challenges and a developing synthesis of ecological community assembly theory, *Philosophical Transactions of the Royal Society B: Biological Sciences*, 366(1576):2403–2413, 2011.



# Appendix

## A.1 Model's parameters

Categorical parameters		
Parameter		Options
Community assembly model		Neutral / Competition / Environmental filtering / Pairwise competition
<i>In situ</i> speciation model		None/Point mutation/Random fission
Local community initial conditions		Metacommunity sample/Monodominance
Metacommunity component parameters		
Symbol	Meaning of parameter	Type and range
$J_M$	Total number of individuals	Integer $\gg 1$
$S_M$	Total number of species	Integer $> 1$
$\lambda$	Per lineage birth rate (speciation)	Real $\in [0, \infty]$
$\epsilon$	Per lineage death rate (extinction) as proportion of $\lambda$	Real $\in [0, 1]$
$\sigma_M^2$	Trait evolution rate variance (Brownian motion)	Real $> 0$
Local community component parameters		
Symbol	Meaning of parameter	Type and range
$J$	Total number of individuals	Integer $> 1$
$S$	Local species richness*	Integer $> 1$
$\nu$	Per capita per birth speciation rate	Real $\in [0, 1]$
$m$	Immigration rate from metacommunity (per step)	Real $\in [0, 1]$
$\sigma^2$	Trait evolution rate variance (per speciation event)*	Real $> 0$
$z_E$	Optimal trait value in environment*	Real
$s_E$	Strength of ecological filtering	Real $> 0$
$\Lambda$	Fraction of turnover equilibrium*	Real $\in [0, 1]$
$\alpha_{intra}$	Strength of intraspecific competition	Real $> 0$
$\alpha_{inter}$	Strength of interspecific competition	Real $> 0$
Population genetics coalescence component parameters		
Symbol	Meaning of parameter	Type and range
$L$	Sequence length of simulated genomic region	Integer $> 0$
$\mu$	Mutation rate	Real $\in [0, 1]$
$\alpha$	Abundance/Ne scaling factor	Integer $> 0$

Table 3: MESS model parameters

All MESS model parameters, their interpretations and range of possible values. Parameters indicated with an asterisk (\*) are pseudo-parameters which are either emergent, compound, or randomly sampled from a distribution with parameters determined by other ele-

ments of the model.

## A.2 Code availability and simulation results

The code is public and available on GitHub.

Model	Branch name	GitHub link
Pairwise competition	pairwise_competition	<a href="https://github.com/messDiv/MESS/tree/pairwise_competition">https://github.com/messDiv/MESS/tree/pairwise_competition</a>
Interaction matrix	interaction_matrix	<a href="https://github.com/messDiv/MESS/tree/interaction_matrix">https://github.com/messDiv/MESS/tree/interaction_matrix</a>
Continuous model	continuous_model	<a href="https://github.com/messDiv/MESS/tree/continuous_model">https://github.com/messDiv/MESS/tree/continuous_model</a>

Table 4: **Link to the different branches of the project in the GitHub repository**

For a better repeatability of the produced results, all simulations results are also available online. They can also be reproduced locally by using the provided random seeds, or replicated by using the same input parameters. All of these are summarised in the simulations files.

Reference in the report (by order of appearance)	Type	Link
Figures 4, 5, 8 and 9	Simulation results with seed	<a href="https://cloud.biologie.ens.fr/index.php/s/ciaTMDfZss70oix">https://cloud.biologie.ens.fr/index.php/s/ciaTMDfZss70oix</a>
Figures 6, 7 Appendix A.6 and A.7	Simulation results with neutrality test no seed	<a href="https://cloud.biologie.ens.fr/index.php/s/38mqsVBWOGWPTIL">https://cloud.biologie.ens.fr/index.php/s/38mqsVBWOGWPTIL</a>
Inference simulations Appendix A.10	Simulations results with seed	<a href="https://cloud.biologie.ens.fr/index.php/s/FjtxEfA7Vv6nnj0">https://cloud.biologie.ens.fr/index.php/s/FjtxEfA7Vv6nnj0</a>
Inference results Appendix A.11	Predicted values for all empirical data sets	<a href="https://cloud.biologie.ens.fr/index.php/s/8Xw9lypmuQC0jo3">https://cloud.biologie.ens.fr/index.php/s/8Xw9lypmuQC0jo3</a>
Speed assessment Appendix A.3	Brut speed measures	<a href="https://cloud.biologie.ens.fr/index.php/s/GhbjtijNI1k10kT">https://cloud.biologie.ens.fr/index.php/s/GhbjtijNI1k10kT</a>
Implementation validity Appendix A.3	Simulation results no seed	<a href="https://cloud.biologie.ens.fr/index.php/s/A2eerLkbC0kzU1Y">https://cloud.biologie.ens.fr/index.php/s/A2eerLkbC0kzU1Y</a>
Number of species Test of gamma law Appendix A.5.1	Simulation results no seed	<a href="https://cloud.biologie.ens.fr/index.php/s/wp5ojYydaxOAXAm">https://cloud.biologie.ens.fr/index.php/s/wp5ojYydaxOAXAm</a>
Number of species Test of matrix Appendix A.5.2	Simulation results with seed	<a href="https://cloud.biologie.ens.fr/index.php/s/eBRGHR00voGNpyv">https://cloud.biologie.ens.fr/index.php/s/eBRGHR00voGNpyv</a>

Table 5: **Links to the presented results.** Seeds to reproduce the results are available when possible

## A.3 Speed improvement of the model

A pairwise comparison of trait values costs more computation time than a single comparison to the mean, and the simulations were around 4 times slower than with the mean competition model. And yet, the computation time is a critical issue in a context of large simulation experiments. This necessitated code optimization (around 6 times faster after optimisation for the pairwise competition model).

### A.3.1 Speed assessment

To compare all the models, 50 simulations were run for 200 generations for each model with all parameters being otherwise equal and chosen to be close to the values inferred from empirical data by the MESS model (21). To increase the speed of the simulations, I changed

the way the Local Community was managed (from lists to arrays), obtaining the following results : (values are the mean of 50 simulations, +/- standard deviation, in seconds)

Model	Without arrays	With arrays	Ratio	Continuous model
Neutral	$9 \pm 2$	$24 \pm 3$	0.375	$177 \pm 68$
Mean competition	$456 \pm 114$	$89 \pm 11$	5.14	<i>no runs</i>
Pairwise competition	$1944 \pm 258$	$323 \pm 40$	6.02	<i>no runs</i>
Environmental filtering	$80 \pm 18$	$70 \pm 10$	1.14	$141 \pm 49$
Interaction matrix	<i>no runs</i>	$419 \pm 54$	<i>no runs</i>	$440 \pm 93$

Table 6: **Comparison of the speed to the code for different version** (mean value per run, in seconds)

Noticeably, this greatly increased the speed of both competition model thus making further modelling and testing of MESS possible, even if the speed of the neutral mode was on contrary increased (due to the management of arrays not useful in this case). The value for the interaction matrix is here to give a point of comparison. This method was also used to test the continuous community assembly model, as to see whether it was usable.

### A.3.2 Validity of the new implementation

To prove that the change of the code structure does not affect the behavior of the program, experiments were made for both version of the software, for a wide range of parameters (for a total of 160 simulations per model per version). The resulting summary statistics were compared using t-test (with Bonferroni correction for multiple testing). The results are unequivocal (Fig. 11).

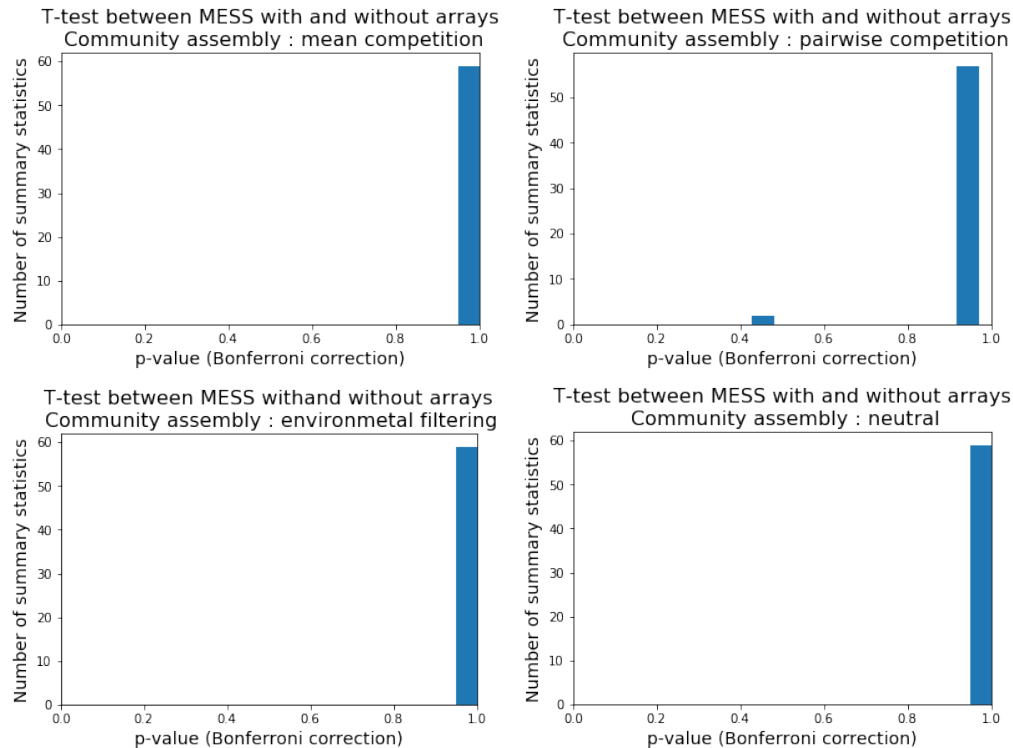


Figure 11: **T-test to compare MESS implementations** (with and without arrays)

This method was also used to compare the interaction matrix with the pairwise competition without the matrix: if it is filled with -1, the results are supposed to be identical to the pairwise competition model without the interaction matrix (Fig. 12)

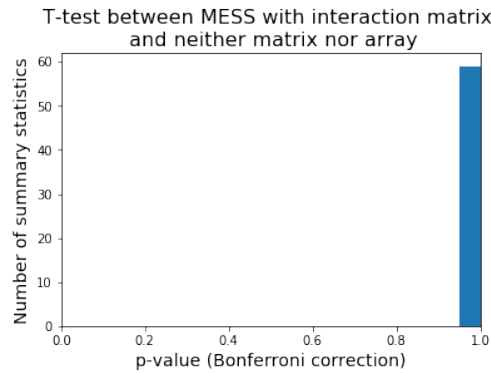


Figure 12: **T-test to compare MESS implementations** (with and without interaction matrix)

Finally, this method also enabled to compared the continuous model selection with previous implementations : giving 100% in one category is expected to produce the same results as the discrete community assembly model (Fig. 13).

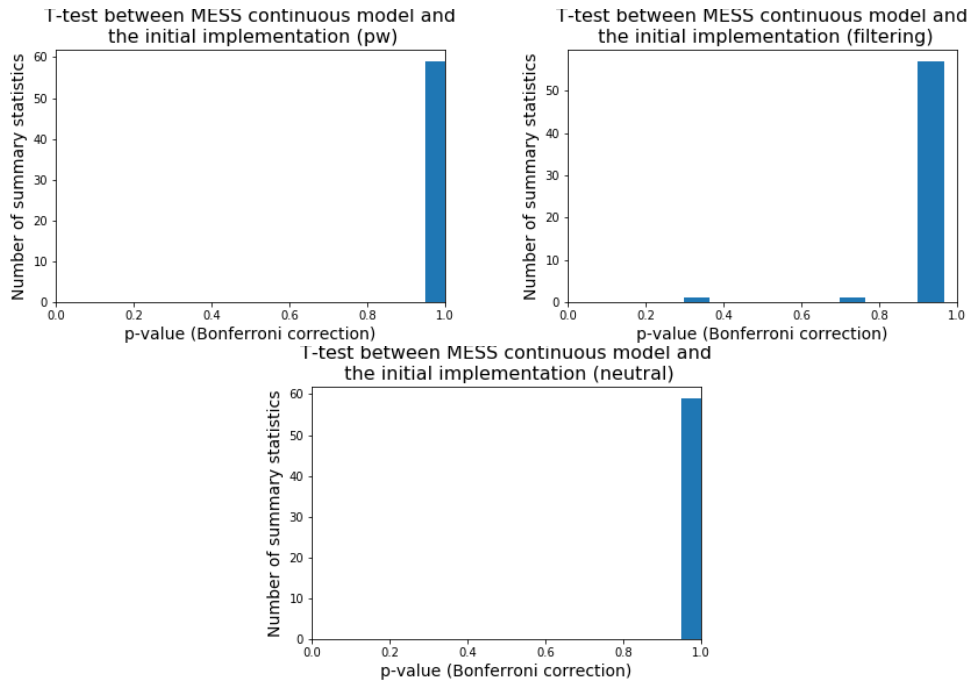


Figure 13: **T-test to compare MESS implementations** (with and without continuous community assembly model selection)

In order to later enable better housekeeping of the code and to provide not only replicable but also repeatable results, I added the possibility to specify a seed at the beginning of the simulation, and to record this seed.

## A.4 Death probabilities per species in later stages of the simulations

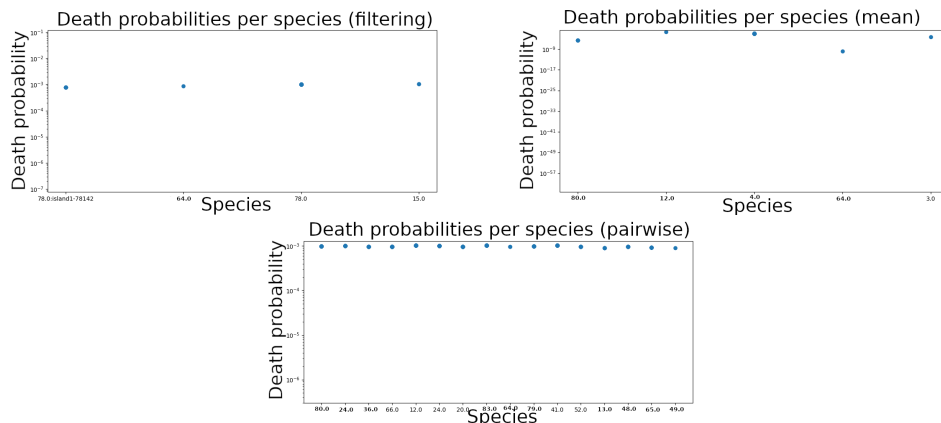


Figure 14: **Death probabilities per species for the 3 non-neutral community assembly models**, after 160 generations.

## A.5 Constraining the interaction matrix as to avoid pathological behaviors

### A.5.1 Exploration of the gamma law distribution for the terms of the interaction matrix

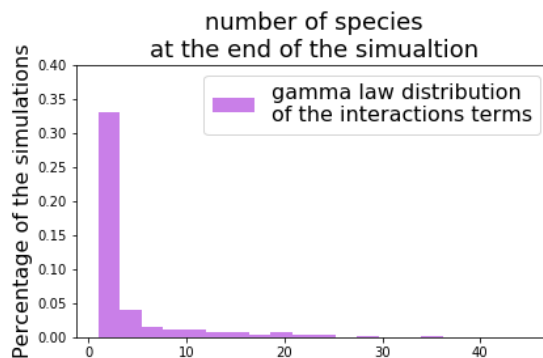


Figure 15: **Number of species after over 400 generation**, with the terms of the interaction matrix being drawn following a broad range of possible gamma laws (variance and mean between 0.01 and 1), for around 360 simulations.

The distribution of the number of species does not correspond to what we expect to observe in nature: in most cases, a single species invades the local community, and the distribution does not look normal as in the other community assembly models.

### A.5.2 Density dependence and relative strength of intra- and interspecific competition

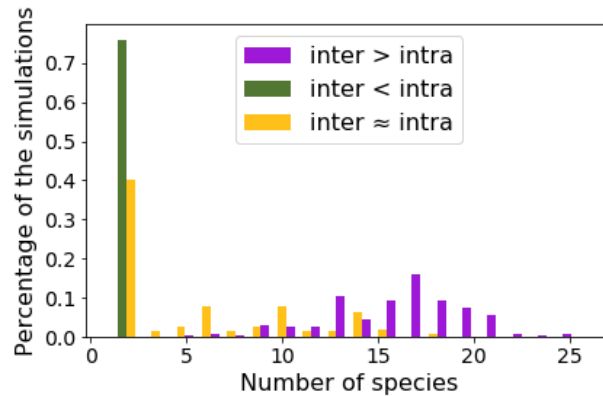


Figure 16: **Number of species after 500 generations** for various ranges of inter and intraspecific competition terms, in the pairwise competition case, for 892 simulations. The comparison for the strength of competition are done in absolute values.

Many combination resulted in pathological behaviors, the local community being totally saturated with a single species. This shows the importance of density dependency in the ecological communities: there is usually an important cost at being a common species.

### A.6 Neutrality tests for the neutral simulations

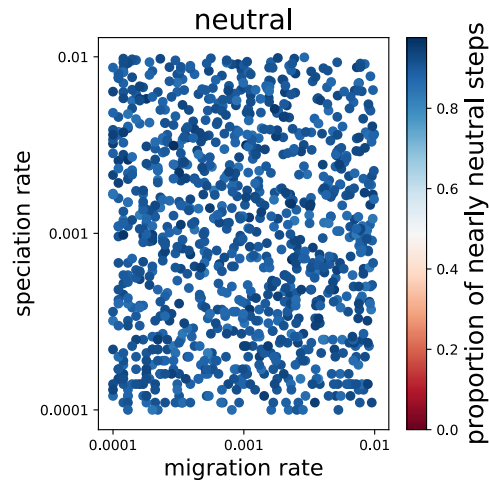


Figure 17: **Assessment of the neutral death step test (Pearson's Chi squared test) for the neutral simulations**, for various migration and speciation rates.

The analyzed steps are predicted neutral, which corresponds to what is expected and validate our test.

## A.7 Neutrality tests through time

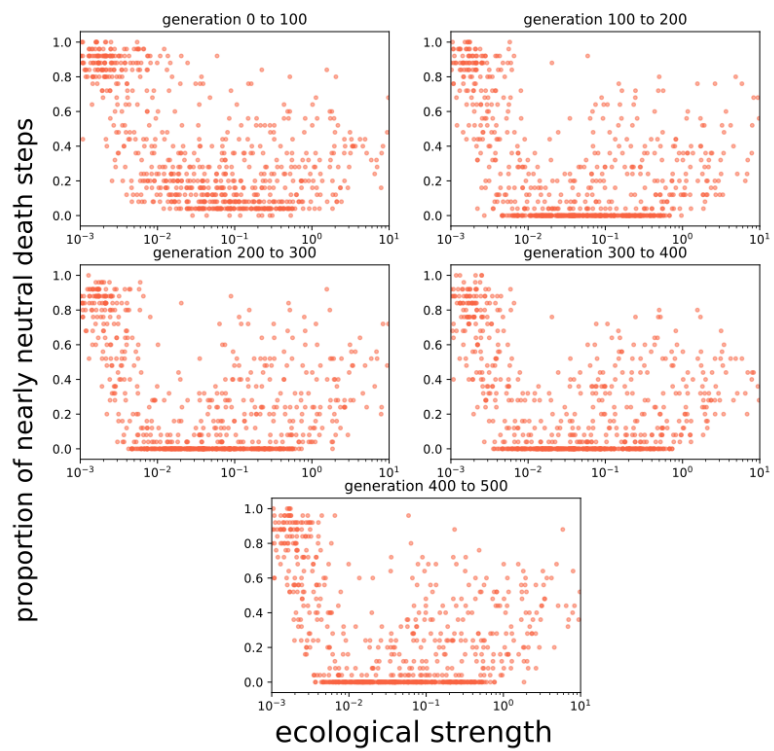


Figure 18: **Evolution of the nearly neutral status across time for the mean competition model** (same simulation set as in Fig 6)

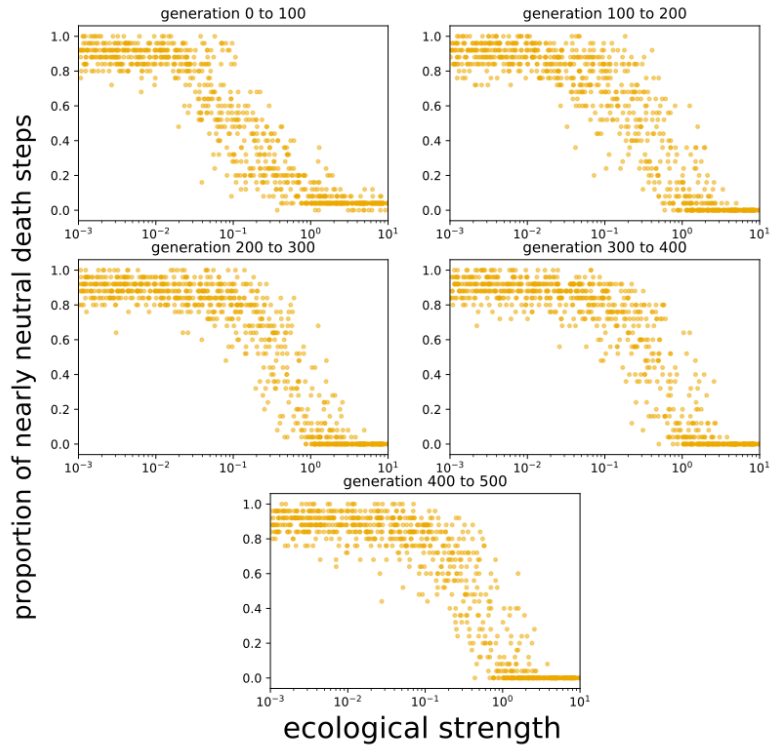


Figure 19: **Evolution of the nearly neutral status across time for the pairwise competition model** (same simulation set as in Fig 6)

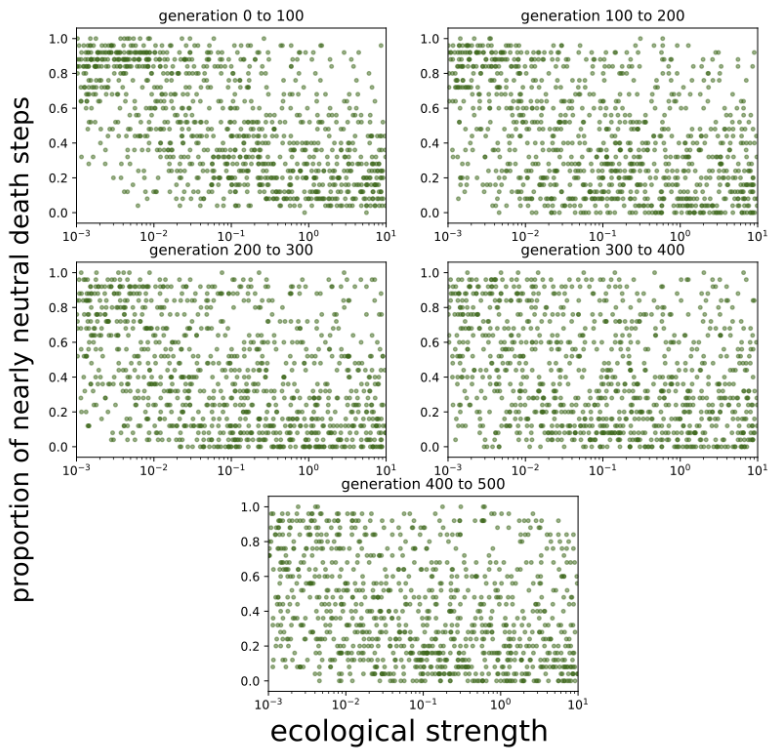


Figure 20: **Evolution of the nearly neutral status across time for the environmental filtering model** (same simulation set as in Fig 6)



## A.8 Trait distribution exploration

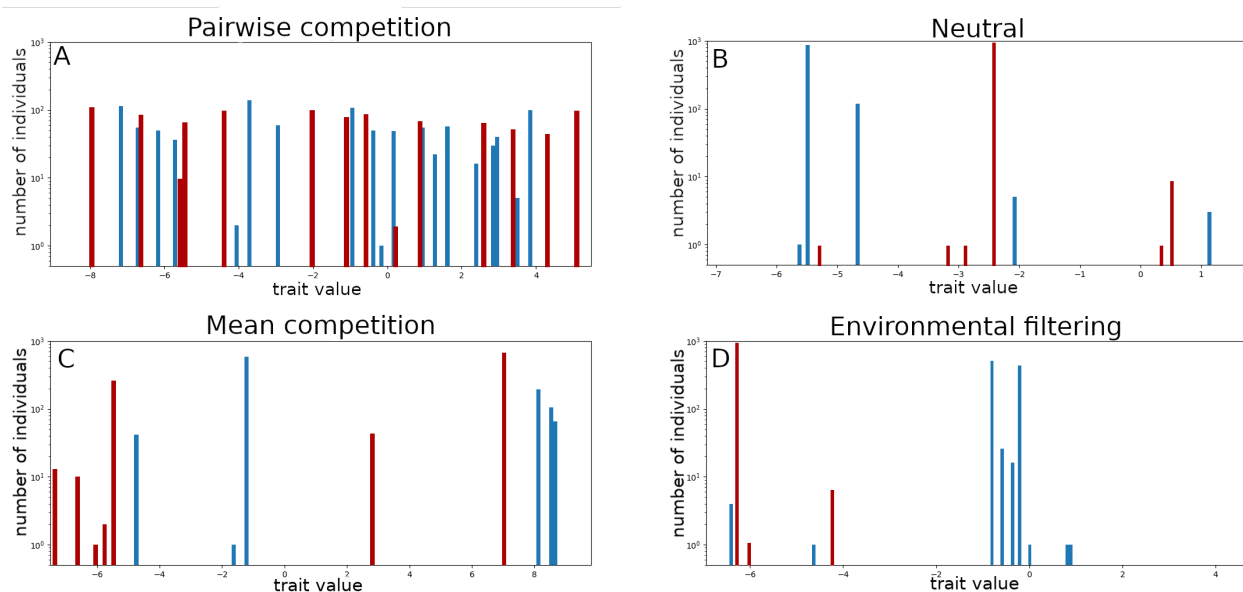


Figure 21: **Typical trait values distribution** for the 4 community assembly models studied. Two examples (red and blue) are given for each model.

This corresponds to our hypothesis on the functioning of the different models of section 2.4. As a matter of fact, two groups of species are distancing themselves in the mean competition model (C), while the species are much more grouped together in the environmental filtering case (D) and evenly distributed in the pairwise competition model (A). In the neutral case, they are random and their abundances follow a typical log-normal distribution. This also shows that we can expect significantly different results in the summary statistics resulting from trait data, but also in the species abundances and their variation and thus in the phylogeny.

## A.9 Parameters range for the inference method

Parameter	Range	Type of sampling
$\Lambda$	0 – 1	uniform
Community assembly model (discrete)	["neutral", "filtering", "pairwise_competition"]	uniform
Community assembly model (continuous)	(0 – 1; 0 – 1) coordinates on a simplex	uniform
$\alpha$	1000 – 10000	uniform
$s_E$	0.01 – 10	loguniform
intraspecific competition	0.01 – 1	loguniform
interspecific competition	0.01 – 1	loguniform
$J$	1000 – 5000	uniform
$m$	0.001 – 0.01	loguniform
$\nu$	0.0005 – 0.005	loguniform

Table 7: **Parameter ranges used for the inferences** (based on OVERCAST et al. 2020 (21))

## A.10 Test of the inference method

### A.10.1 Classification

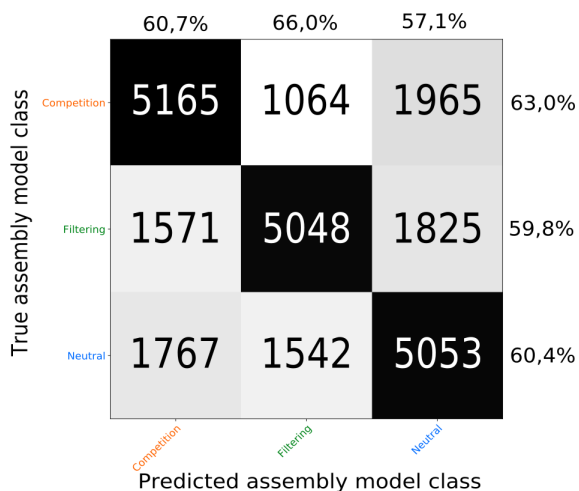


Figure 22: **Confusion matrix**, in the case of the pairwise competition (intraspecific competition = interspecific competition), for a train set of 25 000 simulations (5 -fold cross-validation). Simulations available here: <https://tiny.tobast.fr/sUs5h-Gu>

On average, the second (right) model performs slightly better.

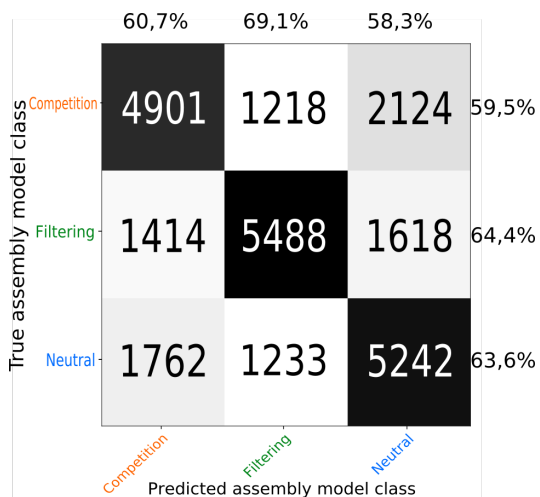


Figure 23: **Confusion matrix**, in the case of the pairwise competition (intraspecific competition > interspecific competition), for a train set of 25 000 simulations (5 -fold cross-validation). Simulations available here: <https://tiny.tobast.fr/kAfAMtg2>

	mean good classification rate	median good classification rate
pairwise competition	61.2 %	60.55 %
<b>interaction matrix</b>	<b>62.6 %</b>	<b>62.15 %</b>

Table 8: **Performance comparison** for the pairwise competition and interaction matrix inferences.

### A.10.2 Regression

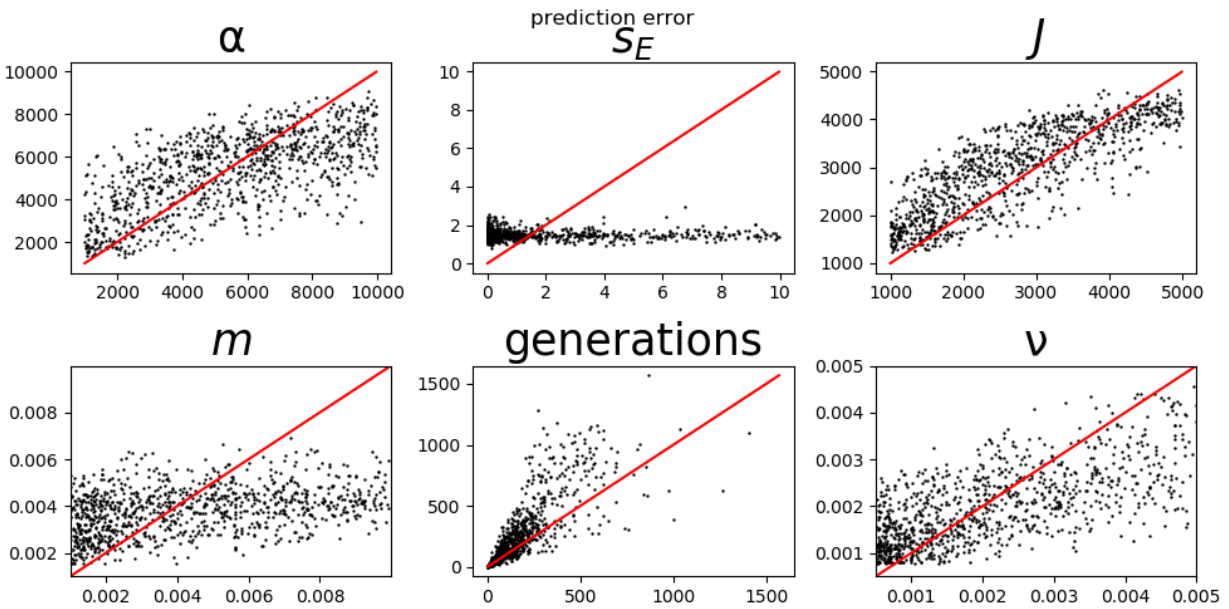


Figure 24: **Predicted against true value for the regressor**, in the case of the pairwise competition (intraspecific competition = interspecific competition), for the **neutral** community assembly. Same dataset as the confusion matrix.

### A.10.3 Test in the interaction matrix case

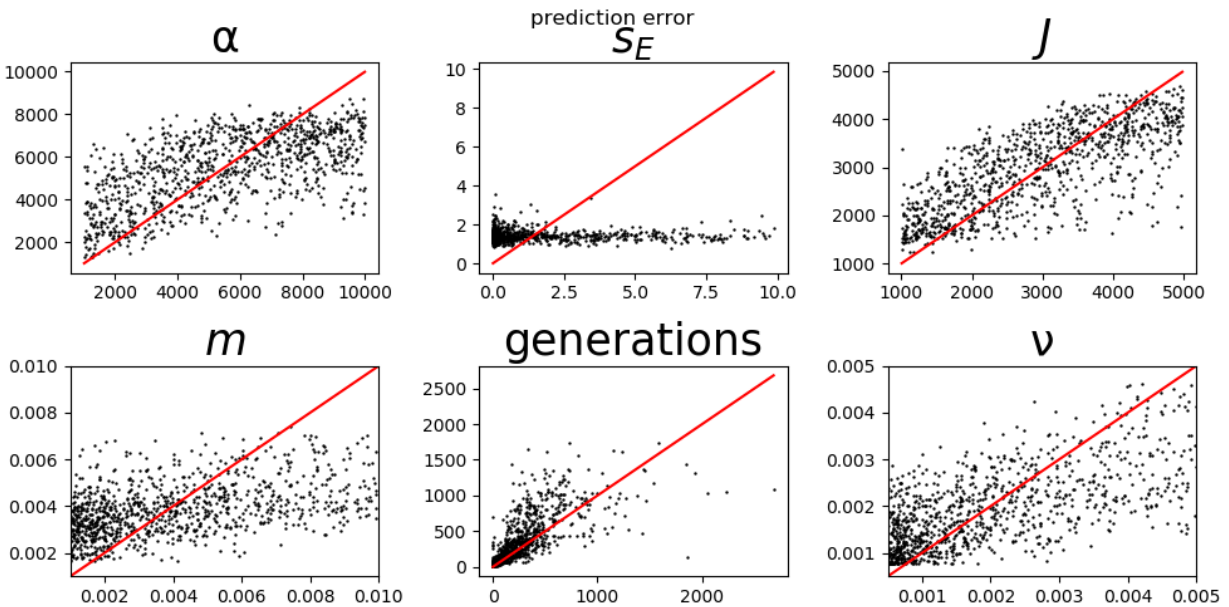


Figure 25: **Predicted against true value for the regressor**, in the case of the interaction matrix (intraspecific competition > interspecific competition), for the **filtering** community assembly. Same dataset as the confusion matrix.

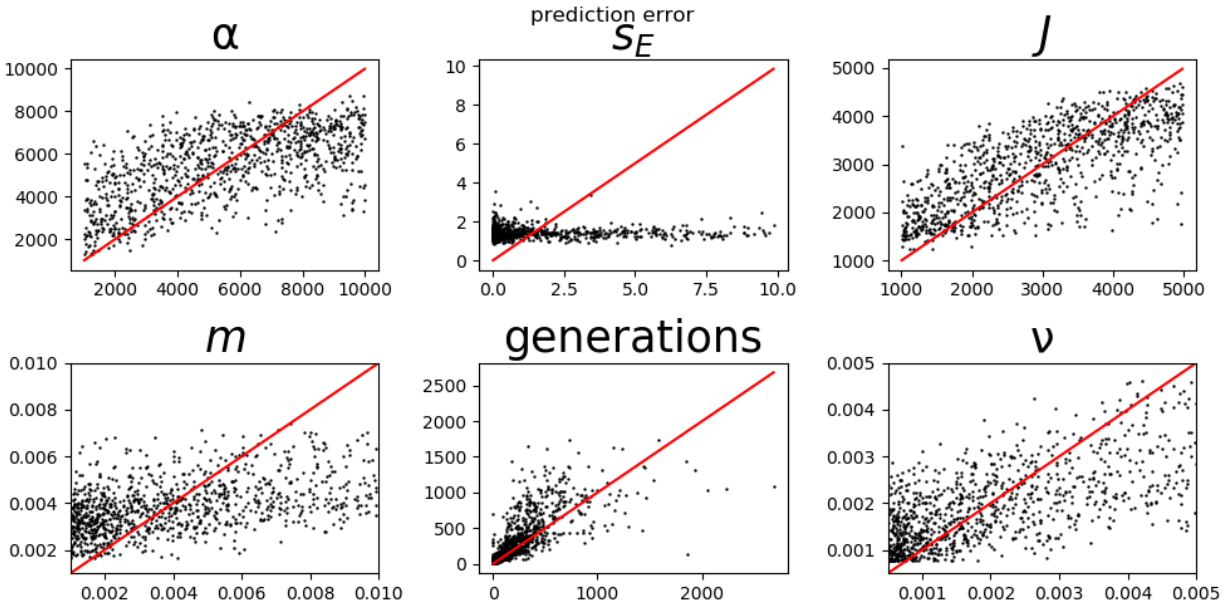


Figure 26: **Predicted against true value for the regressor**, in the case of the interaction matrix (intraspecific competition  $>$  interspecific competition), for the **neutral** community assembly. Same dataset as the confusion matrix.

#### A.10.4 Test in the continuous case

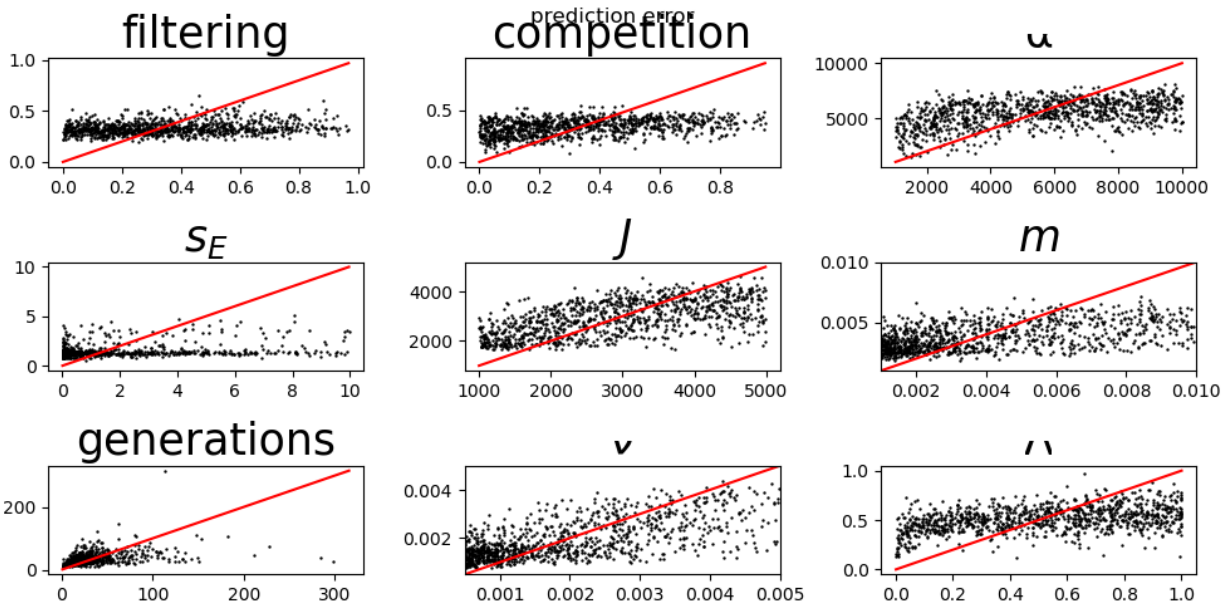


Figure 27: **Predicted VS. true value for the continuous model regressor**, trained on a set of 50607 simulations. Simulation results available here : <https://cloud.biologie.ens.fr/index.php/s/vQQ26TIPKJtGuYM>

### A.11 Full results of the inferences

	alpha	s_E	J	m	generation	nu	Lambda
Reunion spiders	intra=inter	7.94E+03	3.94E+03	4.57E-03	8.46E+02	8.12E-04	8.62E-01
	min	4.51E+03	2.63E+03	1.13E-03	1.03E+03	5.10E-04	7.86E-01
	max	9.90E+03	4.93E+03	9.72E-03	9.63E+02	1.56E-03	9.97E-01
	reduction	59.90%	57.48%	93.32%		48.46%	21.08%
	err	67.87%	58.37%	39.89%		15.68%	24.44%
neutral	intra>inter	8.40E+03	3.78E+03	4.79E-03	8.46E+02	8.53E-04	8.84E-01
	min	5.01E+03	2.25E+03	1.23E-03	1.07E+03	5.10E-04	7.90E-01
	max	9.79E+03	4.82E+03	9.24E-03	9.58E+02	1.79E-03	9.95E-01
	reduction	53.04%	64.42%	87.53%		54.53%	20.46%
	err	56.84%	68.25%	37.74%		17.77%	23.14%
Galapagos snails	cont	7.17E+03	3.77E+03	3.68E-03	2.55E+02	1.51E-03	8.66E-01
	min	3.18E+03	2.03E+03	1.08E-03	1.50E+02	5.20E-04	8.04E-01
	max	9.83E+03	4.87E+03	9.61E-03	7.96E+02	3.88E-03	9.95E-01
	reduction	73.81%	70.96%	94.93%		87.28%	19.13%
	err	92.64%	75.33%	39.00%		30.94%	22.10%
neutral	<b>intra=inter</b>	8.38E+03	1.57E+03	2.02E-03	1.22E+03	7.95E-04	6.94E-01
	min	4.93E+03	1.01E+03	1.00E-03	1.13E+03	5.00E-04	1.20E-01
	max	9.94E+03	2.55E+03	5.04E-03	8.77E+02	1.60E-03	9.78E-01
	reduction	55.64%	38.33%	70.26%		50.51%	85.84%
	err	59.78%	97.43%	26.07%		16.30%	123.69%
filtering	intra>inter	6.74E+03	2.47E+03	2.68E-03	5.46E+02	1.03E-03	8.35E-01
	min	2.31E+03	1.07E+03	1.03E-03	1.20E+02	5.10E-04	2.30E-01
	max	9.87E+03	4.89E+03	8.29E-03	9.30E+01	2.57E-03	1.00E+00
	reduction	84.07%	95.49%	90.57%		70.24%	76.98%
	err	112.27%	154.70%	35.21%		23.51%	92.14%

	alpha	s_E	J	m	generation	nu	Lambda
	cont	7.60E+03	3.64E+00	2.47E+03	2.94E-03	5.03E+02	9.16E-04
	min	4.07E+03	1.67E-02	1.06E+03	1.02E-03	1.51E+02	5.10E-04
	max	9.87E+03	9.49E+00	4.86E+03	9.27E-03	4.92E+02	2.62E-03
	reduction	64.45%	91.85%	95.02%	95.85%		70.99%
	err	76.36%	491.43%	153.86%	37.86%		23.37%
Mascarene weevils	intra=inter	8.26E+03		2.27E+03	4.01E-03	1.67E+03	7.54E-04
Mauritius	min	4.69E+03		1.11E+03	1.08E-03	1.03E+03	5.10E-04
neutral	max	9.91E+03		4.20E+03	9.28E-03	8.93E+02	1.51E-03
	reduction	58.01%		77.26%	93.41%		47.14%
	err	63.20%		135.89%	38.97%		15.10%
	intra>inter	8.61E+03		2.53E+03	3.18E-03	1.37E+03	7.59E-04
	min	6.07E+03		1.33E+03	1.04E-03	1.02E+03	5.00E-04
	max	9.90E+03		3.89E+03	8.43E-03	9.37E+02	1.40E-03
	reduction	42.60%		64.03%	90.86%		44.72%
	err	44.53%		101.36%	36.38%		14.33%
	cont	7.30E+03	8.02E-01	2.07E+03	2.54E-03	5.12E+02	7.87E-04
	min	3.03E+03	1.10E-02	1.04E+03	1.02E-03	1.50E+02	5.10E-04
	max	9.84E+03	7.56E+00	4.45E+03	8.20E-03	4.95E+02	1.54E-03
	reduction	75.71%	94.54%	85.25%	90.52%		48.00%
	err	93.34%	2952.83%	164.41%	34.87%		15.46%
Mascarene weevils	intra=inter	7.98E+03		2.36E+03	3.51E-03	1.24E+03	7.43E-04
Reunion	min	4.74E+03		1.30E+03	1.06E-03	1.01E+03	5.10E-04
neutral	max	9.96E+03		4.19E+03	8.71E-03	9.56E+02	1.51E-03
	reduction	58.03%		72.27%	91.49%		47.14%
	err	65.47%		122.26%	37.27%		15.07%

	alpha	s_E	J	m	generation	nu	Lambda
intra>inter	8.75E+03		2.59E+03	3.23E-03	1.27E+03	7.58E-04	8.52E-01
min	6.31E+03		1.34E+03	1.06E-03	1.04E+03	5.00E-04	7.38E-01
max	9.95E+03		3.74E+03	8.01E-03	9.36E+02	1.40E-03	9.96E-01
reduction	40.47%		60.21%	87.86%		44.72%	25.84%
err	41.63%		93.00%	35.28%		14.33%	30.31%
cont	7.98E+03	1.02E+00	2.39E+03	2.95E-03	5.05E+02	7.81E-04	8.64E-01
min	4.07E+03	1.10E-02	1.04E+03	1.03E-03	1.43E+02	5.10E-04	7.66E-01
max	9.90E+03	7.88E+00	4.45E+03	9.44E-03	4.86E+02	1.54E-03	9.88E-01
reduction	64.83%	95.21%	85.34%	96.22%		48.00%	22.25%
err	73.10%	38017.93%	143.08%	38.02%		15.45%	25.74%
Australian trees	3.72E+03	3.81E+00	3.89E+03	4.24E-03	2.14E+01	3.14E-03	6.45E-01
min	1.09E+03	1.68E-02	2.07E+03	1.11E-03	1.00E+01	8.00E-04	9.86E-02
max	9.14E+03	9.25E+00	4.96E+03	9.61E-03	1.00E+01	4.82E-03	9.98E-01
reduction	89.47%	91.39%	72.26%	93.74%		78.00%	89.98%
err	216.28%	472.11%	74.30%	39.51%		31.16%	139.44%
intra>inter	3.93E+03	5.06E+00	3.59E+03	6.12E-03	2.14E+01	3.66E-03	6.55E-01
min	1.09E+03	5.69E-02	1.96E+03	1.24E-03	1.00E+01	1.56E-03	6.12E-02
max	9.42E+03	9.81E+00	4.93E+03	9.78E-03	1.00E+01	4.88E-03	1.00E+00
reduction	92.55%	74.54%	74.06%	89.71%		49.49%	93.88%
err	212.17%	317.44%	82.56%	40.53%		20.31%	143.31%
cont							
min							
max							
reduction	0.00%	Err:502	0.00%	Err:502		Err:502	0.00%
err	#DIV/0!	Err:502	#DIV/0!	Err:502		Err:502	#DIV/0!

	alpha	s_E	J	m	generation	nu	Lambda
Australian trees	intra=inter	3.86E+03	3.62E+00	3.80E+03	4.45E-03	2.13E+01	3.24E-03
	min	1.06E+03	2.10E-02	2.01E+03	1.11E-03	2.00E+00	9.10E-04
	max	9.26E+03	9.03E+00	4.96E+03	9.58E-03	9.00E+00	4.88E-03
	reduction	91.14%	87.78%	73.88%	93.60%		72.94%
	err	212.70%	471.12%	77.79%	39.80%		179.31%
	intra>inter	4.96E+03	5.05E+00	3.61E+03	6.44E-03	1.76E+01	3.12E-03
	min	1.19E+03	5.68E-02	1.98E+03	1.36E-03	2.00E+00	8.67E-04
	max	9.44E+03	9.78E+00	4.94E+03	9.88E-03	9.00E+00	4.87E-03
	reduction	91.71%	74.54%	73.89%	86.09%		74.96%
	err	166.49%	318.11%	81.92%	39.29%		161.75%
	cont	5.00E+03	4.32E+00	3.24E+03	4.02E-03	1.86E+01	2.88E-03
	min	1.29E+03	2.11E-02	1.49E+03	1.10E-03	1.00E+01	7.00E-04
	max	9.68E+03	9.57E+00	4.92E+03	9.32E-03	9.00E+00	4.85E-03
	reduction	93.25%	88.54%	85.58%	92.82%		84.05%
	err	168.01%	417.69%	105.71%	38.75%		177.52%
Australian trees	intra=inter	3.55E+03	3.80E+00	3.82E+03	4.70E-03	1.42E+01	3.31E-03
	min	1.07E+03	2.12E-02	2.01E+03	1.11E-03	1.00E+00	9.07E-04
	max	9.15E+03	9.06E+00	4.97E+03	9.80E-03	7.00E+00	4.93E-03
	reduction	89.71%	87.70%	74.06%	94.60%		73.52%
	err	227.51%	454.11%	77.48%	40.64%		188.06%
	intra>inter	4.06E+03	5.05E+00	3.65E+03	6.45E-03	1.65E+01	3.77E-03
	min	1.12E+03	5.48E-02	2.05E+03	1.29E-03	1.00E+00	1.58E-03
	max	9.11E+03	9.82E+00	4.92E+03	9.88E-03	9.00E+00	4.88E-03
	reduction	88.76%	75.11%	71.73%	88.44%		49.11%
							98.27%



	alpha	s_E	J	m	generation	nu	Lambda
err	196.88%	320.28%	78.61%	40.38%		20.26%	168.65%
cont							
min							
max							
filtering							
reduction	0.00%	Err:502	0.00%	Err:502		Err:502	0.00%
err	#DIV/0!	Err:502	#DIV/0!	Err:502		Err:502	#DIV/0!

Table 9: Results for all variables

## A.12 Posterior predictive checks

In order to assess the quality of the predicted value, we run 50 simulations using the estimated parameters and plot the results in the space of the 2 first principal components. The empirical data are in red and the simulations in blue.

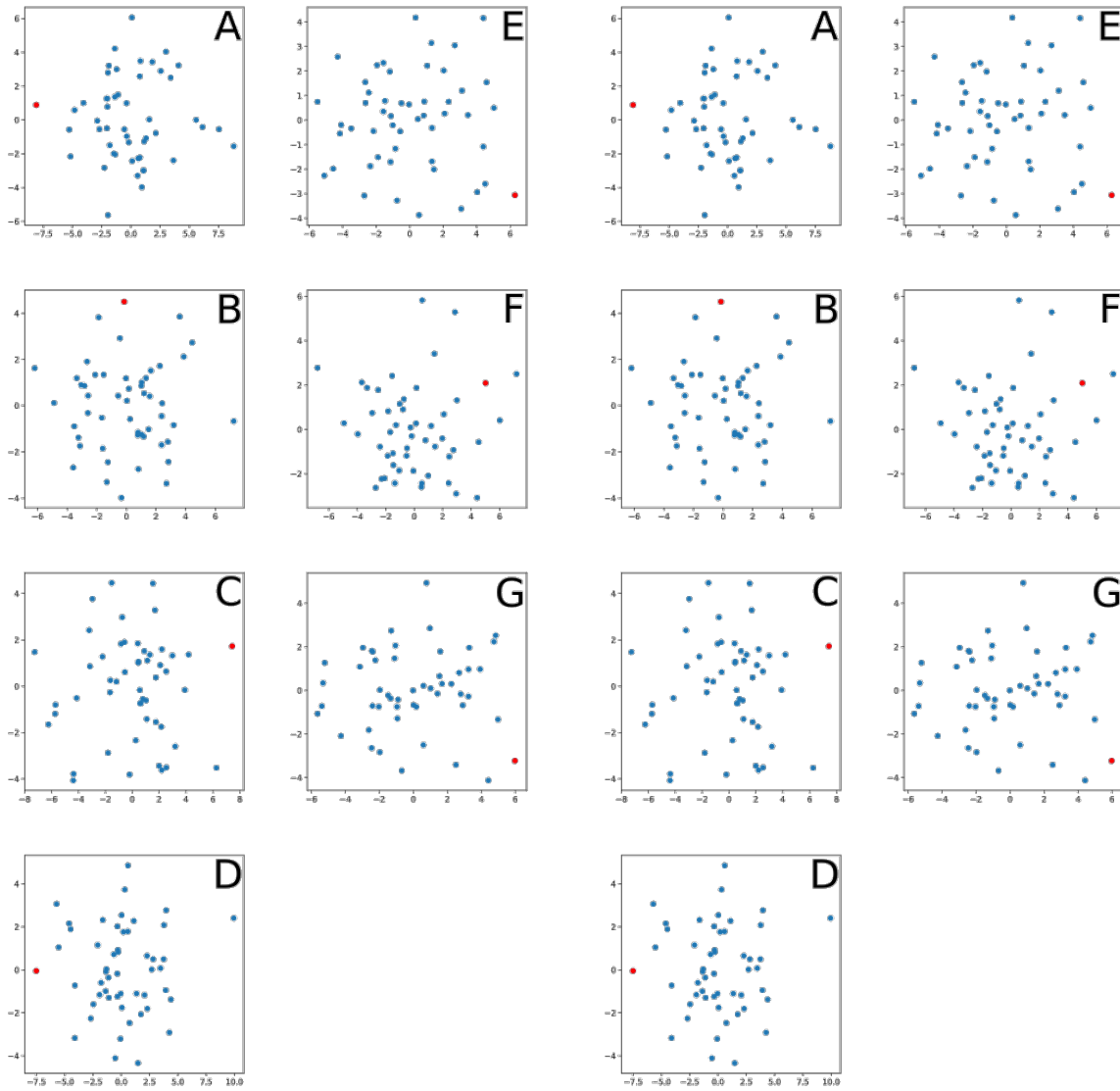


Figure 28: **Posterior predictive check in the pairwise competition case.** The data sets are A: Reunion spiders, B: Galapagos snails, C: Mauritius weevils, D: Reunion weevils, E: Dorrigo trees, F: Nightcap trees, G: Washpool trees

Figure 29: **Posterior predictive check in the interaction matrix case.** The data sets are A: Reunion spiders, B: Galapagos snails, C: Mauritius weevils, D: Reunion weevils, E: Dorrigo trees, F: Nightcap trees, G: Washpool trees

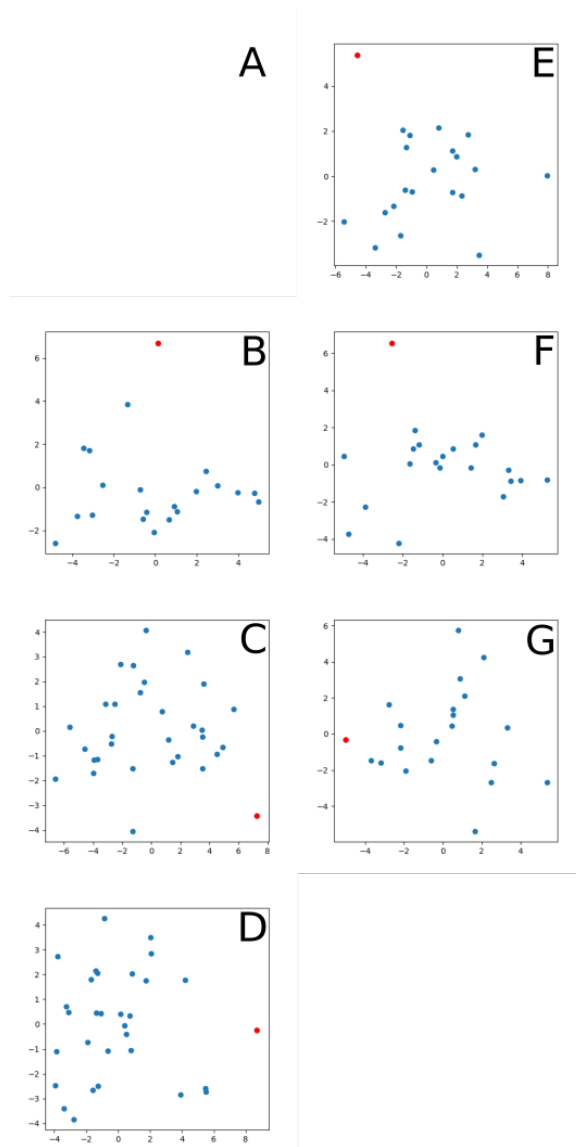


Figure 30: **Posterior predictive check in the continuous model case.** The data sets are A: Reunion spiders, B: Galapagos snails, C: Mauritius weevils, D: Reunion weevils, E: Dorrigo trees, F: Nightcap trees, G: Washpool trees