

# SCIENTIFIC REPORTS



OPEN

## Coffee tree architecture and its interactions with microclimates drive the dynamics of coffee berry disease in coffee trees

Natacha Motisi<sup>1,2</sup>, Fabienne Ribeyre<sup>1,2</sup> & Sylvain Poggi<sup>3</sup>

Coffee berry disease (CBD), which is widespread in Africa, has been responsible for massive yield losses of *Coffea arabica*. In Cameroon, *C. arabica* is mainly planted in agroforestry systems on smallholder farms, where low incomes hinder the use of chemicals to manage the disease. Novel agroecological strategies for controlling CBD are expected to be implemented and even increase in the current context of global changes. In this study, we showed that coffee tree architecture and its interactions with microclimates were important to CBD cluster symptom appearance (CSA), with notable CSA increasing along the tree branch away from the trunk to the tip of the branch. As shade trees can modify microclimates, we further investigated scenarios of various microclimatic conditions under shade to explore the effects of agroforestry systems on CBD dynamics in coffee trees. We showed that shade could result in contrasting effects on disease dynamics, decreasing CSA along the branch and increasing epidemic duration. We suggest that the contrasting effects of shade on disease dynamics need further evaluation of the possible trade-offs among the variables at play, and we recommend a combination of epidemiological and architectural modelling to help design more cost-effective and environmentally friendly CBD management strategies.

Coffee berry disease (CBD), caused by the fungal pathogen *Colletotrichum kahawae*, is widespread in Africa and, depending on climatic conditions, has been responsible for massive coffee yield losses, with up to 90% loss recorded<sup>1</sup>. In many African countries, *C. arabica* is mainly cropped on smallholder farms, where low incomes hinder the use of chemicals to manage the disease. Novel agroecological strategies for controlling CBD are expected to be implemented and even increase in number due to the current context of global changes.

Coffee tree architecture may be a possible lever for managing CBD. Mouen *et al.* suggest that coffee tree architecture, in terms of branch and leaf arrangement on a tree, can influence the distribution of CBD on a tree. In fact, the authors found homogeneous disease incidence in trees showing umbrella architecture (where almost all the branches of a tree are in the same stories, preferentially the top of the tree), while they found an ascending gradient of disease incidence from the top to the base of coffee trees showing pyramidal architecture (where the branches and the stories are well differentiated from the top of a tree to its base)<sup>2</sup>. As *C. kahawae* mainly disperses by rain splash<sup>3,4</sup>, Mouen *et al.* suggest that conidia transported by gravity and raindrops would be intercepted by the branches in the canopy. However, Calonnec *et al.*, who studied powdery mildew on grapevines<sup>5</sup>, argue that architecture is an important variable affecting disease spread in terms of “availability” of susceptible hosts (number, location and susceptibility of hosts). In the case of coffee trees, Vaast *et al.* show that pruning coffee berries increases berry maturation rate<sup>6</sup>. Nonetheless, susceptibility to CBD is assumed to be negatively and nonlinearly linked to berry maturation rate<sup>2,4,7–9</sup>. Consequently, manipulating berry distribution within the tree architecture, hereafter referred to as the arrangement of berry clusters (ensembles of berries borne by a branch node), may affect the “availability” of susceptible hosts by altering (i) distances (locations) between infected and susceptible berries, (ii) the fruit load (number) of susceptible berries, and potentially (iii) the period of berry susceptibility by modifying the maturation rate. Tree architecture is thus likely to influence the movement of inoculum within the tree, potentially providing control of the disease.

<sup>1</sup>CIRAD, UPR Bioagresseurs, F-34398, Montpellier, France. <sup>2</sup>Bioagresseurs, Univ Montpellier, CIRAD, Montpellier, France. <sup>3</sup>IGEPP, INRA, AgroCampus Ouest, Université Rennes 1, F-35653, Le Rheu, France. Correspondence and requests for materials should be addressed to N.M. (email: [natacha.motisi@cirad.fr](mailto:natacha.motisi@cirad.fr))



**Figure 1.** Coffee trees and coffee berry disease characteristics. **(a)** A coffee tree showing an intermediate pyramidal architecture (where branches and stories are well differentiated from the top of a tree to its base), with an umbrella architecture (where branches are almost in the same stories in a tree). **(b)** Berry clusters distributed along a branch. **(c)** Symptoms of coffee berry disease caused by *Colletotrichum kahawae*. All photographs were taken by the author Natacha Motisi.

Modifications to microclimates to hinder *C. kahawae* spread and infection are another possible lever for managing CBD. Nonetheless, knowledge on the relationships between the epidemiology of CBD and microclimates is scarce. The primary inoculum is supposed to remain dormant during the dry season in flower buds, branch bark or mummified berries left on the branch<sup>7,10</sup>; however, the dormant form (either spores or mycelium) has not formally been identified or localized within a tree<sup>9</sup>. Primary infections occur at the beginning of fruit expansion (i.e., four to ten weeks after flowering), supposedly favoured by conditions of relatively low temperatures (15–25 °C) and high relative humidity (>98%) that are prevalent during the rainy season<sup>3,9,11</sup>. The latent period (time between host infection and appearance of symptoms) is highly variable, ranging from six days (recorded under optimal controlled conditions)<sup>12,13</sup> to four weeks (recorded under outdoor field conditions)<sup>8</sup>. The secondary inoculum, conidia formed on the berry surface, is dispersed by rain splash, which is the main dispersal mode of *C. kahawae* on a tree (autoinfection), as well as between trees (alloinfection)<sup>3,4</sup>. Intensification of disease, caused by secondary infections, occurs during the green berry stage (between approximately 7 and 18 weeks after flowering), which is supposed to be the stage of highest susceptibility to CBD<sup>2,4,7–9</sup>.

In Cameroon, coffee trees are mainly cropped under shade trees, providing conditions similar to those of their original habitat. However, shade affects microclimates by buffering temperature<sup>14</sup> and likely increasing wetness<sup>15</sup>, which could be favourable to *C. kahawae* fruit penetration and infection. In addition, Vaast *et al.*<sup>6</sup> showed that a rather dense shade level of 45% delays coffee berry ripening by as much as one month, which could increase the period of berry susceptibility to *C. kahawae*. At the same time, numerous authors argue that agroforestry systems have the potential to control CBD<sup>4,16–20</sup>, notably because shade trees provide protection for coffee trees against rainfall, thereby decreasing pathogen dispersal via splashing<sup>4,19</sup>. Thus, the usefulness of growing coffee trees in agroforestry systems to control CBD is not clear. In particular, if shade can modify CBD propagation through the modification of microclimates, then more information is needed to understand the possible trade-offs between the effects of rainfall and microclimates under shade.

In this study, we investigated whether coffee tree architecture can influence CBD cluster symptom appearance (CSA, Fig. 1c) and analysed tree architecture effects relative to those of microclimates. For this analysis, we carried out a field experiment on a smallholder coffee farm in Cameroon over two years (2012 and 2013) on coffee trees cropped under the partial and non-permanent shade of banana trees, hereafter referred to as the banana shading system (Fig. 2a). In terms of the effect of tree architecture on CBD, we considered various covariates (Table 1): the branch story within a tree (top, middle or lower story of a coffee tree, Fig. 1a), the spatial configuration of berry clusters (Fig. 1b) resulting from prior manual pruning of clusters along the branch (cluster arrangement, Fig. 3), the position of each cluster on a branch starting from the trunk moving to the tip of the branch, the fruit load per cluster, and berry physiological age, which we assumed to be an indicator of berry susceptibility. We monitored



**Figure 2.** Banana shading system vs. kola shading system. (a) The coffee growing area in association with plantain banana trees where original data field were collected (banana shading system). (b) The forestry kola tree where the microclimatic data were collected over two years to investigate scenarios of modifications of coffee berry disease dynamics under agroforestry systems. All photographs were taken by the author Natacha Motisi.

a set of microclimatic variables, namely, precipitation, temperature and relative humidity, to define the microclimate covariates. Precipitation, defined as cumulative rainfall, was assumed to mediate spore dispersal. Suitable conditions for infection (SCI) were computed as the number of hours beyond 5 hours of suitable temperature (15–25 °C) and relative humidity (>98%). The CBD latent period can vary between one week and one month; thus, both precipitation and suitable conditions for infection were considered within three time windows, namely, 7 to 14 days (time window 1, W1), 15 to 21 days (time window 2, W2) and 22 to 29 days (time window 3, W3) before CSA. We applied a statistical method from the field of machine learning, boosted regression trees (BRT), which enables researchers to address questions relative to complex biological systems involving many variables and characterized by multiple interactions between processes. We used this approach to assess the relationship between CSA and the previously mentioned explanatory variables by fitting a model to the field data containing disease dynamics on coffee trees cropped in the banana shading system. We subsequently studied whether agroforestry systems can modify disease dynamics by investigating *in silico* the hypothesis that the microclimate induced by forestry tree can decrease CBD levels. For this purpose, we applied a new set of microclimatic variables (precipitation, temperature and relative humidity) measured over two years under a kola shade tree (Fig. 2b) present on the coffee farm, hereafter referred to as the kola shading system, to the previous fitted model used as the reference situation. Our results provide information on (i) coffee tree architecture management as part of agroecological strategies for controlling CBD and (ii) the manner in which agroforestry systems can affect CBD dynamics at the tree scale. We encourage the use of mechanistic modelling to help design novel, more cost-effective and environmentally friendly management strategies at both the tree scale and plot scale.

## Results

**Field data.** The field data contained disease dynamics on coffee trees cropped under the shade of banana trees. Disease dispersal was investigated over three different cluster arrangements obtained by manually eliminating 0, 25 and 50% of the clusters along all the branches of the coffee trees (Fig. 3). The patterns of CSA were similar over the two years and among the three categories of cluster removal rates, except for 25% removal in 2013 (Fig. 4). In fact, regardless of the year and cluster removal rate, CSA intensified between 10 and 20 weeks after flowering, and within this period, CSA increased moving away from the trunk along the branch to its tip.

**Relative influences and effects of the explanatory variables.** Using a gradient boosting regression, we fitted a model with good prediction performance (Annex 1), demonstrating the substantial role of tree architecture in symptom appearance and a strong interaction with microclimate.

The most important variables influencing CSA were (i) architectural covariates, namely, cluster fruit load, berry physiological age, and cluster position along the branch, and (ii) microclimate covariates, namely, the suitable conditions for infection (SCI) in terms of temperature and relative humidity in two time windows, SCI W3 (22 to 29 days before CSA) and SCI W2 (15 to 21 days before CSA) (Fig. 5).

Cluster fruit load was the strongest predictor of CSA, with a relative importance of 23.4%, and was characterized by a positive relationship with CSA, which reached a plateau of 25 berries per cluster (Fig. 6). Berry physiological age accounted for 14.2% of the total effect on CSA and was nonlinearly and negatively related to CSA. Microclimate, through SCI W3 and SCI W2, contributed 12.1% and 10.8%, respectively, to CSA, with a counter-intuitive decreasing relationship with CSA followed by a plateau for both time windows. Cluster position along the branch was the fifth most influential predictor (relative contribution of 10.7%), characterized by an increase in CSA along the branch moving away from the trunk.

Short name	Description	Type
CSA	Cluster symptom appearance (response variable)	Qualitative (binary)
Year	2-level factor: 2012, 2013	Qualitative
<i>Architectural covariates</i>		
Branch story	Branch position in the coffee tree: top, middle, lower stories	Qualitative
Cluster arrangement	Spatial configuration of berry clusters resulting from the removal of clusters along the branch, 0, 25 and 50% removal (Fig. 3)	Qualitative
Cluster position	Position of each cluster along the branch from the trunk (position 1) to the last possible position on the branch	Qualitative
Cluster fruit load	Number of berries per cluster	Quantitative
Physiological age	Sum of temperatures (°C days) from flowering assumed to be linked to berry susceptibility to coffee berry disease	Quantitative
<i>Microclimate covariates</i>		
Precipitation W1	Cumulative rainfall in time window 1 (W1), i.e., from 7 to 14 days before the observation of CSA	Quantitative
Precipitation W2	Cumulative rainfall in time window 2 (W2), i.e., from 15 to 21 days before the observation of CSA	Quantitative
Precipitation W3	Cumulative rainfall in time window 3 (W3), i.e., from 22 to 29 days before the observation of CSA	Quantitative
SCI W1	Number of hours beyond 5 hours of suitable conditions for infection (temperatures ranging from 15 to 25 °C and relative humidity > 98%) in time window 1 (W1), i.e., from 7 to 14 days before the observation of CSA	Quantitative
SCI W2	Number of hours beyond 5 hours of suitable conditions for infection (temperatures ranging from 15 to 25 °C and relative humidity > 98%) in time window 2 (W2), i.e., from 15 to 21 days before the observation of CSA	Quantitative
SCI W3	Number of hours beyond 5 hours of suitable conditions for infection (temperatures ranging from 15 to 25 °C and relative humidity > 98%) in time window 3 (W3), i.e., from 22 to 29 days before the observation of CSA	Quantitative

**Table 1.** Names and descriptions of the variables used for boosted regression trees (BRT) analyses.

The other variables showed smaller than expected influences due to chance (i.e., smaller than 8.3%, Fig. 5), namely, precipitation in time windows W1, W2 and W3; cluster arrangement along the branch (0%, 25% and 50% of clusters removal); branch story in the tree (top, middle and lower story); and SCI in W1 and year.

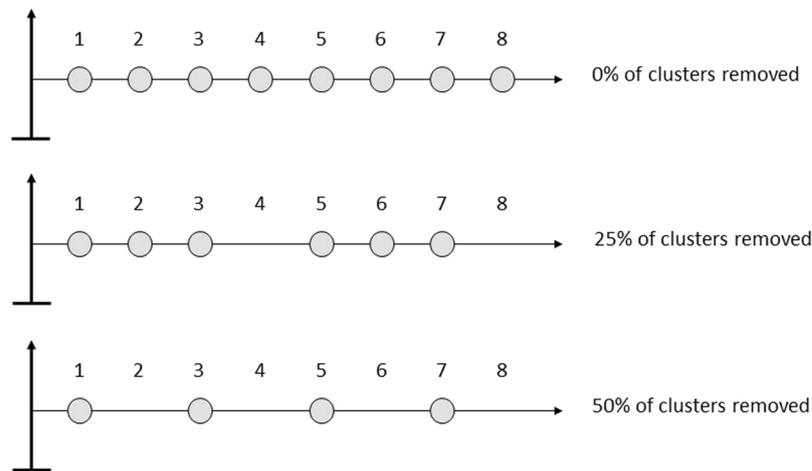
Interestingly, SCI W3 and cluster position were involved in interactions. The most important interaction involved SCI W3 and berry physiological age (interaction strength 0.26), and the second most important involved cluster position and SCI W1 (interaction strength 0.25).

***In silico* experiments.** In the *in silico* experiments, we used the model (Annex 1) fitted with the field data (banana shading system) to predict CSA under new microclimatic conditions, namely, precipitation, temperature and relative humidity, measured under the shade of a forestry kola tree present on the farm (kola shading system) for two years, 2012 and 2013. These two years constituted two scenarios of disease dynamics in agroforestry systems relative to the field data used as the reference situation. The two scenarios led to two contrasting situations (Fig. 7). In 2012, except for disease onset, which appeared to start earlier than in the reference situation, the microclimate induced by the kola shade did not appear to affect the disease course within a tree. However, in 2013, the microclimate induced by the kola shade decreased CSA, notably at the tip of the branch, but the microclimate delayed disease extinction relative to the reference situation.

## Discussion

For the first time, to the best of our knowledge, our findings formally establish that coffee tree architecture and its interactions with microclimates drive CBD dynamics in coffee trees.

We demonstrated that the fruit load per cluster, berry physiological age and cluster position along the branch were the main architectural variables influencing CSA (Fig. 5). Notably, our results showed that during the period of disease intensification (between 10 and 20 weeks after flowering), CSA is likely to increase along the branch moving away from the trunk to the tip of the branch (Figs 4 and 6). The relation between berry position along the branch and the infection of adjacent clusters has been previously observed and suggested to be of significance by Nutman and Roberts<sup>21</sup>; however, to our knowledge, the significance of the relation has never been statistically evidenced. Two epidemiological mechanisms could explain this process: (i) autoinfections within a tree, through contamination between neighbouring clusters, and (ii) late alloinfections between trees, through contamination from neighbouring trees. However, our results showed a weak contribution of cluster arrangement along the branch (obtained by elimination of clusters by manual pruning) to CSA, suggesting that contamination between neighbouring clusters was not significant. This result may indicate that the dispersal distance of conidia within a tree is lower than the internode distances encountered in the coffee trees in this study. If this hypothesis is correct, then the use of dwarf varieties of coffee trees with short internodes and strong aggregation of fruits could favour disease dispersal within a tree as suggested by Avelino<sup>22</sup>. Nevertheless, our results are consistent with previous work by Mouen *et al.*<sup>4</sup>, who showed that alloinfections through inoculum splashing from tree to tree induced higher disease incidences than did autoinfections within trees. Given the relevance of disease dispersal



**Figure 3.** Schematic description of cluster position (with an example of 8 clusters numbered 1 to 8 along the branch) and cluster arrangement with 0% (control trees), 25% and 50% of the clusters removed along the branch.

between trees, creating natural barriers between coffee trees could provide control of CBD spread. In particular, as smallholder farmers in Cameroon usually intercrop food-producing species with coffee trees, we recommend associating tall intercrop species with coffee trees. However, our results showed a weak influence of branch story in the tree on CSA (Fig. 5). These results suggest that the architecture of the trees used for this study, topped trees showing intermediate architecture between umbrella and pyramidal architectural types, behaved more as the umbrella types described by Mouen *et al.*<sup>2</sup>. It would be interesting to compare our results by modelling CSA in umbrella and pyramidal types to test this hypothesis and provide recommendations on coffee tree pruning as an agroecological strategy for controlling CBD.

Our study sheds light on the complex relationships between berry susceptibility, microclimate and eco-physiological functioning of a plant.

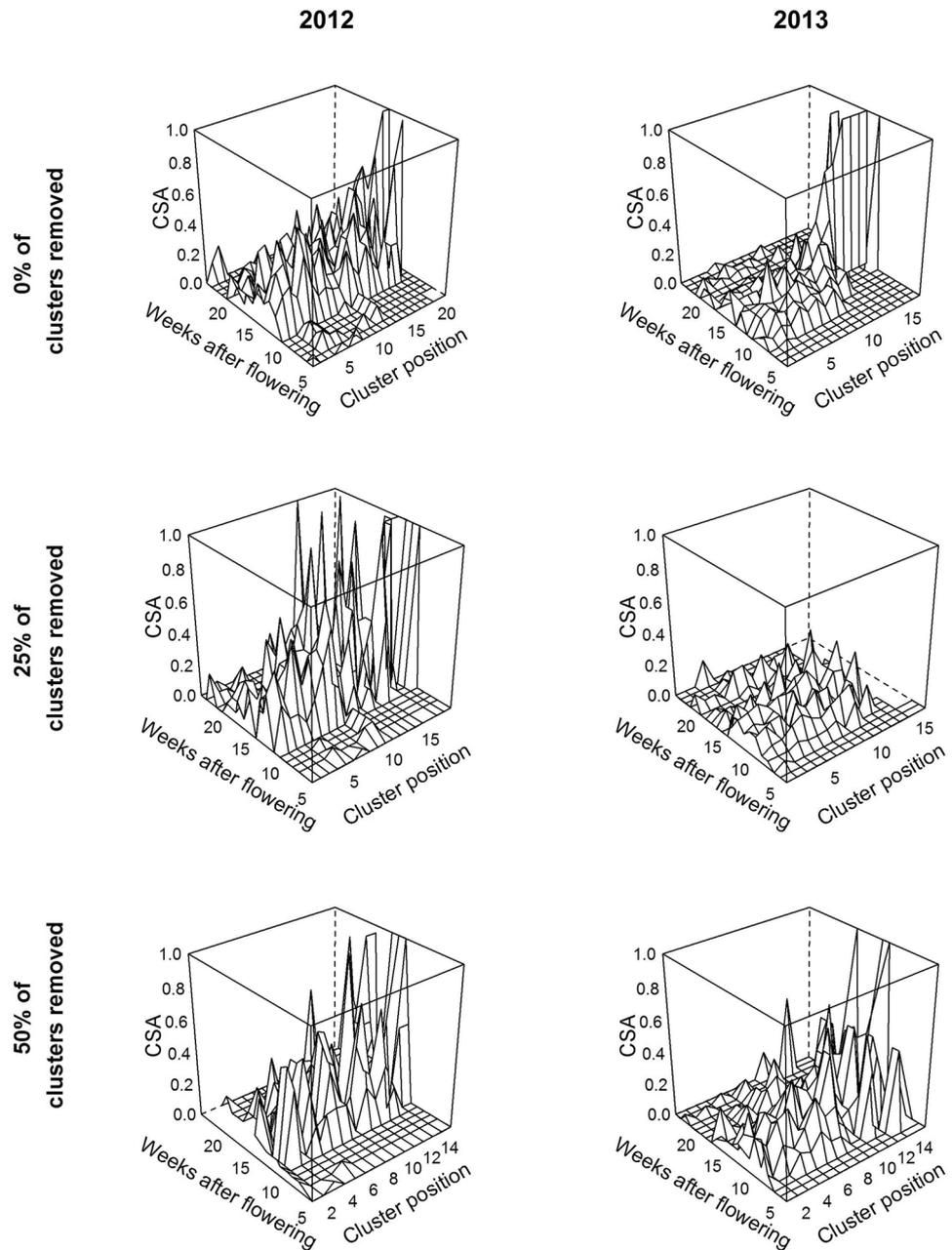
First, the weak contribution of cluster arrangement in changing disease levels suggested that pruning berries neither affects disease spread by modifying distances between clusters nor decreases berry susceptibility by increasing the berry maturation rate<sup>6</sup>. However, berry weight, an important yield-determining factor resulting from differences in berry pruning, may have increased. In fact, modifications of tree architecture alter the source-sink relationships within the plant<sup>23</sup>. The eco-physiological functioning of the plant leads to trade-offs between tree architecture and production, playing on the components of yield (amount of fruit, seed weight, chemical composition of seeds).

Second, an inspection of our data suggested that there is a relative decrease in berry susceptibility over time, as shown by the counterintuitive relationships between CSA and suitable conditions for infection (Annex 2): most symptoms appeared when suitable conditions for infection were at low levels (before 15 weeks after flowering), while the disease died out when suitable conditions for infection were more favourable (after 17 weeks after flowering). This scenario was also observed by Muller who assigned the phenomenon to a decrease in berry susceptibility over time<sup>9</sup>. While the mechanistic relationship between berry maturation stage and susceptibility has still not been formally established<sup>9</sup>, our data suggest that host ontogenic resistance, or age-related resistance, may hinder disease development over time.

Third, our results emphasized that the suitable conditions for infection three and four weeks before CSA (SCI W2 and SCI W3, respectively) were the main climatic variables influencing CSA (Figs 3 and 4), and the suitable conditions for infection four weeks and two weeks before CSA (SCI W3 and SCI W1, respectively) were involved in substantial interactions with berry physiological age and cluster position, respectively. These results suggest that the latent period of the pathogen in host tissues could be highly variable in the field, between two to four weeks. The interaction between suitable conditions for infection and berry physiological age suggests that the latent period increases with ontogenic resistance, and overall, these results increase the need for mechanistically linking latent individuals and berry ontogenic resistance, which was not explicitly introduced in the current model.

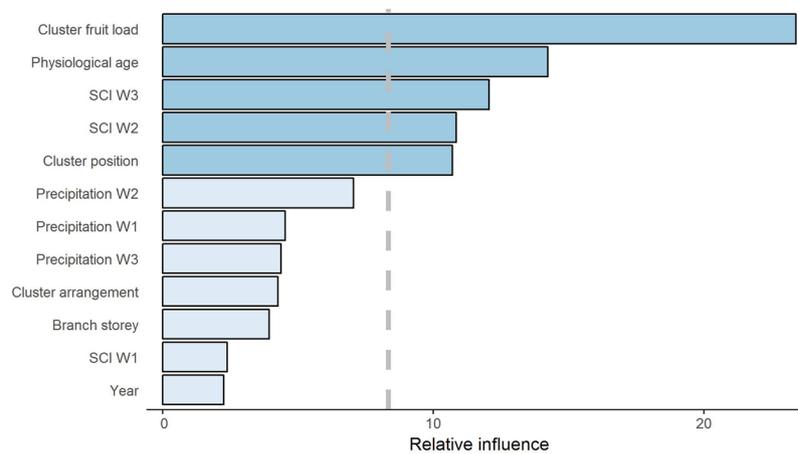
We believe that the complex relationships between berry susceptibility, microclimate and eco-physiological functioning of a plant could be achieved by combining functional structural plant models, such as Greenlab<sup>23</sup>, with a CBD epidemiological model, such as susceptible-infected-exposed-removed models (SEIR models). These models have the advantage of explicitly considering latent individuals (“exposed” compartment) and ontogenic resistance in terms of eco-physiological functioning of a plant.

Given that coffee tree architecture in interaction with microclimates strongly affects CBD epidemics, strategies for coffee crop protection may consist in manipulating these features in harmony with the natural environment of the coffee tree to design naturally induced management practices. Such strategies may be developed in the context of smallholder farms in particular, where low incomes hinder the use of chemicals to manage diseases. Agroecological strategies for controlling CBD are scarce, and agroforestry systems have been recently put forward

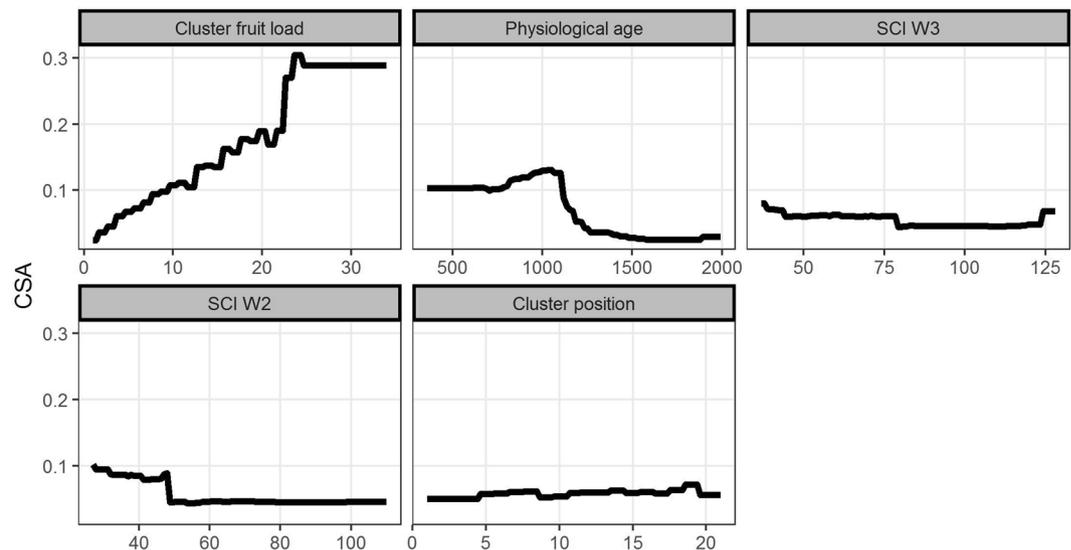


**Figure 4.** Field data showing the weekly dynamics of cluster symptom appearance (CSA) along the branch, moving away from the trunk (cluster position 1) to the tip of the branch. The rows are separated by the modalities of cluster arrangement, with 0, 25 and 50% of the clusters removed, and the columns are separated by the two years studied, 2012 and 2013.

by numerous authors as efficient ways to manage CBD<sup>4,16–20</sup>. In this context, we explored the effects of the shade of a forestry kola tree on CSA through shade modifications of microclimates (precipitation, temperature and relative humidity). As the range of variation in the microclimatic data obtained under the kola shading system was reasonably within the range of variation in the data obtained in the banana shading system (data not shown; dataset available in the Supplementary Information files), there is good confidence in the scenario outcomes. For the two scenarios investigated using climatic data recorded over two years under the kola shading system (Fig. 2b), the microclimates had contrasting effects on CSA. In the 2012 scenario, the microclimate under the kola shading system did not change disease levels relative to those in the reference situation of the banana shading system (Fig. 7). In the 2013 scenario, however, the microclimate under the kola shading system decreased disease levels, notably, at the tip of the branch, but the microclimate increased the epidemic duration. The important variations in the epidemic behaviours predicted by the two scenarios suggest that CBD is governed by a subtle balance between microclimate and architectural variables. This exploratory work emphasizes the need to be careful when



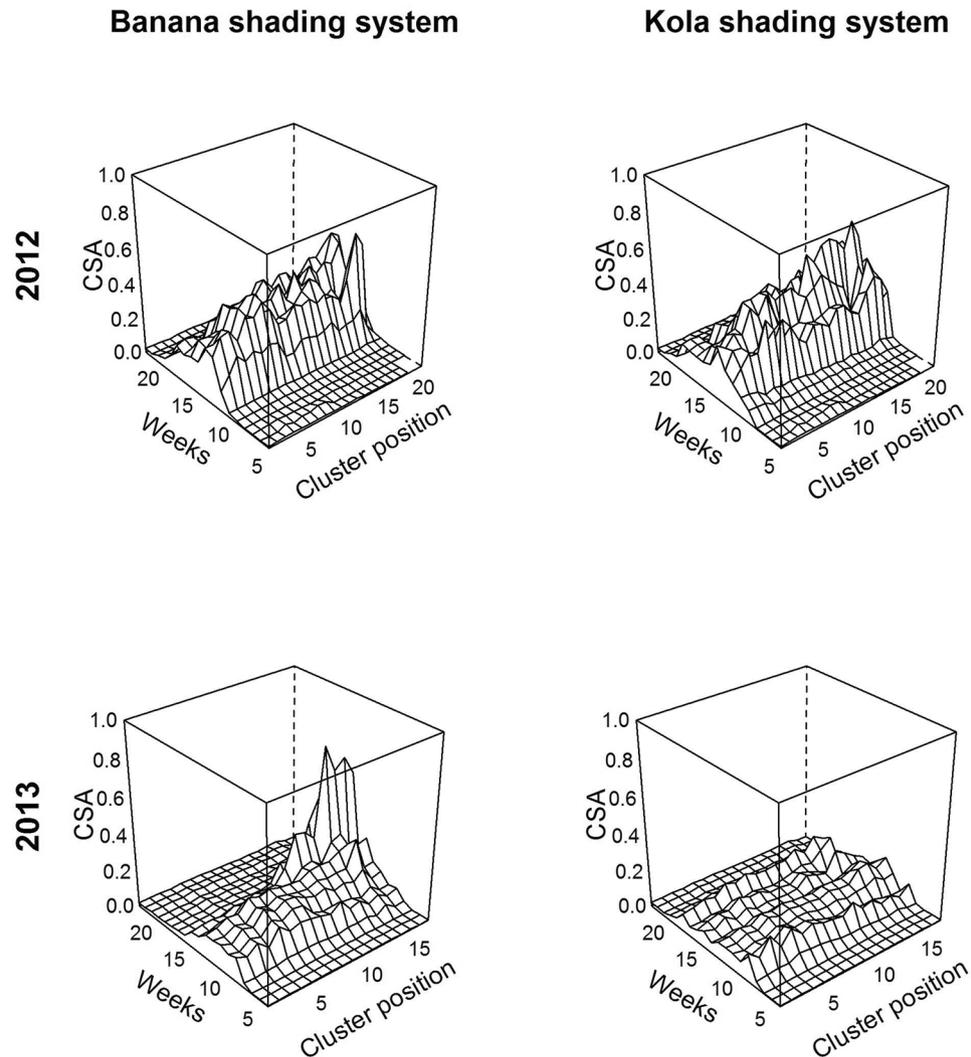
**Figure 5.** Relative influence of the predictor variables on cluster symptom appearance (CSA). The most important variables influencing CSA (dark blue) show higher relative influence than that expected due to chance (i.e., smaller than 8.3%, dashed vertical grey line). Details about the variables are provided in Table 1. SCI are the suitable conditions for infection in terms of temperature and relative humidity in the time windows W1 (7 to 14 days before the observation of CSA), W2 (15 to 21 days before CSA) and W3 (22 to 29 days before CSA).



**Figure 6.** Dependence plots showing the relationships between cluster symptom appearance (CSA) and the most important variables influencing CSA (effect of the other variables fixed to their mean). The variables are described in Table 1. SCI W2 and SCI W3 are the suitable conditions for infection in terms of temperature and relative humidity in the time windows W2 (15 to 21 days before the observation of CSA) and W3 (22 to 29 days before CSA), respectively.

considering management of CBD in agroforestry systems and calls for further studies to improve our knowledge of microclimates generated under varying shading systems to determine their effects on the disease course.

In terms of microclimates, a possible improvement of this study would be to use rainfall intensity (which requires specific equipment for monitoring) rather than cumulative rainfall. In fact, while Waller<sup>3</sup> showed a relation between cumulative rainfall and the quantity of splashed *C. kahawae* spores, our statistical analyses showed that cumulative rainfall was not important for CSA. These results suggest that under natural conditions, the relationship between rainfall and spore splashing may not be as simple as Waller<sup>3</sup> has indicated, and the kinetic energy of rain droplets would likely be more accurate to consider<sup>24,25</sup>. Similarly, it is important to note the need to evaluate the modification of other variables induced by shade. For example, the amount of light intercepted by the coffee tree and wind speed have been shown to be altered under shade<sup>17</sup>, and these variables can potentially affect the disease course through modifications of the host (fruit load due to reduced floral initiation, berry maturation and susceptibility) and modifications of pathogen dispersal, respectively. These variables could be controlled by adequate management of shade cover and the selection of shade species (canopy height, shape and size) and then



**Figure 7.** *In silico* experiment: prediction of cluster symptom appearance (CSA) dynamics along the branch in two different microclimates induced by shade. We applied the model (Annex 1) to the architectural field data, either with the microclimatic variables measured in the banana shading system (left column) or with a new set of microclimatic variables (precipitation, temperature and relative humidity) measured under the shade of a forestry tree (kola shading system) present in the coffee farm (right column), for two years, 2012 and 2013.

used as regulation instruments to adapt recommendations, as previously conducted for coffee leaf rust<sup>24</sup>. We suggest that contrasting the effects of shade trees on disease dynamics needs a more in-depth evaluation to identify the possible trade-offs among the environmental variables at play.

Finally, our study highlights the substantial effect of coffee tree architecture on CBD dynamics and disentangles the differential effects of architecture and microclimate on CBD. We promote the combination of epidemiological and architectural modelling, as has already been undertaken for grapevines<sup>26</sup>. This time-saving strategy would facilitate the design of novel, more cost-effective and environmentally friendly management strategies at both the tree scale and plot scale.

## Methods

**Field data.** *Architecture experiment.* Data were collected on a coffee farm located in Bamendjou (Alt. 1600 m), West Cameroon (5°24'0"N; 10°19'0"E) over two consecutive years (2012 and 2013) from coffee trees (*Coffea arabica*) with a mixture of varieties, Java and Jamaïque, grown in association with plantain banana trees that provide partial and non-permanent shade to coffee trees. Coffee trees were planted on average at 1200–1500 trees/ha as is often conducted in traditional Cameroonian plantings<sup>27</sup>. Banana trees were intercropped with the coffee trees at approximately the same density. These coffee trees were approximately 15 years old and topped, showing an intermediate architecture between pyramidal architecture (branches and stories are well differentiated from the top of a tree to its base), and umbrella architecture (branches are almost in the same stories in a tree) (Fig. 1a).

A factorial trial was carried out on 345 branches from 80 coffee trees (3 branches on 45 trees in 2012 and 6 branches on 35 trees in 2013). First, to address berries of the same age, all late berries were removed manually. To

evaluate disease dispersal, three distances between berry clusters were created by manually eliminating a part of the clusters on each branch of the trees. The three different cluster arrangements were obtained by removing (i) 0 clusters (0% removal corresponding to the control trees), (ii) 1 cluster out of 4 (25% removal), and (iii) 1 cluster out of 2 (50% removal) along all the branches of the coffee trees (Fig. 3).

The total number of berries, the number of diseased berries and the number of newly diseased berries (newly diseased berries were identified by a mark on each observation date) were monitored on all the sampled branches in 2012 and 2013 at the top, middle and lower stories of each studied tree. Berries were monitored weekly from fruit onset, at 5 and 4 weeks after flowering for 2012 and 2013, respectively, to 29 weeks (4 weeks before harvest) after flowering.

Temperature and relative humidity were measured hourly from the beginning to the end of the observation period using Tinytag® (Gemini, UK) dataloggers located in two randomly chosen coffee trees; one datalogger was placed in the middle story of each coffee tree, under leaves to avoid direct sunlight. To measure the quantity of daily rainfall during the observation period, two rain gauges were positioned in the plot without surrounding vegetation to avoid any rain capture perturbation.

*In silico experiments.* Numerical simulations aimed at identifying prospects for how agroforestry systems can affect CBD dynamics at the tree scale. We investigated whether the microclimates induced by the shade of a forestry kola tree decreased CBD levels as cited by numerous authors<sup>4,16–20</sup>. Those simulations involved microclimate data collected in 2012 and 2013 at the farm, under the canopy of a kola (*Cola sp.*) tree, namely, the kola shading system. Temperature and relative humidity were measured hourly during each observation period using Tinytag® dataloggers located in the middle story of one coffee tree grown under the kola tree. This coffee tree was not part of our experimental setup but was used only to obtain microclimate data. The quantity of daily rainfall under the kola tree was estimated by positioning a rain gauge under the kola canopy without surrounding coffee trees.

**Definition and construction of the variables of interest.** *Response variable.* We marked diseased berries each week to detect new symptoms on a berry cluster, referred to as CSA (Table 1). CSA was defined as a binary variable valued 1 if any new diseased berry appeared and valued 0 otherwise.

*Explanatory variables.* Years (2012, 2013) were introduced as a 2-level factor (Table 1).

*Explanatory variables related to tree architecture.* Berry physiological age: Berry physiological age was used as an indicator of berry maturation that we assumed to be linked to berry susceptibility and was obtained by computing the sum of degree days from flowering (Table 1) using a base temperature of 10 °C, below which the coffee tree does not grow<sup>28,29</sup>.

Cluster position: The position of the berry clusters along the branch was defined as an ordinal variable ranging from position 1 (the first cluster immediately at the trunk) to the last position at the tip of the branch (Fig. 3). The last position of a cluster was variable among branches, with the maximum found at position 21.

Cluster arrangement: Cluster arrangement was defined as a factorial variable corresponding to the rates of cluster removal (0, 25 or 50% removal) (Fig. 3). Removing a cluster was achieved by removing all the berries of the cluster.

Branch story: The branch story in the coffee tree was defined as a factorial variable describing the story of the branch in the tree, namely, top, middle and lower stories.

Cluster fruit load: Finally, the fruit load variable was defined as the number of berries per cluster.

*Explanatory variables related to microclimate.* Time windows considered: The CBD latent period (time between host infection and appearance of symptoms) can vary between 6 and 15 days under controlled laboratory conditions according to Van der Vossen and Waweru<sup>12</sup> and Pinard *et al.*<sup>13</sup> and between 14 and 30 days in the field according to Mulinge<sup>8</sup>. Thus, in this study, we considered that the latent period could vary between 6 days and 1 month. We computed the microclimate covariates in 7-day time intervals corresponding to time windows of 7 to 14 days (W1), 15 to 21 days (W2) and 22 to 29 days (W3) before the appearance of cluster symptoms.

Precipitation: Rainfall is acknowledged as a key variable related to CBD development<sup>3,4</sup>. In our study, precipitation was defined as the cumulative rainfall (in millimetres) within each time window, W1, W2, and W3, as previously described.

Suitable conditions for infection: Berry infection (formation of germ tubes and penetration of host tissues) was found to occur at temperatures ranging from 15 to 25 °C<sup>11</sup>, with an optimum temperature of 20 °C for hypocotyl inoculation tests<sup>12</sup>. Inhibition of mycelium growth was found to occur at 30 °C<sup>30</sup>. At least 5 hours of wetness close to saturation (at least 98% relative humidity) at temperatures above 15 °C is required for infection to occur<sup>3</sup>. Based on these findings, we computed the number of hours beyond 5 hours of suitable temperature and relative humidity within each time window, W1, W2, and W3.

**Statistical analyses.** Our study focused on the link between the presence (or absence) of CSA and the explanatory variables (or predictors) defined in the previous section. All statistical analyses were carried out using R version 3.4.1<sup>31</sup>, namely, R packages gbm v2.1.1<sup>32</sup> and dismo v1.1.4<sup>33</sup>

*Model description.* We applied a statistical method from the field of machine learning, namely, BRT<sup>34,35</sup>, to build a stochastic, nonlinear regression model from our dataset. This method is attracting increasing interest in ecology and epidemiology<sup>36–40</sup>. BRT combines the advantages of regression trees (capable of handling qualitative and quantitative predictors and accommodating missing data, and prior information or data transformation is not needed) and boosting (improves prediction accuracy by combining many simple models)<sup>36</sup>. Regression

trees find the best binary split of the data, recursively partitioning the training data set one predictor at a time. The best split is the one that maximizes the homogeneity of the response variable within partitioned groups. The boosting algorithm proceeds by iteratively fitting a new regression tree to the prediction residuals of the preceding tree; thus, the final model is an additive expansion of the regression trees. To avoid overfitting, shrinkage and cross-validation techniques were used<sup>41</sup>. The optimal number of boosting trees was determined using a ten-fold cross-validation procedure, retaining the number that minimized the average holdout residual deviance<sup>33</sup>. In our study, the response variable was the presence or absence of CSA for each week. Accordingly, we selected the Bernoulli distribution in our BRT models. Moreover, data were divided into a training set and a test set (one branch among 6 was randomly selected per tree and per week for the test set). The test set was used to evaluate the model's predictive performance.

The main parameters of the BRT models are tree complexity (tc), which controls the interaction depth; the learning rate (lr), which controls shrinkage<sup>42</sup>; and the bag fraction (bf), which introduces some randomness into the model fit<sup>41</sup>. We assessed the appropriate model settings by testing combinations of these three parameters chosen as follows: tc ranging between 2 and 8, lr in the range {0.01, 0.005, 0.001} and bf in the range {0.75, 1}. We selected the set of values that minimized the model predictive deviance measured on the excluded folds of a cross-validation procedure. The regression model was built using the optimal setting and provided the probability of symptom appearance inside a cluster as the output.

The last step consisted in transforming the quantitative output (probability of symptoms) into a binary variable describing the presence or absence of a disease symptom, i.e., simplifying the regression model into a binary classifier. The receiver operating characteristic (ROC) curve illustrates the performance of a binary classifier as its decision threshold is varied. The curve is obtained by plotting the true positive rate, or probability of detection, against the false positive rate, or probability of a false alarm, at various threshold settings. The area under the curve (AUC) provides a metric ranging between 0 and 1 for classifier performance: the closer the AUC of a model is to 1, the better it is. We computed the AUC using the R package pROC v1.8<sup>43</sup>.

**Relative influence of variables.** BRT provided a measure of the relative influence of predictors on the response variable. BRT is based on the number of times a variable is selected for splitting, weighted by the squared improvement to the model as a result of each split and averaged over all trees<sup>34,44</sup>. The measure is scaled such that all relative influences sum to 100; high values indicate a strong influence.

**Partial dependence plots.** To visualize the marginal effect of each covariate on the response variable, a partial dependence function was calculated, averaging all the effects of all other covariates. In BRT, an approximation is calculated. Along the range of each predictor, the average of the model predictions is computed. Partial dependence plots are extremely useful for knowledge discovery in data, especially when low-order interactions and main effects predominate.

**Relative strength of interaction effects.** The strength of interaction between two predictors was evaluated using the magnitude of departure from a purely additive model. Using the R package gbm, we computed the Friedman's H-statistic<sup>45</sup> to determine the relative strength of the interaction effects. H is in the interval [0, 1], with high values indicating large interaction effects.

**In silico experiments.** For *in silico* experiments, we applied a new set of microclimatic variables (suitable conditions for infection and precipitation) measured under a kola shade tree (the kola shading system) in years 2012 and 2013 to the fitted model (Annex 1) as described in the section "Field data".

## Data Availability

The datasets generated or analysed during this study are included in this article (and its Supplementary Information files).

## References

1. Bieysse, D. M. *et al.* Coffee berry disease: a potential threat to Arabica coffee growing worldwide. *Plant. Rech. Dév. Fr.* **P** 144–156 (2002).
2. Mouen Bedimo, J. A., Bieysse, D., Cilas, C. & Nottéghem, J. L. Spatio-temporal dynamics of arabica coffee berry disease caused by *Colletotrichum kahawae* on a plot scale. *Plant Dis.* **91**, 1229–1236 (2007).
3. Waller, J. M. Water-borne spore dispersal in coffee berry disease and its relation to control. *Ann. Appl. Biol.* **71**, 1–18 (1972).
4. Mouen Bedimo, J. A., Bieysse, D., Nyassé, S., Nottéghem, J. L. & Cilas, C. Role of rainfall in the development of coffee berry disease in *Coffea arabica* caused by *Colletotrichum kahawae*, in Cameroon. *Plant Pathol.* **59**, 324–329 (2010).
5. Calonne, A., Cartolaro, P., Naulin, J.-M., Bailey, D. & Langlais, M. A host-pathogen simulation model: powdery mildew of grapevine. *Plant Pathol.* **57**, 493–508 (2008).
6. Vaast, P., Bertrand, B., Perriot, J.-J., Guyot, B. & Genard, M. Fruit thinning and shade improve bean characteristics and beverage quality of coffee (*Coffea arabica* L.) under optimal conditions. *J. Sci. Food Agric.* **86**, 197–204 (2006).
7. Nutman, F. J. & Roberts, F. M. Investigations on a disease of *Coffea arabica* caused by a form of *Colletotrichum coffeanum* noack: II. Some factors affecting germination and infection, and their relation to disease distribution. *Trans. Br. Mycol. Soc.* **43**, 643–659 (1960).
8. Mulinge, S. K. Development of coffee berry disease in relation to the stage of berry growth. *Ann. Appl. Biol.* **65**, 269–276 (1970).
9. Muller, R. A. Contribution à la connaissance de la phytomycoécénose constituée par *Coffea arabica*, *Hemileia vastatrix*, et *Hemileia coffeicola*. (Inst. Français Café Cacao, 1980).
10. Van der Graaff, N. A. Coffee berry disease. In *Plant Diseases of International Importance* **6**, 202–230 (A. N. Mukhopadhyay, J. Kumar, U. S. Singh, and H. S. Chaube, 1992).
11. Mwang'ombe, A. W., MuKunya, D. M. & Gathuru, E. M. Effects of temperature on appressorium formation and pathogenicity of *Colletotrichum coffeanum* strains. *J. Plant Prot. Trop. Malays* (1991).

12. Van der Vossen, H. A. & Waweru, J. A temperature controlled inoculation room to increase the efficiency of preselection for resistance to coffee berry disease. *Kenya Coffee* **41**, 164–167 (1976).
13. Pinard, F., Omondi, C. O. & Cilas, C. Detached berries inoculation for characterization of coffee resistance to coffee berry disease. *J. Plant Pathol.* **94**, 517–523 (2012).
14. Mbow, C., Smith, P., Skole, D., Duguma, L. & Bustamante, M. Achieving mitigation and adaptation to climate change through sustainable agroforestry practices in Africa. *Sustain. Chall.* **6**, 8–14 (2014).
15. Schroth, G., Krauss, U., Gasparotto, L., Duarte Aguilar, J. A. & Vohland, K. Pests and diseases in agroforestry systems of the humid tropics. *Agrofor. Syst.* **50**, 199–241 (2000).
16. Phiri, N. A., Hillocks, R. J. & Jeffries, P. Incidence and severity of coffee diseases in smallholder plantations in northern Malawi. *Crop Prot.* **20**, 325–332 (2001).
17. Beer, J., Muschler, R., Kass, D. & Somarriba, E. Shade management in coffee and cacao plantations. *Agrofor. Syst.* **38**, 139–164 (1997).
18. Bedimo, J. A. M. *et al.* Effect of cultural practices on the development of arabica coffee berry disease, caused by *Colletotrichum kahawae*. *Eur. J. Plant Pathol.* **119**, 391 (2007).
19. Bedimo, J. M. *et al.* Effect of shade on Arabica coffee berry disease development: toward an agroforestry system to reduce disease impact. *Phytopathology* **98**, 1320–1325 (2008).
20. Bedimo, J. M., Cilas, C., Nottéghem, J. L. & Bieysse, D. Effect of temperatures and rainfall variations on the development of coffee berry disease caused by *Colletotrichum kahawae*. *Crop Prot.* **31**, 125–131 (2012).
21. Nutman, F. J. & Roberts, F. M. Investigations on a disease of *Coffea arabica* caused by a form of *Colletotrichum coffeanum* noack: III. The relation between infection of bearing wood and disease incidence. *Trans. Br. Mycol. Soc.* **44**, 511–521 (1961).
22. Avelino, J. Rationalizing control of coffee tree diseases. in *21st International Conference on Coffee Science* 1250–1259 (ASIC, 2006).
23. Mathieu, A., Cournède, P.-H., Barthélémy, D. & De Reffye, P. Rhythms and alternating patterns in plants as emergent properties of a model of interaction between development and functioning. *Ann. Bot.* **101**, 1233–1242 (2007).
24. Boudrot, A. *et al.* Shade effects on the dispersal of airborne *Hemileia vastatrix* uredospores. *Phytopathology* **106**, 572–580 (2016).
25. Avelino, J., Willocquet, L. & Savary, S. Effects of crop management patterns on coffee rust epidemics. *Plant Pathol.* **53**, 541–547 (2004).
26. Mammeri, Y., Burie, J. B., Langlais, M. & Calonne, A. How changes in the dynamic of crop susceptibility and cultural practices can be used to better control the spread of a fungal pathogen at the plot scale? *Ecol. Model.* **290**, 178–191 (2014).
27. Bouharmont, P. La plantation du caféier Arabica au Cameroun. *Café Cacao Thé* **26**, 3–16 (1982).
28. Pezzopane, J. R. M., Salva, T., de, J. G., de Lima, V. B. & Fazuoli, L. C. Agrometeorological parameters for prediction of the maturation period of Arabica coffee cultivars. *Int. J. Biometeorol.* **56**, 843–851 (2012).
29. Rodríguez, D., Cure, J. R., Cotes, J. M., Gutierrez, A. P. & Cantor, F. A coffee agroecosystem model: I. Growth and development of the coffee plant. *Ecol. Model.* **222**, 3626–3639 (2011).
30. Manga, B. Etude de la diversité de *Colletotrichum kahawae* responsable de l'anthracnose des baies et caractérisation de la résistance du caféier Arabica à cet agent pathogène. (Montpellier II, 1999).
31. Team, R. C. R. A language and environment for statistical computing [Internet]. Vienna, Austria; 2014 (2017).
32. Ridgeway, G. With contributions from others. *gbm: Generalized Boosted Regression Models* (2017).
33. Hijmans, R. J., Phillips, S., Leathwick, J., Elith, J. & Hijmans, M. R. J. Package 'dismo'. *Circles* **9**, 1 (2017).
34. Friedman, J. H. Greedy function approximation: a gradient boosting machine. *Ann. Stat.* 1189–1232 (2001).
35. Hastie, T., Tibshirani, R. & Friedman, J. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. *Biometrics* (2002).
36. Elith, J., Leathwick, J. R. & Hastie, T. A working guide to boosted regression trees. *J. Anim. Ecol.* **77**, 802–813 (2008).
37. Artois, J. *et al.* Clade-level Spatial Modelling of HPAI H5N1 Dynamics in the Mekong Region Reveals New Patterns and Associations with Agro-Ecological Factors. *Sci. Rep.* **6**, 30316 (2016).
38. Allen, T. *et al.* Global hotspots and correlates of emerging zoonotic diseases. *Nat. Commun.* **8**, 1124 (2017).
39. De'Ath, G. Boosted trees for ecological modeling and prediction. *Ecology* **88**, 243–251 (2007).
40. Poggi, S. *et al.* Relative influence of climate and agroenvironmental factors on wireworm damage risk in maize crops. *J. Pest Sci.* **91**, 585–599 (2018).
41. Natekin, A. & Knoll, A. Gradient boosting machines, a tutorial. *Front. Neurobotics* **7** (2013).
42. Buston, P. M. & Elith, J. Determinants of reproductive success in dominant pairs of clownfish: a boosted regression tree analysis. *J. Anim. Ecol.* **80**, 528–538 (2011).
43. Robin, X. *et al.* pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics* **12**, 77 (2011).
44. Williams, G. J., Aeby, G. S., Cowie, R. O. & Davy, S. K. Predictive modeling of coral disease distribution within a reef system. *PLoS One* **5**, e9264 (2010).
45. Friedman, J. H. & Popescu, B. E. Predictive learning via rule ensembles. *Ann. Appl. Stat.* 916–954 (2008).

## Acknowledgements

This study was part of the program *Dispositif en Partenariat “Agroforesterie”* carried out in partnership with CIRAD and IRAD in Cameroon. The study was supported by the *ATP Agroforestry*, financed by CIRAD and IRD, through the project *“Recherche de compromis entre productions et services écosystémiques fournis par les systèmes agroforestiers (SAFSE)”*. We thank Dr. Joseph Mouen for his helpful contribution in the search of experimental fields and Dr. Gerben Martijn ten Hoopen and Dr. Daniel Bieysse for their careful, final review of the manuscript.

## Author Contributions

N.M. conceived, designed and performed the experiments. S.P., N.M. and F.R. designed and conducted the statistical analyses. N.M. and S.P. wrote the manuscript and prepared the tables and figures. All authors reviewed and approved the final manuscript.

## Additional Information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s41598-019-38775-5>.

**Competing Interests:** The authors declare no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019