

Influence de la pollinisation sur le remplissage des cabosses du cacaoyer

Influence of pollination on the number of beans per cacao pod

F. Ribeyre (a), C. Cilas (a), N. Motisi (a), Ph. De Reffye (b)  
(a) CIRAD, UPR Bioagresseurs, F-34398 Montpellier, France.  
Bioagresseurs, Univ Montpellier, CIRAD, Montpellier, France.  
(b) CIRAD, UMR AMAP, F-34398 Montpellier, France.  
AMAP, Univ Montpellier, CIRAD, CNRS, INRA, IRD, Montpellier, France.

### Résumé

La production du cacaoyer, en nombre de fruits produits, et en nombre de graines par cabosse est un caractère très variable et peu héritable chez le cacaoyer. Des facteurs environnementaux influencent donc ces caractères. Dans ce travail, nous cherchons à déterminer quels sont les mécanismes qui expliquent la distribution du nombre de fèves par cabosse pour un clone donné.

Nous utilisons un modèle à trois facteurs adapté des travaux de Ph de Reffye (1978) : le nombre de grains de pollen efficaces, le nombre d'ovules fertiles et l'avortement des cabosses en fonction de leur remplissage. L'approche bayésienne choisie permet de considérer les cinq paramètres du modèle comme des variables aléatoires. Le modèle est appliqué à 10 clones plantés au Togo dans un même site (Zozokondji). Le nombre de graines a été compté pour chacune des trois cents cabosses récoltées par clone.

Les distributions observées du nombre de graines par cabosse présentent des formes différentes selon les clones. Le modèle permet d'ajuster correctement la distribution du nombre de graines par cabosse pour chaque clone. Les résultats montrent que les paramètres estimés pour chacune des distributions sont différents en fonction des clones. Les trois facteurs limitent la production de graines par cabosse mais ont une importance variable selon les clones.

Le modèle permet toutefois d'accéder à des paramètres très difficiles à mesurer comme la rareté du pollen. Les distributions en forme de loi binomiale correspondent à des clones qui ont peu de cabosses mal remplies et pour lesquels le pollen n'apparaît pas comme un facteur limitant. D'autres distributions, plus étalées, correspondent à des clones pour lesquels le modèle estime que le pollen efficace est rare. Notre modèle met en évidence que la pollinisation est un facteur limitant le remplissage des cabosses pour 8 des 10 clones étudiés.

Mots clés: rendement, modèle hiérarchique bayésien, nombre de fèves par cabosse.

### Summary

The production of cocoa, in number of fruits produced, and in number of seeds per pod is variable characters and rarely inheritable in cacao. Environmental factors therefore influence these characters. In this work, we aim to determine the mechanisms that explain the distribution of the number of beans per pod for a given clone.

We used a three-factor model adapted from Ph de Reffye's (1978) work: the number of effective pollen grains, the number of fertile eggs and the abortion of pods according to their filling. The selected Bayesian approach allows to consider the five parameters of the model as random variables. The model is applied to 10 clones planted in Togo in the same site (Zozokondji). The number of beans per pod was counted for each of the 300 pods harvested per clone.

The observed distributions of the number of beans per pod have different shapes depending on the clones. The model allows to correctly adjust the number of seeds per pod for each clone. The results show that the estimated parameters for each of the distributions are different depending on the clones. The three factors limit seed production per pod but are of variable importance depending on the clones.

The model provides access to parameters that are difficult to measure, such as the scarcity of pollen. The binomial distributions correspond to clones that have few pods poorly filled and for which pollen does not appear to be a limiting factor. Other distributions, more spread out or bimodal, correspond to clones for which the model estimates that effective pollen is rare. Our model shows that pollination is a factor limiting the filling of pods for 8 of the 10 clones studied.

Mots clés: production, modèle hiérarchique bayésien, nombre de fèves par cabosse.

## Article

### Introduction

La production du cacaoyer, en nombre de fruits produits, et en nombre de graines par cabosse est un caractère très variable et peu héritable chez le cacaoyer. Le cacaoyer est un arbre qui produit de nombreuses fleurs, jusqu'à 125 000 (Toledo-Hernandez et al., 2017) et peu de cabosses par rapport au nombre de fleurs. Le taux de pollinisation est faible en raison principalement d'un faible nombre d'insectes visiteurs et d'un transport de pollen peu efficace (Toledo-Hernandez et al., 2017). Le taux de pollinisation influence le nombre de graines par cabosse (Falque et al., 1995). Le nombre d'ovules par ovaire est élevé et variable (de Reffye et al., 1978). De nombreux facteurs environnementaux influencent ces mécanismes (Toledo-Hernandez et al., 2017).

Dans ce travail, nous nous intéressons au nombre de graines par cabosse en tant que composante du rendement du cacaoyer. En effet, certains auteurs considèrent que maximiser le nombre de graines par cabosses est un enjeu majeur (Toledo-Hernandez et al., 2017). Le nombre de graine par cabosse est souvent estimé par la moyenne de la distribution observée mais de Reffye et al. (1978) ont montré que les formes des distributions étaient parfois bimodales et que la moyenne n'avait pas de signification biologique. Nous cherchons donc à déterminer quels sont les mécanismes qui expliquent la distribution du nombre de fèves par cabosse pour un clone donné et quelle importance ils ont selon les clones.

### Matériel et méthode

Le matériel végétal provient de champs clonaux situés à Zozokondji au Togo. Pour chacun des 10 clones sélectionnés, 173 à 300 cabosses ont été prélevées et le nombre de graines dans chaque cabosse a été compté (figure 1). La fonction de distribution du nombre de graines par cabosses est ainsi déterminée pour chaque clone.



Figure 1: Le nombre de graines par cabosse est une des composantes du rendement du cacaoyer (Cabosses ouvertes et graines). Photo : F. Ribeyre

Pour modéliser la distribution du nombre de graines, nous utilisons un modèle à trois facteurs adapté des travaux de De Reffye et al.(1978) : le nombre de grains de pollen efficaces, le nombre d'ovules fertiles et l'avortement des cabosses en fonction de leur nombre de graines (figure 2). Le nombre d'ovules fertiles est modélisé à l'aide une loi binomiale  $B(N, p)$ . Le nombre de grains de pollen efficaces est modélisé par une loi de Pareto discrétisée de paramètres  $x_0 = 0.5$  et  $a$ . Lorsque  $a$  est petit, le pollen est plus abondant ; lorsque  $a$  est grand, le pollen est plus rare. ( $a$ ) est donc un paramètre de rareté du pollen. Le nombre de cabosses potentielles qui avortent est modélisé à partir d'une fonction sigmoïde  $f(x) = 1 - 1/(1 + \exp(-c_2 * (x - c_1)))$ . ( $c_1$ ) correspond au point d'inflexion de la sigmoïde et ( $c_2$ ) est un indicateur de la pente.

A partir de ce modèle conceptuel, un modèle hiérarchique bayésien a été construit. L'approche bayésienne choisie permet de considérer les 5 paramètres du modèle comme des variables aléatoires. Les distributions a priori des paramètres sont choisies à partir de connaissances extraites de la littérature. La distribution du nombre observé de cabosses ayant  $x$  graines,  $x$  variant de 1 à 100 est considérée comme une loi multinomiale.

Le modèle est testé sur six jeux de données simulées selon les lois précédentes. Nous utilisons 4 chaînes et 20000 itérations dont 10000 sont conservées pour estimer les distributions a posteriori des paramètres. Le modèle fournit les valeurs estimées des paramètres et leur intervalle de confiance à 95% a posteriori. Les estimations du modèle réalisées à partir des données simulées sont comparées aux valeurs des paramètres qui ont permis de réaliser les simulations. Les valeurs des jeux de paramètres issus des itérations permettent de calculer des corrélations entre paramètres. Le modèle est ensuite appliqué aux 10 clones.

Les modèles sont implémentés dans R (R Core Team, 2017) en utilisant le package R2jags version 0.5-7 (Su and Yajima, 2015). Les graphiques sont réalisés à l'aide des packages ggplot2 version 2.2.1 (Wickham, 2009) et mcmcplots (McKay Curtis, 2015).

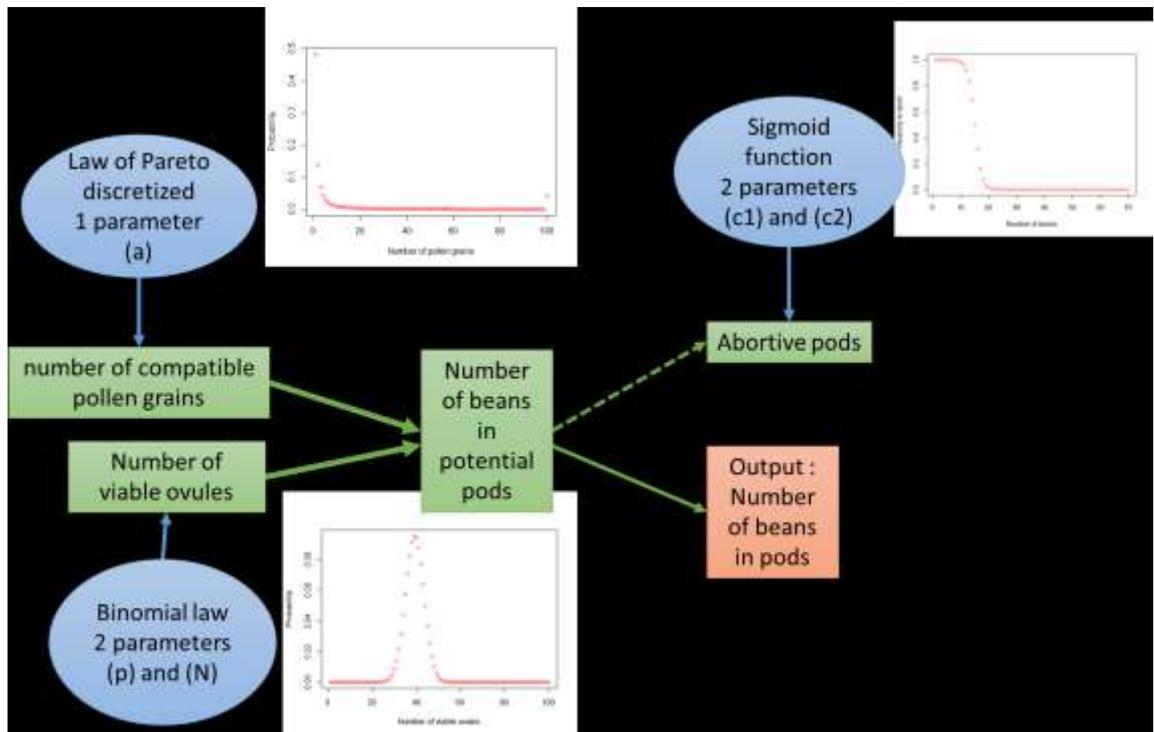


Figure 2: Modèle conceptuel reliant le nombre de grains par cabosse au nombre de grains de pollen efficaces, au nombre d'ovules fertiles et au nombre de cabosse tombées

## Résultats

Dans un premier temps, le modèle est testé à partir de simulations. L'intervalle de confiance de la distribution a posteriori est resserré et centré sur la valeur vraie pour le paramètre de rareté du pollen ( $a$ ) (figure 3). Le nombre maximal d'ovules ( $N$ ) a généralement un intervalle de confiance large contrairement au paramètre ( $a$ ). ( $p$ ) et ( $c1$ ) sont relativement bien estimés pour la plupart des valeurs des paramètres. Pour 5 des 6 simulations, l'estimation du nombre maximal d'ovules est extrêmement corrélée avec l'estimation de la probabilité pour un ovule d'être fertile ( $p > 0.9$ ).

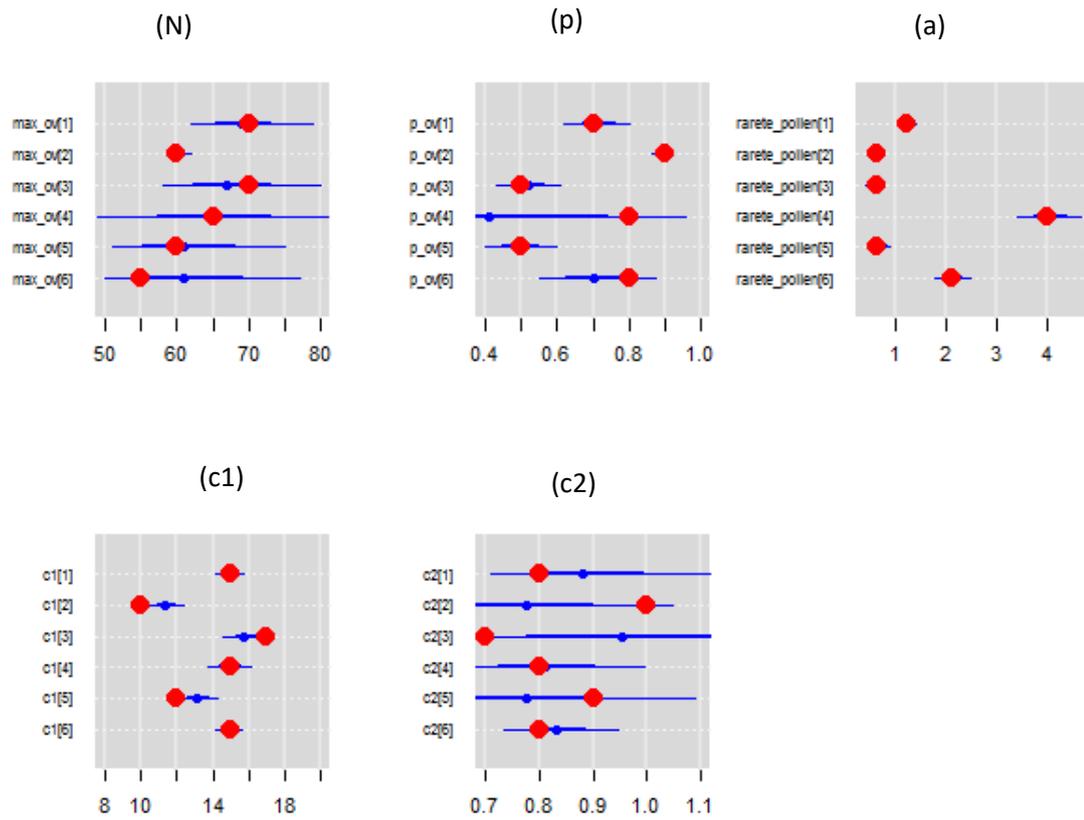


Figure 3: Paramètres utilisés pour la simulation (rouge) et estimation de ces paramètres par le modèle bayésien avec intervalle de confiance à 95% (bleu) pour les 5 paramètres du modèle : (N) nombre maximal d'ovules ; (p) probabilité d'avoir une ovule fertile ; (a) rareté du pollen ; (c1) et (c2) paramètres de la sigmoïde d'avortement des cabosses

Le modèle hiérarchique bayésien est appliqué aux 10 clones et permet d'obtenir des estimations des paramètres (tableau 1) et les distributions prédites (figure 4).

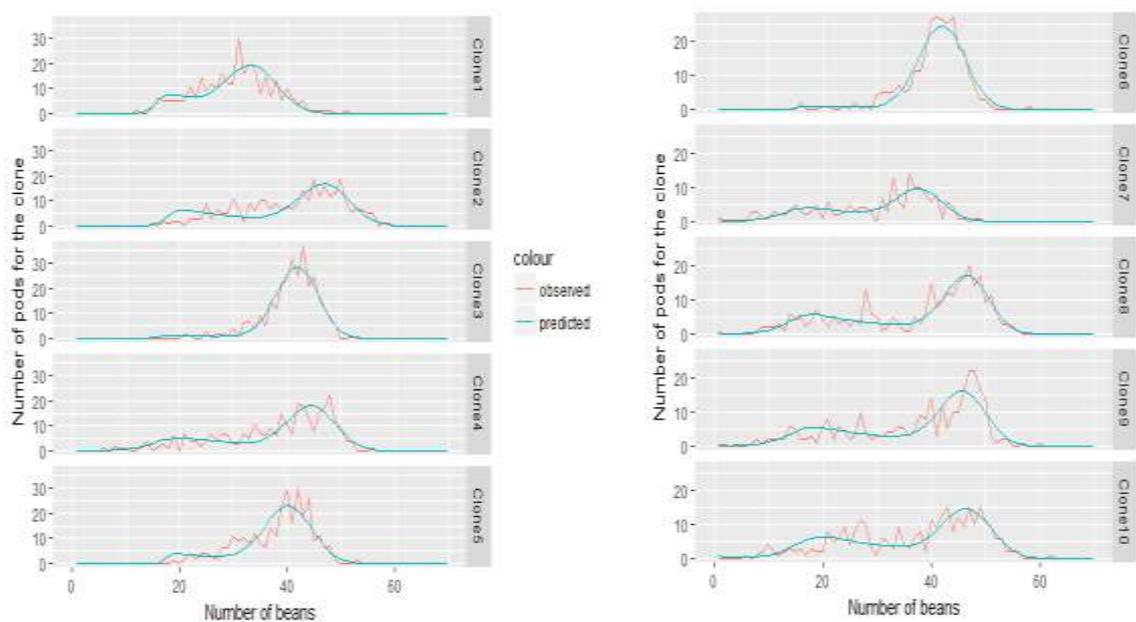


Figure 4 : Distributions observées et prédites pour les 10 clones étudiés

Tableau 1: Estimation des paramètres et intervalle de confiance à 95% a posteriori pour les 10 clones cultivés au Togo

Estimation	Viable ovules ~ Binomial(p, N)		Compatible pollen ~ pareto(a, 0,5)	Aborted pods ~ 1-sigmoid function(c1,c2)	
	p	N	a	c1	c2
Clone 1	0.4 (0.3-0.5)	85 (74-99)	0.5 (0.4-0.7)	16 (15-17)	1.2 (0.7-2.0)
Clone 2	0.5 (0.5-0.6)	86 (75-97)	0.5 (0.4-0.6)	18 (17-20)	1.1 (0.6-2.0)
Clone 3	0.6 (0.6-0.7)	66 (60-72)	0.1 (0.0-0.1)	16 (14-18)	1.5 (0.5-3.0)
Clone 4	0.6 (0.5-0.7)	77 (67-89)	0.4 (0.3-0.6)	16 (15-18)	0.5 (0.4-0.6)
Clone 5	0.5 (0.5-0.6)	74 (63-87)	0.3 (0.2-0.4)	18 (16-19)	1.7 (0.7-3.3)
Clone 6	0.6 (0.5-0.6)	73 (65-82)	0.1 (0.0-0.1)	15 (14-17)	1.5 (0.5-3.1)
Clone 7	0.5 (0.6-0.6)	70 (60-82)	0.6 (0.4-0.8)	15 (13-16)	0.5 (0.4-0.6)
Clone 8	0.6 (0.6-0.7)	72 (65-82)	0.5 (0.4-0.6)	15 (13-16)	0.6 (0.4-0.7)
Clone 9	0.6 (0.5-0.7)	74 (66-84)	0.5 (0.4-0.6)	15 (14-17)	0.5 (0.4-0.6)
Clone 10	0.5 (0.5-0.6)	81 (72-91)	0.6 (0.5-0.7)	17 (15-18)	0.5 (0.4-0.6)

Les distributions observées du nombre de graines par cabosse présentent des formes différentes selon les clones qui sont bien ajustées par le modèle (figure 4) sauf pour le clone 2 pour lequel la première partie de la courbe est mal ajustée. Les clones 3 et 6 ont des distributions qui se rapprochent de lois binomiales. Les clones 1, 2, 4 et 5 ont des distributions étalées. Les clones 7, 8, 9 et 10 ont des distributions bimodales.

Les paramètres estimés pour chacune des distributions diffèrent en fonction des clones. Les clones 3 et 6 ont un paramètre de rareté du pollen très faible donc un fort nombre de grain de pollen disponibles. Les autres clones ont un paramètre de rareté du pollen plus élevé. Le clone 1 a une faible probabilité d'avoir des ovules fertiles. Le clone 3 a un nombre maximal d'ovules faible.

Les clones 4, 6, 7, 8 et 9 ont un point d'inflexion de la sigmoïde (c1) inférieur aux clones 2 et 5. Les clones 4, 7, 8, 9 et 10 ont une pente de sigmoïde (c2) moins inclinée que les clones 1, 2 et 5. Les pentes des clones 3 et 6 sont très mal estimées.

### **Discussion-Conclusion**

Le modèle permet à partir de données de nombre de graines par cabosses de remonter aux phénomènes qui se sont produits lors de la formation des fruits. L'approche hiérarchique permet d'avoir des intervalles de confiance sur les paramètres. Mais la corrélation entre le nombre maximal d'ovules (N) et la probabilité (p) induit une grande variabilité sur l'estimation de ses 2 paramètres. La rareté du pollen est par contre bien estimée.

Comme pour de Reffye et al. (1978), les clones étudiées ont des distributions aux formes variées et souvent complexes. Les clones avec une distribution en forme de loi binomiale correspondent à des clones qui ont peu de cabosses mal remplies et pour lesquels le pollen n'apparaît pas comme un facteur limitant (Clones 3 et 6). D'autres distributions, plus étalées ou bimodales, correspondent à des clones pour lesquels le modèle estime que le pollen efficace est plus rare.

Le deuxième paramètre qui semble influencer fortement le nombre de graines par cabosse est (c1), le point d'inflexion de la sigmoïde. Il correspond au nombre de graines par cabosse en dessous duquel peu de cabosses se sont formées (la plupart ont avortées) et au-dessus duquel les cabosses avortent peu. Ce point d'avortement a lieu à des niveaux de remplissage différents selon les clones. Les clones 2 et 5 ne conservent que des cabosses ayant un nombre élevé de graines contrairement aux clones 7 et 8 qui conserve des cabosses avec très peu de graines.

Notre modèle met donc en évidence que pour certains clones, la pollinisation est un facteur essentiel limitant le remplissage des cabosses. La pollinisation (au sens pollen compatible) est affectée par l'auto-compatibilité, par l'attractivité pour les pollinisateurs mais aussi par les pratiques culturales qui influencent le nombre de pollinisateur (Toledo-Hernandez et al., 2017). C'est donc un paramètre sur lequel pour 8 des 10 clones étudiés, le cacaoculteur pourrait jouer pour améliorer sa production.

Le modèle construit demande à être amélioré notamment pour l'estimation des paramètres de fertilité de l'ovule. De plus la loi d'avortement n'est probablement pas symétrique. Le modèle permet toutefois d'accéder à des paramètres très difficiles à mesurer comme la rareté du pollen mais il demande un effort de mesure important puisqu'il faut suffisamment de cabosses étudiées pour obtenir une distribution fiable.

Lorsque le pollen n'est pas limitant, la distribution est de type binomiale et la moyenne a un sens. Lorsque le pollen est limitant, la distribution est extrêmement étalée voire bimodale. La moyenne correspond à la composition de la loi de plusieurs phénomènes (fertilité de l'ovule et avortement) et n'a donc pas de sens biologique. La décomposition du nombre de cabosses en 3 phénomènes régulés par 5 paramètres pourrait permettre de mieux séparer les effets génétiques et environnementaux qui influencent le nombre de graines par cabosse.

## **Références**

Falque M., Vincent A., Vaissiere B.E., Eskes A.B. (1995). Effect of pollination intensity on fruit and seed set in cacao (*Theobroma cacao* L.). *Sex. Plant Reprod.*, 8 : 354-360

McKay Curtis S. (2015). mcmcplots: Create Plots from MCMC Output. R package version 0.4.2. <https://CRAN.R-project.org/package=mcmcplots>

R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.

de Reffye P., Parvais J.P., Mossu G., Lucas P. (1978). Influence des aléas de la pollinisation sur les rendements du cacaoyer. Modèle mathématique et simulation. *Café, Cacao, Thé*, 22 (4) : 251-274.

Su Y.S. and Yajima M. (2015). R2jags: Using R to Run 'JAGS'. R package version 0.5-7. <https://CRAN.R-project.org/package=R2jags>

Toledo-Hernández M., Wanger T. C. , Tschardt T. (2017). Neglected pollinators: Can enhanced pollination services improve cocoa yields? A review, In *Agriculture, Ecosystems & Environment*, 247 : 137-148, ISSN 0167-8809.

Wickham H. (2009). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.