

INSTITUT UNIVERSITAIRE TECHNOLOGIQUE

Département STID (Statistique et Informatique Décisionnelle)



Thème du stage : Régressions logistiques appliquées à des données binaires de phénologie d'arbres au nord-ouest du Bénin.

STAGE REALISE DU 18 AVRIL 2011 AU 30 JUIN 2011

Fait par Florian NICOLAS
Etudiant en DUT Statistique et Informatique Décisionnelle
A l'IUT d'Avignon
Promotion 2011

Tuteur HSM : Josiane SEGHERI
Tuteur Université : Delphine BLANKE



UNIVERSITÉ D'AVIGNON
ET DES PAYS DE VAUCLUSE

IUT d'Avignon
Université d'Avignon et des Pays du Vaucluse
337 chemin des Meinajaries, BP 1207, 84911 Avignon cedex 9
+33 (0)4 90 84 14 00



REMERCIEMENTS

Je remercie tout d'abord l'IRD de m'avoir donné la chance d'intégrer une de ses unités de recherche et d'avoir pu financer mon stage.

Puis, je remercie Eric SERVAT qui a accepté mon intégration dans le laboratoire dont il est le directeur.

Je remercie beaucoup ma directrice de stage : Josiane SEGHERI, chargée de recherche à l'IRD (Institut de Recherche pour le Développement) et pour sa grande disponibilité et ses conseils éclairés.

J'adresse aussi mes remerciements à Julie CARREAU, qui m'a fait découvrir une méthode pour combler les données manquantes.

Je remercie également toutes les personnes qui m'ont aidé au cours de ce stage, particulièrement les stagiaires du bureau n°130 dont Lhassan AMARSID, Kohomlan Gbénakpon AWESSOU, Bil ISSOUFOU HASSANE) pour leurs conseils et leur contribution à la bonne ambiance générale du bureau.

SOMMAIRE

REMERCIEMENTS	I
SOMMAIRE	II
TABLE DES FIGURES	IV
TABLEAUX	V
TABLE DES ANNEXES	VI
INTRODUCTION	1
I. HSM, UNITE MIXTE DE RECHERCHE DE L'IRD	2
1) L'IRD (INSTITUT DE RECHERCHE POUR LE DEVELOPPEMENT)	2
2) HYDROSCIENCES MONTPELLIER (HSM)	2
II. ETUDE DES FACTEURS ENVIRONNEMENTAUX INFLUENÇANT LA PHENOLOGIE D'ESPECES LIGNEUSES AU BENIN	4
1) CONTEXTE	4
2) SITES D'ETUDE ET TYPES DE COUVERTS	5
3) LES VARIABLES ANALYSEES	8
A) LES VARIABLES ENVIRONNEMENTALES	8
B) LES DONNEES PHENOLOGIQUES	8
4) TRAITEMENT PRELIMINAIRE DES DONNEES : RANGEMENT PAR QUINZE JOURS ET COMPLEMENT DES DONNEES MANQUANTES	9
A) ARRANGEMENT DES DONNEES	9
B) COMPLEMENT DES DONNEES MANQUANTES	10
5) TRAITEMENTS ET ANALYSES STATISTIQUES	11
III. RESULTATS	12
1) LES MODELES RETENUS	12
2) LES MEILLEURS PREDICTEURS	12

A) STATION « BELEFOUNGOU FORET »	12
B) STATION SAVANE ARBOREE DE BELEFOUNGOU	13
C) MOSAÏQUE DE CHAMPS ET DE JACHERES DE NALOHOU	14
D) SAVANE ARBUSTIVE DE BIRA	15

IV. DISCUSSION	16
-----------------------	-----------

V. CONCLUSION	18
----------------------	-----------

REFERENCES	19
-------------------	-----------

TABLE DES FIGURES

FIGURE 1 : SIEGE DE L'IRD	2
FIGURE 2 : LOCAUX D'HSM	3
FIGURE 3 : FENETRE D'OBSERVATION DU PROGRAMME DE RECHERCHE INTERNATIONAL SUR L'ANALYSE MULTIDISCIPLINAIRE DE LA MOUSSON AFRICAINE (AMMA)	4
FIGURE 4 : OCCUPATION DES SOLS DU BASSIN DE LA DONGA (BLANCHARD M ET AL. 2007)	6
FIGURE 5 : PHOTOGRAPHIE DE L'ESPECE <i>ISOBERLINIA DOKA</i> (SERVIERE ET AL. 2010)	7
FIGURE 6 : PHOTOGRAPHIE DE L'ESPECE <i>VITELLARIA PARADOXA</i> (KARITE) (SERVIERE ET AL. 2010)	7

TABLEAUX

TABLEAU 1 : MEILLEURS PREDICTEURS ENVIRONNEMENTAUX POUR L'ESPECE <i>ANNONA SENEGALENSIS</i> DANS LA FORET DE BELEFOUNGOU (« FE » = FEUILLAISSON , « FL » = FLORAISON, « FR » = FRUCTIFICATION, « + » = AUGMENTATION, « - » = BAISSSE, VARIABLES = VOIR PARTIE DU DESSUS)	12
TABLEAU 2 : MEILLEURS PREDICTEURS ENVIRONNEMENTAUX POUR L'ESPECE <i>ISOBERLINIA DOKA</i> DANS LA FORET DE BELEFOUNGOU. (« FE » = FEUILLAISSON , « FL » = FLORAISON, « FR » = FRUCTIFICATION, « + » = AUGMENTATION, « - » = BAISSSE, VARIABLES = VOIR PARTIE DU DESSUS)	13
TABLEAU 3 : MEILLEURS PREDICTEURS ENVIRONNEMENTAUX POUR L'ESPECE <i>OPILIA AMENTACEA</i> DANS LA FORET DE BELEFOUNGOU (« FE » = FEUILLAISSON, « FL » = FLORAISON, « FR » = FRUCTIFICATION, « + » = AUGMENTATION, « - » = BAISSSE, VARIABLES = VOIR PARTIE DU DESSUS)	13
TABLEAU 4 : MEILLEURS PREDICTEURS ENVIRONNEMENTAUX POUR L'ESPECE <i>PAVETTA CRASSIPES</i> DANS LA SAVANE ARBOREE DE BELEFOUNGOU (« FE » = FEUILLAISSON, « FL » = FLORAISON, « FR » = FRUCTIFICATION, « + » = AUGMENTATION, « - » = BAISSSE, VARIABLES = VOIR PARTIE DU DESSUS)	13
TABLEAU 5 : MEILLEURS PREDICTEURS ENVIRONNEMENTAUX POUR L'ESPECE <i>VITELLARIA PARADOXA</i> DANS LA SAVANE ARBOREE DE BELEFOUNGOU (« FE » = FEUILLAISSON, « FL » = FLORAISON, « FR » = FRUCTIFICATION, « + » = AUGMENTATION, « - » = BAISSSE, VARIABLES = VOIR PARTIE DU DESSUS)	14
TABLEAU 6 : MEILLEURS PREDICTEURS ENVIRONNEMENTAUX POUR L'ESPECE <i>XERODERRIS STUHLMANNII</i> DANS LA SAVANE ARBOREE DE BELEFOUNGOU (« FE » = FEUILLAISSON, « FL » = FLORAISON, « FR » = FRUCTIFICATION, « + » = AUGMENTATION, « - » = BAISSSE, VARIABLES = VOIR PARTIE DU DESSUS)	14
TABLEAU 7 : MEILLEURS PREDICTEURS ENVIRONNEMENTAUX POUR L'ESPECE <i>BURKEA AFRICANA</i> DANS LA MOSAÏQUE DE CHAMPS ET DE JACHERES DE NALOHOU (« FE » = FEUILLAISSON, « FL » = FLORAISON, « FR » = FRUCTIFICATION, « + » = AUGMENTATION, « - » = BAISSSE, VARIABLES = VOIR PARTIE DU DESSUS)	14
TABLEAU 8 : MEILLEURS PREDICTEURS ENVIRONNEMENTAUX POUR L'ESPECE <i>PARINARI CURATELLIFOLIA</i> DANS LA MOSAÏQUE DE CHAMPS ET DE JACHERES DE NALOHOU (« FE » = FEUILLAISSON, « FL » = FLORAISON, « FR » = FRUCTIFICATION, « + » = AUGMENTATION, « - » = BAISSSE, VARIABLES = VOIR PARTIE DU DESSUS)	15
TABLEAU 9 : MEILLEURS PREDICTEURS ENVIRONNEMENTAUX POUR L'ESPECE <i>PARKIA BIGLOBOSA</i> DANS LA MOSAÏQUE DE CHAMPS ET DE JACHERES DE NALOHOU (« FE » = FEUILLAISSON, « FL » = FLORAISON, « FR » = FRUCTIFICATION, « + » = AUGMENTATION, « - » = BAISSSE, VARIABLES = VOIR PARTIE DU DESSUS)	15
TABLEAU 10 : MEILLEURS PREDICTEURS ENVIRONNEMENTAUX POUR L'ESPECE <i>BURKEA AFRICANA</i> DANS LA SAVANE ARBUSTIVE DE BIRA (FE » = FEUILLAISSON, « FL » = FLORAISON, « FR » = FRUCTIFICATION, « + » = AUGMENTATION, « - » = BAISSSE, VARIABLES = VOIR PARTIE AU DESSUS)	15
TABLEAU 11 : MEILLEURS PREDICTEURS ENVIRONNEMENTAUX POUR L'ESPECE <i>COMBRETUN COLLINUM</i> DANS LA SAVANE ARBUSTIVE DE BIRA (FE » = FEUILLAISSON, « FL » = FLORAISON, « FR » = FRUCTIFICATION, « + » = AUGMENTATION, « - » = BAISSSE, VARIABLES = VOIR PARTIE AU DESSUS)	16
TABLEAU 12 : MEILLEURS PREDICTEURS ENVIRONNEMENTAUX POUR L'ESPECE <i>DANIELLIA OLIVERI</i> DANS LA SAVANE ARBUSTIVE DE BIRA (FE » = FEUILLAISSON, « FL » = FLORAISON, « FR » = FRUCTIFICATION, « + » = AUGMENTATION, « - » = BAISSSE, VARIABLES = VOIR PARTIE AU DESSUS)	16

TABLE DES ANNEXES

ANNEXE 1 : EXEMPLE DE SCRIPT R - UTILISATION DU PACKAGE MIX	A
ANNEXE 2 : SCRIPT R QUI PERMET LE PASSAGE DES COMPOSANTES PRINCIPALES SIGNIFICATIVES AUX VARIABLES D'ORIGINE	B
ANNEXE 3 : RESULTATS DES MODELES LOGISTIQUES OBTENUS STIMULANT LES EFFETS DES VARIABLES PSEUDO PREDICTIVES. OBSERVATIONS = NOMBRE D'OBSERVATIONS ANALYSEES, PSEUDO R²=COEFFICIENT DU PSEUDO R², AIC= AKAIKE INFORMATION CRITERION, SIGN = NIVEAU DE SIGNIFICATIVITE, NON SIGNIFICATIF NS ($P>0.05$), * $P<0.05$, ** $P<0.01$, ***$P<0.001$), FE=FEUILLAISSON, FL=FLORAISSON, FR=FRUCTIFICATION. C	
ANNEXE 4 : CLASSIFICATION DES PREDICTEURS ENVIRONNEMENTAUX POUR LES ESPECES <i>A. SENEGALENSIS</i>, <i>I.DOKA</i>, <i>O. AMENTACEA</i> DANS LA FORET DE BELEFOUNGOU	E
ANNEXE 5 : CLASSIFICATION DES PREDICTEURS ENVIRONNEMENTAUX POUR LES ESPECES <i>P. CRASSIPES</i>, <i>V. PARADOXA</i>, <i>X. STUHLMANNII</i> DANS LA SAVANE ARBOREE DE BELEFOUNGOU	F
ANNEXE 6 : CLASSIFICATION DES PREDICTEURS ENVIRONNEMENTAUX POUR LES ESPECES <i>B. AFRICANA</i>, <i>P. CURATELLIFOLIA</i>, <i>P. BIGLOBOSA</i> DANS LA MOSAIQUE DE CHAMPS ET JACHERES DE NALOHOU	G

INTRODUCTION

Dans le cadre de notre formation (DUT Statistique et Informatique Décisionnelle), je réalise un stage de fin d'étude de dix semaines. Ce stage a pour objectif de mettre en application, dans le milieu de l'entreprise, les connaissances que j'ai acquises au cours de ma formation.

Après plusieurs demandes de stages adressées ailleurs, Josiane SEGHIÉRI m'accepte finalement pour un stage du 18 Avril 2011 au 27 Juin 2011 au sein de l'Unité Mixte de Recherche HSM (HydroSciences Montpellier).

Les objectifs de ce stage sont 1) d'établir les meilleurs modèles expliquant la phénologie des espèces d'arbres présents au nord du Bénin par des variables environnementales continues supposées être parmi les plus déterminantes 2) d'identifier les variables continues les plus significatives.

Ce rapport de stage décrit l'IRD et plus particulièrement le laboratoire Hydrosiences Montpellier (HSM). Il fournit également des informations sur le déroulement du stage (objectifs, lieu de collecte des données, types de couvert étudiés, méthodes statistiques utilisées). Il énonce ensuite les résultats obtenus, puis les discute en termes de critique des méthodes choisies par rapport aux objectifs fixés (avantages, limites...). Pour finir, une conclusion est faite sur le déroulement du stage et les enseignements retenus (impressions personnelles ...).

I. HSM, Unité Mixte de Recherche de l'IRD

1) L'IRD (Institut de Recherche pour le Développement)

L'institut de Recherche pour le Développement (IRD, ex-ORSTOM) est un établissement public français à caractère scientifique et technologique qui intervient depuis plus de 60 ans dans les pays du Sud. Il est placé sous la double tutelle des ministères chargés de la Recherche et de la Coopération. Ses activités de recherche, d'expertise, de valorisation et de formation ont pour objectif de contribuer au développement économique, social et culturel des pays du Sud.



Figure 1: Siège de l'IRD

En étroite collaboration avec leurs partenaires, 858 chercheurs, 973 ingénieurs et techniciens et 341 personnels locaux interviennent dans une cinquantaine de pays et participent à de nombreux programmes nationaux, européens et internationaux.

Depuis septembre 2008, le siège de l'IRD est localisé à Marseille. L'institut dispose par ailleurs de 30 implantations dont 2 en France métropolitaine (Bondy, Montpellier), 5 dans les régions et collectivités d'outre-mer (la Réunion, Guyane, Martinique, Nouvelle-Calédonie, Polynésie française) et 23 dans des pays situés dans la zone intertropicale, en Afrique, en Méditerranée, en Asie et en Amérique Latine.

L'IRD compte 60 unités : 2 unités de service (US), 5 unités de recherche propres (UR), 2 unités mixtes internationales (UMI) et 51 unités mixtes de recherche (UMR) dont mon unité d'accueil.

2) Hydrosiences Montpellier (HSM)

Le Laboratoire HydroSciences Montpellier est l'Unité Mixte de Recherche 050 de l'IRD. C'est une unité mixte car elle réunit des personnels du CNRS, de l'Université Montpellier 1, de l'Université Montpellier 2, en plus de ceux de l'IRD.



Figure 2 : Locaux d'HSM

HSM est fortement impliquée dans plusieurs structures de formation en Sciences de la Terre, de l'environnement et de l'eau : Département d'Enseignement des Sciences de la Terre et de l'Environnement de Montpellier ; Polytech Montpellier (EPUM) ; Institut Universitaire Professionnel (IUP) « Ingénierie de la Santé » ; Ecole doctorale des Sciences de la Terre, de l'Eau et de l'Espace. Au sein des universités Montpellier 1 et 2, HSM participe à de nombreuses formations et en pilote certaines comme : Licence Géosciences « Prévention et Traitement des Pollutions » (Licence GPTP), Licence « Science de la Terre et de l'Environnement » (Licence STE) ; Master Professionnel « Gestion et Evaluation des Ressources en Eau » (Master GERE) ; Master Recherche Eau et Environnement (Master R2E) ; Master Professionnel « Gestion de Projet en Environnement Santé » Master (GPES) ; Doctorats.

Les activités scientifiques d'Hydrosciences Montpellier se traduisent par des collaborations nombreuses avec diverses structures de recherche et d'enseignement supérieur, principalement autour du bassin méditerranéen et en Afrique subsaharienne. Ses implantations dans plusieurs pays (Tunisie, Liban, Bénin, Mali) contribuent efficacement à renforcer les équipes locales et à développer des partenariats équilibrés. La participation active d'HSM aux réseaux scientifiques FRIEND de l'UNESCO conforte son engagement méditerranéen et ouest-africain.

HSM a un effectif totale de 130 personnes dont 90 permanents et 30 doctorants et posts doctorants. HSM cible son activité de recherche sur les ressources en eau. Le laboratoire développe ses travaux autour de trois thèmes clés : l'eau dans l'environnement et les risques associés, l'eau ressource mobilisable et exploitable, l'eau et la variabilité du climat. Hydrosciences Montpellier s'intéresse prioritairement aux régions méditerranéennes et tropicales.

II. Etude des Facteurs Environnementaux influençant la Phénologie d'Espèces ligneuses au Bénin

1) Contexte

L'étude se déroule sur le super site méso-échelle du programme de recherche international sur l'Analyse Multidisciplinaire de la Mousson Africaine (AMMA) au Nord Ouest du Bénin (voir **Figure 3**) : le bassin versant de la Donga (voir **Figure 4**). Ce bassin versant est situé entre les parallèles 9°40' et 9°54'N et les méridiens 1°34' et 1°58' E, dans le département du même nom, dont le chef lieu est Djougou. C'est un sous-bassin de l'Ouémé qui s'étend sur une superficie de 586 km². Le climat est du type tropical sec avec une saison sèche et une saison des pluies. Cette zone est classée par Aubréville (1950) comme soudano-guinéenne (1000 à 1400 mm de pluie par an en moyenne). L'activité socio-économique, hormis la ville de Djougou (20 000 ha) est essentiellement rurale et basée sur l'agriculture pluviale (igname, manioc, arachide) et l'élevage.

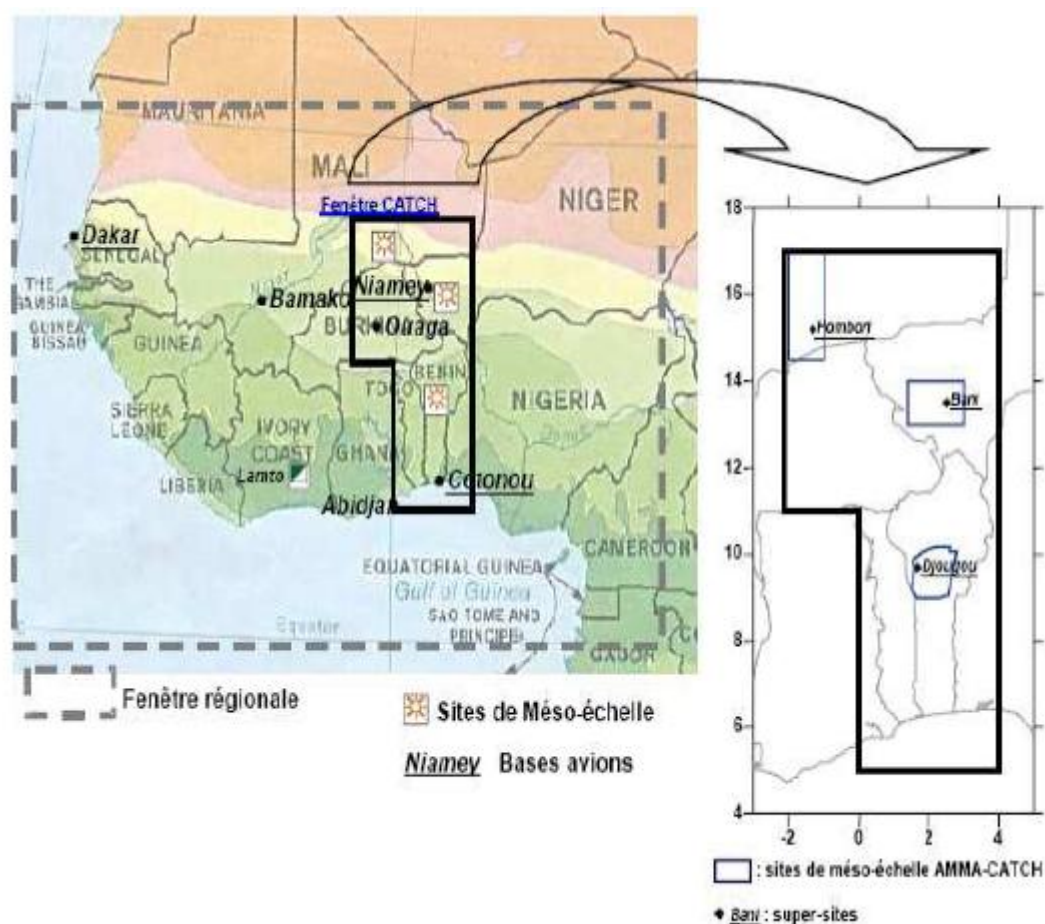


Figure 3 : Fenêtre d'Observation du programme de recherche international sur l'Analyse Multidisciplinaire de la Mousson Africaine (AMMA) (Blanchard M *et al.* 2007)

La phénologie des plantes est l'étude des rythmes de feuillaison, floraison et fructification (appelés phénophases) et de leurs causes (Lieth 1974; Haugaasen & Peres 2005). Le besoin d'identifier les déterminants qui prédisent les réponses spécifiques des plantes est devenu de plus en plus important durant la dernière décennie dans le contexte de changement climatique (van Vliet & Schwartz 2002 ; Chapman *et al.* 2005).

Dans le présent travail, nous allons quantifier statistiquement le pouvoir prédictif de quelques facteurs environnementaux sur les phénophases des espèces ligneuses dominantes au nord du Bénin. La répartition des pluies au cours de la saison des pluies varie entre les années, mais la périodicité des pluies est reconnue par de nombreux auteurs comme étant l'un des facteurs les plus déterminants de l'occurrence des phénophases (Seghieri *et al.* soumis)

Les deux hypothèses testées dans cette étude sont 1) la majorité des plantes tropicales ne répondent pas à un seul prédicteur mais à une combinaison de prédicteurs (Wright & Cornejo 1990), et 2) la dépendance des phénophases vis-à-vis des prédicteurs varie avec les espèces (Singh & Kishwaha 2006). La phénologie des espèces ligneuses dominantes a été observée sur quatre types de couvert dominant le site d'étude. Nous utiliserons les régressions logistiques multi-variées pour identifier les modèles les plus pertinents pour prédire les phénophases d'une espèce donnée, puis pour en déduire les variables prédictives les plus significatives.

2) Sites d'Etude et types de Couverts

Les quatre stations sont situées sur les quatre types de couverts végétaux représentatifs de la majorité des surfaces non cultivées de la zone d'étude. Le choix des sites s'est fait sur trois principaux critères :

- Sa représentativité par rapport à l'ensemble du bassin de la Donga, en prenant en compte le substrat, la densité du peuplement ligneux et l'intensité de l'impact anthropique (activités pastorales et agricoles).
- L'homogénéité de la végétation et de son environnement sur une surface suffisante.
- L'accessibilité en saison des pluies (état des pistes principalement).

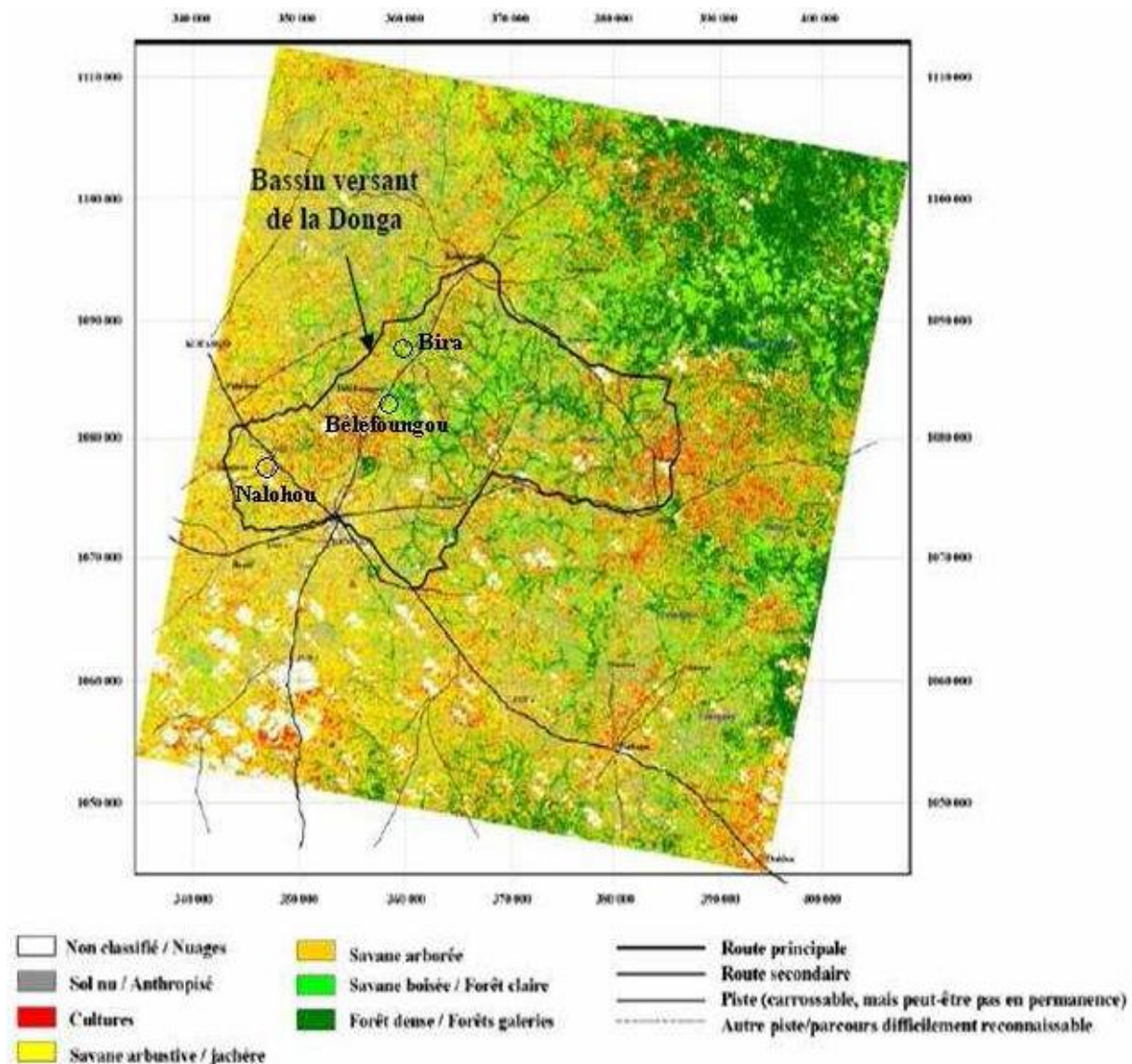


Figure 4 : Occupation des Sols du Bassin de la Donga (Blanchard M et al. 2007)

Il s'agit de :

- La forêt protégée de Béléfoungou

Dans la forêt protégée de Béléfoungou, qui subit également pâturage et prélèvement de bois, le couvert est typiquement forestier avec des individus atteignant jusqu'à 35 m de hauteur. Aujourd'hui, c'est une forêt claire à *Isoberlinia doka* (Aubréville 1950). Structurellement parlant, il s'agit d'une formation mixte, ligneuse et herbacée, mais la strate herbacée souffre de la compétition pour la lumière et reste de faible hauteur comparée aux autres stations (moins d'1 m). On y retrouve les graminées pérennes du groupe *Hyparrhenia* et *Andropogon*. Les espèces ligneuses dominantes sont *Isoberlinia doka* et *Isoberlinia tormentuosa*. L'espèce ligneuse co-dominante est *Pterocarpus erinaceus*.



Figure 5 : Photographie de l'espèce *Isobertinia doka* (Serviere et al. 2010)

- La savane arborée de Béléfoungou

La savane arborée de Béléfoungou est une jachère agée de 10 à 15 ans, après une culture d'igname et/ou de manioc. Actuellement l'exploitation de la zone se fait uniquement sous forme de prélèvements du bois de chauffe et du pâturage de transhumance de bovins. Les formations végétales s'y développent sur un sol mince et partiellement rocheux. La strate herbacée peut atteindre jusqu'à plusieurs mètres de hauteur en saison pluvieuse, avec une contribution importante des graminées pérennes typiques des savanes soudaniennes à *Hyparrhenia involucrata* et *Andropogon gayanus*.

- La savane arbustive de Bira

Ce site est une jeune jachère relativement homogène et traduit ici une mise en culture relativement récente. Selon la nomenclature de Yangambi (Trochain, 1957), ce site est une savane arbustive. Sa strate herbacée est dominée par des graminées pérennes et des karités (*Vitellaria paradoxa*) Le site est parsemé de quelques arbres de hauteur supérieure à 30 m conservés volontairement par les paysans pour leur intérêt économique (karité, néré). La pression anthropique actuelle s'y exerce par du pâturage et le prélèvement de bois de chauffe.



Figure 6 : Photographie de l'espèce *Vitellaria paradoxa* (karité) (Serviere et al. 2010)

- La mosaïque de champs et jachères annuelles de Nalohou

Du point de vue de la physionomie le site de Nalohou peut être considéré comme un « parc à Karité et à Néré », car ces espèces sont toutes deux conservées par les paysans dans les champs. On y trouve également une dizaine de Baobab. Ce site est soumis à une forte pression culturelle. Il présente une très forte hétérogénéité constituée de champs et de jeunes jachères dont la répartition évolue d'une année sur l'autre. Les herbacées pérennes y sont minoritaires et les annuelles dominent dans les parcelles en jachère.

Les trois sites présentés précédemment sont soumis aux feux de brousse contrôlés sauf la forêt protégée de Béléfoungou (sauf accident)

3) Les variables analysées

a) Les variables environnementales

Les variables environnementales testées sont :

- humidité relative de l'air
- humidité du sol jusqu'à 1m de profondeur
- température de l'air (en °C)
- pluie journalière cumulée (en mm)
- durée du jour (en minutes)

Dans le cadre du programme AMMA, une station météorologique a été implantée dans la ville de Djoungou à proximité des trois sites. Les variables mesurées avec un capteur météorologique Vaisala sont l'humidité relative de l'air et la température de l'air, toutes les 15 min au dessus de la canopée.

Les données de pluie ont été mesurées toutes les 15 min avec un pluviomètre automatique installé sur chaque type de couvert, sauf pour la savane de Béléfoungou à cause de sa proximité avec la forêt.

Concernant l'humidité du sol, des mesures automatiques utilisées ont été réalisées avec des capteurs TDR à 6 profondeurs différentes, jusqu'à 1 m en dessous du sol, à une fréquence horaire, en haut et milieu de pente pour les couverts de forêt de Béléfoungou, Bira et Nalohou. La durée du jour a été calculée à partir du site web « Rise and for the Sun » du département d'applications astronomiques de l'Observatoire Naval des EU à Washington. (http://aa.usno.navy.mil/data/docs/RS_OneYear.php).

Les variables environnementales utilisées dans l'analyse statistique sont les valeurs journalières calculées (cumul ou moyennes) à partir de ces données brutes.

b) Les données phénologiques

Les variables à expliquer sont issues d'un jeu de données binaires (occurrence/non occurrence) sur la phénologie (feuillaison, floraison fructification respectivement) des espèces ligneuses dominantes observées sur différentes périodes sur chacun des quatre types de couvert :

- d'avril 2004 à mars 2007 pour la forêt de Béléfoungou, la savane arborée de Béléfoungou et la mosaïque de champs et de jachères annuelles de Nalohou.
- d'avril 2006 à mai 2007 pour la savane arbustive de Bira

Les observations étaient faites toutes les deux semaines en saison pluvieuse et tous les mois en saison sèche. Les espèces se répartissent très inégalement dans les quatre types de couvert végétal. Les arbres dont la phénologie était suivie étaient répartis le long de transects de 500m à 1km de long sur chaque type de couvert. Pour chaque individu, deux états étaient notés :

0 = « non occurrence », correspondant à l'absence de phénophase, où à son commencement (bourgeons foliaires ou floraux, stades les plus précoces pour les fruits, organes matures occupants au maximum 10% de la canopée), ou à sa fin (à partir de 50% de la canopée couvert par des organes sénescents, début de dissémination des graines).

1 = « occurrence » (à partir de 10% de la canopée couvert par des organes matures jusqu'à ce que 50% soit couvert par des organes sénescents)

De cette façon, les débuts de phénophase qui avortent et les longues périodes en fin de phénophase pendant lesquelles les arbres montrent des organes secs (cas des fruits) ne sont pas considérés comme occurrence. 12 individus par espèces ont été observés dans chaque station.

4) Traitement préliminaire des données : rangement par quinze jours et comblement des données manquantes

Cette partie présente le traitement préalable aux analyses statistiques. L'objectif est d'établir pour chaque espèce dans chaque type de couvert, un jeu de données réunissant à la fois les variables phénologiques à expliquer et les variables environnementales prédictives correspondant à une fréquence de 15 jours. Les dates de début et de fin sont les mêmes que celles des observations phénologiques. (un fichier Excel correspond à un site d'étude et une feuille de classeur correspond à une espèce).

a) Arrangement des données

Je suis parti de plusieurs fichiers correspondant chacun à une variable environnementale. Un dernier fichier correspondait aux données phénologiques par espèces et par site. Toutes les données disponibles ne correspondant pas aux mêmes périodes ni aux mêmes fréquences au cours de ces périodes pour la phénologie, il en est résulté des données manquantes dans les jeux de données finaux, par exemple la phénologie a commencé à être observée en avril 2004 et les mesures d'humidité du sol ont commencées début 2005. Des périodes de plusieurs mois sont restées sans observations également.

Pour répondre à l'objectif, j'ai construit d'abord deux tables de données :

- Une table de valeurs phénologiques tous les 15 jours à partir de la première date d'observation. Les variables dans cette table de valeurs sont : *Fe* pour Feuillaison, *Fl* pour Floraison et *Fr* pour Fructification. Si une observation phénologique du jeu de données brut initial était située 3 à 4 jours avant/après la date de la nouvelle table, je conservais la valeur de l'observation dans la nouvelle table. En effet, l'inertie phénologique est de l'ordre de la semaine. Nous appelleront cette nouvelle table la « Table 1 ».
- Un jeu de variables environnementales journalières avec les variables *Date*, *Pluie_Cumul* (pluie cumulée), *T_Air* (= température de l'air), *HR_Air* (= humidité relative de l'air), *HS0_100* (=humidité du sol de 0 à 100cm). Pour les données d'humidité du sol, la température de l'air, l'humidité relative de l'air (autres fichiers),

ayant des données horaires, j'ai du faire moi-même la moyenne journalière de ces données.

- Ayant ainsi une table de données environnementales journalières pour chaque site, j'ai ensuite créé une nouvelle table en prenant les données environnementales correspondant aux mêmes dates que la « Table 1 ». Puis j'ai ajouté les valeurs des mêmes prédictors mais 1, 2, 4 et 8 semaines avant. Cela nous permet de tenir compte de l'inertie de la réaction de la végétation à son environnement. J'appelle ce jeu de données la « Table 2 »

Ainsi, les variables prédictives constituant la « Table 2 » sont :

- *Pluie_Cum* (pluie cumulée)
- *Pluie_Cum_-1* (pluie cumulée 1 semaine avant)
- *Pluie_Cum_-2* (pluie cumulée 2 semaines avant)
- *Pluie_Cum_-4* (pluie cumulée 4 semaines avant)
- *Pluie_Cum_-8* (pluie cumulée 8 semaines avant)
- *T_Air* (= température de l'air)
- *T_Air_-1* (= température de l'air 1 semaine avant)
- *T_Air_-2* (= température de l'air 2 semaines avant)
- *T_Air_-4* (= température de l'air 4 semaines avant)
- *T_Air_-8* (= température de l'air 8 semaines avant)
- *HR_Air* (= humidité relative de l'air)
- *HR_Air_-1* (= humidité relative de l'air 1 semaine avant)
- *HR_Air_-2* (= humidité relative de l'air 2 semaines avant)
- *HR_Air_-4* (= humidité relative de l'air 4 semaines avant)
- *HR_Air_-8* (= humidité relative de l'air 8 semaines avant)
- *HS_Sol* (=humidité du sol de 0 à 100cm)
- *HS_Sol_-1* (=humidité du sol de 0 à 100cm 1 semaine avant)
- *HS_Sol_-2* (=humidité du sol de 0 à 100cm 2 semaines avant)
- *HS_Sol_-4* (=humidité du sol de 0 à 100cm 4 semaines avant)
- *HS_Sol_-8* (=humidité du sol de 0 à 100cm 8 semaines avant)

La date constitue la colonne commune entre la « Table 1 » et la « Table 2 ».

J'ai finalement ajouté les colonnes correspondant à la durée du jour à chaque date et 1 à 8 semaines avant.

J'ai alors combiné les données phénologiques de la « Table 1 » et les données environnementales de la « Table 2 » dans un nouveau fichier « Table 3 » pour chaque type de couvert.

L'ensemble du rangement de données m'a pris 7 semaines.

b) Comblement des données manquantes

J'ai utilisé le package *mix* du logiciel R qui impose certaines conditions :

1. Les premières colonnes du fichier de données doivent être obligatoirement occupées par les variables qualitatives. Il faut en préciser le nombre dans les arguments de la fonction *mix*.

2. Les valeurs qualitatives doivent être obligatoirement en 1/2 (et non en 0/1). Je devais donc opérer un remplacement des valeurs de la « Table 3 ».
3. Aucune ligne ne doit être complètement vide (les données de la longueur du jour étant calculées à chaque date, cette condition était satisfaite) sans autre traitement.
4. La distribution des données environnementales doit suivre celle d'une loi normale.

On installe tout d'abord le package *mix* sur R.

Ensuite on génère la librairie du package avec la commande `library(mix)`. On importe ensuite le fichier qui contient les données manquantes avec la fonction `read.delim`. On crée l'objet sur lesquels les packages vont travailler avec la commande `s <- prelim.mix(X, 3)`. Pour trouver les meilleurs paramètres en fonction du jeu de données et du critère de performance, on utilise la commande `em.mix(s)`. Après on extrait les paramètres du modèle par la ligne de commande `getparam.mix(s, thetahat, corr=TRUE)`. La commande `rngseed(1234567)` est un générateur de nombre aléatoire (ici de 1 jusqu'à 7). C'est enfin avec les commandes `newtheta <- da.mix(s, thetahat, steps=0, showits=TRUE)` et `ximp1 <- imp.mix(s, newtheta)` que les données manquantes sont complétées.

Le package n'a fonctionné sur aucune des espèces de la station Bira. J'ai donc décidé d'éliminer les lignes qui contenaient au moins une donnée manquante avec la fonction `na.omit`. En travaillant sur la matrice réduite, le problème a été éliminé.

5) Traitements et analyses statistiques

Les modèles de régressions logistiques estiment la probabilité d'apparition de la phénophase à partir d'une combinaison linéaire de variables prédictives. Les coefficients du modèle sont calculés en utilisant la probabilité maximale estimée. La qualité globale du modèle est donnée par le critère AIC (Akaike Information Criterion) qui compare le fit d'un modèle donné incluant les prédicteurs au fit d'un modèle excluant les prédicteurs. Le pseudo R^2 de Mc Fadden indique la proportion de variance d'une phénophase expliquée par le modèle. Le niveau de significativité des coefficients du modèle est obtenu par le test de Wald.

Les traitements et analyses statistiques se déroulent en trois étapes :

- Pour éviter les biais dus à la dépendance entre les variables environnementales, j'applique une ACP (Analyse en Composantes Principales) aux variables environnementales pour générer des pseudos variables prédictives non corrélées entre elles. Je standardise (moyenne=0 et écart type = 1) ces variables dans l'ACP. Je retiens les composantes principales qui expliquent au moins 90 % de la variabilité des variables initiales. J'utilise le logiciel statistique R pour faire l'ACP
- Je teste ensuite des modèles de régression logistique. Je l'applique pour chaque phénophase de chaque espèce de chaque type de couvert en incluant tout d'abord toutes les pseudo-variables prédictives avec le logiciel Stata dans lequel je peux avoir le Pseudo R^2 et le critère AIC avec l'option Robust. Stata 9.1 permet de calculer facilement l'erreur standard avec l'option « robust » avec des biais statistiques correctes causé par l'hétéroscédasticité. Puis je teste d'autres modèles en enlevant progressivement les pseudos variables non significatives. Un compromis entre un petit AIC et un grand pseudo R^2 me permet de sélectionner le modèle optimal. Ce dernier

peut donc inclure des variables prédictives non significatives si l'enlèvement de celles-ci provoque une augmentation trop forte de l'AIC ou une diminution trop importante du Pseudo R². Le choix du meilleur modèle se fait

- Une régression linéaire multiple permet enfin de revenir aux variables environnementales initiales en reliant les pseudos variables dans l'espace de l'ACP aux variables environnementales d'origine dans leur espace d'origine. Les coefficients obtenus fournissent une information sur les plus forts prédicteurs environnementaux (voir **Annexe 2**).

Les résultats de ces analyses sont décrits ci-dessous

III. Résultats

1) Les modèles retenus

J'ai pris en exemple trois espèces pour chaque type de couvert.

La plupart des modèles logistiques testés ont été très fortement significatifs ($p < 0.001$) (voir Annexe 3). Dans la plupart des modèles, plus de 50 % de la variance est expliquée (Pseudo R²), bien que dans d'autres (fructification de *B. africana*, *D. oliveri*, *B. africana*) moins de 30% de la variance est expliquée.

Lorsque un seul état de la phénophase a été observé (0 ou 1), je n'ai pas pu exécuter les modèles, faute de variation. Cela a été le cas concernant la feuillaison de *D. oliveri* et *B. africana* pour la station Bira. Parce qu'il y avait plusieurs prédicteurs environnementaux, je les ai classé suivant les espèces et les catégories de valeurs (Voir **Annexe 4**, **Annexe 5**, **Annexe 6**).

2) Les meilleurs prédicteurs

a) Station « Belefoungou Forêt »

Les résultats sur les facteurs environnementaux influençant la phénologie des plantes de la forêt protégée de Béléfoungou sont décrits ci-dessous. Les espèces étudiées sont *Annona senegalensis*, *Isobertinia doka* et *Opilia amentacea*.

- Espèce *Annona senegalensis*

Tableau 1 : Meilleurs prédicteurs environnementaux pour l'espèce *Annona senegalensis* dans la forêt de Béléfoungou (« Fe » = Feuillaison, « Fl » = Floraison, « Fr » = Fructification, « + » = augmentation, « - » = baisse, variables = voir partie du dessus)

	<u>Prédicteurs par ordre décroissant de pouvoir prédictif</u>
<u>Fe</u>	+ T_Air_-4 ; - T_Air_-1 ; + HR_Air ; + HR_Air_-1 ; + HR_Air_-2 ; + HR_Air_-4 ; - Pluie_Cum_-8 ; + Long_Jour_-1 ; + Long_Jour ; + Pluie_Cum
<u>Fl</u>	- Pluie_Cum_-2 ; - Pluie_Cum_-4 ; - Pluie_Cum_-1 ; + T_Air ; - Pluie_Cum - T_Air_-8 ; - HR_Air_-8 ; + HS_Sol_-1 ; - HS_Sol_-8 ; + T_Air.1
<u>Fr</u>	- T_Air_-4 ; + T_Air_-8 ; - T_Air_-2 ; - HS_Sol_-8 ; - Pluie_Cum ; - Pluie_Cum_-1 + Long_Jour_-1 ; + Long_Jour ; + Long_Jour_-2 ; + Long_Jour_-4

- Espèce *Isoberlinia doka*

Tableau 2 : Meilleurs prédicteurs environnementaux pour l'espèce *Isoberlinia doka* dans la forêt de Béléfoungou. (« Fe » = Feuillaison, « Fl » = Floraison, « Fr » = Fructification, « + » = augmentation, « - » = baisse, variables = voir partie du dessus)

	<u>Prédicteurs par ordre décroissant de pouvoir prédictif</u>
<u>Fe</u>	+ Pluie_Cum_-2 ; - HR_Air ; - HS_Sol ; - HS_Sol_-1 ; + Pluie_Cum_-1 ; - HR_Air_-1 ; + Pluie_Cum_-4 ; + T_Air_-8 ; + Pluie_Cum ; - HS_Sol_-2
<u>Fl</u>	- Pluie_Cum_-2 ; - Pluie_Cum_-4 ; - Pluie_Cum_-1 ; + T_Air ; - Pluie_Cum ; - T_Air_-8 ; -HR_Air_-8 ; + HS_Sol_-1 ; - HS_Sol_-8 ; + T_Air_-1
<u>Fr</u>	- T_Air_-4 ; + T_Air_-8 ; - T_Air_-2 ; - HS_Sol_-8 ; - Pluie_Cum ; - Pluie_Cum_-1 ; + Long_Jour_-1 ; + Long_Jour ; + Long_Jour_-2 ; + Long_Jour_-4

- Espèce *Opilia amentacea*

Tableau 3 : Meilleurs prédicteurs environnementaux pour l'espèce *Opilia amentacea* dans la forêt de Béléfoungou (« Fe » = Feuillaison, « Fl » = Floraison, « Fr » = Fructification, « + » = augmentation, « - » = baisse, variables = voir partie du dessus)

	<u>Prédicteurs par ordre décroissant de pouvoir prédictif</u>
<u>Fe</u>	+ T_Air_-4 ; + T_Air_-2 ; + HR_Air_-1 ; - Pluie_Cum_-8 ; + HR_Air ; + Pluie_Cum ; - Pluie_Cum_-4 ; + Pluie_Cum_-1 ; + Pluie_Cum_-1 ; + HS_Sol_-8 ; - Long_Jour
<u>Fl</u>	+ T_Air_-4 ; - T_Air_-8 ; + T_Air_-2 ; + T_Air ; + T_Air_-1 ; - Long_Jour_-4 ; - Long_Jour_-8 ; - Long_Jour_-2 ; - Long_Jour_-1 ; - Long_Jour
<u>Fr</u>	+ T_Air_-4 ; + T_Air_-2 ; - Pluie_Cum_-8 ; + HR_Air_-1 ; + HR_Air ; + Pluie_Cum ; - HS_Sol_-2 ; - Pluie_Cum_-4 ; + Long_Jour ; + Long_Jour_-1

b) Station Savane arborée de Béléfoungou

Les résultats sur les facteurs environnementaux influençant la phénologie des plantes de la savane arborée de Béléfoungou sont décrits ci-dessous. Les espèces étudiées sont *Pavetta crassipes*, *Vitellaria paradoxa* et *Xeroderris stuhlmannii*. (Voir **Annexe 5**)

- Espèce *Pavetta crassipes*

Tableau 4 : Meilleurs prédicteurs environnementaux pour l'espèce *Pavetta crassipes* dans la savane arborée de Béléfoungou (« Fe » = Feuillaison, « Fl » = Floraison, « Fr » = Fructification, « + » = augmentation, « - » = baisse, variables = voir partie du dessus)

	<u>Prédicteurs par ordre décroissant de pouvoir prédictif</u>
<u>Fe</u>	+ HR_Air_-8 ; - T_Air ; + HR_Air_-4 ; + Long_Jour_-8 ; + HR_Air_-2 ; + Long_Jour_-4 ; + HR_Air ; + HR_Air_-1 ; + Long_Jour_-2 ; - T_Air_-1
<u>Fl</u>	+ T_Air_-4 ; + T_Air_-2 ; - HR_Air_-8 ; - HR_Air_-4 ; - Long_Jour_-8 ; + Long_Jour ; - HR_Air_-2 ; + HR_Air ; + HR_Air_-1 ; + Long_Jour_-1
<u>Fr</u>	- T_Air_-4 ; + T_Air ; - HR_Air_-8 ; - T_Air_-2 ; + Long_Jour ; + Long_Jour_-1 ; + Long_Jour_-2 ; + T_Air_-8 ; + Long_Jour_-4 ; - HR_Air_-4

- Espèce *Vitellaria paradoxa*

Tableau 5 : Meilleurs prédicteurs environnementaux pour l'espèce *Vitellaria paradoxa* dans la savane arborée de Béléfoungou (« Fe » = Feuillaison, « Fl » = Floraison, « Fr » = Fructification, « + » = augmentation, « - » = baisse, variables = voir partie du dessus)

	<u>Prédicteurs par ordre décroissant de pouvoir prédictif</u>
<u>Fe</u>	+ T_Air_-2 ; - T_Air ; + T_Air_-4 ; + HR_Air_-1 ; + HR_Air_-8 ; + HR_Air_-4 ; - T_Air_-1 ; + HR_Air ; + Long_Jour_-8 ; + Long_Jour_-4
<u>Fl</u>	-
<u>Fr</u>	+ T_Air_-8 ; + T_Air_-2 ; + T_Air_-1 ; + HR_Air_-4 ; - HR_Air ; + T_Air_-4 ; - HR_Air_-1 ; + Long_Jour ; - HR_Air_-2 ; + Long_Jour_-1

- Espèce *Xeroderris stuhlmannii*

Tableau 6 : Meilleurs prédicteurs environnementaux pour l'espèce *Xeroderris stuhlmannii* dans la savane arborée de Béléfoungou (« Fe » = Feuillaison, « Fl » = Floraison, « Fr » = Fructification, « + » = augmentation, « - » = baisse, variables = voir partie du dessus)

	<u>Prédicteurs par ordre décroissant de pouvoir prédictif</u>
<u>Fe</u>	+ HR_Air_-8 ; + HR_Air_-4 ; - T_Air ; + Long_Jour_-8 ; + Long_Jour_-4 ; + HR_Air_-2 ; + Long_Jour_-2 ; + T_Air_-8 ; + Long_Jour_-1 ; + Long_Jour
<u>Fl</u>	+ T_Air_-4 ; + T_Air_-2 ; - HR_Air_-2 ; - Long_Jour_-8 ; - HR_Air_-4 ; - Long_Jour_-4 ; - Long_Jour_-2 ; - HR_Air_-8 ; - Long_Jour_-1 ; - T_Air_-1
<u>Fr</u>	- HR_Air_-8 ; + T_Air ; + HR_Air_-1 ; + Long_Jour ; + HR_Air ; - HR_Air_-4 ; + Long_Jour_-1 ; + Long_Jour_-2 ; + T_Air_-4 ; - Long_Jour_-8

c) Mosaïque de champs et de jachères de Nalohou

Dans cette partie, les résultats des plus forts prédicteurs environnementaux sur les phénomologies de 3 espèces de Nalohou vont être présentés. Ces trois espèces sont : *Burkea africana*, *Parinari curatellifolia* et *Parkia biglobosa*. (voir **Annexe 6**)

- Espèce *Burkea africana*

Tableau 7 : Meilleurs prédicteurs environnementaux pour l'espèce *Burkea africana* dans la mosaïque de champs et de jachères de Nalohou (« Fe » = Feuillaison, « Fl » = Floraison, « Fr » = Fructification, « + » = augmentation, « - » = baisse, variables = voir partie du dessus)

	<u>Prédicteurs par ordre décroissant de pouvoir prédictif</u>
<u>Fe</u>	+ T_Air_-8 ; - T_Air ; + Long_Jour ; + Long_Jour_-1 ; + T_Air_-4 ; + Long_Jour_-2 ; + HS_Sol_-2 ; + Pluie_Cum_-2 ; 9 Long_Jour_-4 ; - HS_Sol_-8
<u>Fl</u>	- HR_Air_-8 ; - Pluie_Cum ; - Long_Jour_-8 ; -HR_Air_-4 ; - Pluie_Cum_-1 ; - Pluie_Cum_-2 ; - HS_Sol ; + T_Air_-4 ; - HS_Sol_-4 ; - HS_Sol_-1 ; - HR_Air_-2
<u>Fr</u>	+ HS_Sol_-1 ; + T_Air ; + HS_Sol ; + T_Air_-1 ; - HR_Air_-8 ; + T_Air_-2 ; + HR_Air ; - Pluie_Cum_-2 ; - Pluie_Cum_-4 ; - Pluie_Cum_-1

- Espèce *Parinari curatellifolia*

Tableau 8 : Meilleurs prédicteurs environnementaux pour l'espèce *Parinari curatellifolia* dans la mosaïque de champs et de jachères de Nalohou (« Fe » = Feuillaison, « Fl » = Floraison, « Fr » = Fructification, « + » = augmentation, « - » = baisse, variables = voir partie du dessus)

	<u>Prédicteurs par ordre décroissant de pouvoir prédictif</u>
<u>Fe</u>	+ HS_Sol_-4 ; + T_Air_-1 ; - T_Air_-4 ; - HS_Sol_-8 ; - HS_Sol_-2 ; + Long_Jour_-1 ; + HR_Air ; + Long_Jour_-4 ; + Long_Jour_-2 ; + Long_Jour
<u>Fl</u>	+ T_Air_-2 ; + HS_Sol_-2 ; + HS_Sol_-1 ; + T_Air_-8 ; + T_Air_-1 ; + T_Air_-4 ; - HR_Air_-1 ; - Pluie_Cum_-8 ; - HR_Air ; + T_Air
<u>Fr</u>	- T_Air_-1 ; + T_Air_-8 ; - HS_Sol_-4 ; - T_Air ; + HS_Sol_-2 ; - Pluie_Cum_-8 ; + Long_Jour ; - HR_Air ; + HR_Air_-8 ; - T_Air_-2

- Espèce *Parkia biglobosa*

Tableau 9 : Meilleurs prédicteurs environnementaux pour l'espèce *Parkia biglobosa* dans la mosaïque de champs et de jachères de Nalohou (« Fe » = Feuillaison, « Fl » = Floraison, « Fr » = Fructification, « + » = augmentation, « - » = baisse, variables = voir partie du dessus)

	<u>Prédicteurs par ordre décroissant de pouvoir prédictif</u>
<u>Fe</u>	- HS_Sol_-2 ; + Pluie_Cum_-1 ; + Pluie_Cum ; - HS_Sol ; + Pluie_Cum_-2 ; - HS_Sol_-4 ; - HS_Sol_-1 ; + T_Air_-8 ; + Pluie_Cum_-4 ; + T_Air_-4
<u>Fl</u>	+ HS_Sol_-8 ; - HS_Sol_-4 ; - T_Air_-8 ; - Long_Jour_-1 ; - HS_Sol_-2 ; - Long_Jour_-2 ; - Long_Jour_-4 ; - HS_Sol_-1 ; - Long_Jour ; + T_Air
<u>Fr</u>	- Pluie_Cum_-8 ; - HS_Sol_-1 ; - Pluie_Cum_-4 ; + HS_Sol_-8 ; + T_Air_-4 ; + T_Air_-2 ; - HS_Sol ; - Pluie_Cum_-1 ; - Pluie_Cum_-2 ; + HS_Sol_-4

d) Savane arbustive de Bira

Dans cette partie, les résultats des plus forts prédicteurs environnementaux sur les phénologies de 3 espèces de la savane arbustive de Bira vont être présentés. Ces trois espèces sont : *Burkea africana*, *Combretum collinum* et *Daniellia oliveri*.

- Espèce *Burkea africana*

Tableau 10 : Meilleurs prédicteurs environnementaux pour l'espèce *Burkea africana* dans la savane arbustive de Bira (Fe » = Feuillaison, « Fl » = Floraison, « Fr » = Fructification, « + » = augmentation, « - » = baisse, variables = voir partie au dessus)

	<u>Prédicteurs par ordre décroissant de pouvoir prédictif</u>
<u>Fe</u>	-
<u>Fl</u>	- Pluie_Cum ; - Pluie_Cum_-1 ; - Pluie_Cum_-2 ; - Pluie_Cum_-4 ; - HR_Air_-8 ; - Pluie_Cum_-8 ; + Long_Jour ; - HS_Sol_-4 ; + T_Air_-8 ; + T_Air_-1
<u>Fr</u>	+ Long_Jour_-8 ; + HS_Sol ; + HS_Sol_-2 ; + HS_Sol_-1 ; + HR_Air_-4 ; + HR_Air_-8 ; + HR_Air_-2 ; + HR_Air_-4 ; - T_Air ; + Pluie_Cum

- Espèce *Combretum collinum*

Tableau 11 : Meilleurs prédicteurs environnementaux pour l'espèce *Combretum collinum* dans la savane arbustive de Bira (Fe » = Feuillaison, « Fl » = Floraison, « Fr » = Fructification, « + » = augmentation, « - » = baisse, variables = voir partie au dessus)

	<u>Prédicteurs par ordre décroissant de pouvoir prédictif</u>
<u>Fe</u>	- Long_Jour_-8 ; - HR_Air_-2 ; - HR_Air_-1 ; - HR_Air_-4 ; - HR_Air ; - Long_Jour_-4 ; - HS_Sol ; - HS_Sol_-1 ; - Long_Jour_-2 ; - HS_Sol_-2
<u>Fl</u>	- HS_Sol ; - HS_Sol_-1 ; + T_Air_-4 ; - HR_Air_-8 ; - HS_Sol_-2 ; - HR_Air_-4 ; - Long_Jour_-8 ; + T_Air_-2 ; + T_Air_-1 ; - HR_Air_-2
<u>Fr</u>	-

- Espèce *Daniellia oliveri*

Tableau 12 : Meilleurs prédicteurs environnementaux pour l'espèce *Daniellia oliveri* dans la savane arbustive de Bira (Fe » = Feuillaison, « Fl » = Floraison, « Fr » = Fructification, « + » = augmentation, « - » = baisse, variables = voir partie au dessus)

	<u>Prédicteurs par ordre décroissant de pouvoir prédictif</u>
<u>Fe</u>	-
<u>Fl</u>	+ Long_Jour_-8 ; + HR_Air_-2 ; + HR_Air_-1 ; + HR_Air ; + HR_Air_-4 ; + Long_Jour_-4 ; + HS_Sol ; - T_Air_-2 ; - T_Air ; - T_Air_-4
<u>Fr</u>	+ HR_Air ; + HR_Air_-2 ; + HR_Air_-1 ; + Long_Jour_-8 ; + Long_Jour_-4 ; + HR_Air_-4 ; + Long_Jour_-2 ; + Long_Jour_-1 ; + HS_Sol ; - T_Air

IV. Discussion

Avec le package *mix*, les données manquantes sont comblées très rapidement. De plus, les valeurs obtenues sont en général cohérentes vis-à-vis des données initiales. Le package *mix* du logiciel R se base sur l'ensemble de valeurs de la table pour générer la valeur qui manquait. Il a un modèle statistique, et c'est sur ce modèle là qu'il génère la nouvelle valeur. Donc plus il a d'informations, plus le package sera à même de réaliser l'imputation correctement.

Cependant, j'ai observé des incohérences dans certaines valeurs comblées. C'est le cas par exemple des données pour deux arbres situés dans le même type de couvert. S'il manquait une donnée température de l'air au 30 avril 2004, la température estimée pour l'arbre 1 à cette date est de 25.324 alors que la température estimée pour l'arbre 2 est de 30.546. Il a donc fallu modifier manuellement certaines valeurs pour qu'elles aient un sens.

Le programme ne s'appuie pas sur la colonne date, ni ne prend en compte le fait qu'une variable environnementale ne peut changer en fonction de l'arbre ou de l'espèce. Il aurait sans doute fallu prendre un individu après l'autre. J'ai observé aussi pour les données phénologiques comblées que la valeur « 1 » peut apparaître au milieu d'une série de « 0 » ou l'inverse. Il faut donc rectifier cela en se basant sur le phénogramme à partir des valeurs observées seules ; plus précisément il faut modifier la valeur aberrante de sorte que le phénogramme du jeu de données sortant (après exécution du package *mix*) soit cohérent avec le phénogramme initial à partir des valeurs observées seules.

Enfin, sur les données de pluie cumulée, le package n'a pas tenu compte du changement d'année. Les données générées début 2007 étaient de l'ordre de 150mm, 170mm ... alors qu'elles doivent être de 0 avant les premières pluies. Il fallait donc remplacer ces valeurs erronées par des 0 pour le début de l'année 2007.

Le package *mix* nécessite donc de vérifier soigneusement les données générées automatiquement.

Concernant la méthode de régression logistique, le point positif est que celle-ci permet de renseigner sur les pseudo-variables prédictives d'une phénophase donnée. Malgré tout, cette méthode présente des limites : j'étais dans le cas ici d'applications à des séries temporelles avec des variables prédictives dépendantes les unes des autres. C'est pour cela qu'une ACP a été utilisée avant de faire les régressions logistiques.

De plus, pour la plupart des modèles de régression appliqués à la fructification, le pseudo R^2 était assez faible. La fructification ne s'explique donc pas seulement par les facteurs environnementaux testés. Par exemple, les phénophases précédentes influent sur la phénophase en cours d'une plante. La feuillaison influe sur la floraison et la fructification se situe souvent après un intervalle constant de temps après la floraison. Un autre problème se présente enfin : auto corrélation au sein d'une même phénophase. La probabilité de l'apparition ou non d'une feuillaison à la date d'observation d est fonction de la probabilité de feuillaison à la date $d+1$. Selon Beck *et al.* (1998), pour éviter à la fois des tests statistiques imprécis et la perte d'informations dans les données phénologiques, il faudrait dans ces deux cas ajouter deux variables artificielles prédictives (« dummy variables »). La première variable artificielle *dum1* serait le nombre de semaines depuis l'occurrence précédente. Par exemple, « *Fedum1* » (*Fe* pour *feuillaison*, *dum1* pour *dummy variable 1*) serait le nombre de semaines depuis l'occurrence de la feuillaison précédente (et noté comme telle à chaque date d'observation) jusqu'à la fin de la feuillaison suivante. Cette variable artificielle représenterait la durée depuis la phénophase précédente et permet d'intégrer des autos corrélations intra phénophase éventuelles dans les modèles. La seconde variable artificielle (*dum2*) pour une phénophase donnée serait aussi le nombre de semaines séparant l'observation actuelle de l'occurrence précédente, mais contrairement à *dum1*, la valeur de *dum2* reprend la valeur 0 dès le début de l'occurrence suivante. Cela permettrait de prendre en compte l'effet de la phénophase précédente sur une phénophase en cours, autrement dit l'effet de la feuillaison sur la floraison et l'effet de la floraison sur la fructification. En tant que prédicteur d'une phénophase donnée, *dum2* aurait représentée la durée écoulée depuis l'occurrence de la phénophase précédente ou la concomitance entre les deux phénophases. Le chevauchement entre les deux phénophases est indiqué quand *dum2* de la phénophase précédente est égale à 0 et qu'à la même date, la valeur de la phénophase en cours (à expliquer) est égale à 1 (occurrence, présence d'organes).

Après le choix des modèles optimaux, j'ai effectué des régressions linéaires pour repasser aux variables d'origine. J'ai été satisfait de cette méthode car elle permet de voir finalement les variables d'origines corrélées négativement avec la phénophase étudiée. Cela paraît plus clair et plus compréhensible, bien qu'un doute persiste dans l'approximation de la relation entre les composantes principales et les variables de départ.

V. Conclusion

Concernant les méthodes statistiques qui ont été utilisées, ce stage me semble parfaitement en accord avec la formation de DUT Statistiques et Informatique Décisionnelle.

Il m'a permis de me rendre compte de la rigueur de l'approche et de la rédaction, qui est une totale découverte pour moi. Ce stage est aussi une expérience nouvelle en ce qui concerne l'arrangement des données. J'ai aussi découvert un nouveau logiciel statistique : Stata. J'ai pu manipuler un logiciel différent de ceux pratiqués en cours et cela est une expérience supplémentaire à mon actif. J'aurai sans doute plus de facilité à aborder de nouveaux logiciels statistiques par la suite.

Lors de ce stage, j'ai eu la chance aussi de côtoyer d'autres stagiaires qui préparaient un diplôme supérieur au mien. Ils m'ont fait découvrir de nouveaux logiciels statistiques comme SYSTAT par exemple, appris à mieux utiliser le logiciel Microsoft Word 2007 (découvertes de nouvelles options, nouvelles fonctionnalités que je ne savais pas auparavant).

J'ai aussi acquis de nouvelles connaissances dans le domaine phénologique des plantes. Puis j'ai appris beaucoup de choses sur les régions africaines et notamment sur les pays du Bénin mais aussi du Niger et du Mali. Tout cela est et sera très bénéfique pour ma culture générale.

Ce stage m'a fait enfin comprendre que le plus important était de répondre au problème posé par l'entreprise. Il est donc prépondérant de savoir l'utilité d'un logiciel statistique (outil), la réponse que celui-ci peut fournir, et si celle-ci est suffisante compte tenu problème posé.

REFERENCES

- Aubreville, A., 1950. Flore forestière Soudano-Guinéenne AOF.-Cameroun-A.E.F. Éditions Géographiques, Maritimes et Coloniales, Paris.
- Beck, N., Katz, J.N., Tucker, R. 1998. Taking time seriously : time-series-cross-section analysis with a binary dependant variable. *American Journal of Political Science* 42: 1260-1288.
- Blanchard M, Peugeot C, Seghieri J. 2007. Structure et dynamiques saisonnière de la végétation sur quatre types de couvert végétal représentatifs de la diversité du bassin versant de la Donga.
- Chapman, C.A., Chapman, L.J., Struhsaker, T.T., Zanne, A.E., Clark, C.J. & Poulson, J.R. 2005. A long evaluation of fruiting phenology: importance of climate change. *Journal of Tropical Ecology* 21 : 31-45.
- Lieth, H. 1974. Phenology and seasonality modeling. Springer, Berlin, Germany.
- Harris J., 2004, Rapport de campagne Végétation., ORE AMMA_CATCH
- Haugaasen, T. & Peres, C. A. 2005. Tree Phenology in adjacent Amazonian flooded and unflooded forests. *Biotropica* 37: 620–630.
- Seghieri J, Carreau J, Boulain N, Arjounin M, Timouk F. Predictive power of physical factors for phonological responses of woody plants in the central Sahel (Mali). Article soumis.
- Serviere M, Mémoire pour l'obtention du master en ingénierie de l'eau et de l'environnement, option : eau. 2010. Les variations de la transpiration de la strate ligneuse en environnement soudano-guinéen

- Singh, K.P. & Kushwaha, C.P. 2006. Diversity of flowering and fruiting phenology of trees in a tropical deciduous forest in India. *Annals of Botany* 97: 265-276.
- Trochain J.L., Accord inter-africain sur la définition des types de végétation de l'Afrique tropicale. *Bull. Inst. Et. Centrafric.* N°13-14, 55-93.
- Vliet (van), A.J.H. & Schwartz M.D. 2002. Phenology and climate: the timing of life cycle events as indicators of climate variability and change. *International Journal of Climatology* 22: 1713-1714.
- Wright, S. J. & Cornejo, F. H. 1990. Seasonal drought and leaf fall in 1 a tropical forest. *Ecology* 71: 1165-1175.

Annexe 1 : Exemple de script R - Utilisation du Package Mix

```
#####  
#  
# 1ERE PARTIE : COMPLETER DONNNEES MANQUANTES  
# PACKAGE MIX  
#  
#####  
  
ATTENTION :  
  Les valeurs qualitatives doivent etre obligatoirement en 1/2 (et non en 0/1)  
  Les 3 premieres colonnes doivent etre obligatoirement les variables qualitatives en 1/2  
  Aucune ligne ne doit etre completement vide  
  
library(mix, pos=4)  
  
X=read.delim("C:/Users/Florian/Desktop/TRAITEMENT2/BelefontouForet/BeleForet_Annosene.txt",header=T,sep="\t",dec=",")  
X  
colnames(X1)  
X=X[,~c(1:4)]  
X[,1:3]=X[,1:3]+1  
s <- prelim.mix(X,3)  
thetahat <- em.mix(s)  
rngseed(1234567)  
newtheta <- da.mix(s,thetahat,steps=100)  
ximp <- imp.mix(s, newtheta, X)  
ximp[,1:3]=ximp[,1:3]-1  
  
### EXPORTATION DU JEU DE DONNEES COMPLET  
  
write.csv2(ximp,"FullBeleForet_Annosene.csv")  
  
## modifier certaines valeurs pour pluiecumulée car quelques erreurs +  
## enlever la colonne 1 + enregistrer fichier pour étape suivante
```

Annexe 2 : Script R qui permet le passage des Composantes Principales significatives aux variables d'origine

```
X_CENTRED_PCi=read.csv("C:/Users/Florian/Desktop/TRAITEMENT2/Belefou  
ngouForet/Isobdoka/FullData_BeleForet_Isobdoka.csv",header=T,sep=";"  
,dec=".")  
colnames(X_CENTRED_PCi)  
DataFePCCoeff <-  
read.table("C:/Users/Florian/Desktop/TRAITEMENT2/BelefoungouForet/Is  
obdoka/IsobFrPCCoeff.csv",header=T,sep=";",dec=".")  
pc=as.matrix(X_CENTRED_PCi[,c(28,29,31)])  
stand=as.matrix(X_CENTRED_PCi[,c(3:27)])  
regj=lm(stand~pc-1)  
summary(regj)  
t(DataFePCCoeff[,2]) %*% regj$coef  
IsobFe=t(DataFePCCoeff[,2]) %*% regj$coef  
max(IsobFe)  
max(abs(IsobFe))  
IsobFe=t(regj$coef) %*% DataFePCCoeff[,2]  
IsobFe  
write.csv2(IsobFe,"IsobdokaFe_BeleForet_RESULTAT FINAL.csv")
```

Annexe 3 : Résultats des modèles logistiques obtenus stimulant les effets des variables pseudo prédictives. Observations = nombre d'observations analysées, Pseudo R²=coefficient du Pseudo R², AIC= Akaike information criterion, Sign = niveau de significativité, non significatif NS ($p>0.05$), * $p<0.05$, ** $p<0.01$, *** $p<0.001$), Fe=Feuillaison, Fl=Floraison, Fr=fructification.

<i>Especies</i> Site d'Etude Observations	Pheno- phases	Pseudo R ²	AIC	Sign. Modèle	Variables Prédictives dans le modèle logistique et signification
<i>B. africana</i> Nalohou 432obs.	Feuillaison	0.5116	264.7349	***	pc1(***) pc2(***) pc4(***) pc5(***)
	Floraison	0.7385	149.0817	***	pc1 (***) pc2 (***)
	Fructification	0.1855	498.7708	***	pc1 pc2(***) pc3(***) pc4(***) pc 5 pc6
<i>P. curatellifolia</i> Nalohou 216 obs.	Feuillaison	0.4400	122.5345	***	Pc1(**) pc2(***) pc4(**) pc6(***)
	Floraison	0.6670	103.5633	***	pc1(***) pc2(***) pc3(***) pc5(***) pc6(***)
	Fructification	0.1846	208.0835	***	pc1(**) pc2(***) pc5 (NS) pc6 (***)
<i>P. biglobosa</i> Nalohou 576 obs.	Feuillaison	0.1725	576.9735	***	pc3(***) pc4(***)
	Floraison	0.6114	219.0289	***	pc1(***) pc2(***) pc3(***) pc4(***)
	Fructification	0.4428	396.4297	***	pc1(***) pc2(***) pc4 (***) pc5(***) pc6(**)
<i>P. crassipes</i> BeleSavane 720 obs.	Feuillaison	0.4829	434.3336	***	pc1(***) pc2(**) pc4(***) pc5(***)
	Floraison	0.8047	?	***	pc1(***) pc2(***) pc3(**) pc5(***)
	Fructification	0.4223	168.0424	***	pc1(***) pc2(***) pc4(**) pc5(***)
<i>V. paradoxa</i> BeleSavane 720 obs.	Feuillaison	0.4829	434.3336	***	pc1(***) pc2(**) pc4(***) pc5(***)
	Floraison	0.8525	124.8935	***	pc1(***) pc2(***) pc3(***) pc4(**) pc5 (***)
	Fructification	0.6660	280.4375	***	pc1(***) pc2(***) pc3(***) pc5(***)

<i>X. stuhlmannii</i> BeleSavane 720 obs.	Feuillaison	0.5734	343.7117	***	pc1 (***) pc3 (NS) pc4 (**) pc5 (**)
	Floraison	0.5323	366.7316	***	Pc1(***) pc2 (**) pc4 (**) pc5 (NS)
	Fructification	0.1855	741.0011	***	pc2(***) pc3(**) pc4(**) pc5 (**)
<i>A. senegalensis</i> BeleForêt 216 obs.	Feuillaison	0.7080	75.31353	***	pc1(***) pc2(***) pc4(***)pc5(***)
	Floraison	0.4807	124.8107	***	pc1(***) pc2(***) pc3(***) pc4(***)
	Fructification	0.1166	230.0356	***	pc2(***) pc3(**) pc5 (*)
<i>I.Doka</i> BeleForêt 216 obs.	Feuillaison	0.6649	82.9713	***	pc1(***) pc2(***) pc4(*)
	Floraison	0.4807	124.8135	***	pc1(***) pc2(***) pc3(***) pc4(***)
	Fructification	0.1165	230.0373	***	pc2(***) pc3(**) pc5(*)
<i>O. amentacea</i> BeleForêt 576 obs	Feuillaison	0.9892	7.638171	***	pc1 pc2 pc3 pc4 pc5
	Floraison	0.5580	358.7371	***	pc1(***) pc2(***) pc3(***) pc4(***) pc5(***)
	Fructification	0.4239	470.0216	***	pc2(***) pc3(***) pc4(***) pc5(***)
<i>B africana</i> Bira 180 obs.	Feuillaison	Tous les valeurs de Fe correspondent à 1			
	Floraison			***	pc1 (***)
	Fructification	0,1275		***	pc1 (***)
<i>D. oliveri</i> Bira 133 obs	Feuillaison	Toutes les valeurs de Fe correspondent à 1			
	Floraison	0.3995		***	pc1 pc2
	Fructification	0.2255	50.47807	*	pc2
<i>C.coll</i> Bira 240	Feuillaison	0.78	27	***	pc 2 (***)
	Floraison	1.0000	8	***	Pc1 (***) pc2 (***) pc3 (***)
	Fructification	1.0000	7	***	Pc1 (***) pc2 (***) pc3 (***)

Annexe 4 : Classification des prédicteurs environnementaux pour les espèces *A. senegalensis*, *I.Doka*, *O. amentacea* dans la forêt de Béléfoungou

	Feuillaison			Floraison			Fructification		
	Annosene	Isobdoka	Opilamen	Annosene	Isobdoka	Opilamen	Annosene	Isobdoka	Opilamen
T_Air				4(+)	4 (+)	4 (+)			
T_Air_.1	2 (-)			10 (+)	10 (+)	5 (+)			
T_Air_.2			2 (+)			3 (+)	3 (-)	3 (-)	2 (+)
T_Air_.4	1 (+)		1 (+)			1 (+)	1 (-)	1 (-)	1 (+)
T_Air_.8		8 (+)		6 (-)	6 (-)	2 (-)	2 (+)	2 (+)	
HR_Air	3 (+)	2 (-)	5 (+)						5 (+)
HR_Air_-1	4 (+)	6 (-)	3 (+)						4 (+)
HR_Air_-2	5 (+)								
HR_Air_-4	6 (+)								
HR_Air_-8				7 (-)	7 (-)				
HS_Sol		3 (-)							
HS_Sol_-1		4 (-)		8 (+)	8 (+)				
HS_Sol_-2		10 (-)							7 (-)
HS_Sol_-4									
HS_Sol_-8			9 (+)	9 (-)	9 (-)		4 (-)	4 (-)	
Pluie_Cum	10 (+)	9 (+)	6 (+)	5 (-)	5 (-)		5 (-)	5 (-)	6 (+)
Pluie_Cum_-1		5 (+)	8 (+)	3 (-)	3 (-)		6 (-)	6 (-)	
Pluie_Cum_-2		1 (+)		1 (-)	1 (-)				
Pluie_Cum_-4		7 (+)	7 (-)	2 (-)	2 (-)				8 (-)
Pluie_Cum_-8	7 (-)		4 (-)						3 (-)
Long_Jour	9 (+)		10 (-)			10 (-)	8 (+)	8 (+)	9 (+)
Long_Jour_-1	8 (+)					9 (-)	7 (+)	7 (+)	10 (+)
Long_Jour_-2						8 (-)	9 (+)	9 (+)	
Long_Jour_-4						6 (-)	10 (+)	10 (+)	
Long_Jour_-8						7 (-)			

Annexe 5 : Classification des prédicteurs environnementaux pour les espèces *P. crassipes*, *V. paradoxa*, *X. stuhlmannii* dans la savane arborée de Béléfoungou

	Feuillaison			Floraison			Fructification		
	Pavecras	Vitepara	Xerostuh	Pavecras	Vitepara	Xerostuh	Pavecras	Vitepara	Xerostuh
T_Air	2 (-)	2 (-)	3 (-)				2 (+)		2 (+)
T_Air_.1	10 (-)	7 (+)				10 (-)		3 (+)	
T_Air_.2		1 (+)		2 (+)		2 (+)	4 (-)	2 (+)	
T_Air_.4		3 (+)		1 (+)		1 (+)	1 (-)	6 (+)	9 (+)
T_Air_.8			8 (+)				8 (+)	1 (+)	
HR_Air	7 (+)	8 (+)		8(+)				5 (-)	5 (+)
HR_Air_-1	8 (+)	4 (+)		9 (+)				7 (-)	3 (+)
HR_Air_-2	5 (+)		6 (+)	7 (-)		3 (-)		9 (-)	
HR_Air_-4	3 (+)	6 (+)	2 (+)	4 (-)		5 (-)	10 (-)	4 (+)	6 (-)
HR_Air_-8	1 (+)	5 (+)	1 (+)	3 (-)		8 (-)	3 (-)		1 (-)
HS_Sol					-				
HS_Sol_-1									
HS_Sol_-2									
HS_Sol_-4									
HS_Sol_-8									
Pluie_Cum									
Pluie_Cum_-1									
Pluie_Cum_-2									
Pluie_Cum_-4									
Pluie_Cum_-8									
Long_Jour			10 (+)	6(+)			5 (+)	8 (+)	4 (+)
Long_Jour_-1			9 (+)	10 (+)		9 (-)	6 (+)	10 (+)	7 (+)
Long_Jour_-2	9 (+)		7 (+)			7 (-)	7 (+)		8 (+)
Long_Jour_-4	6 (+)	10 (+)	5 (+)			6 (-)	9 (+)		
Long_Jour_-8	4 (+)	9 (+)	4 (+)	5 (-)		4 (-)			10 (-)

Annexe 6 : Classification des prédicteurs environnementaux pour les espèces *B. africana*, *P. curatellifolia*, *P. biglobosa* dans la mosaïque de champs et jachères de Nalohou

	Feuillaison			Floraison			Fructification		
	Burkafri	Paricura	Parkbigl	Burkafri	Paricura	Parkbigl	Burkafri	Paricura	Parkbigl
T_Air	2 (-)				10 (+)	10 (+)	2 (+)	4 (-)	
T_Air_.1		2 (+)			5 (+)		4 (+)	1 (-)	
T_Air_.2					1 (+)		6 (+)	10 (-)	6 (+)
T_Air_.4	5 (+)	3 (+)	10 (+)	8 (+)	6 (+)				5 (+)
T_Air_.8	1 (+)		8 (+)		4 (+)	3 (-)		2 (+)	
HR_Air		7 (+)			9 (-)		7 (+)	8 (-)	
HR_Air_-1					7 (-)				
HR_Air_-2									
HR_Air_-4				4 (-)					
HR_Air_-8				1 (-)			5 (-)	9 (+)	
HS_Sol			4 (-)	7 (-)			3 (+)		7 (-)
HS_Sol_-1			7 (-)	10 (-)	3 (+)	8 (-)	1 (+)		2 (-)
HS_Sol_-2	7 (+)	5 (-)	1 (-)		2 (+)	5 (-)		5 (+)	
HS_Sol_-4		1 (+)	6 (-)	9 (-)		2 (-)		3 (-)	10 (+)
HS_Sol_-8	10 (-)	4 (-)				1 (+)			4 (+)
Pluie_Cum			3 (+)	2 (-)					
Pluie_Cum_-1			2 (+)	5 (-)			10 (-)		8 (-)
Pluie_Cum_-2	8 (+)		5 (+)	6 (-)			8 (-)		9 (-)
Pluie_Cum_-4			9 (+)				9 (-)		3 (-)
Pluie_Cum_-8					8 (-)			6 (-)	1 (-)
Long_Jour	3 (+)	10 (+)				9 (-)		7 (+)	
Long_Jour_-1	4 (+)	6 (+)				4 (-)			
Long_Jour_-2	6 (+)	9 (+)				6 (-)			
Long_Jour_-4	9 (+)	8 (+)				7 (-)			
Long_Jour_-8				3 (-)					