Université de Liège
Faculté des Sciences
Département des Sciences et Gestion de l'Environnement

Assessment of fodder biomass in Senegalese rangelands using earth observation and field data

Abdoul Aziz DIOUF

Thèse présentée en vue de l'obtention
du grade de Docteur en Sciences

Octobre 2016

Composition du jury :

Président :	Dr Pierre OZER (ULg)
Promoteur :	Pr Bernard Tychon (ULg)
Co-promoteur :	Dr, HDR, Jacques André NDIONE (CSE)
Lecteurs :	Dr Bakary Djaby (AGRHYMET)
	Pr Jos Van Orshoven (KU Leuven)
	Dr Pierre Hiernaux (GET)
	Dr Joost Wellens (ULg)

Année Académique 2016-2017

Avec l'appui de :


100%
FSC
www.fsc.org
RECYCLÉ
Papier fait à partir
de matériaux recyclés
FSC® C004423

Université de Liège
Faculté des Sciences
Département des Sciences et Gestion de l'Environnement

Assessment of fodder biomass in Senegalese rangelands using earth observation and field data

Abdoul Aziz DIOUF

Thèse présentée en vue de l'obtention
du grade de Docteur en Sciences

Octobre 2016

Composition du jury :

Président :	Dr Pierre OZER (ULg)
Promoteur :	Pr Bernard Tychon (ULg)
Co-promoteur :	Dr, HDR, Jacques André NDIONE (CSE)
Lecteurs :	Dr Bakary Djaby (AGRHYMET)
	Pr Jos Van Orshoven (KU Leuven)
	Dr Pierre Hiernaux (GET)
	Dr Joost Wellens (ULg)

Année Académique 2016-2017

Avec l'appui de :

To my late father Abdoulaye Diouf;
To my mother, Aïssatou Diouf;
To my wife, Khady Diouf;
And to you;

Acknowledgments

My first thanks go to the members of my thesis committee who framed and followed my work, and who have given me support and encouragement throughout this long and beautiful adventure. Thank you to Bernard Tychon, promoter of this thesis, for the trust he has given me from the beginning of my research and for his promptness to answer my queries both scientifically and administratively. Thank you to Bakary Djaby, who guided my first steps in the use of statistical tools and the modelling. Thank you to Jacques André Ndione, co-advisor of this thesis, for being always available to read and reread my search results from Dakar to Lome, but also for giving me valuable advices which have contributed to the completion of this work and that is now "*ma tasse de Kinkeliba*" ...

Thank you to the General Direction of the Centre de Suivi Ecologique (CSE), especially Assize Touré, General Director, Amadou Moctar Dièye, Technical director and Oumar Sarr, Administrative and Financial director, for their trust to me and for allowing me to do this thesis. Thank you to the AGRICAB project managers and experts at the Flemish Institute for Technological Research (VITO), the CSE and the University of Liège (ULg): Tim Jacobs, Carolien Tote, Mouhamadou Bamba Diop ("segn bi", especial thanks to you for all), Antoine Royer (many thanks to you for contributing vastly to my knowledge in remote sensing) and Julien Minet, for making all this research possible.

Thank you to the members of the jury for the time, interest and constructive comments you have made to this dissertation.

Many thanks go to Abdoulaye Wélé, Aliou Diouf, Moussa Dramé, Moussa Sall, Aliou Ka, Gora Bèye and Ibrahima Diop etc., not forgot all car-drivers and other field assistants, who helped to collect ground data from 1999 to 2015. Without your support this work would not have been possible.

Thank you to my co-authors of the publications: Martin Brandt, Aleixandre Verger, Moussa El Jarroudi, Bakary Djaby, Rasmus Fensholt, Jacques André Ndione, Bernard Tychon, Mouhamadou Bamba Diop, Gayane Faye and Pierre Hiernaux. I learned a lot from the knowledge of everyone. I'm particularly grateful to Martin Brandt who allowed me yet at

the beginning to have my first peer reviewed publication in Remote Sensing. He also made a remarkable contribution to the design of my work, from the codes shared to the proofreading of the dissertation.

Thank you to the Rasmus's team at Department of Geosciences and Natural Resource Management of the University of Copenhagen (not to mention names) for the hospitality during my visit over there in May 2016, the time and the valuable remarks they have made on the first draft of this dissertation. Thanks also to Stefanie M. Herrmann.

Thank you to all colleagues at CSE for the valuable advice and encouragement along this research: special thanks to Taibou Ba, who has been the first to identify me as a good candidate for this thesis by informing me about the proposal. Further thanks go to my office colleague, Marième Diagne and his twins (Fatou Binetou Traore and Batouly Ly). Thanks to the lunch break team, "wa ministère de la santé or wa daradji" for the good times spent together around the meals.

Thank you to all colleagues of the Department of Science and Environment Management (at the Arlon Campus Environment of ULg): especially to the "BT" team, Moussa El Jarroudi, Joost Wellens, Abdoul Hamid Mouhamed Sallah, Marie Lang, Antoine Denis, Sie Pale, Julien Minet, Claire Simonis, Ingrid Jacquemin, Mahamadou Karimou Barke, for their sense of humor (soupé les "jeudredis" soir), and for discussions about science and everyday life during coffee-breaks. Especial thanks to Tarik Ben Abdelouahab for all. Thanks also to Farid Traore for guiding my first steps across the Arlon's city and for advising me about PhD requirements, and to Louis Kouadio Amani for sharing useful statistical codes. Thanks go to Catherine Heyman for its assistance each time I was planning a travel to Belgium. Thanks to the running team (FulguRUN!!!) of the campus, for the good time we spent together during the rally "year 199th of the ULg": I'll be happy to run next year again for the bicentenary of the University...

Many thanks to the Senegalese community in Arlon, especially to Idrissa, Mame Fatou, Ka Rokhaya and Ka Maïmouna for all wonderful things they have done for me. I do not forget Insa Goudiaby and Pape Diouf and their families.

Thank you to managers and godparents of the Non-Governmental Organization “JUDDU” which help children from poor families in Pikine (Senegal) to have a better education. Especial thanks to Cathy Le Baron, Jean Francois Blérot and Françoise Poullet Blérot for effort they are spending for “JUDDU”, and for encouragement and nice moments we shared in Belgium.

Thank you to the European Union which funded the AGRICAB project (282621) within the seventh Framework Programme (FP7). Thanks also to the European Earth observation programme, Copernicus Global Land and the GIOBIO (32-566) project for providing the satellite imagery at no cost. Thank you to the USGS Earth Resources Observation and Science Center (EROS) and the Famine Early Warning Systems Network (FEWS NET) Project for providing satellite imagery and the GeoWRSI software. Many thanks to the “Académie de Recherche et d’Enseignement Supérieur (ARES)” and the “Centre pour le Partenariat et la Coopération au Développement (PACODEL)” of the University of Liège for the fellowship allocated for the completion of this thesis.

Finally, this work would not have been completed without the support and encouragement of my friends, my parents and my wife to whom I further say thank you.

Summary

Senegalese livestock size has largely increased during the last three decades in relation to the population growth. The fodder biomass stock available at the end of the growing season, therefore, becomes increasingly limited to meet feeding needs of pastoral livestock which provides third of the national agricultural wealth. With the reduction of natural grazing lands mostly generated by the expansion of croplands, and the reduction of fodder biomass production due to drought effects, the increase of the livestock size leads to the rangelands overload whose persistence can lead in turn to their degradation. A technique based on a simple linear relationship between the temporal integration of the Normalized Difference Vegetation Index (NDVI) and the ground biomass data, developed in the 1980s, has been operationally applied by the *Centre de Suivi Ecologique* (CSE) of Dakar (Senegal) to assess the fodder biomass available in rangelands at the end of the growing season. The derived map of total biomass production enables to help pastoral livestock managers as well as national stakeholders against food insecurity and natural resources degradation. Carried out annually, this approach comprises unfortunately some uncertainties as: (1) the saturation drawback of NDVI in areas with high biomass productivity, (2) the temporal scale which is restricted to biomass data of the ongoing year not being used again in the following year, (3) the low predictive ability due to the large time gap between data collection and published results, and (4) the high costs for annual data collection. In addition, although the earth observation (EO) data have largely progressed during the last three decades, this technique has not changed over this period and consequently is not state-of-the-art. To tackle these limitations and advance the traditional method, new statistical models that include new earth observations datasets and historical in situ plant biomass data were developed for estimating and / or predicting the forage availability at the end of the growing season in Senegalese semi-arid rangelands. A backward analysis of the linear regression approach currently applied in Senegal provided evidence that nonlinear regression functions such as Exponential and Power are more suited to estimate the end-of-season total biomass in this region using annual data solely. A completely new methodology using multiple-linear models which include various

phenological metrics from the time series of the Fraction of Absorbed Photosynthetically Active Radiation (FAPAR) and 14 years of *in situ* total biomass samples was developed. The proposed approach provided more reliable and accurate estimates as compared to the current CSE biomass product. Multiple-linear models developed with specific metrics adapted to ecosystem properties increased the overall accuracy of the fodder biomass estimates and mitigated the saturation of FAPAR obtained with models run across the whole study area. With this new approach, timely information about possible deficits/surplus of total fodder biomass can be provided to stakeholders using phenological metrics that are available relatively early in the growing season. Another new approach based on a machine learning algorithm (i.e., Cubist) was developed, as never done before, to assess herbaceous biomass in Senegalese Sahel. Three Cubist models using FAPAR seasonal metrics and/or agrometeorological variables (i.e., soil water status indicators) were established and compared. The Cubist model including both FAPAR and agrometeorological variables provided the best estimation performance. This model enabled to mitigate the saturation affecting optical remotely sensed vegetation data in areas of high plant productivity as well as the discrepancy between herbaceous biomass and greenness, and corrected therefore for herbaceous biomass underestimations observed with the sole FAPAR based model, particularly in sparsely vegetated areas. In contrast to the date of the growing season onset retrieved from FAPAR seasonal dynamics, the rainy season onset was significantly related to the herbaceous biomass and its inclusion in models could constitute a significant improvement in forecasting risks of fodder biomass deficit. The methods developed in this research provide tools to assess Senegalese forage resources at two levels: herbaceous and total fodder biomass (Herbaceous + woody leaf biomass). They require limited data and free available software and therefore can be easily replicated in other countries of the West African Sahel.

Keywords: fodder biomass, models, FAPAR, phenological metrics, growing season, herbaceous, forecast, food security, Senegal, Sahel.

Résumé

La taille du cheptel sénégalais a connu une forte augmentation au cours des trois dernières décennies en relation avec la poussée démographique. La production fourragère en fin de saison se rapproche de plus en plus des limites de satisfaction des besoins alimentaires du cheptel essentiellement pastoral qui fournit le tiers de la richesse agricole nationale. Avec la réduction des parcours naturels liée globalement à l'augmentation des terres de culture, et la réduction de la production de biomasse causée par la sécheresse, l'accroissement du cheptel entraine une surcharge des parcours, dont la persistance conduit à leur dégradation. Afin d'estimer le disponible fourrager des parcours à la fin de la saison de croissance, le Centre de Suivi Ecologique (CSE) de Dakar applique de manière opérationnelle, une technique utilisant une régression linéaire simple entre l'indice de végétation NDVI (*Normalized Difference Vegetation Index*) cumulé au cours de la saison et les données de biomasse végétale collectées sur le terrain. La carte de production fourragère élaborée annuellement est très utile pour les gestionnaires du système pastoral ainsi que les décideurs nationaux en matière de lutte contre l'insécurité alimentaire et la dégradation des ressources naturelles. Malheureusement, cette technique comporte un certain nombre de contraintes liées à : (1) la saturation du NDVI dans les zones à forte production fourragère, (2) les données de biomasse limitées à l'année en cours et qui ne sont pas utilisées les années suivantes, (3) la faible capacité de prévision en raison du temps important entre la collecte des données sur le terrain et la publication des résultats, et (4) le coût élevé requis pour la collecte annuelle des données. Par ailleurs, bien que les données d'observation de la terre (satellitaires) aient largement évolué au cours des trois dernières décennies, cette technique n'a pas changé et reste à améliorer par rapport à l'état de l'art actuel. Afin d'améliorer cette méthode « traditionnelle », de nouvelles approches statistiques ont été proposées dans cette étude. Ces méthodes intègrent de nouvelles données d'observation de la terre et des données historiques de production fourragère pour estimer et/ou prévoir les quantités de fourrages disponibles à la fin de la saison au niveau des parcours du Sénégal. Une analyse comparative de l'approche par régression simple actuellement utilisée au Sénégal, a montré que les

fonctions non-linéaires, Exponentiel et Puissance sont plus adaptées à l'estimation du disponible fourrager de fin de saison en utilisant les données d'une seule année. Une approche entièrement nouvelle avec des modèles de régression multilinéaire a été développée. Elle utilise différentes métriques saisonnières issues de séries chronologiques de la fraction absorbée du rayonnement photosynthétiquement actif (FAPAR ou *Fraction of Absorbed Photosynthetically Active Radiation*) et 14 ans de données de biomasse fourragère totale. Cette approche fournit des estimations plus précises telles que comparées avec les sorties de la méthode du CSE. Les modèles de régression multiple développés avec des métriques saisonnières spécifiques et adaptées aux propriétés des écosystèmes ont permis d'améliorer la précision globale des estimations de la biomasse fourragère totale mais aussi d'atténuer la légère saturation du FAPAR observée avec les modèles appliqués à l'échelle de la zone d'étude. Avec cette nouvelle approche, l'information sur les déficits/surplus de biomasse fourragère totale peut être très tôt transmise aux décideurs en utilisant des variables saisonnières du FAPAR disponibles relativement tôt au cours de la saison de croissance. Une autre approche incluant un algorithme d'apprentissage automatique appelé Cubist, a été développée pour l'estimation de la biomasse fourragère herbacée au Sénégal. Trois modèles Cubist qui utilisent des métriques saisonnières du FAPAR et/ou des variables agrométéorologiques (indicateurs de l'état de l'eau dans le sol), ont été développés et comparés. Le modèle Cubist qui utilise à la fois les métriques du FAPAR et les variables agrométéorologiques, s'est montré plus performant. Ce modèle a permis d'atténuer la saturation qui caractérise les données de télédétection optique de la végétation dans les zones à forte densité végétale mais aussi, de réduire la différence souvent observée entre la masse de végétation herbacée et la valeur dérivée des indices satellitaires. Par conséquent, elle a permis de corriger la sous-estimation de la production de biomasse notée avec le modèle utilisant le FAPAR uniquement, particulièrement dans les zones à faibles couvert végétal. Contrairement à la date de démarrage de la saison de croissance calculée à partir du FAPAR, celle du démarrage de la saison des pluies a été significativement liée à la biomasse herbacée et son utilisation dans les modèles pourrait nettement améliorer la prévision des risques de

déficits fourragers. Les méthodes développées dans cette étude constituent des outils d'estimation des ressources fourragères du Sénégal à deux niveaux : herbacé et total (herbacé et ligneux). Elles font appel à un nombre réduit de données et à des logiciels disponibles gratuitement, et donc peuvent facilement être reproduites dans d'autres pays du Sahel Ouest Africain.

Mots-clés : Biomasse fourragère, modèles, FAPAR, métrique, saison de croissance, herbacé, prévision, sécurité alimentaire, Sénégal, Sahel.

Abbreviations and Symbols

AET	Actual EvapoTranspiration
AGRHYMET	Centre régional de formation et d'applications agronomique, hydrologique et météorologique
AGRICAB	Agriculture Capacity Building
AIC	Akaike Information Criterion
AMMA	Analyse Multidisciplinaire de la Mousson Africaine
AMSU	Advanced Microwave Sounding Unit
ANNs	Artificial Neural Networks
APAR	Absorbed Photosynthetically Active Radiation ($MJ.m^{-2}$)
ARC	African Rainfall Climatology
AVHRR	Advanced Very High Resolution Radiometer
BS	Bootstrap Sample
C	Circumference (cm)
CART	Classification And Regression Tree
CCD	Cold Cloud Duration
CESBIO	Centre d'Etudes Spatiales de la BIOsphère
CFS	Committee on World Food Security
CILSS	Comité inter-États de lutte contre la sécheresse au Sahel
CPC	Climate Prediction Center
CSE	Centre de Suivi Ecologique
CSWB	Crop Specific Water Balance
CV	Coefficient of variation
DMP	Dry Matter Productivity
ECOeast	Ferruginous Pastoral Region
ECOnorth	Northern Sandy Pastoral Region
ECOsouth	Eastern Transition Region
ECOwest	Mixed Pastoral-Agricultural Region
ENVISAT	ENVIronment SATellite
EO	Earth Observation
EOS	End Of Season
EOSp	End Of Season from rainfall
ETm	Maximum evapotranspiration in the decadal period
EWS	Early Warning Systems
FAO	Food and Agriculture Organization of the United Nations
FAPAR	Fraction of Absorbed Photosynthetically Active Radiation ($MJ.m^{-2}$)

FEWS Net	Famine Early Warning Systems Network
fr	Relative frequency of the stratum along the transect
GCOS	Global Climate Observing System
GEOV1	Geoland Version 1
GeoWRSI	Geo-spatial Water Resource Satisfaction Index
GPP	Gross primary production ($g.m^{-2}$)
HAPEX	Hydrologic Atmospheric Pilot Experiment
ILCA	International Livestock Centre for Africa
ILRI	International Livestock Research Institute
ISRA	Institut Sénégalais de Recherches Agricoles
Kc	Crop coefficient
LAC	Local Area Coverage
LAI	Leaf Area Index
LOS	Length Of Season
LUE	Light Use Efficiency
MAE	Mean Absolute Error
METOP	Meteorological Operational
MODIS	Moderate Resolution Imaging Spectroradiometer
ms	Dry matter rate (%)
Mse	Dry matter rate (%)
MSG	Meteosat Second Generation
MVC	Maximum Value Composite
NDVI	Normalized Difference Vegetation Index
iNDVI	Seasonal integrated of NDVI
NDVIpk	Seasonal peak value of NDVI
NGOs	Non-Governmental Organizations
NIR	Near-infrared portion of the electromagnetic spectrum (%)
NMAE	Normalized Mean Absolute Error
k-NN	k-Nearest Neighbors
NOAA	National Oceanic and Atmospheric Administration
NPP	Net primary production ($g.m^{-2}$)
OLCI	Ocean and Land Color Instrument
OLS	Ordinary Least-Squares
PAR	Photosnthetically Active Radiation received by the canopy ($MJ.m^{-2}$)
PAW	Plant Available Water
Pe	Primary foliage production of one species (kg)
PET	Potential Evapotranspiration

Pf	Foliar production (kg·DM.ha^{-1})
Ph	Herbaceous dry matter production (kg·DM.ha^{-1})
Pi	Primary foliage production (kg)
PLS	Partial Least Squares
pm	Average green weight (g·m^{-2})
PM	Passive Microwave
PPT	Annual Rainfall Amount
PROBA-V	Project for On-Board Autonomy-Vegetation
*Ps*0	Standard average dry weight of foliar biomass of 10 twigs (kg)
Pve	Average weight of fresh foliar biomass of 10 twigs (kg)
R	Red portion of the electromagnetic spectrum (%)
R^2	Coefficient of determination
RD	Root Depth
RF	Random Forest
RFE	Rainfall Estimate
RG	Global Radiation (MJ.m^{-2})
RMSE	Root Mean Square Error
RPCA	Réseau de prévention des crises alimentaires
S	Total area of four sampling plots (ha)
SD	Standard Deviation
SG	Savitzky-Golay
SOS	Start Of Season
SOSp	Start Of Season from rainfall
SPIRITS	Software for the Processing and Interpretation of. Remotely sensed Image Time Series
SPOT	Satellite Pour l'Observation de la Terre
SSM/I	Special Sensor Microwave Imager
STEP	Sahelian Evaporation Transpiration and Production
SW	Soil Water Content
SWC	Soil Water Level
TAMSAT	Tropical Applications of Meteorology using SATellite
TIMESAT	Software package to analyse time-series of satellite sensor data
TIR	Thermal infrared
ULg	Université de Liège
USAID	United States Agency for International Development
VGT	VEGETATION
VIF	Variance Inflation Factor

VIP	Variable Importance in the Projection
VIs	Vegetation Indices
VITO	Flemish Institute for Technological Research
V-test	Wilcoxon signed rank test
W_d	Amount of water stored in the soil at the end of the dekad
W_{d-1}	Amount of water stored in the soil at the end of previous dekad
WDEF	Water Deficit
WHC	Water Holding Capacity
WP	Work Package
WRSI	Water Requirement Satisfaction Index
WSUR	Water Surplus
WWF	World Wide Fund
$\mathcal{E}$	Environmental stress scalar
$\mathcal{E}_c$	Net conversion efficiency
$\mathcal{E}_p$	Maximum biological efficiency of PAR conversion to dry matter
$\mathcal{E}_s$	Climatic efficiency

Table of Contents

List of Figures

List of Tables

1 Introduction

1.1. Research framework

1.1.1. General context and research justification

Livestock is the first renewable resource in Sahel (Dicko *et al.*, 2006), and in particular West Africa. Livestock is dominated in this region by pastoral systems with large herds of cattle and small ruminants (RPCA, 2010). For this farming type, more than 90% of dry matter consumed by livestock comes from natural pastures (Carrière, 1996). Then, the rangelands form an indispensable component for meeting the needs of animal production in the West African Sahel. Also, they play an important ecological role, through soil fixing, carbon uptake and conservation of biodiversity. The grazing lands represent approximately 21% of the total area of Senegal (ISRA, 2003). They are largely dominated by natural pastures, and are found mainly in the eco-geographical zones of the Senegal River, the sylvo-pastoral zone (Ferlo), the eastern Senegal and the Upper Casamance.

As in the whole Sahel, the Senegalese farming systems have undergone the effects of drought and agricultural influence over grazing lands, since early 1970s. These actions drove firstly, to a significant reduction of the herbaceous cover, the dieback of woody plants (Wispelaere, 1980) and the change in species composition (Akpo, 1990), and secondly, to the abatement of natural rangelands available for pastoral livestock. The pasture lands varied from 58,232 km² in 1983 to 57,224 km² in 2011 with an area decrease about 2% (Figure 1.1). Following this dynamics, the rangelands quality also decreases in the long term, because the best lands are mostly reserved for agriculture (Carrière, 1996). The livestock size (i.e. Cattle, Sheep, Goats,

Horses, Asses and Camels) has largely increased over the past decades and therefore the fodder stock available at the end of the growing season is increasingly limited to meet the livestock feeding needs. This increase of animal population can be explained by the important role of livestock in the socioeconomic development of the Senegalese people that also highly increased in the same period. In Senegal, livestock provides about a third of the national agricultural wealth (Magrin, 2008; Cesaro *et al.*, 2010). The animal production affects in particular a significant part of the rural population for which they provide food security, savings, labor force and field fertilization (ISRA, 2003). With the reduction of natural grazing lands, mostly generated by the expansion of croplands, and the decrease in pastures productivity due to drought effects, the increase of the livestock size leads to rangelands overload whose persistence can lead to their irreversible deterioration (CSE, 2010). Figure 1.2 shows a scheme of the overall context of the rangelands exploitation and their consequences.

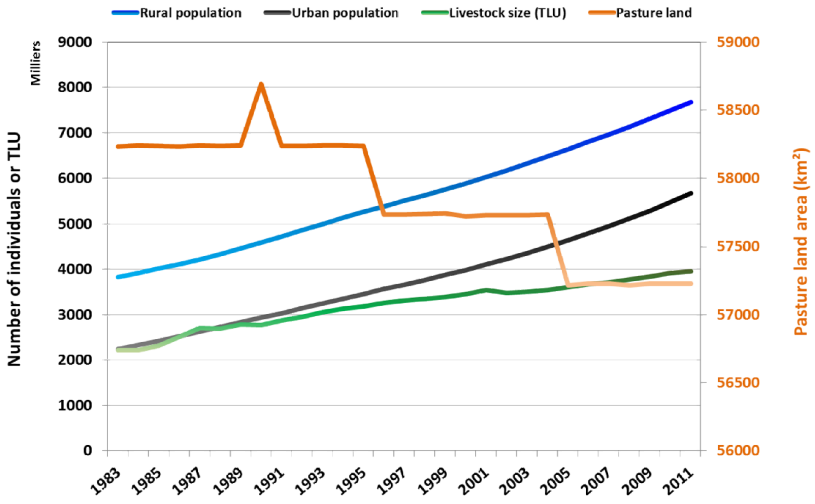

Figure 1.1 – Evolution of human and livestock population (in Tropical Livestock Unit or TLU) as well as grazing area from 1983 to 2011 in Senegal. Source: (FAOSTAT, 2016).

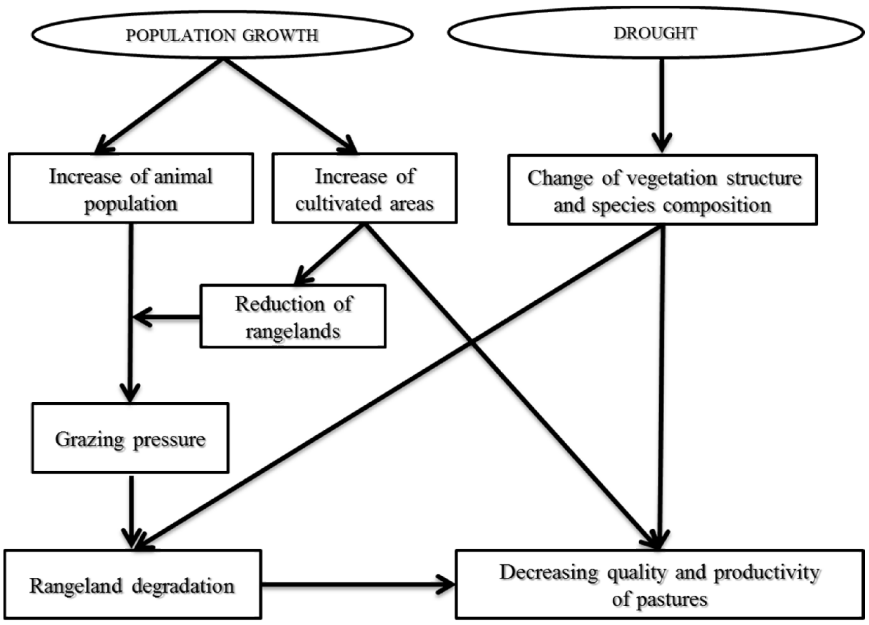

Figure 1.2 – Schematics of the general context of the natural resources use and change in Senegalese rangelands

In this context, many studies have been conducted in the Sahel to establish methods for assessing the forage resources (Tucker *et al.*, 1983; Tucker *et al.*, 1985; Prince, 1991; Mougenot *et al.*, 2000). Among these methods we can cite the one proposed by Tucker *et al.* (1983) and Tucker *et al.* (1985), including the temporal integration of the NDVI (Normalized Difference Vegetation Index) and the aboveground plant biomass produced during a growing season. This technique was revolutionary in 1980's since it was the first time this new NDVI was related to field biomass data, through a simple linear regression. The overall approach has been operationally applied by the *Centre de Suivi Ecologique* (CSE) of Dakar (Diallo *et al.*, 1991; Diouf and Lambin, 2001) (and also the Department of Livestock and Animal Industries of Niger) for evaluating the fodder biomass available in pasture lands at the end of the growing season. In Senegal, the sampling of herbaceous and woody foliar biomass is performed within ground sites distributed across the Senegalese pastoral domain that covers approximately fifteen administrative districts for an area of about 125,000 km². The ground biomass data collected each year have been typically used to calibrate by

regression the NDVI derived from the low resolution satellite imageries of the NOAA / AVHRR (from 1987 to 2002), SPOT / VEGETATION (from 2003 to 2013) and Proba-V (from 2014 to present). The output product is a map of the total plant biomass yield (kg.DM/ha) with 1 km resolution across the whole Senegal. This information carrier of the fodder biomass availability at the end of the growing season helps for assessing the balance between the amount of fodder biomass production and the livestock. It constitutes, therefore, a helpful guide tool for implementing the annual backup strategies and protection of pastoral flocks. It is also a useful tool for the identification of areas with very high dry matter production which could be a starting point of bushfires, and so to prevent against degradation of natural resources. However, there exists a significant time from the data collection and the publication of model results. Several studies also supported that the simple linear relationships developed annually depend highly on the ongoing growing season and the studied region (Cornet, 1984; Bégué, 2002). These facts of course, limit the operational capacity of the method specially to deal with the current needs of the early warning systems (EWS) on fodder biomass availability and of the public institutions in charge of environmental protection against bushfires. Despite these limitations, the approach has not changed over the last 30 years whereas the earth observation (EO) data have largely progressed. The EO data are more reliable with more cloud free images. With new methods implemented in various analysis tools (e.g. TIMESAT, SPIRITS, GeoWRSI etc.), seasonal profiles can be extracted from EO time-series and seasonal metrics computed accordingly either for vegetation products such as NDVI, the Fraction of Absorbed Photosynthetically Active Radiation (FAPAR), the Leaf Area Index (LAI) or for satellite based rainfall data. Making use of the new data and statistical methods which have emerged especially in the past 10 years, can allow improving the 1980’s approach to a newer state-of-the-art. Definitely, this enables to develop more suited tools for assessing more accurately the plant biomass production at the end of the growing season, and to make warnings as early as possible in the season applying specific metrics.

1.1.2. Main features of the studied region

The Sahel name comes from two Arabic words, (i) *Es-Sahel* that means "shore" as "the shore of south Sahara" and (ii) *S'hel ou Sahil* meaning "plain" or "flatlands" (Tracol, 2004; Hiernaux and Le Houérou, 2006). Located between 12 °N and 20 °N latitudes, the Sahel is a biogeographic entity defined generally by its arid to semi-arid tropical climate (Hiernaux and Le Houérou, 2006), with an unimodal rainfall regime. The rainfall regime is mainly controlled by the monsoon from the Gulf of Guinea and the Sahara Harmattan. The Sahel is a transition zone between the arid to hyperarid Sahara areas in the north and the humid tropical areas of the south Sudan savanna (Brooks, 2004). It extends about 6000 km between the Atlantic Ocean (west) and the Red Sea (east) for a width between 400 and 600 km from north to south (Le Houerou, 1989). The Sahel belt passes over the Cape Verde, Mauritania, Senegal, Mali, Niger, Chad, Sudan and crosses the north of Burkina Faso, Nigeria and Cameroon, for a total area of 3 million km² (Tracol, 2004). Figure 1.3 shows the major land cover classes of the Sahel as proposed by the Global Land Cover 2000 (GLC2000) map (Bartholomé and Belward, 2005).

Part of the Sahel belt and member of the Comité inter-États de lutte contre la sécheresse au Sahel (CILSS), the Senegal covers an area of 196,722 km² and is located in the extreme west of Africa, between 12° 20' N and 16° 40' N latitude and 11° 20' W and 17° 30' W longitude. It is bounded by the Atlantic Ocean to the west, the Islamic Republic of Mauritania to the north and northeast, Mali to the east, Guinea Bissau and Guinea Conakry to the south. It also has a border with Gambia which draws an enclave of about 300 km long and 20 km width for an area of 10 300 km².

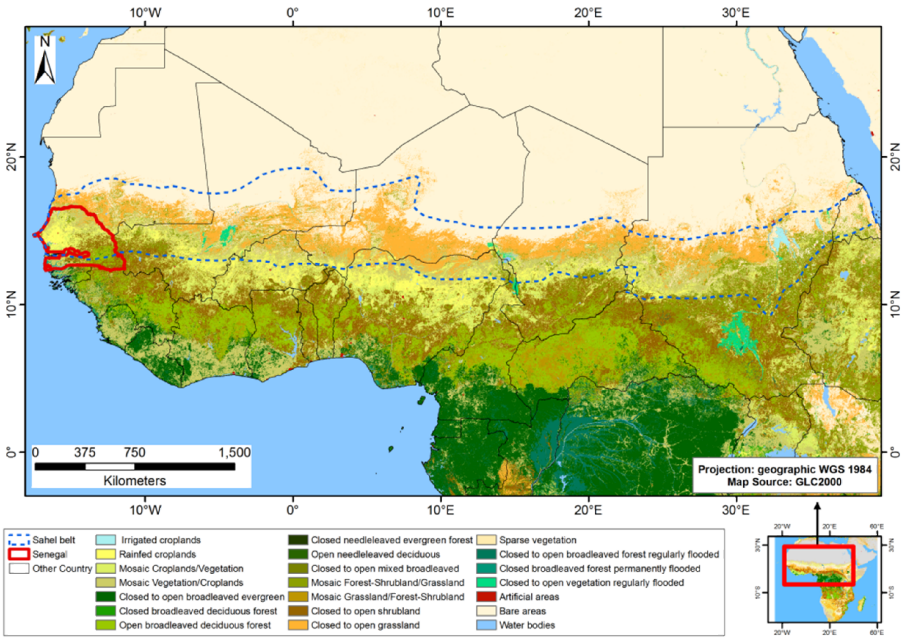

Figure 1.3 – Major Land cover classes of the Sahel provided by the GLC2000 map. The Sahel limits are based on annual average precipitation (African Rainfall Climatology Version 2 1983–2013).

The study area covers 15 districts belonging to five administrative regions (i.e., Saint-Louis, Matam, Kaffrine, Louga and Tambacounda) in Senegal, with a total area of 125,000 km² (Figure 1.4). The area lies in the Sahelian and northern Sudano-Guinean zone of Senegal between 16.69°N and 12.63°N latitude and 16.74°W and 11.86°W longitude. It includes all natural grazing areas of the pastoral domain as defined by (Stancioff *et al.*, 1986), as well as some croplands, including fallows. The mean annual precipitation varies between 200 and 980 mm from north to south, with reference to the FEWS Net rainfall estimates for the 2000-2015 period (Herman *et al.*, 1997; Xie and Arkin, 1997). The rainy season, driven by the West African monsoon, is unimodal, occurring over 3-5 months between June and October. The studied area includes different ecoregions, with a prevalence of red-brown sandy soils, ferruginous tropical sandy soils, leptic, gley and vertic soils (Maignen, 1965; Tappan *et al.*, 2004). Typical of the Sahel, the herbaceous vegetation is particularly dependent on the intra-seasonal rainfall distribution (Valenza, 1977) and is dominated by annual plants with C4-type photosynthesis (Hiernaux and Le Houérou, 2006).

According to the Centre de Suivi Ecologique (CSE, 2013; CSE, 2014), the northern zone (<= 300 mm annual rainfall) is characterized by Poaceae such as *Chloris prieurii, Aristida mutabilis* and *Dactyloctenium aegyptium*, as well as legumes such as *Alysicarpus ovalifolius* and *Zornia glochidiata*, whereas in the central zone (between 300 and 500 mm) the characteristic species are *Zornia glochidiata* (Fabaceae) and *Schoenfeldia gracilis, Pennisetum pedicellatum* and *Eragrostis tremula* (all Poaceae). Towards the south, Andropogoneae species such as *Andropogon pseudapricus* and *A. amplectans* are the most common species. Other species, such as *Spermacoce stachydea* (Rubiaceae), *Cassia obtusifolia* (Fabaceae) and *Fimbristylis exilis* (Cyperaceae) occur irregularly, depending on the terrain morphology (e.g., depressions) and human and livestock influence.

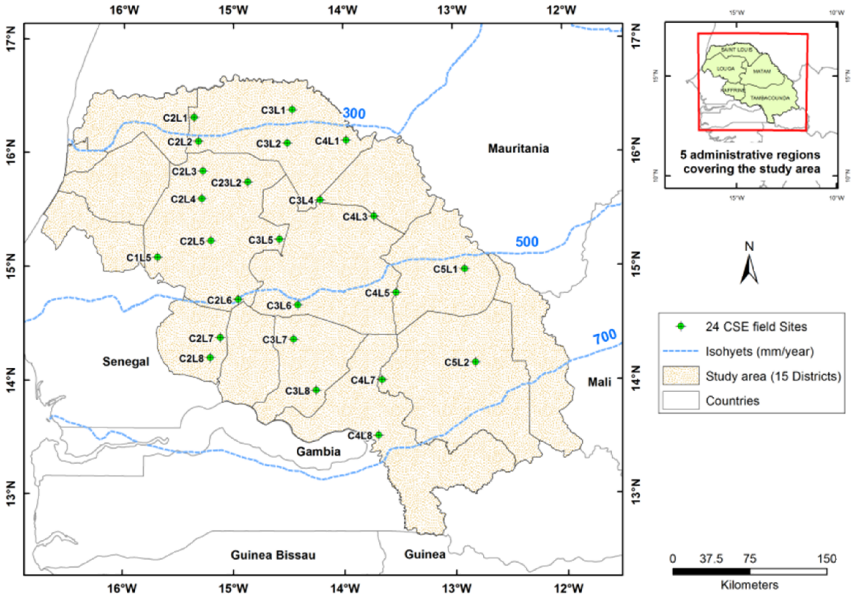

Figure 1.4 – Location of the study area and the ground control sites

Note: All information provided about the study area has adapted from the publication (Diouf *et al.*, 2016).

1.2. Monitoring sites and plant biomass data

The CSE established 36 ground control sites in 1987 for monitoring biomass at the end of the growing season in Senegalese pastoral areas. Among these sites, only 24 were monitored until now (see Figure 1.4) and used in this study. Sites are located in homogeneous vegetated zones of 3×3 km², making them ideal for comparison with moderate/coarse resolution remote sensing data. They are representative of the main geomorphological forms of sampled landscapes (Diallo *et al.*, 1991). Sites were selected, indeed, based on a number of criteria as their accessibility, their uniformity in ecological conditions, their distance from the cropping areas and boreholes as well as their representativeness of the surrounding landscape. They are arranged in a grid (virtual) identified by its columns (C) and lines (L), hence their identification in Cx Ly, where x and y are numbers varying between 1 and 9 in relation to their position on the intersection of columns and lines into the grid (see sites name in Figure 1.4).

The *in situ* plant biomass data used in this research were obtained from the CSE database and covered the 1999–2015 periods, apart from 2004 when no data was collected. We have personally participated at data collection in 2013 and 2014 for the purposes of this thesis. The measurements were conducted annually at the end of the growing season (October), separately for the herbaceous and woody layers. The herbaceous and woody leaf biomass values were subsequently added together to provide an estimate of total plant biomass production, if any. The *in situ* data were not regularly collected, however, for all monitoring sites due to occasional lack of logistics or the early passage of bush fires before planned conductance of field campaigns.

1.2.1. Collection of herbaceous biomass

The collection technique used for the herbaceous layer was the stratified sampling line developed by the International Livestock Centre for Africa (ILCA) for monitoring pastoral ecosystems in Mali's Gourma region (Cissé, 1980). Taking a 1000 m transect, each meter on one side of the measuring

line is allocated to one of four density/production strata, ranging from 0 to 3: 0 = bare soil; 1 = low production; 2 = medium production; and 3 = high production. Then, between 35 and 100 plots of one square meter are chosen randomly along the transect, taking account of the variability of different strata. The plant biomass in each plot is cut close to the ground and weighed with a precision scale. After three strata (low, medium, and high production) are re-sampled, only three samples of about 200 g of fresh material are taken back to the laboratory and dried in an oven (three samples for each stratum = nine samples for the site). They are dried for 48 h at 110 °C in order to obtain the dry matter.

The dry matter rate, obtained by dividing the dry weight of the sample by the green weight, is then integrated into an equation for calculating the herbaceous biomass production. The herbaceous dry matter production of the site is given by adding together the dry matter production of all three strata (low, medium, and high). For each stratum, the calculation equation is written as follows:

$$Ph = fr \times pm \times ms \times 10 \quad (1)$$

where Ph is the herbaceous dry matter production ($kg \cdot DM.ha^{-1}$) in the stratum, fr is the relative frequency of the stratum along the transect, pm is the average green weight ($g \cdot m^{-2}$) measured in the field, ms is the dry matter rate, and 10 is a conversion factor for translating $g \cdot DM/m^2$ into $kg \cdot DM/ha$.

1.2.2. Collection of woody leaf biomass

The leaf biomass of trees and shrubs was sampled for each site in two steps, one repeated every 2 years and one annually. In the first step, every 2 years, four circular plots were delineated and centered at 200, 400, 600, and 800 meters of the 1000 m-long transect. The plot size depended on vegetation density and varied from 1 to $^1/_{16}$ ha. The plots tended to be larger in the open land that characterizes the Sahelian area in the north of the study area and relatively smaller in the North Sudanian domain where the woody plant cover is denser. Within these plots, the parameters measured were

individual height, number of live trunks (some species, such as *Guiera Senegalensis* and *Boscia Senegalensis*, can have several trunks), plant cover, circumference at the base of the trunk(s), and phenological state (flowers, fruits, *etc.*). In the second step, the most representative species were sampled annually and, for each species type, 10 twigs were defoliated and fresh leaf biomass weighed. About 200 g of fresh leaves were dried and weighed in order to determine the leaf dry matter content. The primary foliage production of a given site reflected the total amount of leaf biomass produced for all trees and shrubs. Therefore, for each plant belonging to a given species, the primary foliage production (Pi) in kg was reached via an allometric relationship that integrates its circumference at the base. This relationship has been established for certain species of trees in the Sahel (Diallo *et al.*, 1991; Diouf and Lambin, 2001) as a result of the work done by (Cissé, 1980) and (Hiernaux, 1980) in Mali. The expression is written as follows:

$$Pi = a \times C^b \tag{2}$$

where *a* and *b* are two constants, depending on the species, and *C* is the base circumference of the trunk in cm.

The primary foliage production in kg of one species (Pe) within the four sampling plots of the site is obtained using the following formula where n represents the number of individual plants inventoried:

$$Pe = \sum_{i=1}^{n} Pi \tag{3}$$

The correction of this primary foliage production (including fruits, if any) (Cissé, 1980) into foliar production (Pf) (i.e., material that is available for the ongoing season and can be eaten by livestock as fodder) is done using the formula:

$$Pf = Pe \times \frac{Pve \times Mse}{Ps0} \times \frac{1}{S} \tag{4}$$

where *Pve* is the average weight of fresh foliar biomass (kg) of the 10 twigs, *Mse* is the dry matter rate in %, *Ps*0 is the average dry weight of foliar biomass (kg) of 10 twigs (reference value), and *S* is the total area of the four sampling plots in hectares.

Finally, the leaf biomass of one site was obtained by adding together the foliar production of all the inventoried woody species.

1.2.3. Plant biomass filtering

The ground measurements conducted by different people over 15 years were subject to uncertainty in sampling, measuring, and post-processing, which inevitably resulted in some unrealistic outliers. The post-processing of data (typing handwritten values into digital values) was never quality-checked, thus, in order to filter out obviously erroneous observations related to typos, etc., the datasets went through a two-step filtering: (1) first, the data were examined through an exploratory analysis of all observations using a boxplot with boundaries that were ±1.5 times the interquartile range and (2) since biomass production in Sahel is highly dependent on rainfall (Figure 1.10), a second filtering of biomass estimates was conducted with outliers identified in the first step to remove unrealistic values that showed no relation to the rainfall estimates obtained from FEWS Net imagery (Xie and Arkin, 1997). This rainfall data has a resolution of 8 km and is thus able to capture the heterogeneous pattern of the study area. Only those observations for which the difference between the anomalies of plant biomass and anomalies of annual rainfall was less than 65% were selected. This threshold value was determined by referring to the total biomass anomalies explained by the linear regression with rainfall anomalies (Figure 1.5b). The filtering removed 33 observations in the 1999–2013 periods, leaving 263 observations in the dataset for further analysis. This dataset was used into Chapter 2, and 3 while for Chapter 4, 34 observations from the years 2014 and 2015 were added to the dataset after they have been checked with same procedure described above.

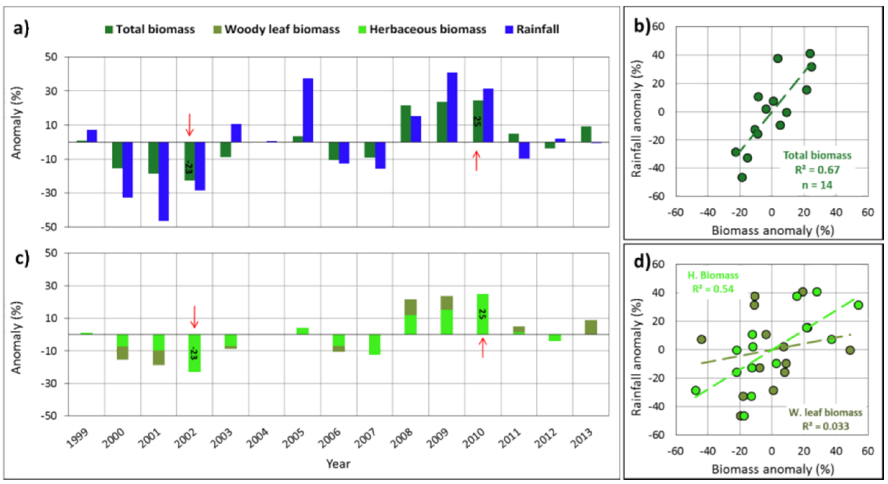

Figure 1.5 – Inter-annual evolution of total biomass and rainfall anomalies (a), and comparison of total biomass (b), herbaceous and woody leaf biomass contribution in global anomalies (c), and herbaceous and woody leaf biomass (d) with annual rainfall anomalies averaged in the dataset between 1999 and 2013 (ground biomass data are missing for 2004). Red arrows in (a) and (c) show the highest total biomass anomalies in the time series (*i.e.*, years 2002 and 2010). Note that the relationship between rainfall and biomass is strengthened when woody leaf and herbaceous biomass are added together.

Note: All information provided in this Section 1.2.2 have been adapted from the paper (Diouf *et al.*, 2015).

1.3. Remote sensing data

After the launch of the first Landsat satellite in 1972, many studies had demonstrated the advantages of using satellite remote sensing data for vegetation monitoring, with indices using the reflectance recorded by the satellite sensors. That is why the United Nations Conference on the Human Environment, held in Stockholm in 1972, suggested use of the remote sensing as a tool of "direct global monitoring" (Grainger, 2013). For monitoring the vegetation cover across the world, various satellite from medium to low spatial resolutions were launched and several indices and biophysical products developed for many application. The overall characteristics of the most used satellites in the past 40 years are shown in Table 1.1.

Table 1.1 – The main medium and low spatial resolution satellites and their overall characteristics

Satellite		**NOAA**	**SPOT**	**TERRA/AQUA**	**MSG**	**ENVISAT**	**METOP**	**PROBA**	**SENTINEL3**
Sensor		AVHRR	VGT	MODIS	SEVERI	MERIS	AVHRR	V	OLCI
Frequency		daily	daily	1/2 days	15mn	3/5 days	daily	daily	1/2 days
Spatial resolution		1.1km	1km	250/500m	3km	300/1000m	1km	300/1000m	300m
Swath width		2800km	2,200km	2300km	hemisphere	1150km	2500km	2500km	1270Km
Wavelength (nm)	Blue		430-470			437.5 - 447.5			
	Red	580-680	610-680	620 - 670	560-710	677.5 - 685.0	580-680		
	NIR	725-1000	780-890	841 - 876	740-880	855 - 875	725-1000		
Service start		1979	1998	1999	2002	2003	2007	2013	2016
Planned service end		-	2013	2015	2015	2013	2016	2017	2023

1.3.1. Vegetation indices products

There exist a large variety of quantitative indices of vegetation conditions using remote sensing instruments. These vegetation indices (VIs) measure the green vegetation that has the particularity through the chlorophyll to absorb solar energy for use in the photosynthesis process. The green vegetation has a variable spectral signature related to the chlorophyll activity in plants. In favorable conditions (healthy canopies), the chlorophyll in green vegetation absorbs red portion (Red) of the electromagnetic spectrum, while the near-infrared portion (NIR) is strongly scattered by the spongy structure of the mesophyll in leaves, due to the presence of numerous intercellular spaces. In contrast when conditions are unfavorable, for example in situations of water deficit, the NIR reflectance is low, while the return of Red (poorly absorbed) to the sensor is higher. It is this contrast between Red and NIR reflectance by canopies of green vegetation that has been considered to develop the large set of existing VIs nowadays. The VIs can be classified into two groups after (Baret and Guyot, 1991; Jackson and Huete, 1991): (1) ratios (or slope-based) and (2) linear combinations (or distance-based) VIs. Figure 1.6 shows a schematic distribution of pixels within a bi-dimensional plot (scattergram) of Red against NIR reflectance of a given scene, and allows easy distinction of these two groups.

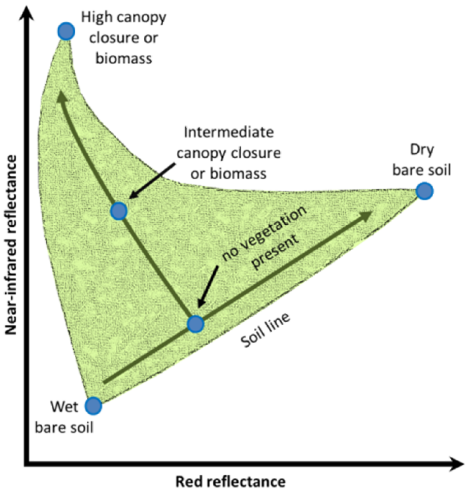

Figure 1.6 – Distribution of all pixels of a scene into the red and near-infrared bi-spectral space. Adapted from Silleos *et al.* (2006).

The **slope-based VIs** are simple ratio, or the ratio of sums, differences or products of the Red and NIR spectral bands (Jackson and Huete, 1991; Pettorelli, 2013). These VIs focus on the contrast between the spectral response patterns of green vegetation in the red and near-infrared portions of the electromagnetic spectrum and their values indicate both the state and abundance of green vegetation cover and plant biomass (Silleos *et al.*, 2006). The position of each point in the 2-dimensional Red-NIR space is geometrically equivalent to the slope of the line connecting the origin of reference and this particular point on the scattergram (Mróz and Sobieraj, 2004). Table 1.2 shows the main slope-based VIs used in the literature.

Table 1.2 – Slope-based vegetation indices

Name	Formula	Author
Ratio Vegetation Index (standard)	$RATIO = \frac{NIR}{R}$	Birth and McVey (1968)
Normalized Difference Vegetation Index	$NDVI = \frac{NIR - R}{NIR + R}$	Rouse *et al.* (1974)
Soil-Adjusted Vegetation Index	$SAVI = \frac{NIR-R}{NIR+R}(1 + L)$	Huete (1988)
Transformed Vegetation Index	$TVI = \sqrt{\frac{NIR - R}{NIR + R}} + 0.5$	Deering *et al.* (1975)
Ratio Vegetation Index (inverse)	$RVI = \frac{R}{NIR}$	Richardson and Wiegand (1977)
Enhanced Vegetation Index	$EVI = G * \frac{NIR - R}{NIR + C_1R - C_2B + L}$	Liu and Huete (1995)

Where NIR = near-infrared, R = red, B = blue, L = soil adjustment factor, C_1 and C_2 = constants and G = a gain factor.

The **distance-based** VIs are functionally different to the slope-based ones and are computed from linear combinations of the Red and NIR bands. These VIs are calculated by taking into account the difference of any pixel's reflectance from the reflectance of bare soil (Baret and Guyot, 1991; Silleos *et al.*, 2006). This requires establishing the "soil line" that corresponds to the linear regression line of the NIR band against the Red band for a sample of bare soil pixels. Then it enables to remove the effect of soil brightness in cases where vegetation is sparse and pixels are supposed to be contaminated by the soil background. Pixels located near the soil line are assumed to represent the soil, while those far away are assumed to represent vegetation. The distance-based VIs require the slope and intercept of the soil line as inputs for their calculation as well. Table 1.3 shows examples of distance-based VIs frequently cited in the literature.

Table 1.3 – Distance-based vegetation indices

Name	Formula	Author
Perpendicular Vegetation Index 1	$PVI_1 = \frac{(aNIR - R) + b}{\sqrt{a^2 + 1}}$	Perry Jr and Lautenschlager (1984)
Difference Vegetation Index	$DVI = aNIR - R$	Richardson and Wiegand (1977)
Weighted Difference Vegetation Index	$WDVI = NIR - aR$	Clevers (1988)
Transformed Soil-Adjusted Vegetation Index 1	$TSAVI_1 = \frac{a(NIR - aR - b)}{R + aNIR - ab}$	(Baret *et al.*, 1989)
Modified Soil-Adjusted Vegetation Indices 1	$MSAVI_1 = \frac{NIR - R}{NIR + R + L}(1 + L)$	Qi *et al.* (1994)

Where NIR = near-infrared, R = red, a and b are slope and intercept of the soil line respectively,

L = 1- 2*a*NDVI*WDVI (Weighted near-infrared-red Difference Vegetation Index)

Among all VIs, the NDVI remain the most commonly used in the Sahel for the quantification and temporal monitoring of the vegetation covers (Bénié *et al.*, 2005). Indeed, it gives a quantitative measure that only yields relative estimates of vegetation amounts, but it can be used to better reflect the actual changes in primary production, as well as for quantification of its absolute value (Seaquist *et al.*, 2003). NDVI time series were used in several studies to depict information on the spatial distribution of bioclimatic zones (Jönsson and Eklundh, 2004) and also on their cover change/trend over time (Anyamba and Tucker, 2005; Wei *et al.*, 2012; Brandt *et al.*, 2014).

1.3.2. Vegetation biophysical products

The Fraction of Photosynthetically Active Radiation absorbed by vegetation (FAPAR) and the Leaf Area Index (LAI) are important biophysical variables for quantifying interactions between the vegetation surface and the atmosphere (Myneni *et al.*, 2002; Tian *et al.*, 2004; Demarty *et al.*, 2007; Baret *et al.*, 2013). LAI and FAPAR represent two biophysical complementary ways of describing the earth's vegetated surfaces (Fensholt *et al.*, 2004) and they play a key role in several surface processes, including

photosynthesis, transpiration rates, rainfall interception and gas exchange. For this raison, they have been identified by the GCOS (2006) to be essential terrestrial climate variables in the context of global change studies.

Defined as half of the total intercepting area per unit ground surface area (m^2 leaf area per m^2 ground area) (Chen and Black, 1992), the LAI characterizes the functioning surface area of a vegetation canopy (Myneni *et al.*, 2002) and it quantifies the thickness of the green leaf area of terrestrial vegetation (Fensholt *et al.*, 2004). Satellite remote sensing enables retrieval of LAI with algorithms based on the physics of radiative transfer (Hanes, 2014). FAPAR is defined as the fraction of photosynthetically active radiation (400–700nm) absorbed by the vegetation canopy and expresses, thus, a canopy's energy absorption capacity (Fensholt *et al.*, 2004; Fan *et al.*, 2014; Hanes, 2014). In order to avoid the local-specific disadvantage of empirical methods (i.e., statistical) which use spectral vegetation indices (e.g., NDVI) or LAI to calculate the FAPAR, physical inversion methods are commonly used nowadays based on radiative transfer models (Knyazikhin *et al.*, 1998; Myneni *et al.*, 2002; Baret *et al.*, 2007; Gobron *et al.*, 2007; Fan *et al.*, 2014; Li *et al.*, 2015). These latter models which describe the transfer of solar radiation in vegetation canopies are generally based on the energy conservation law (transmittance, reflectance and absorptance of the canopy) and retrieve the FAPAR products using satellite remote sensing observations (e.g., canopy spectral and directional characteristics) as constraints (Knyazikhin *et al.*, 2004; GCOS, 2006). Plant biomass production is closely related to light interception (van Wijk and Williams, 2005), which is determined by FAPAR and LAI. These products, therefore, were used as key variables in many models of net primary production (NPP) to calculate the surface ecosystems productivity (Potter *et al.*, 1993; Field *et al.*, 1995; Running *et al.*, 1999). Both FAPAR and LAI products are globally available now, through remotely sensed imagery, at different spatial resolutions (250–1000m) and temporal frequency (10–30days).

1.3.3. Rainfall estimates dataset

The number of rain gauges throughout Africa is small and unevenly distributed, and the gauge network is deteriorating (Dinku *et al.*, 2007), explaining accordingly the difficulty to access on field rainfall data. Satellite rainfall estimates constitutes valuable products to ensure continuity of environmental studies in these areas. Several rainfall products have been elaborated and are generally freely available in various temporal and spatial levels. The product used in this research called FEWS Net RFE is produced by NOAA's Climate Prediction Center (CPC) specifically for United States Agency for International Development (USAID) Famine Early Warning Systems (FEWS) to assist in drought monitoring activities over Africa (Dinku *et al.*, 2007). The first version (RFE v1.0) was produced from 1995 to 2000 (Herman *et al.*, 1997) and was replaced from 2001 onwards by the RFE v2.0. This latest version is a blended product based on cold cloud duration (CCD) derived from Meteosat thermal infrared (TIR), estimates from the Special Sensor Microwave Imager (SSM/I) and the Advanced Microwave Sounding Unit (AMSU), and daily station rainfall data (Toté *et al.*, 2015). The main difference between these two versions is that RFE v2.0 uses passive microwave (PM) estimates while RFE v1.0 includes a procedure to estimate warm orographic rain (Dinku *et al.*, 2007). Particularly, RFE v2.0 product was shown to be in good agreement with ground rainfall data (Jobard *et al.*, 2011). Linear regression between annual FEWS Net RFE and field data collected in the period 1996-2012 for Linguere, Podor, Matam and Tambacounda gauge stations gave a coefficient of determination of 0.788. The data are available over Africa and freely accessible at (http://earlywarning.usgs.gov/fews/datadownloads) in dekadal time step and 8 km spatial sampling grid.

1.4. Plant biomass modelling

The warning shot of the years 1967-1968 went unnoticed, quickly erased by the wet year 1969, and then came the "Great Drought" from 1972 which surprised everyone (Leroux, 1995). During this period of drought, thousands of people and millions of animals died because of the famine (Glantz, 1976). The lack of information on available food resources had not provided time to effectively assist the affected populations, particularly in the Sahel, where the drought effects on natural resources were particularly adverse. With this experience, the analysis of the hunger causes have been multiplied and the new concept of "food security" was developed with the creation in 1975 of the committee CFS on world food security (FAO, 2000). In the same context, the United Nations has inserted into the Charter of Human Rights, the right to sufficient food in 1976 (Bruegel and Stanziani, 2004). However, while the assessment of crop yields has always been the focus for food security, the availability of pastoral resources were actually taken into account in the control systems against hunger in the Sahel, only towards 1970's. Then, the assessment of plant biomass productivity in rangelands became an important research area for many agro-pastoralists who carried out several studies in Senegal e.g. (Bille, 1975; Bille, 1977; Breman and Cissé, 1977; Boudet, 1984). The use of remote sensing data for monitoring the plant production in the Sahel rangelands really began after the launch of the United State weather satellite NOAA-AVHRR. Subsequently, numerous studies have demonstrated the benefits of satellite remote sensing for vegetation monitoring (Tucker *et al.*, 1983; Asrar *et al.*, 1985; Tucker *et al.*, 1985; Seaquist *et al.*, 2003; Fensholt *et al.*, 2006), as well as the good correlations between plant biomass and satellite variables, expressed as models. The question we should ask, therefore, is: "What is a model?" Several definitions are proposed in the literature:

- " *the models are simple representations of systems, defining a system as a coherent part of the real world* " (Djitèye and Penning de Vries, 1991)
- "*a model is a simplified representation of a complex reality expressing the relationships between elements of the system envisaged by the use of mathematical or computational expressions, providing enough operating*

analogy with the studied system to allow forecasts" (Daget and Godron, 1995)

Considering the two definitions, it is clear that the role of models is not always to estimate only a given situation but in some cases they should be able to make forecasting (i.e., estimation in a distant future). To estimate and / or forecast the production of plant biomass with remotely sensed data, two types of models are generally used worldwide: the parametric models and the non-parametric models.

1.4.1. Parametric models

Parametric models are computing tools that rely on assumptions about the shape of the distribution in the underlying samples and about the parameters (i.e., means and standard deviations) of the assumed distribution (Tanya, 2009). Parametric modelling can be divided into two sub-categories such as:

(1) stochastic (or probabilistic) regression modelling that consists to define accurate tools to model the observed/sampled data, taking into account their randomness (Wikistat, 2013). Stochastic models focus more on static patterns than on the dynamic processes that produce them (Bolker, 2008). Such statistical approach, either linear or nonlinear, has the advantage of being simple and quick to set up, focusing directly on relevant indicators. The generated regression models can be single-based variable (i.e., including only one explanatory variable) or multiple-based variables (i.e., including several explanatory variables).

(2) deterministic modelling where the dynamic patterns of input variables are well known as well as the processes that drive these patterns, without any randomness. Deterministic models can be divided into two groups: the mechanistic models for which input variables have a direct correspondence to the underlying mechanisms of the dependent variable being modeled (O'Reilly, 2016) and semi-mechanistic models which use fewer input variables with a lower complexity of their relationships to the dependent variable. In contrast to stochastic models where randomness is present, the deterministic models are completely dependent on the data to which they are trained, and will more easily reveal the fundamental causality

between predictors and the dependent variable (Immunetrics, 2016). Figure 1.7 presents a simple classification of the main existing parametric models.

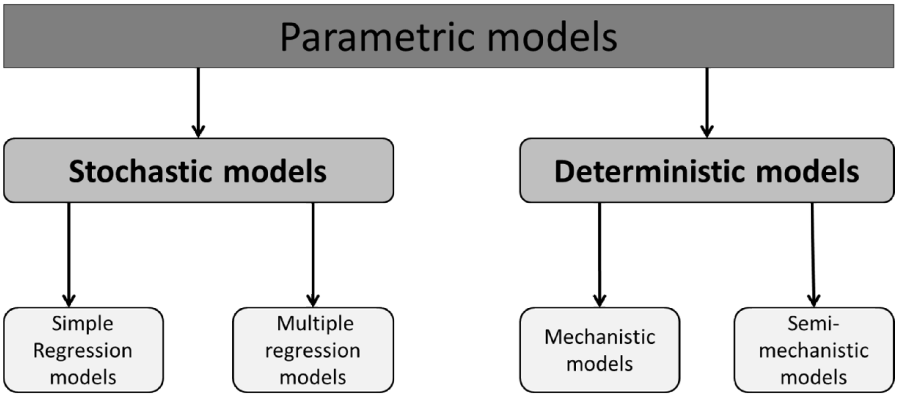

Figure 1.7 – Schematic classification of the main groups of parametric models. This scheme is not exhaustive and can be improved.

1.4.1.1. Simple regression models

To assess the plant biomass productivity, the monocriterion models are generally obtained by a simple regression between ground sampled vegetation mass (dependent variable) and a satellite-derived proxy (independent variable) considered being relevant enough to explain it. The linear regression is the most commonly used function and the basic equation can be written as:

$$\boldsymbol{Y = aX + b + e} \quad (5)$$

where Y indicates the plant biomass production; X is the explanatory variable; a and b are the slope and offset of the regression line respectively and e is the error term.

The model coefficients (a and b) are commonly determined using the ordinary least-squares (OLS) approach with the assumption that the satellite-derived variable is accurately measured (Zhang and Ni-meister, 2014). The herbaceous yield mass estimation has been one of the first remote sensing

applications in the West African Sahel using simple regression models (Jarlan *et al.*, 2008). Those were mainly the linear and exponential models established from field data and seasonal accumulated NDVI (Tucker *et al.*, 1983; Tucker *et al.*, 1985; Diallo *et al.*, 1991; Prince, 1991). The *Centre de Suivi Ecologique* (CSE) in Dakar (Senegal) used this method, operationally over thirty years, to nationally estimate the total production of plant biomass at the end of the growing season. It is used also by the Department of Livestock and Animal Industries of Niger to provide the annual map of fodder biomass in Niger.

1.4.1.2. Multiple regression models

Unlike the simple regression models, the multiple regression ones allow to better account for the interactions between plant biomass (predictand or dependent variable) and various explanatory variables (predictors or independent variables). Among those, metrics of the seasonal profile computed from remote sensing data can be applied to fit the different vegetation phenology. Thus the traditional forecasting models can be improved especially when they are built based on historical data including information from several years. The basic model equation for a given year can be written as follows:

$$\boldsymbol{Y_i = b_0 + b_1 X_{i,1} + b_2 X_{i,2} + \ldots + b_k X_{i,k} + e_i} \qquad (6)$$

where Y_i indicates the plant biomass yield in year i; b_0 is the regression constant; $X_{i,k}$ is the value of k^{th} predictor in year i; b_k is the coefficient on the k^{th} predictor and e_i is the error term.

Multiple regression models from OLS approach takes all predictors into account (Hanes, 2014) and is widely used for estimating forest aboveground biomass e.g. (Foody *et al.*, 2003; Zheng *et al.*, 2004; Lu, 2005; Powell *et al.*, 2010; Günlü *et al.*, 2014). However, coefficients of established models could be instable if explanatory variables themselves are significantly correlated or have a weak relationship with the dependent variable. For this reason,

correlation tests (Pearson's correlation, Kendall rank correlation or Spearman rank correlation) are often used to identify the highly and significantly correlated variables to keep for modelling. Another way used is the Variance Inflation Factor (VIF) retrieved during the regression and indicating that a variable is causing collinearity effect when VIF > 10 (Belsley *et al.*, 1980). In cases where there are several potential explanatory variables, a prior selection of relevant predictive variables can be very useful. To deal with, dedicated stepwise search procedures as the sequential forward selection, backward elimination and stepwise selection (combining the two previous procedures) are generally applied (Hastie *et al.*, 2009; Kuhn and Johnson, 2013). For biomass modelling, these diagnosis steps can be useful before computing multiple regression models, since these models assume basically that the independent variables (remotely sensed predictors) are uncorrelated and that a linear relationship exists between them and the plant biomass (Hanes, 2014).

It is worth to note that the assumption requiring a normal distribution in regression models, applies only to the error term (i.e., the random error in the relationship between the independent variables and the dependent variable) and not to the variables (Statistics Solutions, 2013). In addition, when the sample size is large enough (i.e., >200), the normality assumption is not needed to fit a linear regression as the Central Limit Theorem ensures that the distribution of error term will approximate normality (Lumley *et al.*, 2002; Statistics Solutions, 2013).

1.4.1.3. Mechanistic models

The plant biomass modelling has been the subject to numerous studies in different countries of the world using especially the so-called mechanistic model (Di Bella, 2002). International programs such as the Hydrologic Atmospheric Pilot Experiment in the Sahel (HAPEX-Sahel) initiated in 1991 (Prince *et al.*, 1995) and the *Analyse Multidisciplinaire de la Mousson Africaine* (AMMA) launched in 2001 (Redelsperger *et al.*, 2006), show undertaken effort by the international scientific community in using this kind of models. The mechanistic models help to explain plant biomass production from very complex interactions between factors that influence this

production such as plant physiology, phenology and environmental conditions (i.e., rainfall, soil moisture, texture...). Although more realistic, these very complex models are generally difficult to implement because requiring frequent (daily or even hourly) and detailed data that are not readily accessible in Sahelian countries (Bénié *et al.*, 2005).

As an example, the Sahelian Evaporation Transpiration and Production (STEP) model is by far the most famous among all mechanistic models tested in the Sahelian semi-arid areas. It is an eco-physiological functioning model based on an approach originally proposed by (Rambal, 1980) to monitor the Tunisian semi-arid areas. Developed by the Centre d'Etudes Spatiales de la BIOsphère (CESBIO) in France (Mougin *et al.*, 1995), the model has been used in several thesis research performed typically in the Sahel: Tracol (2004) and Baup (2007) in the Malian Gourma, Zine (2004) in the agricultural region of Fakara in Niger and Faye (2013) in the Senegalese Ferlo. This model is well-adapted to the Senegal environment since its first release tool has been validated in the Senegalese Ferlo and Malian Gourma (Lo Seen *et al.*, 1995). In the last two decades, STEP has experienced significant improvements including its coupling with optical and radar microwave remote sensing data (Tracol, 2004; Baup, 2007; Faye, 2013). For application, the STEP model combines spectral remote sensing measurements with a model describing the functioning processes of the Sahelian annual herbaceous species, taking into account the soil-plant-atmosphere interactions. It includes two main interrelated modules: a vegetation growth module and a water balance module. Within these main modules, other sub-modules are involved. Figure 1.8 shows the STEP model operation with its different components.

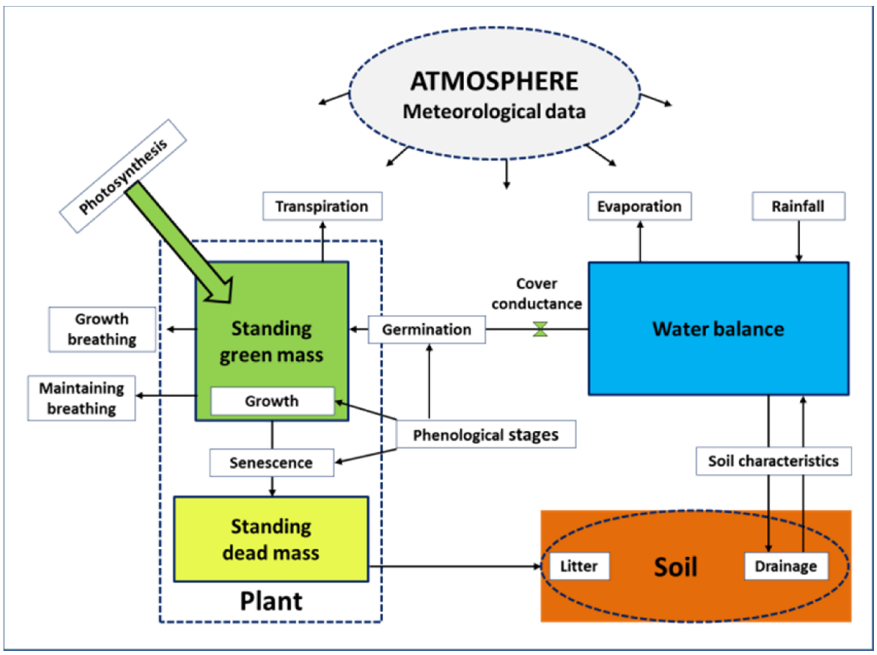

Figure 1.8 – STEP operation diagram representing the relationships between the water balance and vegetation growth/senescence main modules. Adapted from Mougin *et al.* (1995).

1.4.1.4. Semi-mechanistic models

The Monteith's model is the most commonly used as semi-mechanistic model for assessing the plant production. This model is based on the concept of solar light use efficiency (LUE) that can be expressed as a conservative ratio between the absorbed photosynthetically active radiation (APAR) by a canopy and the net primary production (NPP) (Monteith, 1972). The solar radiation is absorbed, indeed, by plants to provide the useful energy for photosynthesis, while moisture deficits control the LUE (Seaquist *et al.*, 2003). Thus, the use efficiency of solar light is maximum when the canopy cover is wet and it is minimum when the canopy water content is low. On this basis, Monteith's model calculates the net primary production as the product of the photosynthetically active radiation absorbed by the canopy and the efficiency with which plants of the canopy convert this energy into dry matter. This efficiency is the product of seven efficiency types which are

difficult to quantify in general. For this reason, the model with seven efficiencies was subsequently amended by (Kumar and Monteith, 1981) and (Varlet-Grancher *et al.*, 1982) in a much simpler form with only three efficiencies as presented in equation 7.

$$NPP = \sum_{i=1}^{n} \varepsilon_c \, f_{APAR} \, \varepsilon_s \, GR \tag{7}$$

where *NPP* is the net primary production (g.m^{-2}) produced during n days; n is the running time step; $\mathcal{E}_c$ is the net conversion efficiency; f_{APAR} is the absorption efficiency; $\mathcal{E}_s$ is the climatic efficiency and *GR* indicates the daily global radiation (MJ.m^{-2}).

The above formula, simulates the NPP that should be differentiated to the gross primary production (GPP) also commonly cited in the literature. The NPP indicates the net carbon flux incorporated by plants from the atmosphere by photosynthesis minus the carbon flux lost by autotrophic respiration, while the GPP corresponds to the carbon retrieved by photosynthesis before any losses of energy (Running *et al.*, 1999). Using the Monteith's concept, Seaquist *et al.* (2003) proposed the GPP model presented in equation 7, based on the Normalized Difference Vegetation Index (NDVI) for grassland biomes across the West African Sahel.

$$GPP = \sum_{i=1}^{n} \varepsilon_p \, \varepsilon \, (aNDVI + b) \, PAR \tag{8}$$

where *GPP* is the gross primary production (g.m^{-2}) produced during n days; *n* is the time step; $\mathcal{E}_p$ is the maximum biological efficiency of *PAR* conversion to dry matter; $\mathcal{E}$ is the environmental stress scalar; *NDVI* is the Normalized Difference Vegetation Index; a and b the regression coefficients and *PAR* indicates the photosynthetically active radiation received by the canopy (MJ.m^{-2}).

Furthermore, the absorbed PAR seems better linked to gross than net primary production (Running *et al.*, 1999) but some authors have found a very good correlation between the absorbed PAR and the net primary production (Asrar *et al.*, 1984; Olofsson *et al.*, 2007). This could justify the

use of Monteith's simplified version to estimate the total dry matter produced during a given period as currently done by the Flemish Institute for Technological Research (VITO) to compute the Dry Matter Productivity (DMP) satellite product on a decadal basis and 1km spatial resolution (Smets *et al.*, 2010). This DMP product is currently used in operational manner to assess plant biomass in West Sahel (Ham and Fillol, 2010).

1.4.2. Nonparametric models

The nonparametric models are statistical tools that make no assumptions regarding the distribution of the input data (Altman, 1990; Tanya, 2009). These models perform recursive partitioning of data sets and solve complex non-linear relationships between the response and predictor variables (Zhang and Ni-meister, 2014). An interesting application example of the nonparametric approaches is given now by machine learning algorithms that are an outgrowth of the intersection of Computer Science and Statistics (Mitchell, 2006). Unlike both simple and multiple regression models, nonparametric tools can handle a very large number of variables from satellite and ancillary data (Zhang and Ni-meister, 2014). These techniques were widely used in recent ecological and environmental studies (Powell *et al.*, 2010; Herrmann *et al.*, 2013; Zhu *et al.*, 2015; Graves *et al.*, 2016; Ke *et al.*, 2016). Table 1.4 presents some examples of machine learning algorithms organized by topic with related packages in the open source software R. More information on these algorithms implemented in the R software are provided by Hothorn (2016).

Table 1.4 – Examples of machine learning algorithms usually applied in ecological and environmental studies.

N°	Main topics	Algorithms	References	Package in R software
1	Neural Networks	Multi-Layer Perception	(Bishop, 1995); Venables and Ripley (2002)	nnet
2	Ensemble learning	Random Forest	Breiman *et al.* (1984)	randomForest

3	Rule system	Cubist	RuleQuest Research (2015)	Cubist
4	Regularization and shrinkage	Least absolute shrinkage and selection operator (lasso)	Tibshirani (1996)	penalized
5	Instance based	K-Nearest Neighbors	Cover and Hart (1967)	caret
6	Support vector machines	Support Vector Machines	Boser *et al.* (1992)	e1071
7	Bayesian	Bayesian Additive Regression Trees	Chipman *et al.* (2010)	BayesTree

After a large review of applied statistical methods for predicting forest inventory attributes among which the plant biomass production, Brosofske *et al.* (2014) concludes that no analytical technique examined - specially the nonparametric ones - emerged as superior for all cases of classification and regression tree (CART), artificial neural networks (ANNs), random forest (RF) and k-nearest neighbors (k-NN). In addition, all these methods have particular strengths and weaknesses and should be determined carefully by identifying the specific goals of an analysis, the type of response to be modeled, the characteristics of the input data, and the resources available.

1.4.3. Semi-parametric models

Semi-parametric regression is a fusion between parametric and nonparametric regressions that integrates low-rank penalized splines, mixed model and hierarchical Bayesian methodology – thus allowing more streamlined handling of longitudinal and spatial correlation after Ruppert *et al.* (2009). This approach was not developed in this document but it was provided only for information and to dress the state-of-art.

1.5. Research focus

1.5.1. Research objectives

The research presented in this dissertation was conducted in the framework of the AGRICAB (Agriculture Capacity Building) project and through the "livestock systems" work package (WP) that was led by the International Livestock Research Institute (ILRI) with participation of the University of Liège (ULg) and the CSE. This WP aims at improving the current approaches or developing news methodologies to estimate the fodder biomass in rangelands and monitoring livestock by using models fed by biophysical variables computed from earth observation (EO). In this context, the main objective of this research is to develop models that include remote sensing data and field biomass data, for estimating and / or forecasting forage availability at the end of the growing season in Senegalese semi-arid areas. This overall research objective can be separated into three specific objectives as follows:

i) To evaluate the existing annual estimation method of total fodder biomass by linear regression (currently applied in Senegal) in comparison to nonlinear approaches;
ii) To develop a multiple regression approach (i.e., parametric models) that integrate FAPAR phenological metrics computed from EO time series for estimating and forecasting total fodder biomass at the end of the growing season;
iii) To develop machine learning approach (i.e., non-parametric models) including agrometeorological variables and FAPAR phenological metrics to estimate the herbaceous biomass at the end of the growing season.

In general, approaches developed in this research are expected to provide diagnosis and prognosis tools to the scientific community, research and development projects and NGOs that are working on pastoral livestock monitoring and management of natural resources (e.g., land degradation, carbon sequestration...) especially in the West African Sahel countries.

1.5.2. Dissertation outline

This dissertation comprises five chapters that can be declined as follows:

Chapter 1 presents the research framework, the study area, the ground biomass data and the remote sensing data used, a review of the statistical modelling approaches for assessing the plant biomass production, and finally the research objectives and the dissertation outline.

Chapter 2 addresses the first objective and starts with a comparative analysis of the linear regression equation with five nonlinear regression functions (cubic, power, exponential, logarithmic and quadratic), following the same approaches currently applied in Senegal on an annual basis. This comparison lies on the hypothesis that the nonlinear models are more suited for plant biomass estimation in Sahelian countries because of (1) the influence of soil background on NDVI values in areas with low plant cover and (2) the saturation problem in areas with very high plant productivity.

Chapter 3 deals with the second objective through the development of multiple linear regression models based on FAPAR metrics and historical field data. The main purpose of the study was to establish end-of-season models of total fodder biomass production (i) for the overall study area, (ii) within four different ecological regions (i.e., ecoregions) and (iii) for early warning across the study area. A comparison of outputs from these new developed models with the CSE biomass product based on NDVI shows if improvements are made with the new multiple linear models in term of precision for fodder biomass assessment in Senegalese Sahel regions.

Chapter 4 addresses the objective three and provides a framework to develop suited nonparametric regression models for herbaceous biomass estimation. Additionally to models establishment, this chapter aims to verify if the saturation problem affecting also the FAPAR product is attenuated when including non-optical data as the rainfall estimates (RFE) from the Famine Early Warning Systems Network (FEWS Net) and related water status indicators. This chapter investigates also the spatial and temporal relationship between the onset/end of the growing season retrieved from FAPAR and those of rainy season calculated from RFE data. An analysis of influence of the growing season onset and rainy season onset on the final

herbaceous biomass yield is carried out to highlight their potential for early monitoring of herbaceous biomass in Senegalese semi-arid areas.

Chapter 5 presents a general conclusion for the different chapters and outlook.

2 Evaluation of simple regression approaches to estimate the total biomass

2.1. Introduction

The rainfall is well known for its very high spatial and temporal variability in the Sahel. It is the main climatic driver of vegetation fluctuations (Hickler *et al.*, 2005; Ali, 2010; Huber *et al.*, 2011; Anyamba *et al.*, 2014). For this reason, the forage availability at the end of the season is particularly variable and difficult to predict (CSE, 2010) in this part of the West African Sahel. The livestock is essentially pastoral with most part of dry matter, consumed by flocks, coming from natural pastures (Carrière, 1996). The plant production of rangelands is crucial to the economy of local people and the satisfaction of their food requirements.

The potential of remote sensing for estimating the annual production of herbaceous pasture in the Senegalese Sahel has been demonstrated by Tucker *et al.* (1983) using the Normalized Difference Vegetation Index (NDVI) from the National Oceanic and Atmospheric Administration-Advanced Very High Resolution Radiometer (NOAA-AVHRR). Subsequently, several studies have been conducted in the Sahel semi-arid areas to better understand the relationship between plant biomass and NDVI applying a simple linear regression function (Tucker *et al.*, 1985; Diallo *et al.*, 1991; Prince, 1991). Currently, the Centre de Suivi Ecologique in Senegal and the Ministry of Livestock and Animal Industries in Niger use this linear regression approach to operationally assess forage availability in Sahelian rangelands. The variables used by these institutions are respectively the cumulated NDVI during the season identified visually and the seasonal maximum of NDVI.

However, the relationship between the above-ground biomass and the NDVI is not exactly linear because of NDVI values that easily saturates

when the vegetation becomes too dense (Box *et al.*, 1989; Pettorelli *et al.*, 2006; Vescovo *et al.*, 2012). In addition, this empirical relationship depends on local conditions of the region where it was developed as well as on the vegetation type (Hiernaux and Justice, 1986; Jarlan *et al.*, 2008). Such empirical models, therefore, should be checked or fitted for each ecological zone (i.e., semi-arid, arid, etc.) before application (du Plessis, 1999). For this reason, other fitting functions have been tested in a variety of ecosystems and found to be better than the linear one: the Exponential function for the Malian Gourma (Hiernaux and Justice, 1986), the power function for the steppe ecosystems of China (Jin *et al.*, 2014) and the logarithmic and quadratic regression function for rangelands located in the northeast of the Tibetan plateau (Xiaoping *et al.*, 2011). For a given region, also, it is essential to test such a vegetation index before using it as vegetation productivity indicator particularly in an operational system (Santin-Janin *et al.*, 2009). This chapter aims: 1) to investigate five nonlinear fitting functions against the linear regression model used in West Africa Sahelian areas to date, and to show, if any, their limitation or suitability to better handle the relationship between the NDVI from the SPOT-VGT images and the ground data of total biomass (herbaceous + woody leaf); 2) to verify if the use of field sampling data from several successive years enables improvement in estimates accuracy.

2.2. Materials and methods

2.2.1. Monitoring sites and plant biomass data

As described in Chapter 1 (Section 1.1.2), the study area corresponds to the Senegal rangelands regularly frequented during the year by livestock belonging particularly to the transhumance. It covers all or part of eco-geographical areas that are the river valley, the pastoral zone (or Ferlo), the groundnut basin, the Niayes and the eastern Senegal. The choice of this area is justified firstly by the fact that it contains most of the natural pastures of the Senegalese pastoral area and, secondly, because it contains all the ground control sites of the CSE for which total biomass production data are

available (Figure 1.4). The 24 monitoring sites used in this study are well described in Chapter 1 (Section 1.2) and shown in Figure 1.4.

2.2.2. Seasonal NDVI integration

For this integration, the period of the plant growing season was determined visually by analyzing the dekadal NDVI images within the May-October period of each year. Then the growing season was considered to occur for between the first dekad of August and the third decade of October (3 months) due to the fact that significant NDVI values are approximately observed for the whole study area from the end of July. To compute the seasonal cumulated NDVI (i.e., iNDVI) from SPOT-VGT images, a weighted average of the nine dekadal images of the considered growing period was applied as the CSE during the last three decades. The basic equation of the integration procedure can be written as follows:

$$iNDVI = (\sum_{j=1}^{9} NDVI * X_j) / P \quad (9)$$

Where *iNDVI* corresponds to the seasonal cumulation of NDVI value; *NDVI* indicates the index value of the 10-day images; X_j indicates the number of days used to compute the 10-day images and *P* corresponds to the number of days into the integration period (i.e., growing period).

In the specific case of this study, the covered period by each SPOT-VGT S10 image was considered equal to ten since the number of daily images used in the synthesis process was unknown. So the retrieved iNDVI for a given pixel is a simple seasonal average of dekadal NDVI values within the august-october period.

2.2.3. Seasonal NDVI maximum

The plant growing period was determined for each pixel in the study area using the TIMESAT software (Eklundh and Jönsson, 2015). For this reason, the beginning and end of season were estimated to be respectively at 20 and 50% of the seasonal amplitude of the NDVI (Diouf *et al.*, 2015). As required by the tool to retrieve this metrics, the profile of the annual NDVI was

smoothed by the Savitzky-Golay method and the key parameters used are: seasonal parameter = 0.5 (to fit one season per year), number of envelope iterations = 2 (number of iterations for upper envelope adaptation), adaptation strength = 2 (strength of the envelope adaptation, maximum = 10), and window size = 4 dekads. The seasonal peak of NDVI (NDVIpk) is given by the maximum value of the index recorded between the beginning and the end of the growing season.

2.2.4. Statistical fitting functions

To model the vegetation mass in Senegalese rangelands, six fitting functions including the commonly applied linear regression one were used in this study. All functions are easily accessible in various statistical analysis tools. The fitting was made with total biomass data collected over a period of fifteen years between 1999 and 2013 (except for 2004). Table 2.1 shows the basic equations of the functions used.

Table 2.1 – Basic equations of fitting functions

Noms	Equations
Cubic	$Y = b0 + (b1 * X) + (b2 * X^2) + (b3 * X^3)$
Power	$Y = b0 * (X^{b1})$
Exponential	$Y = b0 * (e^{(b1 * X)})$
Linear	$Y = b0 + (b1 * X)$
Logarithmic	$Y = b0 + (b1 * \ln (X))$
Quadratic	$Y = b0 + (b1 * X) + (b2 * X^2)$

Y= plant biomass production (kg.DM/ha) and X= iNDVI or NDVIpk

2.2.5. Assessment of model's performance and coherence

According to Djitèye and Penning de Vries (1991), "the models are simple representations of systems, defining a system as a coherent part of the real world". So a model is only good when it reflects more or less faithfully the process for which it was calibrated. The NDVI is an indicator of the vegetation greenness and somehow of the plant biomass production. This theoretically means that the higher the index value, the higher the plant production is. The calibrated models (by regression between the total plant biomass data and the NDVI variables) have been tested on NDVI values ranged between 0.1 and 0.7 to check if they express the reality or not. This interval contains the typical NDVI values that reflect the presence of vegetation especially when using the SPOT-VGT satellite data (Jarlan *et al.*, 2008). Thereafter a model was considered to be consistent when estimates of total plant biomass are increasing or decreasing along with the NDVI with "reasonable" values of plant biomass (i.e., which can be justified by the historical field data).

The number of observations was generally small for a single year and varied between 14 and 24. This was due to the fact that plant biomass data were not regularly collected for all 24 monitoring sites because of the occasional lack of logistics or the early passage of bush fires before field campaigns. Because the coefficient of determination (R^2) of the calibration is not an adequate measure to compare nonlinear models (Spiess and Neumeyer, 2010), the predictive capability of the yearly models was assessed in this study instead by their goodness-of-fit to the total biomass data used. The models validation was achieved using the bootstrap technique (Efron, 2004). Then, the criteria used to measure the predictive ability of the models were the R^2 between observed and predicted total biomass (i.e., after bootstrap) and the Root Mean Square Error (RMSE), which was calculated using the following formula:

$$RMSE = \sqrt{\frac{\sum(O_i - P_i')^2}{N}} \qquad (10)$$

Where O_i = observed total biomass, P_i= predicted total biomass by models and N = total samples in the validation set.

Multiple year data-based models were compared to the single year data-based models of the same function. The Wilcoxon signed-rank test was computed at the 95% confidence interval and applied to appreciate if there is difference (or not) in accuracy of the total plant biomass estimates of these models. This comparison was done in order to check if the use of successive years data enables improving models from one year of samples.

2.3. Results

2.3.1. Statistical performance of fitted functions

The coefficient of determination (R^2) and the validation root mean squared error (RMSE) of the fitted models varied strongly from a year to the next for both iNDVI and NDVIpk variables (Figure 2.1). The Cubic and Quadratic functions showed the higher values of R^2 (Figure 2.1a, b) and consequently the lower values of RMSE (Figure 2.1c, d) as shown by the averaged values of these statistics between 1999 and 2013 in Figure 2.2. The lower R^2 and the higher RMSE were provided by the Linear and Logarithmic models which constitute the less accurate models. Meanwhile, the Exponential and Power fitted functions showed intermediary performance in comparison to the two previous groups of models. The Exponential models, however, showed substantially better performance than the Power models. The iNDVI was more suited than the NDVIpk for estimating the total biomass in Sahelian Senegalese rangelands. Considering the models one by one, indeed, the statistical performances were better when using iNDVI than NDVIpk (see Figure 2.2).

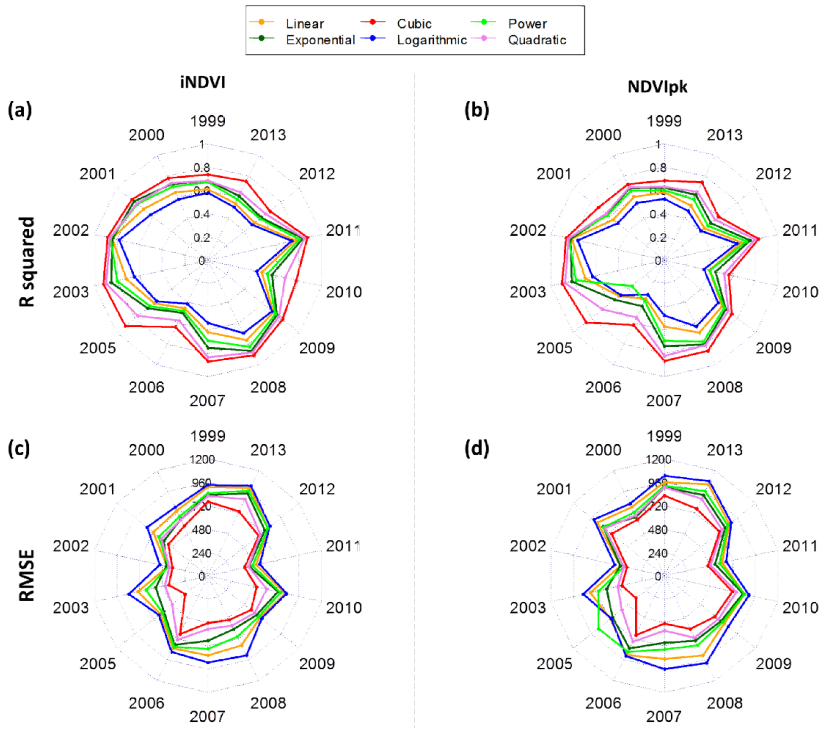

Figure 2.1 – Statistical performance of the six models using the iNDVI and NDVIpk predictors: (a, b) the coefficient of determination (R^2) between observed and predicted total biomass and (c, d) the root mean squared error of the validation (RMSE) from 1999 to 2013 (no data collection in 2004).

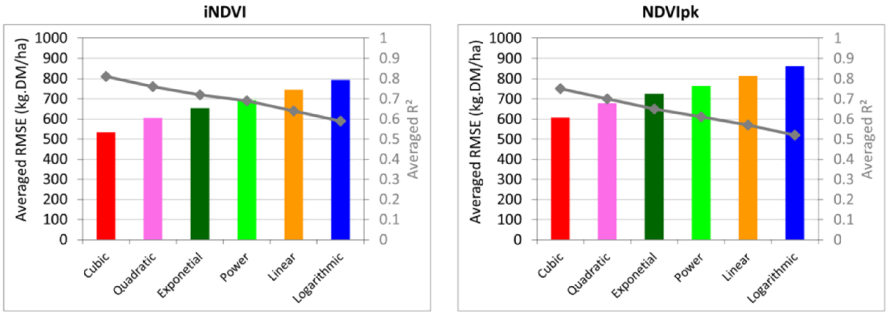

Figure 2.2 – Averaged statistical performance of the six models using the iNDVI and NDVIpk predictors for the period 1999-2013.

2.3.2. Biophysical consistency of model estimations

Simulations of the six models in the range of NDVI values between 0.1 and 0.7 showed similar profiles of total plant biomass for both iNDVI and NDVIpk variables (Figure 2.3). The Exponential and Power models presented the most consistent profiles looking to our basic theory where the predicted value of total biomass must increase if the NDVI increase or vice versa. Simulations by the Quadratic indeed, did not always progress in the same direction than the NDVI values. The Cubic fitted function has given a sinusoidal profile while the Quadratic model showed a parabolic curve. This means that these latter models could predict decreasing total biomass while the NDVI values are increasing. This demonstrates their inconsistency to the reality as well. The Logarithmic and Linear models provided null estimations of the total biomass when the NDVI values were generally about 0.25 for the Linear function and 0.3 for the Logarithmic one. With NDVI values below these limits, the two models predicted negative values of total biomass which is also contrary to reality. Furthermore, for the same value of NDVI, the total biomass predictions were generally lower with NDVIpk than iNDVI for all models.

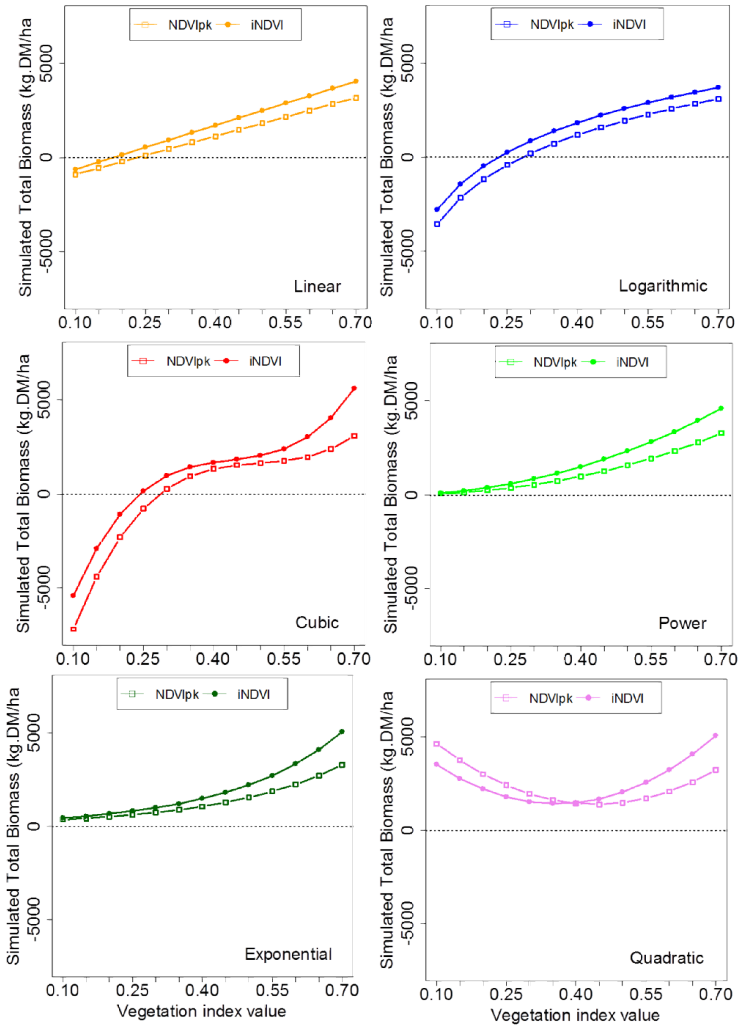

Figure 2.3 – Estimates of total plant biomass within the 0.1-0.7 range of iNDVI and NDVIpk, averaged from 1999 to 2013.

2.3.3. Multiple year data-based estimates of total biomass

Only Exponential and Power models were considered in this analysis, since they are the only ones among the six models that have been proved to be statistically significant and to enable consistent estimation of total plant biomass for all NDVI values within the theoretical range (0.1-0.7) corresponding to vegetation presence. The Linear model was included for

comparison purpose. Calibrated with multiple year data, the Exponential model was substantially the most accurate (higher R^2 and lower RMSE), followed by the Power fitted function and the Linear model (lower R^2 and higher RMSE) for both iNDVI and NDVIpk variables (Table 2.2). This performance order confirms the results shown for the same fitted functions with single year data of total biomass (Section 2.3.1). Here again, the models run with NDVIpk (e.g., $R^2 = 0.58$ and RRMSE = 38% for the Exponential model) were generally less accurate than those established with iNDVI (e.g., $R^2 = 0.63$ and RRMSE = 35% for the same model). The Wilcoxon signed rank test (at 0.05 p-level) showed that there is no significant difference between the means of total biomass estimated with the single year data- and the multiple year data-based models within the studied period 1999-2013 (Figure 2.4). For the three models, the p-value of the V-test was higher than 0.05, indicating that the null hypothesis (i.e., the means of total biomass estimations are not different) was accepted.

Table 2.2 – Statistical performance of the Exponential, Power and Linear models calibrated with field sampling data of the period 1999-2013. The relative RMSE (RRMSE) in percentage was calculated using the averaged total biomass from all samples of the bootstrap dataset.

	Model	Equation	R^2	RMSE (kg.DM/ha)	RRMSE (%)
iNDVI	**Exponential**	370.91 * e^ (3.57 * iNDVI)	0.64	806.98	34.6
	Power	8022.57 * (iNDVI^1.78)	0.62	831.36	35.7
	Linear	-1234.75 + (iNDVI * 7407.31)	0.59	863.14	36.9
NDVIpk	**Exponential**	277.79 * e^ (3.49 * NDVIpk)	0.58	868.12	37.2
	Power	6642.51 * (NDVIpk^ 2.03)	0.56	891.90	38.3
	Linear	-1455.46 + (NDVIpk * 6596.07)	0.52	930.64	39.82

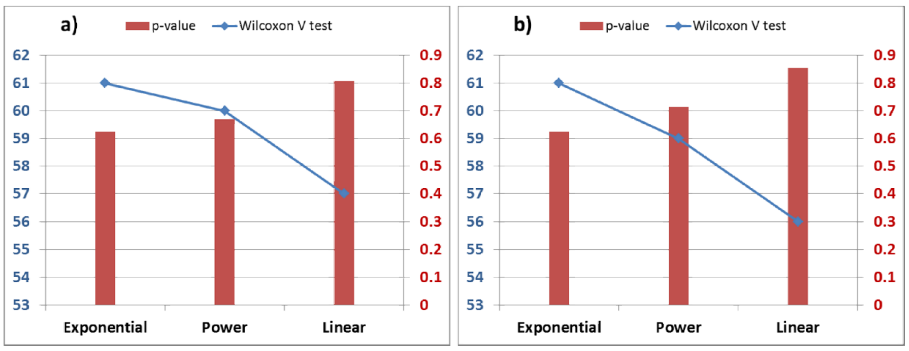

Figure 2.4 – Statistical difference of means of total biomass estimates made by Exponential, Power and Linear models using the Wilcoxon signed rank test at 0.05 p-level.

2.4. Discussion

The assessment of fodder biomass is currently done in Senegal and Niger from the simple linear regression between ground biomass data and seasonal integration of NDVI (iNDVI) and seasonal maximum of NDVI (NDVIpk) respectively. Taking into account the saturation effect of NDVI in areas with high plant biomass productivity mentioned by many authors (Santin-Janin *et al.*, 2009; Bégué *et al.*, 2011; Mbow *et al.*, 2013; Tian *et al.*, 2015), one should expect a nonlinear relationship between NDVI values and total biomass data, instead. However for a given region, the most suited fitting function is not known in advance since the empirical relationship between plant biomass and NDVI depends on the local conditions in which it was developed (Hiernaux and Justice, 1986; Jarlan *et al.*, 2008). For these reasons, five simple nonlinear functions (Exponential, Power, Logarithmic, Quadratic, and Cubic) were tested against the usual Linear model to determine the most accurate and consistent to estimate total biomass from NDVI variables in Sahelian rangelands of Senegal.

The results showed a high inter-annual variability of the relationships between NDVI and plant total biomass. As supported by Prince (1991), this variability materialized by inter-annual variation of R^2 and RMSE values is characteristic to the Senegalese Sahelian areas. Among the six models, the Logarithmic model was the lesser accurate followed by the Linear function while the most accurate were the Cubic and Quadratic fitted functions. The

Exponential and Power models showed intermediary accuracies which were, however, generally higher when using the iNDVI than NDVIpk variable. These two latter functions provided the most consistent behavior of total biomass estimates which increased, as expected, in the same way as the NDVI values. Meanwhile, the Cubic and Quadratic models even statistically good did not simulate consistently the total biomass within the given interval of vegetation presence (0.1 – 0.7). Particularly, the Quadratic fitted function showed a parabolic profile of estimated total biomass which means that this model could provide decreasing total biomass even if the NDVI values are increasing. This indicates that these two models are not suited for the Senegalese Sahelian rangelands at the risk to make erroneous estimations of total biomass at the end of the growing season. In addition to be the less accurate among the six studied functions, the Logarithmic and Linear models did not work very well with the low NDVI values. They provided null estimation of the total biomass with NDVI values about 0.3 even those do not necessarily mean absence of vegetation in Sahel areas. Below this value, the Logarithmic and Linear fitted functions predicted negative values of total biomass. These malfunctioning could be a limitation for the two models to be applied in Sahelian areas with low vegetation cover and in extreme years characterized by an important deficit of total biomass as in 2002 (Figure 2.4). In the year 2002, for example, pixels with negative NDVI values were masked out since they could impede the use of the corresponding images to establish a fodder balance (i.e. ratio between fodder production and livestock size into a delineated area). Particularly for the Logarithmic model, the exclusion of the negative estimates has led to an important loss of information in the northern part of the study area. For the Exponential and Power models, however, the distribution of the total biomass was spatially complete even if the Exponential fitted function provided higher predictions for a given NDVI variable (e.g., predictions above 6,000 kg.DM/ha to the south of the study area for iNDVI).

From a multiple year data basis, the Exponential model was the most accurate followed by the Power function that was in turn more accurate than the Linear model as in the case of single year data calibration. This order of models performance was observed both for iNDVI and NDVIpk. No

significant difference was observed between means of total biomass estimated by models calibrated using single year data and multiple year data for the studied period 1999-2013. This means that there is statistically no improvement in calibrating the models with multiple year data. Models based on single year biophysical information were good enough to provide total biomass estimates at the end of the season. However, the advantage of using models based on several successive years of field sampling data remains in the fact that they can effectively enhance the stability of the estimates (Jin *et al.*, 2014) and in their applicability on a given year without collecting the corresponding field biomass data. Multiple years models can be very useful when there is problem to conduct field data collection (e.g., lack of logistics, early passage of bushfire in sites…). The iNDVI was shown, in general, to be more suited than the NDVIpk for estimating the total biomass in Senegalese Sahelian rangelands. This contrasts with results obtained by Mougenot *et al.* (2000), where the NDVIpk was found to be the more suited for such Sahelian areas.

Furthermore, in addition to their monocriterion character (they include only part of the reality with a single explanatory variable), the simple regression approaches apply the same equation for all pixels of the studied area. This constitutes another drawback of these methods with the well-known high spatial variability of ecological conditions (i.e., soil type, rainfall, woody cover, and species patterns) across the Sahel. For this reason, different models which include multiple variables better describing the ecological characteristics (e.g., phenological metrics from satellite datasets) of different Sahelian ecoregions are needed to improve this approach.

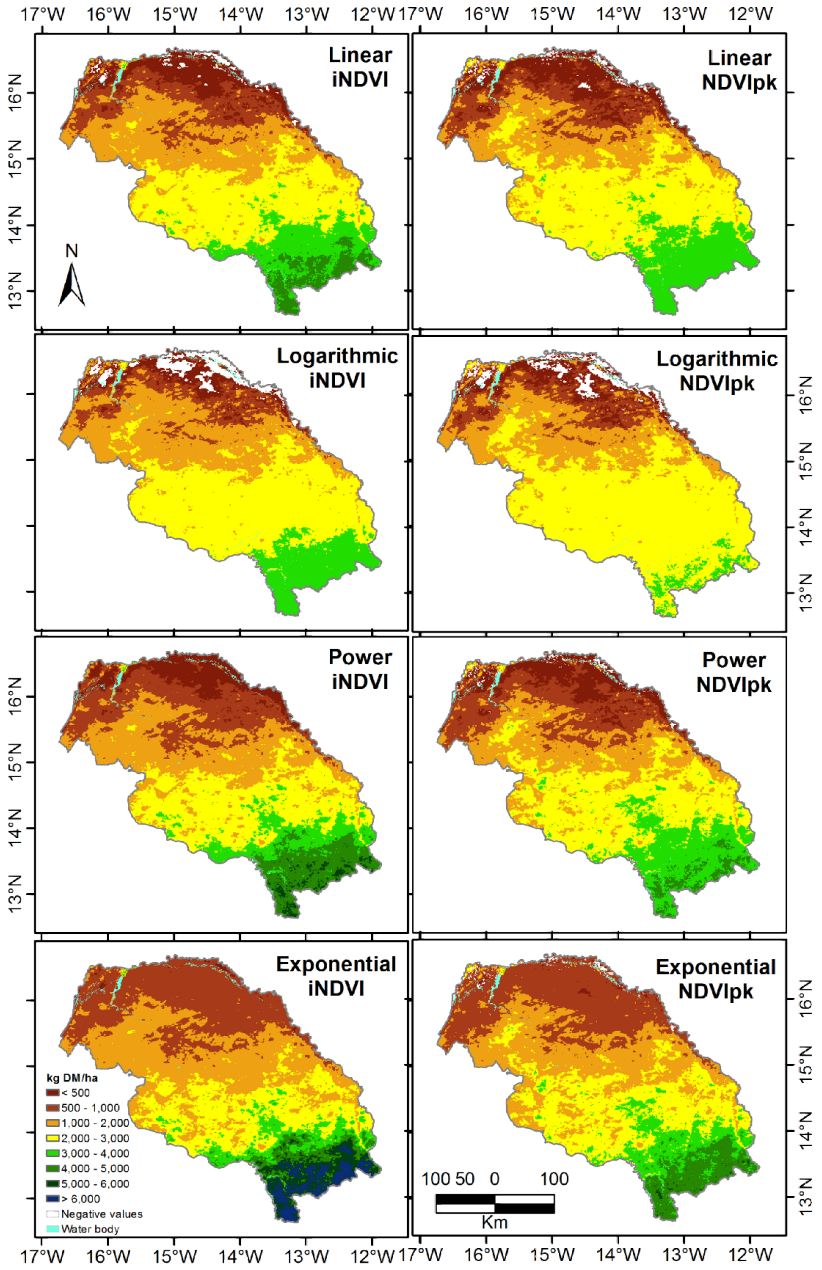

Figure 2.5 – Example of total biomass predictions for the year 2002 with iNDVI and NDVIpk. Negative predictions were masked out and correspond to the white pixels within the study area.

2.5. Conclusion

Instead of the Linear fitting function, the Exponential and Power models can be applied to fit the field sampling data with the seasonal NDVI data for estimating accurately and consistently the end of season total biomass in Senegalese Sahel rangelands. The Quadratic, Cubic and Logarithmic models show inconsistency and malfunctioning for estimating the total biomass and, therefore are not recommended for this purpose in the studied area. Unlike the seasonal maximum, the more suited variable is the seasonal integration of NDVI values within the August-October period. No significant difference exists between the means of total biomass estimated with models using single year data and those using several successive years of field sampling data. The multiple year data-based models, however, offer the advantage of more stable coefficients and their usability for predicting the total plant biomass on unseen years. To improve the 1980's estimating method, a new approach including specific models for the different Sahelian ecoregions has required and investigated in the next chapter.

3 Total biomass estimation using multiple-linear regression models and phenological metrics from FAPAR time Series

This Chapter was adapted from the following publication:

Diouf, A., Brandt, M., Verger, A., Jarroudi, M., Djaby, B., Fensholt, R., Ndione, J., & Tychon, B. (2015). Fodder Biomass Monitoring in Sahelian Rangelands Using Phenological Metrics from FAPAR Time Series. Remote Sensing, 7, 9122-9148. Doi:10.3390/rs70709122.

3.1. Introduction

Livestock farming is the most widespread human activity and most dominant form of land use in rangeland ecosystems (Alkemade *et al.*, 2013). Worldwide, it contributes 40% of the agricultural gross domestic product, and provides income for more than 1.3 billion people and nourishment for at least 800 million food-insecure people (Herrero and Thornton, 2013). In the West African Sahel in particular, livestock is the primary renewable resource (Dicko *et al.*, 2006) and the region is characterized by an extensive breeding through rangelands. Rainfall here is considered to be the main driving factor responsible for fluctuations in natural vegetation (Hickler *et al.*, 2005; Huber and Fensholt, 2011; Anyamba *et al.*, 2014) and, therefore, fluctuations in grazing stock, for which growth is limited to 3–4 months of the rainy season (between July and October), are closely linked to rainfall. Because of unpredictable intra-annual variations in rainfall, the vegetation is often exposed to a water shortage in the growing season, sometimes leading to severe droughts, food shortages, and production deficits (Fensholt *et al.*, 2004). Given the random and recurring environmental stress, estimating plant biomass in the rangelands is important in assessments of livestock fodder availability (Mbow *et al.*, 2014). Depending on the needs of the

agricultural monitoring systems, however, estimates should be provided as early as possible in the growing season so that stakeholders can take early decisions and identify areas with large variation in (and potential for) vegetation productivity (Atzberger, 2013).

The estimation of biomass production using remote sensing data has been the subject of many studies in the Sahel (Tucker *et al.*, 1983; Tucker *et al.*, 1985; Diallo *et al.*, 1991; Prince, 1991; Rasmussen, 1992; Mougenot *et al.*, 2000; Diouf and Lambin, 2001). The Normalized Difference Vegetation Index (NDVI) is the most common satellite index used in the region for the temporal monitoring of vegetation (Bénié *et al.*, 2005). Another widely used indicator is Fraction of Absorbed Photosynthetically Active Radiation (FAPAR), recognized as a key variable in the assessment of vegetation productivity (Prince, 1991; Meroni *et al.*, 2013). FAPAR is defined as the fraction of radiation absorbed by the green vegetation elements in the 400–700 nm spectral domains under specified illumination conditions (Baret *et al.*, 2013). It is directly linked to photosynthesis and therefore expresses a canopy's energy-absorption capacity (Fensholt *et al.*, 2006). Many authors have studied the relationship between FAPAR and NDVI (Goward and Huemmrich, 1992; Bégué, 1993; Hanan *et al.*, 1995; Myneni *et al.*, 1995; Le Roux *et al.*, 1997; Lind and Fensholt, 1999; Fensholt *et al.*, 2004), which has been shown to be generally linear for green vegetation, particularly in the semi-arid environment of the Sahel (Fensholt *et al.*, 2006).

The *Centre de Suivi Ecologique* (CSE) has been estimating total annual biomass in Senegal on an operational basis since 1987 in order to monitor fodder availability in the national pastoral rangelands. The method used is inspired by an approach proposed in 1983 by (Tucker *et al.*, 1983) and is based on an empirical relationship between the satellite seasonal integrated NDVI and in situ measured biomass at the end of the growing season. The same approach is used by Niger's Ministry of Livestock and Animal Industries to assess forage availability in the rangelands at the end of the season. Although very useful for the spatial management of pastoral resources, this approach based on a simple linear regression has several limitations, including: (i) its temporal scale is restricted to biomass data of the ongoing year not being used again in the following year, which leads to

high costs for annual data collection; (ii) its low predictive potential is due to the large time gap between data collection (mid-October) and published results (late December); and (iii) this is a single-predictor model, which generally omits details of reality (Neville, 2013).

The satellite-derived seasonal metrics of plant phenology and seasonal dynamics in reflectance and greenness (Hanes *et al.*, 2013) could be very useful for establishing multiple-predictor models for estimating plant biomass production. Many authors have endorsed the applicability of these phenological metrics for crop (Atzberger, 2013) and grassland (Colombo *et al.*, 2011) phenological monitoring, particularly for analyzing African farming systems (Vrieling *et al.*, 2011). Recently, phenological metrics based on the FAPAR time series were recommended by (Meroni *et al.*, 2014b) for providing relevant early information on biomass production in the Sahel. When associated with ground biomass data, such variables would enable early warning models to be established and would improve the use of early warning systems (EWS) in planning emergency responses and food aid interventions in the case of a forthcoming crisis (Meroni *et al.*, 2014a).

A significant relationship ($p < 0.001$) between the field measurements of herbaceous biomass and the cumulated FAPAR over the growing season was found by Meroni *et al.* (2014b) in Senegal, although only 34% of the in situ biomass was explained by the linear regression. The study focused only on herbaceous biomass, however, and recently (Brandt *et al.*, 2015) highlighted the importance of woody plants in the satellite time series of vegetation indices in the Sahelian ecosystems.

In this context, the present study sought to develop an operational system for monitoring total biomass, including both herbaceous and woody leaf biomass, rather than focusing exclusively on using remote sensing data to forecast vegetation productivity (Funk and Brown, 2006; Meroni *et al.*, 2014a). The proposed method is based on multiple linear regression models using phenological variables derived from the seasonal dynamics of the FAPAR SPOT-VGT time series and ground measurements of total biomass production collected in different Sahelian ecosystems in Senegal over 15 years. The specific objectives of the study were to: (i) determine the proper metrics and establish the best model for total biomass estimation across the

study area; (ii) verify whether or not the disaggregation of the overall dataset into ecoregions and related metrics improved the estimates; and (iii) validate the multiple-predictor models developed, including an early warning model, based on the comparison with ground measurements as well as with the CSE biomass product that is currently used for fodder-monitoring in Senegal.

3.2. Materials and methods

3.2.1. Study area and limit of studied ecoregions

The study area is part of the one described in Chapter 1 (see Section 1.1.2). It covers the northern and eastern parts of Senegal between 12.9°N and 16.8°N latitude and 12.3°W and 16.5°W longitude (an area of about 98,556 km²; Figure 3.1) and includes all pastoral areas of the country. The mean annual rainfall varies between 250 and 750 mm from north to south, based on the average rainfall estimates (RFE) time series (1999–2013) of the Famine Early Warning Systems Network (FEWS Net) with 8 km spatial resolution (Herman *et al.*, 1997). The terrain is relatively flat, with a maximum altitude of 163 m in the southeastern part of the study area.

In this study, the 24 ground control sites were grouped into four ecological zones (ecoregions): Northern Sandy Pastoral Region (ECOnorth), Ferruginous Pastoral Region (ECOeast), Mixed Pastoral-Agricultural Region (ECOwest), and Eastern Transition Region (ECOsouth). The ECOwest area corresponds to a combination of the Southern Sandy Pastoral Region (north) and Agricultural Expansion Region (south) of the initial classification proposed by Tappan *et al.* (2004). Detailed information on the ecoregions used is provided in Table 3.1 (see also (Tappan *et al.*, 2004; Brandt *et al.*, 2015)).

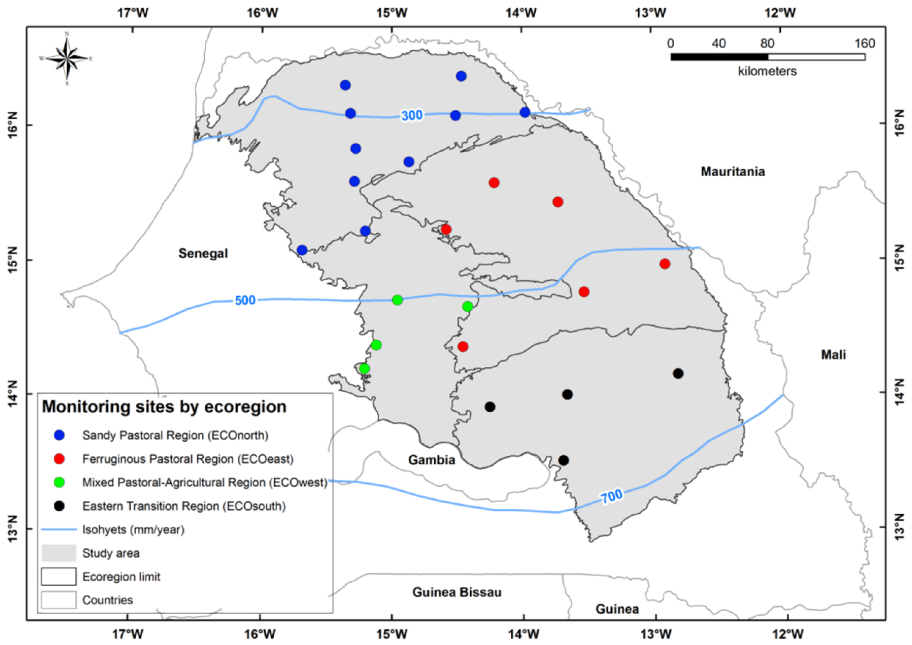

Figure 3.1 – Location of monitoring sites in the study area covering a range of Sahelian ecosystems in Senegal. The isohyets are based on average rainfall estimates provided by FEWS Net between 1999 and 2013.

Table 3.1 – Description of ecoregions: vegetation type and main woody species, annual rainfall (based on average values of RFE of FEWS Net data for the period 1999–2013), biomass, and woody cover based on ground measurements (1999–2013). The values for rainfall, woody leaf biomass, herbaceous biomass, and woody cover correspond to the average from sites in each ecoregion.

Ecoregion	Main Vegetation Type and Woody Species	Annual Rainfall (mm)	Woody Leaf Biomass (kg·DM/ha)	Herbaceous Biomass (kg·DM/ha)	Woody Cover (%)
ECOnorth (Sandy Pastoral)	Pseudo-steppe: *Boscia Senegalensis*, *Balanites aegyptiaca*, *Guiera Senegalensis*, *Calotropis procera*, *Combretum glutinosum*, *Sclerocarya birrea*	345.0	490	905	5.5
ECOeast (Ferruginous Pastoral)	Shrub savannah: *G. Senegalensis*, *C. glutinosum*, *Pterocarpus lucens*, *Grewia bicolor*, *B. Senegalensis*, *Adenium obesum*	488.0	1219	1257	17.7
ECOwest (Pastoral-Agricultural)	Shrub/tree savannah: *C. glutinosum*, *G. Senegalensis*, *G. bicolor*, *B. Senegalensis*, *C. micranthum*, *Commiphora Africana*	524.0	1204	1867	16.9
ECOsouth (Eastern Transition)	Tree savannah/woodland: *C. glutinosum*, *Strychnos spinosa*, *Acacia macrostachya*, *Crossopteryx febrifuga*, *Terminalia avicennioides*, *Maytenus Senegalensis*	633.0	2611	1937	33.1

3.2.2. Ground Biomass Data

The *in situ* biomass data used in this Chapter were obtained from the CSE database and covered the 1999–2013 period, apart from 2004 when no data was collected. More details on the collection of these data are provided in Chapter 1 (see Section 1.2).

3.2.3. Satellite Data

3.2.3.1. CSE biomass product

The total biomass estimated by the CSE based on a NDVI single-predictor model was used here for comparison purposes. The CSE biomass estimates are derived annually at the end of the growing season by a simple linear regression between the integrated NDVI over the growing season and ground biomass measurements for the corresponding year. The growing season is determined by visual analysis of dekadal (10-day) images throughout the May–October period. The start of the season corresponds roughly to the dekadal in May that displays generalized plant growth for the whole study area and the end of the season is fixed at the third dekade of October. Between 1987 and 1999, this method was implemented using the seasonal integrated NDVI (i.e., seasonal weighted average) from the Advanced Very High Resolution Radiometer (AVHRR) of the National Oceanic and Atmospheric Administration (NOAA) satellites acquired in Local Area Coverage (LAC) format at the CSE receiving station in Dakar. Since 2000, the 1-km NDVI S10 products derived from atmospherically corrected SPOT-VEGETATION reflectances based on the maximum composition value (MVC) algorithm at 10-day intervals have been used (VITO, 2005). Here, we used only the CSE biomass estimates from 2000, which are based on the same SPOT-VEGETATION satellite sensor as our biomass estimates. It has to be noted that the aim of the research presented here was not a direct one-to-one comparison between models, since different datasets were used for their calibration. On the contrary, the comparison is done at the level of biomass output estimates and the CSE biomass product represents the reference product used operationally for fodder-monitoring in Senegal.

3.2.3.2. Phenological metrics from FAPAR time series

We used the 1999–2013 time series of the GEOV1 Copernicus Global Land FAPAR product derived from the SPOT-VEGETATION instrument. The products are freely available at a 1-km resolution and 10-day intervals on http://land.copernicus.eu/global. The principles for generating this FAPAR dataset are based on the use of neural networks that were first trained with CYCLOPES reflectances and existing MODIS and CYCLOPES FAPAR products (Verger *et al.*, 2008). In order to take benefit from their complementarities, MODIS and CYCLOPES products were fused by assigning higher weights to the MODIS (CYCLOPES) product for the high (low) FAPAR values, with weight being defined by previous validation studies (Weiss *et al.*, 2007 ; McCallum *et al.*, 2010). The trained neural network algorithm was then applied using the directionally normalized top of canopy reflectance in the red, near infrared, and shortwave infrared bands and the cosine of the sun zenith angle at the time of satellite overpass (i.e., about 10:30 for VEGETATION). For details on the algorithm used to estimate GEOV1 products, we refer to Baret *et al.* (2013). Recent validation studies indicated that GEOV1 outperformed MODIS, CYCLOPES, and JRC-SeaWIFS FAPAR products in both accuracy and precision (Atkinson *et al.*, 2012).

Phenological metrics were derived from the GEOV1 FAPAR time series. Although compositing techniques were used for synthesizing dekadal GEOV1 products, these could have been affected by artifacts related to the presence of clouds (Fensholt *et al.*, 2007) and residual atmospheric effects (Chen *et al.*, 2004). Prior to using them for phenology detection, therefore, filtering the noise was essential (Atkinson *et al.*, 2012). This is particularly important in the Sahel zone where the vegetation has a rapid phenological cycle associated to the short rainy season when most of the optical satellite observations are missing or affected by noise (Vintrou *et al.*, 2014). GEOV1 FAPAR time series were filtered using the Savitzky-Golay (SG) fitting method available via TIMESAT software (Jönsson and Eklundh, 2004). SG is a simplified least squares fit convolution that can be understood as a weighted moving average filter with weighting provided as a polynomial of

a given degree (Chen *et al.*, 2004). This filtering method was demonstrated to improve other existing temporal filters for reconstructing satellite time series in terms of the accuracy as compared to the original data by ensuring robustness to noise and missing data, while preventing over-smoothing (Kandasamy *et al.*, 2013). The accuracy of the estimation of the timing of phenological stages for the start, maximum, and end of the season as derived from the Savitzky-Golay reconstructed time series is about 10 days for the sites with a percentage of missing data <10% (which is mostly the case in the study area, except in some regions mostly located in the south close to the Gambia River; Appendix Figure 5.1-A1 and Figure 5.2-A2) but shows a rapid decrease with the percentage of missing data and the length of the period with missing data (Verger *et al.*, 2011; Kandasamy *et al.*, 2013; Verger *et al.*, 2013b). The key parameters used in the TIMESAT SG filter are: seasonal parameter = 0.5 (to fit one season per year), number of envelope iterations = 2 (number of iterations for upper envelope adaptation), adaptation strength = 2 (strength of the envelope adaptation, maximum = 10), and window size = 4 (half-window for SG filtering, also defining the degree of smoothing) (Eklundh and Jönsson, 2011). The beginning and end of the growing season were estimated at 20% and 50% of the FAPAR seasonal amplitude. These latter values were chosen after a qualitative analysis was conducted of the growth profile for different pixels throughout the study area. The result was a smoothed curve fitted to the upper envelope of the FAPAR values of the time series. From the smoothed profiles, 11 phenological metrics were calculated (Table 3.2 and Figure 3.2). All metrics were computed at the pixel scale (1-km spatial resolution) for each year between 1999 and 2013. The annual values were then averaged for each site over a 3 × 3 pixel window to match, as far as possible, the spatial sampling of ground data.

Table 3.2 – Phenological metrics derived from the FAPAR time series (extracted using TIMESAT software).

No.	Variables	Abbreviation	Short Definition
1	Start of season	SOS	Time when the left edge has increased to 20% of the amplitude
2	End of season	EOS	Time when the right edge has decreased to 50% of the amplitude
3	Length of season	LOS	Time from the SOS to the EOS
4	Middle of season	PMID	Computed as the mean value of the times when the signal is higher than 80% of the amplitude
5	Base value	BVAL	Averaged left and right minimum values over the annual cycle
6	Maximum value	PEAK	Highest FAPAR value over the season
7	Amplitude	AMPL	Difference between the maximum and BVAL
8	Large integral	LINTG	Integral of the signal from the SOS to the EOS
9	Small integral	SINTG	Integral of the signal above the BVAL from the SOS to the EOS
10	Left derivative	LDERIV	Rate of FAPAR increase at the SOS, between the left 20% and 80% of the amplitude
11	Right derivative	RDERIV	Rate of FAPAR decrease at the EOS, between the right 20% and 80% of the amplitude

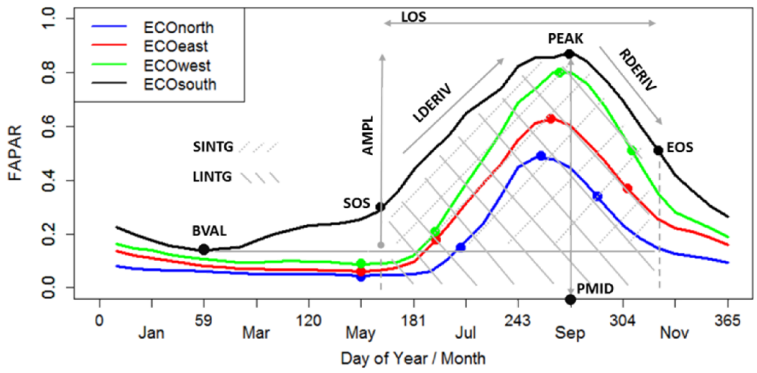

Figure 3.2 – Mean annual GEOV1 FAPAR time series for the four ecoregions in the study area. Phenological metrics are represented on the ECOsouth curve. For acronyms see Figure 1.4 and Table 2.3.

3.2.4. Modelling of the total plant biomass production

3.2.4.1. Reduction of explanatory variables and model development

In order to understand the relationship between phenological metrics and total biomass across the Senegalese rangelands, a Pearson correlation analysis was performed on the whole dataset (n = 263). Only those variables that were significantly correlated to the total biomass at the 99% significance level were considered for the study. The significant phenological metrics were sorted according to their importance in predicting total biomass using the Partial Least Squares (PLS) regression method (Wold, 2001). This technique is generally used to build predictive models when the number of explanatory variables is large or presents multi-collinearity. The usefulness of such a method to determine the most important predictor variable in predicting a response variable was discussed by (Afanador *et al.*, 2013). The ranking was performed using the statistical criterion Variable Importance in the Projection (VIP) (Mehmood *et al.*, 2012). Only the most important phenological metrics with a VIP $\geq$ 0.8 (Desbois, 1999) related to total biomass estimation were selected so as to reduce the number of variables for model development and therefore avoid the computational problems.

In the development of statistical models, single selection parameters could be biased and only a combination of different unrelated parameters should be used to assess model performance. Several performance criteria for measuring the quality of the models have been proposed in the literature based on minimizing a penalty parameter. In this study, we used the Akaike Information Criterion (AIC) and the adjusted coefficient of determination (Adj. R^2), which constitutes a good indicator of model robustness (Johnson *et al.*, 2006). The Adj. R^2 is a corrected value of the R^2, taking account of the influence related to the number of predictor variables in a given model (Kouadio *et al.*, 2012) and also considering the sample size.

In order to detect the influence of observations on the model's adjustment and to verify the homoscedasticity and normality of residuals, we analyzed the Cook's Distance, the Studentized Residuals, and the QQ-Plot (Confais and Le Guen, 2006). The Variance Inflation Factor (VIF) indicator was also

used for detecting multi-collinearity between the variables in the models. Only those models in which all included variables had a VIF below 10 were selected (Belsley *et al.*, 1980). Table 3.3 presents the main criteria and indicators used for model selection.

Table 3.3 – Criteria for the variable reduction, model selection, and validation.

Criterion	Annotation	Formula	Decision Rule
Ajusted coefficient of determination	Adj. R^2	$1-(1-R^2)\left(\frac{n-1}{n-p}\right)$	Performance increases with |Adj. R^2|
Akaike Information Criterion	AIC	$n.\log\left(\frac{SSE}{n}\right)+2(p+1)$	Performance increases with lower AIC
Variation Inflation Factor	VIF	$1/(1-R^2j)$	VIF > 10 points to co-linearity
Variable Importance in the projection	VIP	$\sqrt{p\sum_{a=1}^{A}[SS_a(w_a/\|\|w_a\|\|)^2]/\sum_{a=1}^{A}(SS_a)}$	Variable is important if VIP ≥ 0.8
Mean Absolute Error	MAE	$\frac{\sum_{i=1}^{n}\|O_i-P_i\|}{n}$	A low MAE shows higher reliability
Normalized Mean Absolute Error	NMAE	$\frac{MAE}{Bm}\times 100$	A low NMAE shows higher reliability

Where n is the number of observations; p is the number of predictor variables; R^2j is the correlation coefficient for regression of Xj with the $(p-1)$ other variables; SSE is the sum of squared errors; Oi is the observed biomass in the field, Pi is the predicted biomass by model; Bm is the mean observed biomass; A is the number of relevant components for prediction; wa is the loading weight; SSa is the sum of squares explained by the a^{th} component; $(wa/\|wa\|)^2$ is the importance of the corresponding variable (Mehmood *et al.*, 2012).

3.2.4.2. Bootstrap resampling and model verification

The small sample size of the ground dataset does not allow a reliable assessment of the bias and variance report of the evaluated models and extraction of a validation sample of adequate size simultaneously. In order to overcome this, we used the bootstrap validation technique (Efron, 2004) to

assess the predictive ability of the models (Borra and Di Ciaccio, 2010). This technique involves using multiple randomized subsets of the observations in the original sample (Afanador *et al.*, 2013). It provides random samples with the same size as the original sample, each of which is referred to as the bootstrap sample (Acquah, 2012). The number of bootstrap samples was set at 200, with a sample rate of 0.75. Statistical calculations were performed using these bootstrap samples, for which the standard deviation of the re-sampled statistics was the empirical standard error of the statistics generated by the original sample (Cole, 1999).

The parameters used to assess the predictive ability of the final models were the Mean Absolute Error (MAE) and the Normalized Mean Absolute Error (NMAE) (Table 3.3). The MAE was chosen for the model verification because it provides an unambiguous measure of the magnitude of the average error and is therefore more appropriate than the Root Mean Square Error (RMSE) for dimensioned evaluations of average model-performance error (Willmott and Matsuura, 2005). Finally, only models that included, at most, three variables with the highest Adj. R^2, while minimizing the AIC, MAE, and NMAE criteria, were selected.

3.3. Results

3.3.1. Relationship between total biomass and phenological variables

The results of the Pearson correlation analysis (Table 3.4) showed that the phenological variables extracted were all significantly correlated to the total biomass ($p < 0.0001$), except for the increasing rate during the green-up stage (LDERIV). The large FAPAR integral (LINTG) had the highest correlation value, while the lowest correlation was produced by LDERIV. Therefore, only LDERIV was removed, and all the remaining phenological variables were used to assess total biomass across the Senegalese rangelands.

Table 3.4 – Mean values, standard deviation (SD), and Pearson correlation statistics of the phenological variables with total biomass collected in the 1999-2013 period (n = 263). For acronyms, see Table 3.2.

Variable	Simple Statistics		Pearson Correlation Statistics	
	Mean	**SD**	**R**	***p*-value**
LINTG	6.09	2.88	0.79	<0.0001
PEAK	0.65	0.19	0.77	<0.0001
SINTG	5.25	2.34	0.76	<0.0001
AMPL	0.59	0.17	0.72	<0.0001
LOS	109 (days)	30	0.63	<0.0001
BVAL	0.06	0.05	0.59	<0.0001
SOS	196 (days)	20	−0.52	<0.0001
RDERIV	14.03	6.10	0.45	<0.0001
EOS	304 (days)	21	0.38	<0.0001
PMID	258 (days)	14	0.23	0.0002
LDERIV	20.85	6.21	−0.11	0.0825

3.3.2. Importance of the explanatory variables in total biomass prediction

Key variables were identified by VIP in order to further reduce the number of variables for model development. At the scale of the whole study area, the results confirmed that LINTG was the most important phenological metric, whereas at the ecoregion scale, the maximum FAPAR (PEAK) was the most important (Figure 3.3). Variables of minor importance (VIP < 0.8) were removed and the remaining variables for model selection differed among the ecoregions. For example, SOS was the least important variable

for ECOnorth and ECOwest, whereas in ECOeast and ECOsouth, the least important variable was middle of season (PMID).

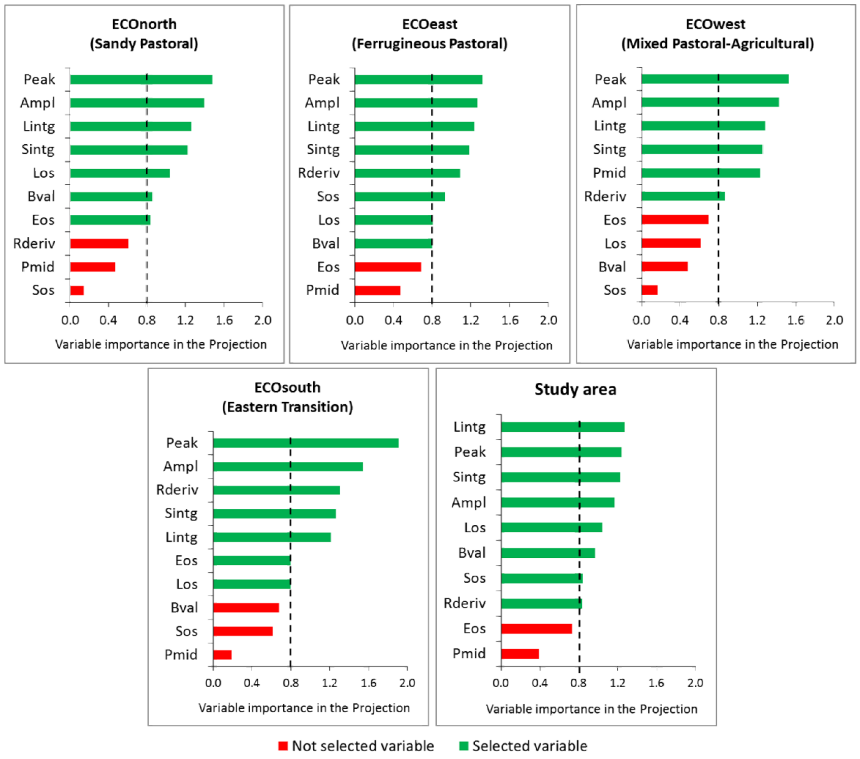

Figure 3.3 – Phenological variable importance for total biomass estimation across ecoregions and for the whole study area. Dashed dark line shows the threshold (VIP = 0.8) for the selection of key variables used for further model development (ecoregions are shown in Figure 3.1).

3.3.3. Selection and verification of the estimation models

Table 3.5 presents the best-performing models built across the study area and per ecoregion. It shows that the study area model (Model_SA) with the LINTG, LOS, and RDERIV variables could be considered the most suitable ones for estimating total biomass across the study area at the end of the growing season, with an adjusted R^2 (Adj. R^2) of 0.67 and an NMAE of 26%. Although slightly less strong (Adj. R^2 = 0.62) compared with Model_SA, the early warning model (Model_EW) using the PEAK and SOS variables also showed good accuracy (NMAE = 27.3%). For the ecoregions,

the models showed good accuracy, with NMAE usually below 24%, apart from ECOnorth (NMAE = 31%). The Adj. R^2 values were generally low and varied between 0.15 (ECOwest) and 0.49 (ECOeast). The selected phenological metrics varied among ecoregions. According to the VIP, PEAK was ranked as the most important variable in predicting total biomass throughout the four ecoregions (Figure 3.3). However, according to the prediction performance of the ecoregion models, PEAK was selected only in models for ECOnorth and ECOwest, whereas in ECOeast and ECOsouth, the models included LINTG (Table 3.5), suggesting a preferential selection of variables depending on ecoregion. LOS was also an important metric in modelling total biomass across ECOeast and ECOsouth.

Table 3.5 – Calibration and bootstrap verification performances of multiple linear regression models for total biomass estimation across the study area and per ecoregion. "n" is the size of the original sample used for calibration and "n_test" is the size of all bootstrap samples used for statistical calculations of verification. For other acronyms, see Tables 3.2 and 3.3.

Region	Estimation Model of Total Biomass (B)		Adj. R^2	MAE (kg·DM/ha)	NMAE (%)	n/n_test
Study area	Model_SA	B = 424.13 × LINTG − 100.91 × LOS + 39.80 × RDERIV + 293.71	0.67	608	26.0	263/39600
	Model_EW	B = 4594.18 × PEAK − 129.09 × SOS + 1866.17	0.62	641	27.3	263/39600
Ecoregion	ECOnorth	B = 1703.10 × PEAK + 1644.92 × BVAL + 432.94	0.24	427	31.0	121/18600
	ECOeast	B = 463.02 × LINTG − 296.29 × LOS − 152.37 × SOS + 5969.39	0.49	575	23.1	65/9800
	ECOwest	B = 3341.72 × PEAK + 282.87 × PMID − 7125.91	0.15	589	19.1	44/6600
	ECOsouth	B = 603.53 × LINTG + 52.83 × RDERIV − 325.30 × LOS + 1944.04	0.31	513	11.3	33/5000

3.3.4. Comparison with the NDVI-based CSE biomass product

In order to assess the improvement in the quality of biomass estimates provided by the multiple linear models, the estimates from Model_SA, Model_EW, and the ecoregion models were compared with the CSE biomass product (Figures 3.4 and 3.5). The total biomass estimates for 1999 were removed because the CSE estimates were based on LAC AVHRR data for that year (see Section 3.2.3.1). The relationship between the observed and estimated total biomass was plotted using the validation dataset comprising n = 247 samples collected between 2000 and 2013 (except 2004). The relationship between satellite and ground estimates of biomass was highly significant in all cases ($p < 0.01$). Model_SA ($R^2 = 0.68$, Figure 3.4a), Model_ EW ($R^2 = 0.64$, Figure 3.4b), and the ecoregion models ($R^2 = 0.77$; Figure 3.4c) outperformed the CSE product ($R^2 = 0.51$, Figure 3.4d). The CSE product was the least accurate, with an MAE of 818.46 kg·DM/ha, whereas the ecoregion models used throughout the study area, with varying metrics per ecoregion, provided the most reliable estimates (MAE = 489.21 kg·DM/ha). Total biomass estimates from the ecoregion models provided the highest slope (0.78) and the lowest offset (517) values of the linear regression equation with observed data, indicating an improvement in prediction accuracy when disaggregating the study area and applying models related to ecological characteristics. The ecoregion models allowed a better sampling of the space of biomass and corrected the slight saturation effects observed for Model_SA and Model_EW (Figure 3.4a,b) where high biomass values were misrepresented. The validation with ground measurements demonstrated that the new developed models improved the estimation of total biomass as compared with the CSE current product. This was confirmed by the temporal evolution of the estimated total biomass from these models between 1999 and 2013 (Figure 3.5). Along the full time series, estimates provided by Model_SA, Model_EW, and the ecoregion models had similar values, unlike the CSE biomass products that were generally found to be over-estimating total biomass. This over-estimation was clearly apparent in 2010, when estimated total biomass by CSE products exceeded

5000 kg·DM/ha, whereas the observed biomass value was about 3000 kg·DM/ha, similar to the values provided by the multiple-predictor models.

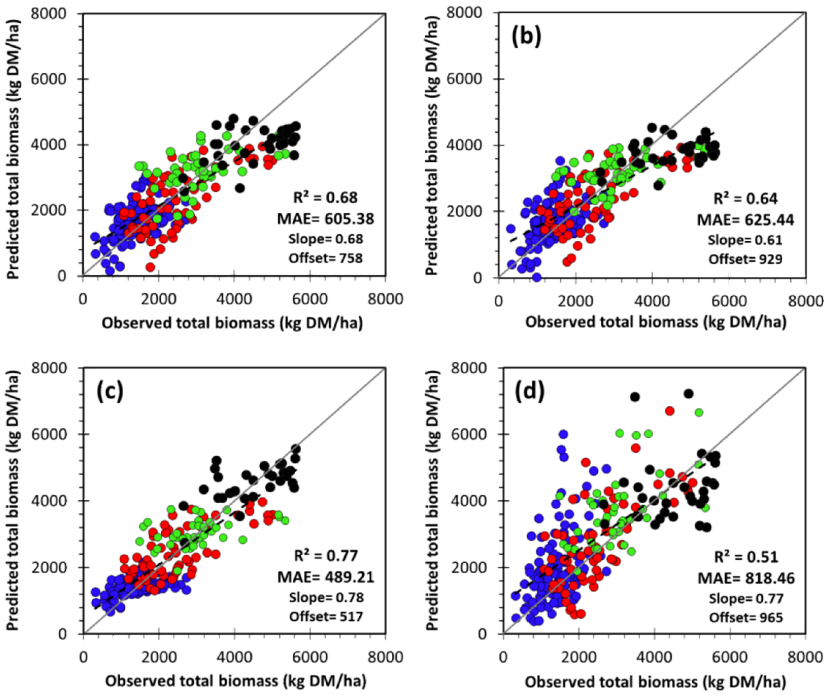

Figure 3.4 – Relationships between observed and predicted total biomass by (a) Model_SA, (b) Model_EW, (c) ecoregion models, and (d) CSE biomass product. Evaluation over the same validation dataset (n = 247 samples) from 2000 to 2013. The given statistics are the coefficient of determination (R^2), the mean absolute error (MAE, kg·DM/ha), and the slope and offset of the linear regression equation. For color correspondence, see Figure 3.1.

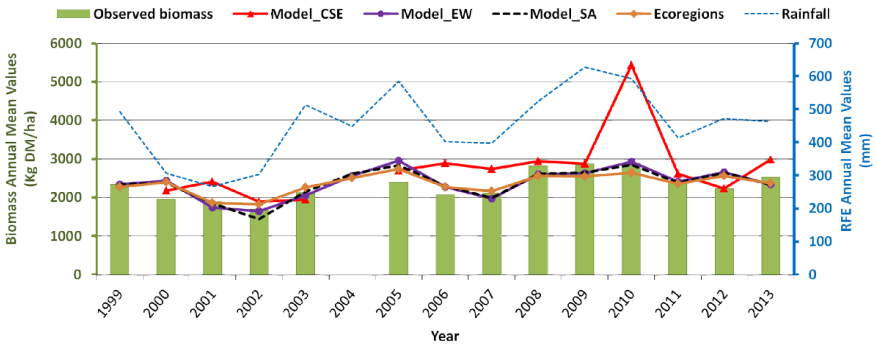

Figure 3.5 – Temporal evolution of the estimated total biomass from CSE and multiple predictor models with regard to observed total biomass, averaged from the monitored sites between 1999 and 2013. Ground data are missing for 2004. Only satellite estimates from SPOT-VEGETATION data are considered (i.e., CSE from the year 2000).

3.3.5. Testing the multiple-predictor model for early warning

The early warning model (Model_EW) was tested and applied in 2002 and 2010 (Figure 3.6a,b), selected because of extreme biomass production in these years compared with the 1999–2013 mean across the study area. Total biomass production was particularly low in 2002, with a deficit of about 23%, whereas in 2010, there was a surplus, with an excess of about 25%. The extrapolated results with Model_EW visually reflected these exceptional anomalies of total biomass production. In 2002, particularly north of Linguère and Ranérou, estimated total biomass was very low, with values generally below 500 kg·DM/ha (Figure 3.6a). Total biomass values higher than 4000 kg·DM/ha were obtained only south of Tambacounda. In contrast, in 2010, about 60% of the study area was characterized by total biomass production above 3000 kg·DM/ha (Figure 3.6b).

In addition, a strong and significant relationship was revealed by the scatterplot of observed and predicted total biomass anomalies from Model_EW (Figure 3.7a). A similar relationship have been noted between anomalies of predicted total biomass and rainfall (Figure 3.7b). This indicated that the early warning model was able to predict total biomass spatially (Figure 3.6) and temporally (Figure 3.7) at 1-km scale resolution

and provided enough information on the spatial distribution of biomass in extreme years, such as 2002 (in deficit) and 2010 (in surplus).

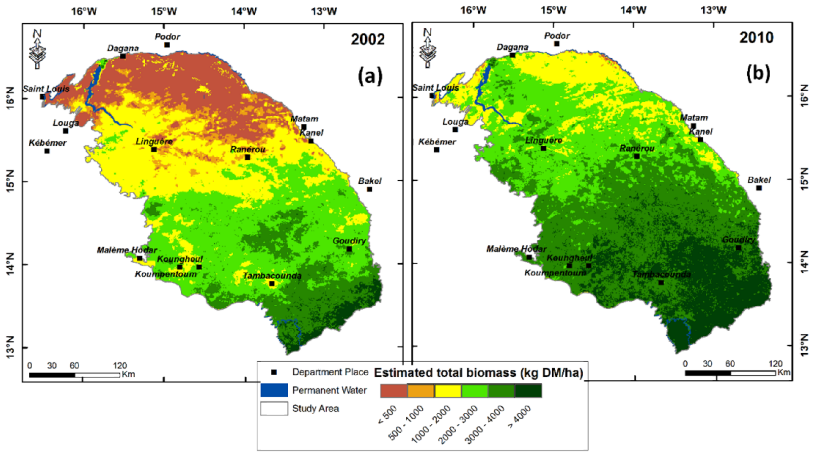

Figure 3.6 – Total biomass estimates for (a) 2002 in deficit and (b) 2010 in surplus, given by the early warning model (Model_EW).

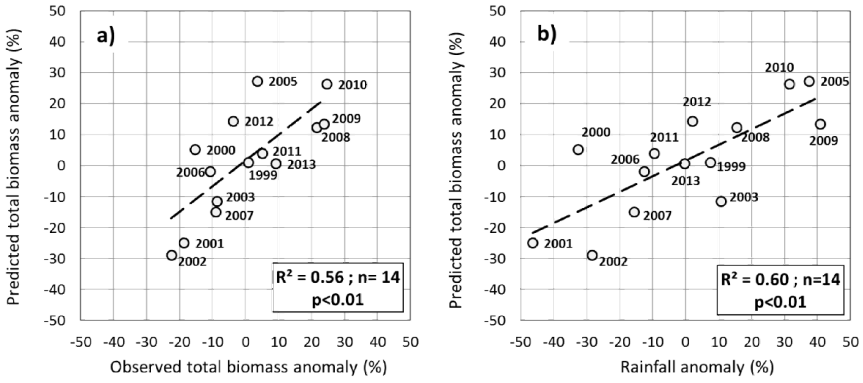

Figure 3.7 – Scatterplots of anomalies of total biomass predicted by the early warning model, with (a) observed total biomass and (b) rainfall from 1999 to 2013.

3.4. Discussion

The seasonal analysis of FAPAR patterns along the 1999–2013 time series enabled us to retrieve 11 phenological metrics for total biomass modelling in the Senegalese rangelands. With the exception of LDERIV, all the phenological metrics were significantly correlated to total biomass and were therefore used to develop prediction models. Model development

involved using a multiple regression approach with a 15-year time series of in situ biomass data. The results showed that the model with the three input variables LINTG, LOS, and RDERIV was the most suitable for estimating total biomass production across the study area, with a high adjusted R^2, while minimizing the MAE, indicating good fit and accuracy of the model. The three variables selected for this model have been described in the literature in relation to their ecological function. The cumulated FAPAR during the growing season was found to be a well-suited proxy of biomass production (Fensholt *et al.*, 2006; Dardel *et al.*, 2014; Meroni *et al.*, 2014b). The length of the season provides information on the timing of vegetation growing start and end. In various terrestrial ecosystems (e.g., savannah and grasslands), the LOS metrics are positively correlated with annual carbon sequestration, and thus with biomass productivity, simply due to the fact that more days are available for carbon assimilation (Richardson *et al.*, 2010). The decay rate is strongly species-dependent (Mbow *et al.*, 2013). Thus, by applying this metric information, the spatial patterns of dominating species are included, as well as other factors like grazing and burning. The performance of the models varied among the Sahelian ecoregions studied, with different metrics selected for each ecological region, adapted to local ecological conditions. The ecoregion models reduced the MAE on the total biomass estimates with 19%, compared with models calibrated over the whole study area. This meant that the subdivision of the study area into ecoregions increased the overall accuracy of estimates, confirming the study by Jin *et al.* (2014) in Chinese rangelands. The adjusted R^2 of the models per ecoregion, however, was generally low (less than 0.50). This can be explained by the reduced number and limited distribution of ground monitoring sites per ecoregion, giving an unequally distributed time series with a low dynamic range of values, in addition to possible remaining uncertainties related to the ground sampling method. This could be mitigated by using wider ecological regions where the distribution of monitoring sites is selected to reflect the spatial distribution of total biomass. This approach is supported by Brandt *et al.* (2015), who found that a wider spatial coverage of biomass data, including different ecological areas over many years, could improve results and reduce data uncertainties in the Senegalese rangelands.

The poor fit for ecoregion models could also be explained by the classification of the ecoregions, based on the integration of various biophysical and socio-economic components of the Senegal landscape (Tappan *et al.*, 2004). Therefore, the ecoregions that comprise several environmental factors might not sufficiently reflect the spatial variation of plant biomass production. From this perspective and according to Moron *et al.* (2006), who agree that vegetation type is the main source of landscape heterogeneity across Senegal, future studies should apply other types of classifications related solely to the vegetation growth cycle, (e.g., "phenoregions" based mainly on phenological parameters, such as SOS, EOS, and LOS; an example is provided by Ivits *et al.* (2013).

In addition, disaggregation to the ecoregion scale provides important information on the most appropriate phenological metrics for monitoring vegetation because metrics are closely linked to specific ecological characteristics, such as soil type, rainfall, woody cover, and species patterns (Table 3.1). These four points are closely intertwined and differ between the ecoregions. They control the biomass production in a different way, resulting in varying production levels of herbaceous and woody leaf biomass for each ecoregion. The selection of different metrics for the best performing ecoregion models indicates that these differences might be reflected in the metrics selected. LINTG was found to be the best proxy for total biomass production in ECOeast and ECOsouth, whereas in ECOnorth and ECOwest, the best proxy was the PEAK metric. The LINTG variable is known to represent the total amount of biomass produced by the entire vegetation cover, including woody and herbaceous vegetation (Olsson *et al.*, 2005). Its selection in ECOeast and ECOsouth, therefore, seems to be related to the presence of woody strata providing a high woody leaf biomass production. Likewise, the PEAK metric was selected as the main proxy for ECOnorth and ECOwest. These regions have a high herbaceous biomass production (Table 3.1), the growth of which is highly dependent on rainfall and its intra-seasonal distribution. The relevance of PEAK for grassland biomass monitoring in northern Senegal has been endorsed by Meroni *et al.* (2014b), supporting our results. LOS was found to be an important variable for ECOeast and ECOsouth, providing information throughout the growing

period. In addition, among the 10 phenological metrics used for developing models, SINTG and AMPL were the only ones not selected in the best-performing models, either for the study area or at the ecoregion scale. Although significantly correlated with ground total biomass, SINTG and AMPL are therefore thought to reflect mainly the seasonal herbaceous cycle signal (Mbow *et al.*, 2013) and are of minor importance compared with LINTG and PEAK.

Until now, fodder biomass in the Sahelian rangelands of Senegal has been estimated using a single-predictor model retrieved annually by simple linear regression between a cumulated vegetation index and ground biomass measurements (Diouf *et al.*, 2014). This method is cumbersome and expensive, with data having to be collected annually and processed before providing information on the available forage. Although the established method gives useful biomass estimations, the multiple-predictor models with phenological variables have proved to improve the available biomass estimates in terms of accuracy. Overall, the advantages and improvements with the proposed approach can be summarized as : (i) the phenological variables used are retrieved more precisely, pixel by pixel, by applying a fixed rule for the detection of the start and end of the growing season, instead of by a visual and general approximation of these two dates across the whole studied area; (ii) the multiple-predictor models are closer to reality because they take account of all or part of the interactions among phenological factors and therefore provide more reliable estimates; (iii) calibrated on a 15-year time series of biomass sampling data, the new approach is more robust than the conventional one, which is based on field sampling data for a single year (Jin *et al.*, 2014), and it allows current year predictions to be made without additional field work; and (iv) the multiple-predictor models allow the use of phenological metrics available early in the growing season to predict fodder biomass at the end of the season. A multiple-predictor model tailored to provide early biomass predictions is potentially very useful for the early warning monitoring systems in rangeland ecosystems in general and in Sahelian countries in particular, where most livelihoods are very dependent on fodder biomass. The usefulness of these phenological metrics for early warning of food insecurity

in the Sahel zone has been noted recently by (Meroni *et al.*, 2014a; Meroni *et al.*, 2014b). It is possible to link the SOS variable with the PEAK one in order to establish a precise early warning model, although the PEAK variable is often detected relatively late in the growing season (i.e., on average, in the second dekadal of September). With these two variables, the model gave valuable results when applied to 2002 and 2010, demonstrating the ability to provide information on a deficit or surplus in fodder production in extreme years. The early warning model outperformed the CSE prediction, which over-estimated actual biomass production in 2010. Such models could have mitigated the effects of the Sahel crisis in 2012, which was caused by late and irregular rains and the prolonged dry spells in 2011 (FAO, 2012). In Senegal, this crisis led to a decline in agropastoral production that threatened the food security of more than 739,000 people (FAO, 2012). In the future, such early warning models should enable stakeholders to take decisions as early as September (current year as biomass shortage) with regard to livestock by triggering protocols designed for livestock management (e.g., Opération de Sauvegarde du Bétail (DEPA, 2012)) in Senegal. The phenological anomalies for a particular year as compared to the long-term baseline characteristics of the seasonal cycle derived from FAPAR time series at the dekadal time step would allow the setup of critical indicators for food security (Verger *et al.*, 2013b; Verger *et al.*, 2015). The operational delivery of near real-time as well as long-time series of biophysical variables in the Copernicus Global land service from PROBA-V (Weiss *et al.*, 2014) and forthcoming Sentinel-3 (Lacaze *et al.*, 2015) will ensure continuity of the early warning biomass monitoring system. A recalibration of the proposed models is required for a year with very particular climatic conditions (Mora *et al.*, 2013) or in case of change in the satellite system (i.e., satellite/sensor drifts/end-of-life) (Richardson *et al.*, 2012). Special attention would be devoted to the early warning model in its future use insofar as Meroni *et al.* (2014b) found a large heterogeneity in the strength of the relation between the cumulated FAPAR over the growing season (i.e., biomass production) and the SOS and PEAK variables. More analyses are needed to better understand the relations between ground biomass and applied metrics.

3.5. Conclusion

The multiple-predictor models using phenological metrics and 15 years of ground observation data showed robust performance and gave accurate estimates of fodder biomass production across the Senegalese rangelands. The phenological variables selected in the predictor models depend on the production level and the ratio between the total woody leaf and herbaceous biomass. The large integral (LINTG) of Fraction of Absorbed Photosynthetically Active Radiation (FAPAR) appears well-suited in pastoral areas characterized by a dense woody cover and a high leaf biomass production, whereas the seasonal maximum (PEAK) metric is preferentially selected in grazing areas with lower woody density and a high herbaceous biomass production. The subdivision of the study area into ecoregions increased the overall accuracy of the multiple-predictor models. The validation with ground measurements shows that the proposed approach improves fodder resource monitoring in Senegal, providing more reliable and accurate estimates as compared to the current CSE biomass product. In addition, it allows reducing time and cost because, upon first model calibration, biomass estimation for a given year can be obtained without additional field work. On the contrary, the CSE model requires a yearly calibration with dedicated ground biomass measurements. Finally, using phenological metrics that are available relatively early in the growing season (i.e., PEAK and SOS), the proposed models can provide timely information on forage availability in rangelands. This allows helping stakeholders to make early decisions about possible livestock production deficits and related food insecurity. It constitutes an important benefit as compared to the current state of biomass estimation in Senegal, which is based on a single-predictor model that ingests the NDVI data at the end of the growing season. In order to enhance the performance of ecoregion models, future studies should apply classification schemes centered on the vegetation growth cycle (e.g., “phenoregions” based mainly on remotely sensed phenology). Further research is required to better understand the relation between satellite-derived phenological metrics and ecosystem properties. In addition, models which use the satellite optical vegetation products are generally hampered by

their saturation in densely vegetated area and the effect of soil brightness in regions with sparse vegetation cover. In assumption, this could be mitigated when adding independant agrometeorological data and justify the study conducted in Chapter 4.

4 Herbaceous biomass estimation using machine learning models, agrometeorological data and metrics from FAPAR time Series

This Chapter was adapted from the following publication:

Diouf, A. A., Hiernaux, P., Brandt, M., Faye, G., Djaby, B., Diop, M. B., Ndione, J. A., & Tychon, B. (2016). Do Agrometeorological Data Improve Optical Satellite-based Estimations of Herbaceous Yield in Sahelian Semi-Arid Ecosystems? Remote Sensing, 8, 668. doi:10.3390/rs8080668

4.1. Introduction

In the Sahel belt south of the Sahara desert, an arid to semi-arid region, the vegetation is composed of a herbaceous layer dominated by annual grasses and scattered woody plants including bushes, shrubs and small trees, among which are several thorny species (Anyamba and Tucker, 2005; Tagesson *et al.*, 2015). These areas provide the bulk of pastoral livestock feeding (Carrière, 1996) and contribute to carbon sequestration, nutrient uptake and cycling, soil fixation and soil biologic activity, as well as water cycle regulation. In this context, an accurate evaluation of herbaceous mass yield at the end of the growing season is essential for ensuring the rational use of available resources and environmental sustainability. Vegetation indices derived from satellite data have been widely used to monitor Sahelian herbaceous productivity for about 40-years. After the severe drought of 1970s, the first applications of herbaceous yield estimation in Sahelian rangelands using the Normalized Difference Vegetation Index (NDVI) from the National Oceanic and Atmospheric Administration-Advanced Very High Resolution Radiometer (NOAA-AVHRR) were published, and were pioneer studies in the field of applied remote sensing

(Tucker *et al.*, 1983; Tucker *et al.*, 1985; Justice and Hiernaux, 1986). In the past two decades and with regard to the technological advances in sensor design for vegetation monitoring (Seaquist *et al.*, 2003), new satellites such as the Satellite Pour l'Observation de la Terre-VEGETATION (SPOT-VGT) and the Moderate Resolution Imaging Spectroradiometer (MODIS)-TERRA/AQUA have been launched and more datasets have become available with higher spatial and temporal resolutions (Dardel *et al.*, 2014; Tian *et al.*, 2015). In addition to the NDVI, the Fraction of Absorbed Photosynthetically Active Radiation (FAPAR) in relation to surface processes such as photosynthesis (Baret *et al.*, 2013) has been recognized as constituting a key variable in the assessment of vegetation status (Prince, 1991; Prince and Goward, 1995). For this purpose, the metrics of the FAPAR seasonal dynamics are commonly used in environmental studies (Meroni *et al.*, 2014a; Meroni *et al.*, 2014b; Brandt *et al.*, 2016a) and their potential for herbaceous mass monitoring in Sahelian rangelands has recently been endorsed by Diouf *et al.* (2015). Most studies, however, rely exclusively on satellite-based vegetation indices (e.g., (Funk and Brown, 2006; Meroni *et al.*, 2014a)), with only a minority combining these indices with rainfall data, soil water status indicators and ground plant mass data.

Rainfall distribution is generally considered as the main driver of plant growth in the West African Sahel (Cornet, 1984; Hickler *et al.*, 2005; Ali, 2010; Huber and Fensholt, 2011), although there are many other local drivers of plants' photosynthetic capacity and spatial variability in the Sahel (Brandt *et al.*, 2015; Ibrahim *et al.*, 2015). Among them, the supply of mineral nutrients such as nitrogen (N) and phosphorus (P) are major determinants of plant growth rate (Penning de Vries and Djitèye, 1982). In the Sahel, N and P are the main limiting factors to plant production when rainfall is sufficient. For this reason, (Breman and de Ridder, 1991) proposed subdividing the Sahel belt according to the 250 mm isohyet, with rainfall as the main limiting factor below 250 mm and the nutrients N and P the main limiting factors above 250 mm annual rainfall. The wide temporal variability of N and P, however, in relation to the soil moisture regime and the absence of reliable information on their spatial distribution across Sahelian regions makes it difficult to include this information in statistical modelling

approaches. In contrast, comprehensive rainfall data have become increasingly accessible for use in the Sahel through the development of satellite-based products such as the Tropical Applications of Meteorology using SATellite (TAMSAT) data and ground-based observations (Tarnavsky *et al.*, 2014), the Famine Early Warning System NETwork (FEWS NET) Rainfall Estimate (RFE) (Xie and Arkin, 1996) and the African Rainfall Climatology (ARC) (Novella and Thiaw, 2013). The use of rainfall data as the overall driver of plant growth is supported by the high variability of Sahelian herbaceous productivity, which is difficult to predict in time and space (Hiernaux, 1983; CSE, 2010). Not all rainfall water, however, is directly available to plants. Rainfall water availability is mediated by the redistribution (i.e., run-off/run-on) on the soil surface in relation to soil physical characteristics (i.e., structure and texture), to topography (Breman and de Ridder, 1991) and to the physical nature of the canopy (Breshears and Barnes, 1999). This could help explain the poor relationship observed between herbaceous yield and annual total rainfall in the Sahelian rangelands of Niger and Mali (Hiernaux *et al.*, 2009a; Hiernaux *et al.*, 2009b). It justifies the concept of a water requirement index (also called a water satisfaction index or a water requirement satisfaction index, WRSI) developed and implemented through a soil water balance model named the Crop Specific Water Balance (CSWB) by the Food and Agricultural Organization (FAO) of the United Nations. With this model, the water balance of a given cropped soil can be calculated in time increments, usually 10 days (i.e., dekad), as presented in Equation 1 (Rojas *et al.*, 2005). Then, when filled with complementary parameters, the CSWB model produces a set of water status indicators generally used to assess the effect of weather conditions on crop development and yield (Doorenbos and Pruitt, 1977; Frère and Popov, 1979).

$$W_d = W_{d-1} + R - ETm - (r + i) \quad (11)$$

where,

W_d: amount of water stored in the soil at the end of the dekad (d)

W_{d-1}: amount of water stored in the soil at the end of previous dekad (d–1)

R: cumulated rainfall during the dekad

ETm: maximum evapotranspiration in the decadal period
r: represents the water losses due to runoff in the decadal period
i: represents the water losses due to deep percolation in the decadal period

For its agricultural monitoring activities, FEWS Net developed a grid cell-based modelling environment from the FAO's CSWB (Verdin and Klaver, 2002). The established geospatial model, GeoWRSI, was then enhanced by Senay and Verdin (2002) for West Africa's Sahelian rangelands (Senay *et al.*, 2011). Plant production models established solely from the absorbed PAR, as mentioned earlier and according to Mougin *et al.* (1995), neglect the direct effects of other factors such as water availability and nutrient shortage on plant growth. Models including both soil moisture (i.e., soil water status indicators) and the metrics of FAPAR seasonal dynamics are therefore expected to overcome FAPAR-related problems (e.g., saturation and cloud contamination) and improve herbaceous standing mass estimations. In such models, agrometeorological data introduce information about soil water availability, whereas FAPAR metrics provide information on herbaceous productivity and species patterns not taken into consideration by the agrometeorological component (Rudorff and Batista, 1990).

In this context, the overall aim of our study was to predict herbaceous yield across the arid to sub-humid areas of Senegal. The novelty of the study was the combination of satellite-based rainfall estimates, agro-ecological data and satellite-derived FAPAR metrics using a machine learning approach. The more specific objectives were to: i) develop three herbaceous models based respectively on FAPAR metrics, on agrometeorological variables and combined FAPAR and agrometeorological variables; ii) conduct a spatio-temporal comparison of model outputs; and iii) analyze the relationship between herbaceous yields and the metrics of onset/end of season calculated from FAPAR and satellite-based rainfall data for early warning purposes.

4.2. Materials and methods

4.2.1. Study Area

Detailed information of the study area is provided in Chapter 1 (see Section 1.1.2). It contains, however, nine land cover classes that, when excluding cropped lands, coincide with woody plant density, which increases from north to south (Figure 4.1, Table 4.1). These nine classes were constituted by aggregating the 60 classes of the original land cover database of Senegal (FAO, 2009a) and more or less corresponded with those used by (FAO, 2009b) for analyzing land cover change in Senegal between 1990 and 2005.

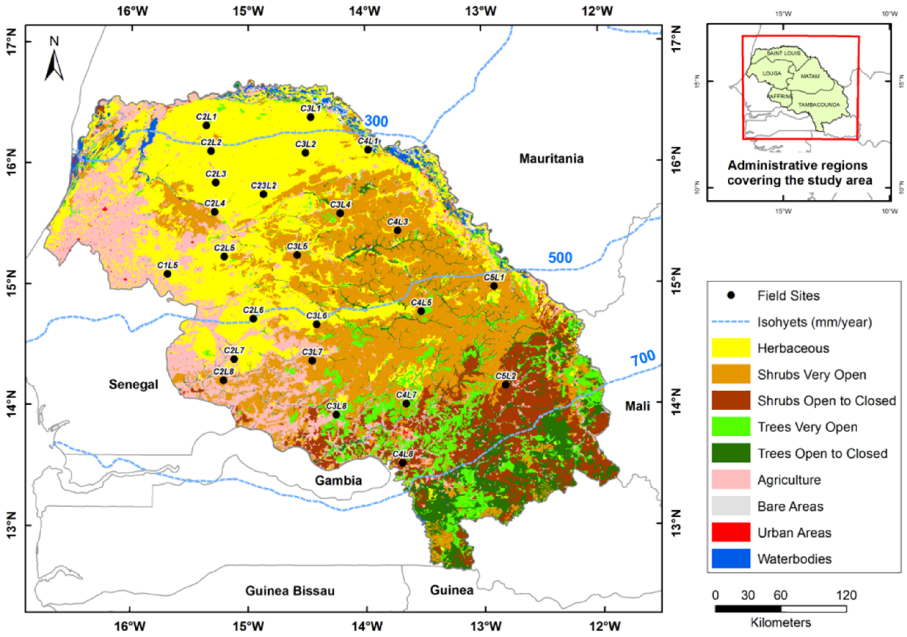

Figure 4.1 – Location of the monitoring sites and the main land cover classes (FAO 2009a). The isohyets are based on average rainfall estimates provided by FEWS Net between 2000 and 2015 (Xie and Arkin 1997).

Table 4.1 – General descriptions of land cover classes (FAO 2009c, d). Woody cover values were obtained from the woody cover map provided by (Brandt et al. 2016) and correspond to the averaged values of pixels covered by classes.

Land cover class	Abbreviation	Short description	Area (km²)	Area (%)	Woody cover (%)	Number of sites
Herbaceous	HER	Open to closed herbaceous vegetation with sparse trees and shrubs	38043	30.53	9	13
Shrubs very open	SVO	Very open shrubs	33724	27.07	17	8
Shrubs open to closed	SOC	Open to closed shrubs	12155	9.75	28	2
Trees very open	TVO	Very open trees, gallery forest	8889	7.13	25	1
Trees open to closed	TOC	Open to closed trees, gallery forest	9274	7.44	25	0
Agriculture	AGR	Large to small tree plantations and rainfed herbaceous crops	19483	15.64	14	0
Other classes	-	Bare areas, urban areas and water bodies	3041	2.44	-	0

Generic height classification:
Herbaceous = 0.03 to 3 m; Shrubs = 0.3 to 5 m; and Trees = 3 to 30m.

4.2.2. Data and Processing

4.2.2.1. Historical Field Herbaceous Yields

The *in situ* herbaceous mass data used in this study for the calibration of the models were collected annually from 2000 to 2015 (except for 2004) from 24 field sites located in free access natural rangelands, as shown in Figure 1 (black dots). The sites were selected away from the main water points and pastoral camps in order to avoid heavy grazing. Each field site covered a 3x3 km² homogeneous area and represented the most common land cover classes. Above-ground herbaceous mass was measured towards the end of the growing season, in early October. The technique used was the stratified sampling line originally proposed by the International Livestock Centre for Africa (ILCA) for monitoring pastoral ecosystems in the Gourma of Mali (Diouf *et al.*, 1998). Along a 1000 m transect, four strata (bare soil, and low, medium and high herbaceous mass production) were identified.

Then, taking into account the variability of herbage mass in the three covered strata, between 35 and 100 plots (each 1 m2) were sampled randomly along the transect line and fresh herbaceous mass was cut and weighed within each plot. After re-sampling, three samples for each stratum (i.e., nine samples per site) were kept for drying. The dry matter rate obtained by dividing the dry herbaceous mass weight by the fresh mass weight, as well as the relative frequency along the 1000 m transect, were then used to calculate the herbaceous mass yield in kg·DM/ha, first for each of the three strata and then for the site by adding them together. Note that the biomass data were not regularly collected in all monitoring sites due to occasional lack of logistics or the early passage of bush fires before field measurement (Diouf *et al.*, 2015). For more detailed information on the overall herbaceous mass collection, see Diouf *et al.* (2015).

4.2.2.2. FAPAR Vegetation Dynamics and Calculated Metrics

GEOV1 Copernicus Global Land FAPAR runs from 24-12-1998 to the present day at a ground sampling distance of 1/112° (about 1 km at the Equator) and 10-day steps (Verger *et al.*, 2015). Derived from the SPOT-VEGETATION (from 2000 to 2013) and Proba-V (for 2014 and 2015) instruments, this product is freely available on http://land.copernicus.eu/global. For detailed information on the principles used to estimate the GEOV1 FAPAR product, see Baret *et al.* (2013).

In order to calculate seasonal metrics, the GEOV1 FAPAR time series were filtered using the Savitzky-Golay (SG) fitting method available via TIMESAT software (Jönsson and Eklundh, 2004; Diouf *et al.*, 2015). Filtering is essential in semi-arid areas such as the Sahel, where many factors (e.g., clouds, aerosols, shadows, surface water) tend to produce noisy and erroneous FAPAR values (Chen *et al.*, 2004; Fensholt *et al.*, 2007). From the filtered FAPAR time series, eight metrics were retrieved with TIMESAT and used in this study (Figure 4.2): the start of the plant growth season (SOS); the end of the season (EOS); the length of growing season (LOS); the maximum FAPAR value over the season (PEAK); the amplitude (AMPL) corresponding to the difference between PEAK and the averaged left and

right minimum values over the ongoing annual cycle (BVAL); the small integrated FAPAR (SINT) from SOS to EOS and above BVAL; the small integrated FAPAR from SOS to PEAK time and above BVAL (GSINT); and the decreasing rate during the senescence phase (RDERIV). All these metrics have been described by Diouf *et al.* (2015) and Brandt *et al.* (2016a), except for GSINT, which was computed to include information during the green-up phase. SOS and EOS are essential for calculating the cumulated FAPAR metrics and were set to occur at 20% and 50% of the seasonal AMPL before and after the peak value, respectively (Diouf *et al.*, 2015). All the metrics were derived for each year between 2000 and 2015 on a pixel basis and 1 km spatial resolution. The annual values were then averaged for each site over an area of about 3x3 km² to match, as far as possible, the spatial sampling of the ground data.

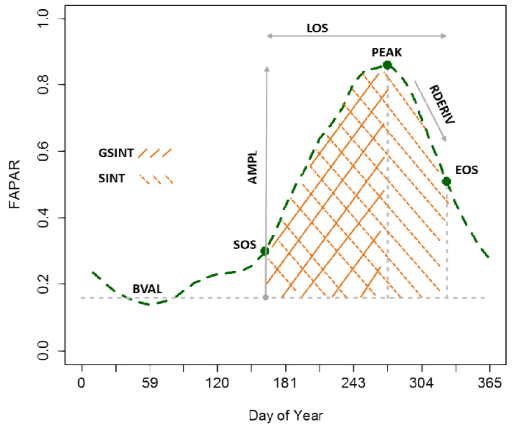

Figure 4.2 – Seasonal FAPAR metrics considered in this study and shown for a single pixel. The base value (BVAL) represents the averaged minimum values over the annual cycle (i.e., before and after the growing season).

4.2.2.3. Obtaining Agrometeorological Data

Agrometeorological data were calculated using the GeoWRSI model of FEWS Net adapted to a grid cell-based modelling environment (Verdin and Klaver, 2002) from the original water balance algorithm of FAO (Doorenbos and Pruitt, 1977). As noted by Senay (2004), the GeoWRSI model was enhanced and extended into an operational mode for Africa, Central America

and Afghanistan after the work by Senay and Verdin (2003). The WRSI indicator (Equation 12) is the ratio of seasonal actual crop evapotranspiration (AETc) to the seasonal crop water requirement, which is equivalent to potential crop evapotranspiration (PETc) (Senay, 2004). AETc represents the actual amount of water withdrawn from the soil water reservoir by vegetation transpiration and soil evaporation (Senay and Verdin, 2001; Senay, 2004) and PETc is the cumulative optimum crop water requirement to be met by rainfall and soil moisture for a given accumulation period (Senay *et al.*, 2011). After Senay (2004), the PETc at a given time in the growing season is calculated by multiplying the potential evapotranspiration (PET) by the crop coefficient (Kc), as in Equation 13. The AETc is determined by a set of functions integrating rainfall amount (PPT), plant available water (PAW), critical soil water level (SWC), soil water holding capacity (WHC), crop root depth (RD) and soil water content (SW) at the end of the study period. SW can be estimated using a soil water balance (Senay, 2004; Senay, 2008) with formula 14.

$$\mathrm{WRSI} = \frac{\sum \mathrm{AETc}}{\sum \mathrm{PETc}} * 100 \tag{12}$$

$$\mathrm{PETc} = \mathrm{Kc} * \mathrm{PET} \tag{13}$$

$$\mathrm{SW_i} = \mathrm{SW_{i-1}} + \mathrm{PPT_i} - \mathrm{AETc_i} \tag{14}$$

Where i corresponds to the time step.

For more information on the development and parameterization of the GeoWRSI model, see (Senay and Verdin, 2003; Senay, 2004). The main parameters needed to run the model are PPT, PET, WHC and Kc values. We ran the model in this study using the rainfall estimate images (i.e., Africa-wide blended product from satellite-gauge data) implemented by the Climate Prediction Center (CPC) of NOAA (Herman *et al.*, 1997; Xie and Arkin, 1997) and the global PET images estimated from the 6-hourly numerical meteorological model output using the Penman-Monteith equation (Shuttleworth, 1992). Rainfall and PET images were downloaded in dekadal time steps with a 0.1 degree (about 10 km) and 1 degree (about 110 km) spatial resolution respectively, from the ftp server of the Climate Hazard

Group for the archive data between 2000 and 2010 (Climate Hazard Group, 2015) and the FEWS NET server for the latest data between 2011 and 2015 (FEWS NET, 2015). The WHC image shows the spatial variation of easily available water capacity in the upper 100 cm, based on soil physical characteristics (Verdin and Klaver, 2002) and was obtained from the FAO digital soil map of the world (FAO, 1994) with a 0.1 degree spatial resolution. Kc values correspond to those proposed by Allen *et al.* (1998) for the extensive grazing pastures in semi-arid climates (Figure 4.3). These values were chosen with the assumption that the study area is uniform in hydro-climatic conditions and after a visual analysis of the WRSI variable which showed the most reasonable distribution for a median year regarding rainfall conditions.

The rainy season onset date (SOSp) was calculated for each modelling grid cell using the criteria of at least 25 mm of rainfall received over a given dekad, followed by a total of 20 mm for the next two dekads as applied by (Verdin and Klaver, 2002) after (AGRHYMET, 1996). The end-of-season date (EOSp) occurred when a climatological ratio between rainfall and potential evapotranspiration (i.e., PPT<PET/2) was observed (Senay, 2004). SOSp and EOSp were in turn used to calculate the herbaceous vegetation WRSI. In addition to WRSI, SOSp and EOSp, we calculated 12 other agrometeorological variables for the growing periods of each year (Table A3), particularly actual evapotranspiration (AET), water deficit (WDEF) and water surplus (WSUR) that accumulated during each of the four herbaceous vegetation growth phases (illustrated in Figure 4.3): initial (i); vegetative (v); flowering (f); and ripening (r). The seasonal fractions assigned to these four phases were 23%, 39%, 23% and 15%, respectively (numbers taken from the Sahelian Transpiration, Evaporation, and Productivity [STEP] model) (Mougin *et al.*, 1995). The computed rainfall variables corresponded to the seasonal amount accumulated from the start to the end of the rainy season (PPTc) and the mean rainfall from the start to the end of the season (PPTm). All the agrometeorological images were re-sampled to a 1 km resolution in order to match the FAPAR metrics using a bilinear interpolation method.

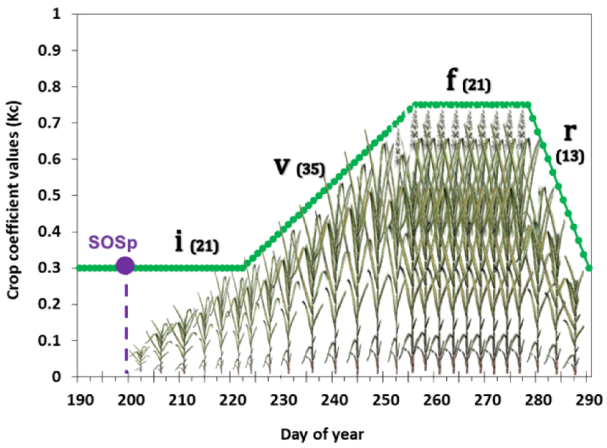

Figure 4.3 – Overall crop coefficient curve for Senegal's Sahelian rangelands during a growing season of 90 days. The growing period was divided into four phases: initial (i); vegetative (v); flowering (f); and ripening (r). SOSp indicates the mean date of the onset of the rainy season in the 2000-2015 period. Numbers in brackets indicate the total days of the phase.

4.2.3. Methods

The overall approach applied for elaborating the estimation models of herbaceous yield is presented in Figure 4.4.

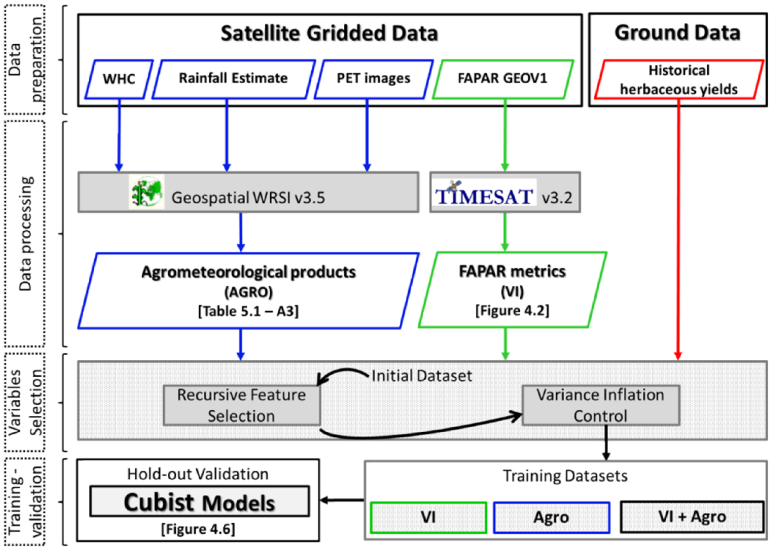

Figure 4.4 – Workflow for the development of the rule-based piecewise regression (i.e., Cubist) models for herbaceous yield estimation.

4.2.3.1. Explanatory Variable Selection for Herbaceous Mass Estimation

Over-fitting is a major drawback of machine learning methods (Domingos, 2012) and can occur when there is a strong relationship between the explanatory variables. In order to avoid over-fitting, only relevant variables with no significant collinearity should be kept in a model. Redundant predictors can trigger substantial instability in a model's coefficients. Wrapper methods (e.g., backward and forward) are widely used in the search procedure for predictors' selection in environmental studies (e.g., (Fraser and Li, 2002; De Benedetto *et al.*, 2013; Zhang *et al.*, 2014)). As reported by Hastie *et al.* (2009) and Kuhn and Johnson (2013), however, the principle of these methods lies in repeated hypothesis tests with the same data, which invalidates many of their statistical properties. In order to overcome this drawback, Guyon *et al.* (2002) proposed the recursive feature elimination procedure. This method is a sequential backward selection approach (Marill and Green, 1963) that avoids refitting many models at each step of the search by including an importance-ranking criterion instead (Kuhn and Johnson, 2013). The recursive feature elimination algorithm available with R software (Kuhn *et al.*, 2014) was used to identify less pertinent variables from the original dataset. In this algorithm, the random forest function (Liaw and Wiener, 2002) was used to compute the importance of the variables (i.e., mean decrease in accuracy when a variable is permuted) and rank them over subsets. The procedure's output with variables ranked from most to least important was then evaluated in order to eliminate interrelated explanatory variables (Brandt *et al.*, 2016a), using the Variance Inflation Factor (VIF) indicator (Equation 5). The least important variables with a VIF ≥ 10 (Belsley *et al.*, 1980) were eliminated one by one. This process was repeated until the VIF for all remaining variables was below 10.

$$VIF = 1/(1 - R^2 j) \quad (15)$$

Where R^2j is the correlation coefficient for the regression of each variable with the remaining ones.

4.2.3.2. Rule-Based Regression Tree and Model Building

The classification and regression trees (CART) approach, also called decision trees, is a data mining method used to identify patterns between a dependent variable and several independent predictor variables (Herrmann *et al.*, 2013). CART models can be used for either classification or regression, where the classification trees predict classes and the regression trees predict a continuous response (Liaw and Wiener, 2002; White *et al.*, 2005; Brosofske *et al.*, 2014). We used the Cubist program (Kuhn *et al.*, 2014), providing rule-based models using regression trees because the modeled herbaceous mass was continuous. Each tree was reduced to a set of conditional rules retrieved from the most important predictor variables (Kuhn *et al.*, 2012). These rules partitioned the independent variables (FAPAR metrics and/or agrometeorological products) into smaller groups, and each of which was linked to a multiple linear regression model that predicted the dependent variable (herbaceous mass) (Kuhn and Johnson, 2013). This algorithm, based on if/then rules, is well suited for Sahelian ecosystems, which are characterized by high spatial heterogeneity in terms of soil type and fertility, rainfall mediated by run-off/run-on, and species composition. The initial dataset was randomly separated into a 'training set' and a 'verification set' containing 70% and 30% of the data, respectively. Then a simple check was made to ensure that all land cover classes are represented in the two sub-datasets. Model learning was done with the training set, using a boosting-like scheme called 'committees' (Kuhn *et al.*, 2012), where each committee member corresponds to one regression tree. The number of committees was tuned by applying the commonly used 10-fold cross-validation (Kohavi, 1995; Refaeilzadeh *et al.*, 2009) in the model's performance estimation. All instances of the initial dataset were used and the cross-validated root mean squared error (RMSE) applied in order to obtain the best value of the committee number. It should be noted that for each committee after learning, a subsequent member (multiple linear regression model) adjusts for inaccuracies in the prediction of the previous one and the final predicted value is a simple average of all predictions from the various members (Kuhn *et al.*, 2012; Herrmann *et al.*, 2013; Kuhn and

Johnson, 2013). For comparative purposes, three cubist models were established: i) the Vegetation Index-model (VI-model) including only FAPAR metrics; ii) the Agrometeorological-model (AGRO-model) computed with agrometeorological variables; and iii) the VIAGRO-model including both FAPAR and agrometeorological data.

4.2.3.3. Model Verification, Error Analysis and Yield Anomaly Computation

Model accuracy verification was performed using an independent set of samples that were not used (i.e., hold-out) for the training of the model. The Cubist models were trained using the training set (207 samples) and then used to estimate the herbaceous yield values of the verification set (90 samples). The validation accuracy was retrieved using the observed and predicted herbaceous yield of the verification set. The quality of the models was assessed by the RMSE and mean absolute error (MAE). The anomaly values were calculated pixel-wise by the ratio (in percentage) of the difference between the actual and long-term average of the herbaceous yield (or rainfall) and the 15-year long-term average.

4.3. Results

4.3.1. Variable Selection and Model Development

After removing all the correlated variables, a final dataset of 15 explanatory variables remained and were used for model development (Figure 4.5c). Three models were developed for estimating herbaceous yields using: (a) only FAPAR metrics (VI-model); (b) agrometeorological data (AGRO-model); and (c) both FAPAR metrics and agrometeorological data (VIAGRO-model). The set of variables used for each model is shown in Figure 4.5 and the model performance is shown in Figure 4.6.

The contribution of the input variables varied with each model. For the VI-model, the PEAK variable was the most important, followed by the right derivative (RDERIV) and EOS (Figure 4.5a). GSINT and SOS were the least

pertinent variables as neither was used in model conditions or equations. In the AGRO-model, the SOSp and water deficit during ripening (WDEFr) were the most important variables, whereas the least important were water deficit (WDEFv) and water surplus (WSURv) during the growth stage. The most important variable in the VIAGRO-model was PEAK, followed by RDERIV. The SOS had little or no importance in the VI-model and VIAGRO-model, unlike the SOSp, which played a major role in AGRO-model and VIAGRO-model. The validation statistics of the three models are given in Figure 4.6. The VIAGRO-model showed the best performance, with $R^2 = 0.69$, RMSE = 483kg·DM/ha, MAE = 355kg·DM/ha and slope = 0.66. The VI-model performed less well ($R^2 = 0.63$; RMSE= 550kg·DM/ha; MAE =388kg·DM/ha; slope = 0.56), followed by the AGRO-model ($R^2 = 0.55$; RMSE= 585kg·DM/ha; MAE = 432kg·DM/ha; slope=0.51).

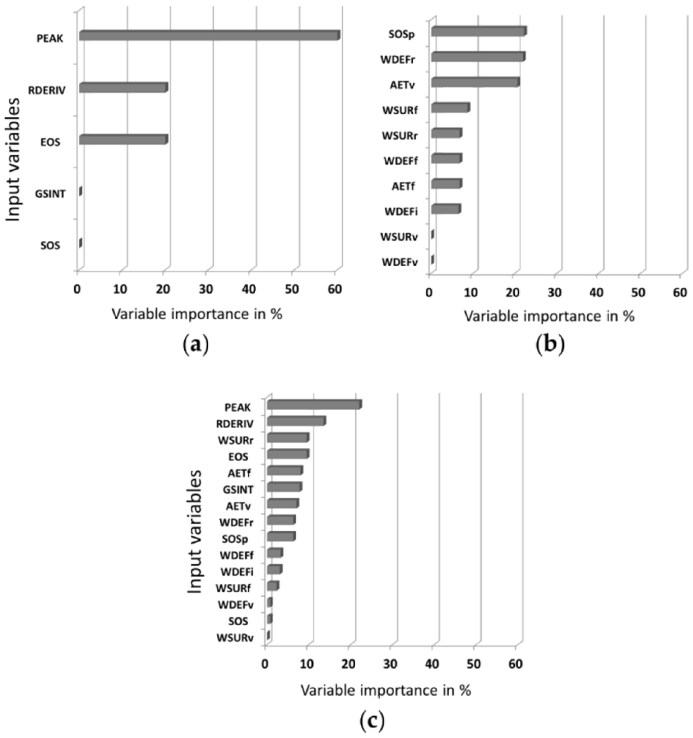

Figure 4.5 – Importance of the predictor variables for the three herbaceous yield estimation models: (a) VI-model, (b) AGRO-model and (c) VIAGRO-model. Single

variable importance initially given as the means of the percentage of use in model conditions and equations were then normalized to sum 100% for each model.

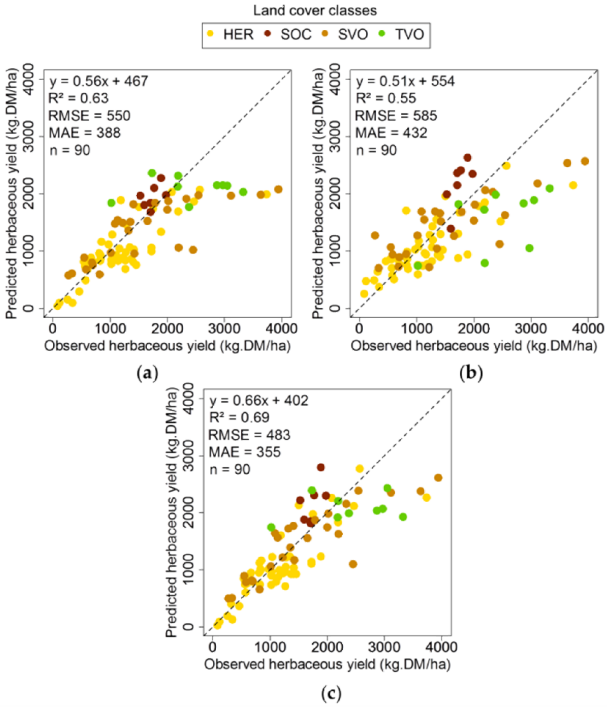

Figure 4.6 – Accuracy assessment of the developed Cubist models: relationship between observed and predicted herbaceous yield for (a) the VI-model, (b) the AGRO-model and (c) the VIAGRO-model.

4.3.2. Spatio-Temporal Comparison of the Models' Output

All the established models provided herbaceous yield estimations with values increasing along a north-south gradient in the 2000-2015 period (Figure 4.7). The VI-model, however, gave generally lower estimations than the two other models. Particular years, such as 2002 and 2014, when there was a considerable deficit in herbaceous mass production, were reflected well in both the VI-model and VIAGRO-model. The year 2010, however, was characterized by a high herbaceous mass production, which was clear in the AGRO- and VIAGRO-model estimations but less so in the VI-model. Overall, the VI-model underestimated high values, especially in the more southern regions; this drawback was corrected in the VIAGRO-model.

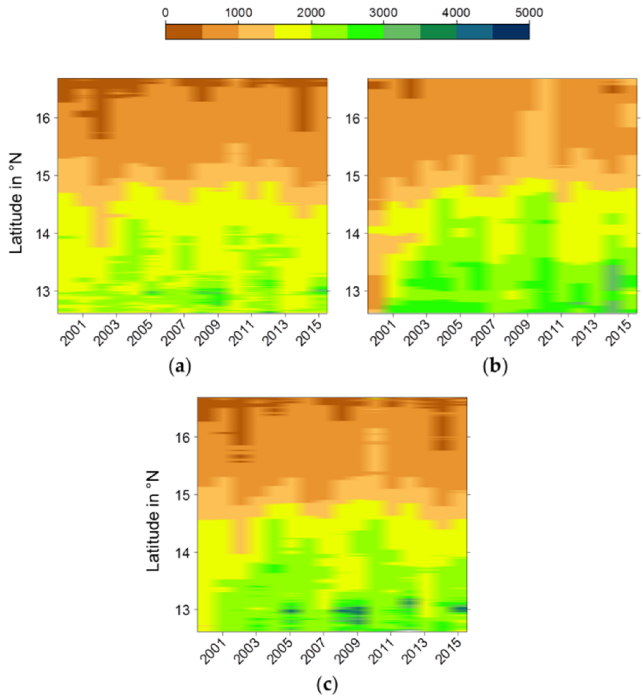

Figure 4.7 – Latitudinal variation of the herbaceous yield estimated by the (a) VI-model, (b) AGRO-model and (c) VIAGRO-model during the 2000-2015 period.

As an indicator of the inter-annual variability of vegetation (Hiernaux *et al.*, 2009b; Jin *et al.*, 2014), the coefficient of variation (CV) of the herbaceous yield estimated by the three models over a 16-year period was used to assess the temporal variations in herbaceous yield across natural vegetation and agricultural land cover classes (Figure 4.8). The inter-annual variations in herbaceous yield differed among the land cover classes. The AGRO-model showed the highest CV values, followed by the VIAGRO-model and the VI-model. For all the models, the HER class had the highest temporal variability, with an average CV of 17%, followed by SVO and AGR, with 14% and 13%, respectively. The lowest inter-annual fluctuations were observed in land cover classes with high herbaceous yield and woody cover, such as TVO (CV = 10%), TOC (CV = 9%) and SOC (CV = 9%).

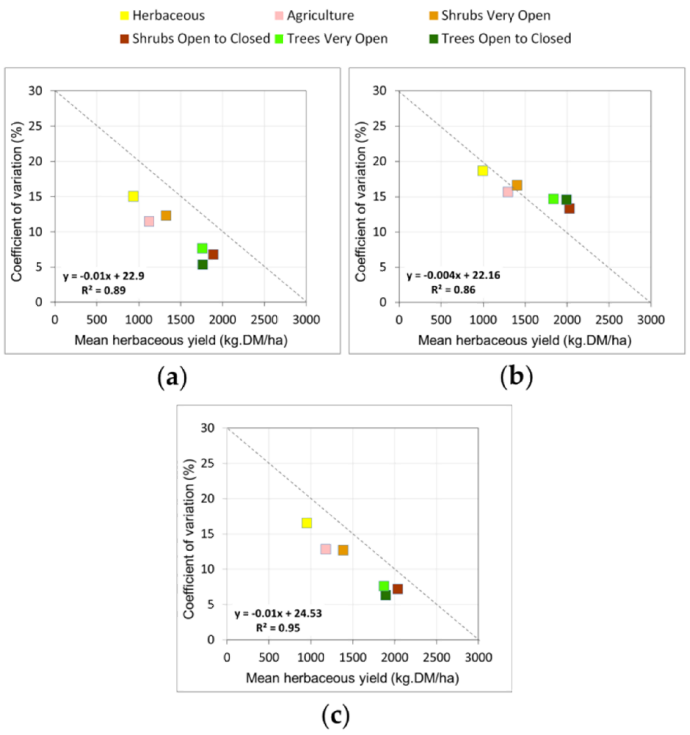

Figure 4.8 – Coefficients of variation in annual herbaceous yield estimated by the three models: (a) VI-model, (b) AGRO-model and (c) VIAGRO-model, according to the 2000-2015 average.

The herbaceous yield anomalies were assessed (Figure 4.9) and were generally in agreement with rainfall anomalies across the studied land cover classes, particularly for extreme years such as 2002 and 2010. Some discrepancies between the VI-model and the other two models were observed in 2000 and 2014. The VI-model showed a positive anomaly for SVO, whereas the AGRO- and VIAGRO models provided negative and no anomalies for the same class in 2000. For 2014, the AGRO-model provided positive anomalies for the SOC, TVO and TOC classes, whereas the VI-model and VIAGRO-model provided negative anomalies. As shown in Figure 4.9, the VIAGRO-model generally provided a blended estimation of the anomalies compared with the VI-model and AGRO-model.

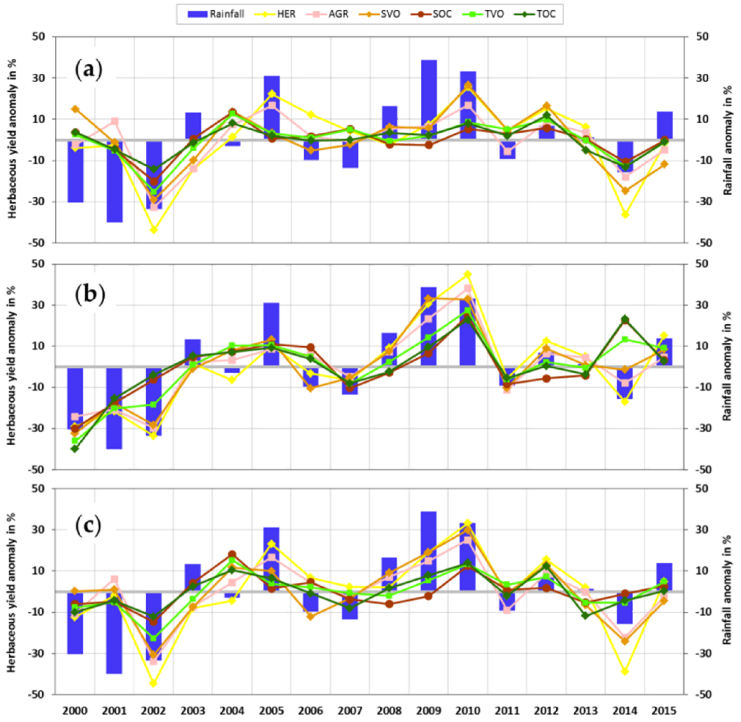

Figure 4.9 – Inter-annual variations in rainfall and estimated herbaceous yield over the whole study area from the (a) VI-model, (b) AGRO-model and (c) VIAGRO-model. Rainfall values were averaged from the 24 monitoring sites and the estimated herbaceous yield was averaged from all pixels covered by a given class. Colors and acronyms are explained in Figure 8.4 and Table 1.4.

4.3.3. Season Onset/End Derived from FAPAR and Rainfall Data

As observed in Section 4.3.1, the onset/end metrics estimated from the FAPAR seasonal curve (indicating the plant growing season) and rainfall data (indicating the rainy season) played different roles in model establishment (Figure 4.5), contrary to what was expected. The onset and end of the growing season (SOS and EOS from FAPAR) and of the rainy season (SOSp and EOSp from rainfall) are illustrated in Figure 4.10. The mean SOS and SOSp dates were delayed from the southern to northern land cover classes along the climatic gradient, reflecting the progression of the West African monsoon. With regard to the median values in boxplots, the growing season started in June for the three classes (SOC, TVO and TOC)

located mainly in the south, whereas for the AGR, SVO and HER classes the growing season started in July. The SOS occurred mainly in the same dekad as the SOSp for all land cover classes, except HER, where it started about a dekad early (Table A4). The end of the growing season was concentrated between late October and November, being later towards the south. On average, the EOS occurred about 2 dekads after the EOSp which occurred about late October. The EOS varied, occurring early for the HER class, but late for TOC in November.

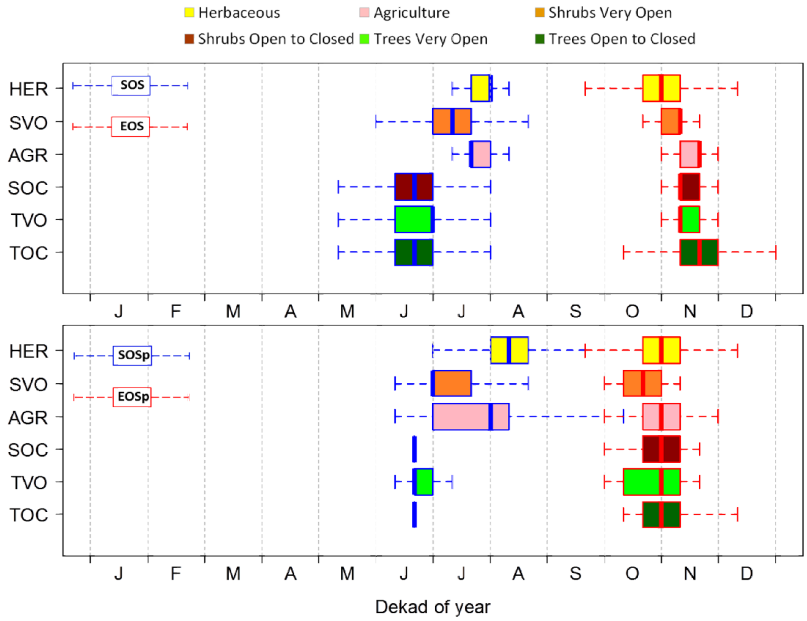

Figure 4.10 – Boxplots of the start and end of season metrics derived from FAPAR (SOS and EOS) and rainfall data (SOSp and EOSp), averaged over the 2000-2015 period for each land cover class.

4.3.4. Linkage between Start of the Growing/Rainy Season and Annual Herbaceous Yield

This section analyzes the relationship between SOS and SOSp and the herbaceous yield. The range of the SOS and SOSp anomalies was narrow (-5 to +10%) compared with the herbaceous yield anomalies (-60 to +60%). A very weak and non-significant relationship ($r = -0.30$ and $p = 0.28$) was

observed between the SOS and *in situ* herbaceous yield anomalies averaged over the study area (Figure 4.11a). In contrast, the meteorological variable SOSp had a highly significant relationship with observed herbaceous yield anomalies ($r = -0.65$ and $P < 0.01$) (Figure 4.11b). This was confirmed over all the land cover classes where SOSp achieved $r = -0.74$ for SVO, whereas the maximum r value for the SOS metrics was -0.47 registered for HER (Figure 4.11c). For all the land cover classes, there was no significant relationship (at the 95% significance level) between SOS and the observed herbaceous yield anomalies. The weakest relationships for the two metrics were observed for the open to closed classes dominated by either shrubs or trees (i.e., SOC and TOC, respectively), being insignificant for both SOS and SOSp.

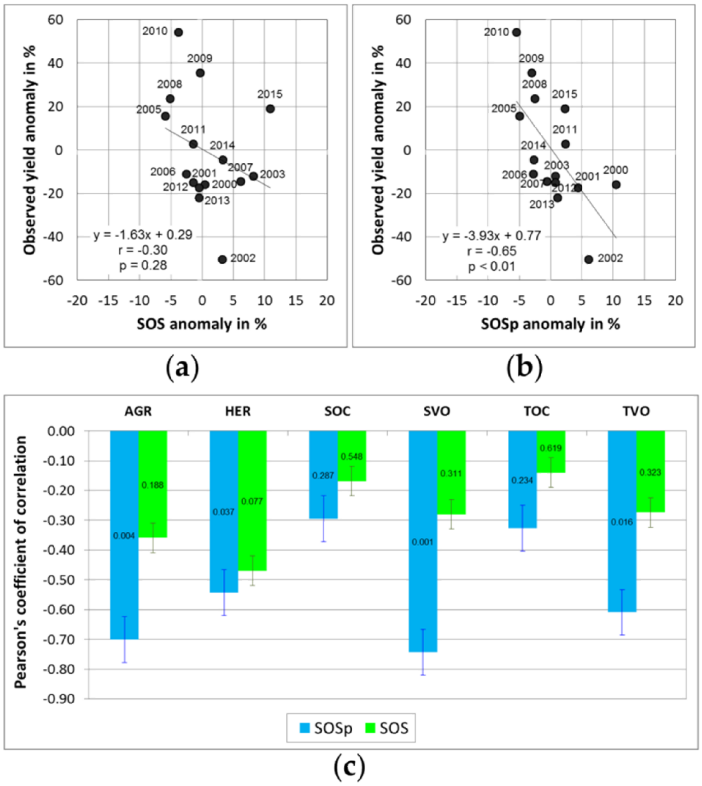

Figure 4.11 – Relationship between anomalies of herbaceous yield mass and onset of (a) the growing season (SOS) and (b) the rainy season (SOSp) for the whole studied area, and (c) for each land cover class. Numbers on bars correspond to p-values.

4.4. Discussion

4.4.1. Model Development and Output Comparison

All three models gave a reasonable performance. The VIAGRO-model (including both FAPAR and agrometeorological variables), however, outperformed the models based solely on FAPAR metrics (VI-model) and agrometeorological variables (AGRO-model). Among the three models, the AGRO model had the weakest performance, which could be related to the fact that all applied variables are independent of internal factors (i.e., the influence of grazing) whereas all FAPAR metrics as well as the field data are influenced by grazing over time. All the models showed a coherent distribution of the estimated herbaceous yield, with values increasing along the north-south gradient, together with decreasing inter-annual variability. The saturation effect has always been a drawback in the optical vegetation products in relation to the high productivity of vegetation mass (Huete *et al.*, 2002; Santin-Janin *et al.*, 2009; Xiaoping *et al.*, 2011; Diouf *et al.*, 2014; Tian *et al.*, 2016) and was consequently observed in the VI-model. The simultaneous use of agrometeorological data and FAPAR metrics mitigated the saturation effect, being more sensitive to high greenness values. However, high biomass values (>3000 kg DM/ha) are still underestimated in the AGRO- and VIAGRO-models, which could be related to few sites sampled in densely vegetated classes within the global dataset (see Table 1). Hence there is need to establish more sites in southern Senegal to achieve a higher accuracy of the herbaceous yield forecasting over the whole country. The agrometeorological data also reduced the discrepancy between herbaceous mass and FAPAR values, with the VIAGRO-model having a smaller scatter of low values and high values closer to the 1:1 line (see Figure 4.6). The contribution and significance of the input variables varied among the models. In the VIAGRO-model and VI-model, the FAPAR PEAK was the most important variable, whereas SOS was the least important. The relevance of PEAK in Sahelian grasslands has been demonstrated by Meroni *et al.* (2014b) and Diouf *et al.* (2015). The most important variable for the AGRO-model, however, was SOSp and the least

important was WDEFv. The latitudinal variation of the herbaceous yield estimated by the three models is mainly related to the dominant role of rainfall for vegetation growth along the north-south gradient. The variation of the rainfall onset as well as the mean annual rainfall along this gradient is a well-known characteristic of the Sahel. This is the basis of the transhumance system where the herds move during the dry season from the north to the south to take advantage of more humid areas with available water and fodder biomass.

The HER land cover class, located mainly in the northern Sahelian regions, had the highest temporal variability in herbaceous yield among all the land cover classes (see Figure 4.8). For this reason, anomalies in herbaceous mass were also the most pronounced in HER, particularly in extreme years (such as 2002 and 2010). Land cover classes with a higher herbaceous yield and annual rainfall, generally located in the south (i.e., SCO, TVO and TCO), had much lower inter-annual fluctuations. These results accord with those reported by Hiernaux *et al.* (2009b), who showed that inter-annual variability in herbaceous yield increases as the climate gets drier with latitude (i.e., to the north). The high temporal variability and magnitude of herbaceous mass anomaly (in percentage) in HER could also be related to the herbaceous species composition, which can vary greatly from one year to another. Species in HER are characterized mainly by annuals (Poaceae as *Aristida mutabilis, Chloris prieurii* and *Dactyloctenium aegyptium*) that are closely related to the annual rainfall and its intra-annual distribution (Valenza, 1977). These Poaceae species are also characterized by lower dry-matter production than species such as *Andropogon amplectens* (perennial) and *A. pseudapricus* (annual long-life cycle) prevalent in SCO and TVO areas, which have lower inter-annual mass fluctuations.

4.4.2. Model Applicability and Uncertainties

The three models can be used to estimate the spatial availability of herbaceous fodder at the end of each growing season across the Sahel in Senegal. The VIAGRO-model combines the advantages of both agrometeorological and FAPAR variables. The value of such a model

compared with traditional assessments (i.e., empirical statistic relationship between plant production and FAPAR metrics) derives from the integration of agrometeorological information that links herbaceous yield with climate and biophysical processes, thus taking account of management intervention, soil water availability and species patterns. The VIAGRO-model is replicable in other Sahelian countries because the data used (i.e., satellite images and programs) are freely available and easy to access (see links in Section 4.2.2.). The models should be applied carefully, however, because uncertainties caused by different collection dates of in situ data could lead to bias in their coefficients. Some dicotyledons, such as *Zornia glochidiata*, *Alysicarpus ovalifolius* and *Tribulus terrestris*, as well as grasses such as *Tragus racemosus* and *Dactyloctenium aegyptium*, shed their leaves before the end of the growing season (i.e., before the in situ data is collected). This can reduce the *in situ* herbaceous mass compared with the annual production peak. In addition, grazing during the growing season can also reduce the in situ herbaceous mass, particularly during dry years where transhumant herds return earlier to the south, crossing the whole study area. The standard deviation of in situ data collected for the 24 monitoring sites during the 2000-2015 period are provided in Figure 5.3. Other sources of uncertainties could also be caused by the field herbaceous mass subsampling for the dry weight assessment and by the same Kc values applied to all land cover classes which may be not appropriate for densely vegetated areas. Future research can thus adjust Kc values to each land cover type based on average crop heights across Senegalese semi-arid areas and may lead to an adaptation of the GeoWRSI including a land cover mask to compute output variables with dedicated Kc values.

4.4.3. Management Implications of Models Results

The herbaceous yield estimated by the VIAGRO-model can be applied for the computation of both herbaceous production and anomaly detection per administrative unit across the whole Senegal. The biomass production could then be linked with the livestock number per unit to assess a prospective fodder balance which is useful to guide the management of

rangelands and livestock movement (Touré *et al.*, 2012). To reduce the pastoral households vulnerability in relation to livestock mortality, the livestock insurance domain is more and more investigated to develop dedicated index based models (Chantarat *et al.*, 2009; Ikegami *et al.*, 2012). However, the classical NDVI based models have many flaws and do not realistically estimate the herbaceous biomass in shrub dominated areas which is important to properly predict livestock mortality. The VIAGRO-model shows good results across all Sahelian land cover classes and could be an important step towards an applicable insurance index.

4.4.4. Comparison of FAPAR and Rainfall-Based Onset/End Metrics

The onset/end of the rainy season marks a vital time for livestock managers, pastoralists and stakeholders in natural resource management in West African regions (Fitzpatrick *et al.*, 2015). Several definitions of the onset of the rainy season have been proposed in the literature in relation to the local onset of persistent rainfall (Sivakumar, 1988; Omotosho *et al.*, 2000; Marteau *et al.*, 2009). With advances in remote sensing technology, other metrics of onset/end dates have been proposed for assessing the plant growing season. These metrics are generally retrieved using specific rules based on thresholds of the amplitude of seasonal vegetation indices and on certain rainfall amounts over given time periods. So far as we know, however, no study has yet investigated the relationship between the onset/end metrics derived from satellite vegetation indices and from rainfall estimates in the West African Sahel. We have shown the dissymmetry of the onset/end of the rainy season along the Sahel bioclimatic gradient, with the onset date staggered over 3 months from May in the south to July in the north. The end date is spread over only 1 month, from late September in the north to early November in the south, which accords with observations by Hiernaux and Le Houérou (2006). Our results also show that FAPAR based on SOS occurred at the same time as rainfall based on SOSp across the whole study area, except for HER land cover class where SOS was observed one dekad earlier (see Table A4). There are many possible explanations for this. The onset of the growing season or greening of plants follows the rains,

except for woody plants that start foliation earlier in response mainly to an increase in air temperature and humidity, enabling them to produce green leaves before the first rains (Hiernaux *et al.*, 1994; Devineau, 1999; Ali *et al.*, 2007). In addition, herbaceous vegetation is sensitive to low rainfall amounts, requiring less than 25 mm for the first dekad and 20 mm for the following two dekads to trigger the growth. The thresholds set for calculating SOS and SOSp could therefore be adjusted to better match the region's temporal dynamics. Apart from these botanical explanations, the reason could be related to data uncertainties. Almost no cloud-free optical satellite image is available at the start of season (Fensholt *et al.*, 2007) and derived satellite data differ greatly from in situ measurements (e.g., NDVI) at this time (Proud *et al.*, 2014). Both the smoothing algorithm used for GEOV1 data (influenced by clouds) and the further smoothing via TIMESAT therefore add uncertainty to the FAPAR-based SOS calculation and our results indicate that this metric should be used with great caution at the annual scale.

Our results also showed that the onset dates of the rainy and growing seasons were more variable than their cessation, confirming the findings reported by Sivakumar (1988). The end date of both seasons occurs between late October and November for all land cover classes. The EOS is not entirely controlled by the end of the rains, but depends also on the wilting of herbaceous vegetation, which is determined by a biological clock (i.e., photoperiodicity) (Brandt *et al.*, 2016a). This was confirmed by our results where EOS dates, defined by the FAPAR reduction by 50% of the seasonal amplitude, occurred later than the cessation dates of the rainy season. The EOS delay is further influenced by crops that stay green longer, depending on their genetic features and soil tillage, and by woody plants remaining green much longer, depending on their phenological behavior.

4.4.5. Early Assessment of Herbaceous Yield from Onset Metrics

For agricultural application, the SOS has attracted more attention from the scientific community and has been investigated in recent studies (i.e., (Mbow *et al.*, 2013; Meroni *et al.*, 2014b; Vintrou *et al.*, 2014)). It has

frequently been used to investigate the land surface phenology trends in relation to climate variability (Begue *et al.*, 2014). After analyzing the SOS metrics against a proxy of biomass production in the Sahel, however, (Meroni *et al.*, 2014b) indicated that vegetation monitoring and biomass production forecasts should not be based on SOS in areas with non-significant correlation. Given that our study showed no significant relationship (at 0.05 p-level) between SOS and in situ herbaceous mass, this statement can be confirmed by our results and we do not recommend using this metric solely at the annual scale. The SOSp metrics, however, were shown to have great potential for assessing herbaceous yield in the Sahelian region of Senegal. This was reflected in the VIAGRO-model where the SOSp variable was very important, confirming its superiority over SOS for detecting unfavorable rainfall conditions at the early stage of a growing season and therefore enabling early warnings about forthcoming risks of herbaceous mass deficit to be given to pastoral livestock managers and national stakeholders. Hence, even though the models presented here require a full season, this finding advances our knowledge towards an applicable early warning prediction (Diouf *et al.*, 2015).

4.5. Conclusion

Using the Fraction of Absorbed Photosynthetically Active Radiation (FAPAR) seasonal metrics computed with TIMESAT software and agrometeorological variables retrieved from the GeoWRSI program, we established three Cubist models for estimating herbaceous yield at the end of the season in the semi-arid areas of Senegal: (a) the VI-model with FAPAR metrics, (b) the AGRO-model with agrometeorological variables and (c) the VIAGRO-model with both FAPAR and agrometeorological variables. All three models gave reasonable estimations of herbaceous yield over time and across land cover classes, among which those herbaceous areas with low woody cover showed the highest inter-annual variability and those in the south with higher woody cover showed lower variability over time. The VIAGRO-model gave the best estimation performance and indicated that the simultaneous use of agrometeorological data and FAPAR metrics improved

the estimation accuracy and mitigated problems encountered with the sole use of FAPAR metrics (i.e., VI-model): (1) the saturation affecting optical remotely sensed vegetation data in areas of high vegetation productivity was attenuated; (2) the discrepancy between herbaceous mass and greenness (caused by species with high greenness and low mass production, or vice versa) was attenuated, with a weaker scattering around the low and high values closer to the 1:1 line (the additional use of agrometeorological data corrected for underestimations with the VI-model, particularly in sparsely vegetated areas); (3) the onset of the rainy season calculated from rainfall data was shown to be well suited for herbaceous mass assessment, in contrast to the onset of the growing season retrieved from FAPAR satellite data, which was not significantly related to herbaceous yields. Nevertheless, these metrics should be further investigated in order to improve our understanding of their temporal patterns and determine future setting parameters for their retrieval across Sahelian ecosystems.

5 General conclusion and outlook

5.1. General conclusion

The overall aim of this research was to develop models that use remote sensing data and field biomass data, to estimate and / or forecast fodder biomass availability at the end of the growing season in Senegalese semi-arid areas. To achieve this, a 17-year dataset of in situ biomass measurements and various satellite-based vegetation products, rainfall and evapotranspiration were used. Then, the global objective was separated into three specific objectives which led to open questions investigated subsequently into three steps (i.e., Chapter 2, 3 and 4 of the dissertation). The following paragraphs summarize our findings for each step and present the answer(s) proposed for related question(s).

5.1.1. Total biomass estimation with simple regression models and NDVI variables

> ***Objective 1:*** *Evaluate the existing annual estimation method of total fodder biomass by linear regression in comparison to nonlinear approaches.*

In this first step of the research, a comparison analysis of the linear regression equation (method currently applied in Senegal) with five nonlinear regression functions was done using the SPOT-VEGETATION time series of NDVI. This analysis showed that Exponential and Power models are more suited than the linear one to estimate accurately and consistently the total biomass at the end of the season in Senegalese rangelands. Another finding was that there is no significant difference between the means of total biomass estimated with models calibrated with

single year data and those retrieved from several successive years of field sampling data. The advantage of the several years-based models is that they have more stable coefficients and are applicable on a given year without including additional field data of the corresponding year.

5.1.2. Total biomass estimation with multiple regression models and FAPAR metrics

> ***Objective 2:*** *Develop a multiple regression approach (i.e., parametric models) that integrate FAPAR phenological metrics computed from EO time series for estimating and forecasting total fodder biomass at the end of the growing season.*

Although nonlinear models have been demonstrated to be well suited to assess fodder biomass in Senegalese rangelands, the simple regression approaches still have some uncertainties due to their monocriterion status and their weak forecasting capability. In this second step, a new methodology using multiple-linear models between phenological metrics computed from the SPOT-VEGETATION time series of the Fraction of Absorbed Photosynthetically Active Radiation (FAPAR) and 15 years of in situ total biomass data was developed. Multiple-predictor models showed a robust performance and gave accurate estimates of fodder biomass production across the Senegalese rangelands. The proposed approach provided more reliable and accurate estimates as compared to the current CSE biomass product (obtained from the traditional linear regression approach). Likewise, disaggregation to the ecoregions scale provided important information on the most appropriate phenological metrics for monitoring fodder biomass because metrics are closely linked to specific ecological characteristics, such as soil type, rainfall, woody cover, and species patterns. For example, the seasonal large integral (LINTG) of FAPAR appeared well-suited to pastoral areas characterized by a dense woody cover and a high leaf biomass production, whereas the seasonal maximum metric (PEAK) was preferentially selected in grazing areas with lower woody density and a high herbaceous biomass production. Although

significantly correlated with ground total biomass, the small integral (SINTG) and the FAPAR amplitude (AMPL) were thought to reflect mainly the seasonal herbaceous cycle signal and therefore are of minor importance compared with LINTG and PEAK to assess total biomass. Developing multiple linear models with specific metrics into ecoregions increased the overall accuracy of the fodder biomass estimates and mitigated the saturation of FAPAR. Using phenological metrics that are available relatively early in the growing season (i.e., PEAK and start of season), timely information (e.g., in September) on forage availability within rangelands can be provided to stakeholders to enable early decisions about possible fodder biomass deficits and related food insecurity. Although their development at ecoregion scale attenuates the saturation in densely vegetated areas, the models which use the satellite optical vegetation products still are hampered by the effect of soil brightness in regions with sparse vegetation cover. In assumption that those can be mitigated when adding independent agrometeorological data, the research in the following step (i.e., Chapter 4) was conducted.

5.1.3. Herbaceous forage estimation using FAPAR metrics and agrometeorological variables

Objective 3: *Develop machine learning approach (i.e., non-parametric models) including agrometeorological variables and FAPAR phenological metrics to estimate the herbaceous biomass at the end of the growing season.*

As never done before, a completely new approach based on a machine learning algorithm (i.e., Cubist) was developed to assess herbaceous biomass in Senegalese Sahel. Three Cubist models using FAPAR seasonal metrics and/or agrometeorological variables (i.e., soil water status indicators) were established. All models gave reasonable estimates of herbaceous biomass over time and across different land cover classes of the studied area. The model including both FAPAR and agrometeorological variables provided the best estimation performance and indicated that the simultaneous use of these two data source improved the estimation accuracy and mitigated problems

encountered with models using FAPAR metrics solely. The saturation affecting optical remotely sensed vegetation data in areas of high vegetation productivity as well as the discrepancy between herbaceous mass and greenness (caused by species with high greenness and low mass production or vice versa) was attenuated. As a positive consequence, the additional use of agrometeorological data corrected for herbaceous biomass underestimations with the sole FAPAR based model, particularly in sparsely vegetated areas. The onset of the rainy season calculated from rainfall data was shown to be well suited for herbaceous biomass assessment, in contrast to the onset of the growing season retrieved from FAPAR satellite data, which was not significantly related to herbaceous biomass.

This PhD dissertation proposes diagnosis and prognosis methods developed to assess accurately the total biomass and the herbaceous biomass in Senegalese semi-arid areas, using limited data and free available software and therefore easily replicable in other West African Sahel countries. The proposed tools constitute an important improvement and benefit as compared to the current state of biomass estimation in Senegal. It allows reducing time and cost because, upon first model calibration, fodder biomass estimation for a given year can be obtained without additional field work and as early as possible, to help stakeholders to make early decisions about pastoral livestock monitoring (e.g., herd guiding and food insecurity prevention) and natural resources management (e.g., prevention of bushfires and land degradation).

5.2. Outlook

In accordance to the objectives of the study, different statistical tools that estimate and / or forecast forage availability at the end of the growing season in Senegalese semi-arid areas were developed and discussed in this PhD dissertation. In overall this study was done to contribute to the challenging task of modelling the plant biomass component of the very dynamic and complex Senegalese Sahelian ecosystem. Focusing more on the future operational usability of developed models, only the Geoland Version 1 (GEOV1) vegetation products were used. Their 1 km spatial resolution,

indeed, matches very well with the biomass data collection method on the field (along a 1 km long transect). According to the length of these satellite imagery datasets, only plant biomass data collected from 24 monitoring sites (out of 36 before 1999) were used. This indicates clearly that several possibilities remain for the refinement and improvement of the proposed methodologies.

A possibility of improvement will be the GEOV2 dataset (expected to be available very soon) which significantly improves GEOV1 in terms of continuity (less than 1% of missing data in GEOV2 as compared to the 20% of gaps in GEOV1) and consistency (smoother products less affected by noise in the data) (Verger *et al.*, 2013a). The SENTINEL-3 satellite launched recently will ensure the delivery of near real-time biophysical variables with finer spatial resolution of 300 m. The total and herbaceous biomass models will be adapted, therefore, to ensure continuity of the developed assessing systems. To find out more improvement possibilities and for the sake of interoperability with the Copernicus Global Land products, different data source such as MODIS-TERRA/AQUA should be tested.

The plant biomass production maps provided each year to the Department of livestock in Senegal through the *Direction de l'Elevage et des Productions Animales* (DEPA) is used only by visual analysis and qualitative interpretation to assess the fodder biomass availability with respect to livestock population. No formal method exists to date for this purpose in Senegal. Establishing an operational approach for the evaluation of fodder biomass balance and anomalies impact analysis should be a logical sequence of the research (findings) presented in this dissertation. The livestock numbers per district provided annually by the DEPA could be used even if these numbers are retrieved, in general, by projection based on growing rate. A research paper is currently being prepared about the supply/demand of fodder biomass in Senegalese rangelands. This study should result to an approach that defines a formal implementation of the results obtained from this thesis.

A longer than 17-year time series of plant biomass maps can be very useful for eco-environmental studies (e.g., on land degradation, carbon uptake…). After a downscaling process with dedicated algorithms, the

Global Inventory Modelling and Mapping Studies GIMMS-3g NDVI imageries (Pinzon and Tucker, 2014) could be used to provide a time-series of plant biomass maps from 1987 to 2012, across the whole Sahel and with about 1/12° spatial resolution. Another potential imagery dataset can be the Vegetation Optical Depth (VOD) that has proven recently by (Tian *et al.*, 2016) to be an efficient proxy for green biomass of the entire vegetation stratum (both herbaceous and woody plant foliage) in the semi-arid Senegalese Sahel, extending from 1992 to 2011 with about 0.25° spatial resolution. For this purpose, biomass data available for Senegal (Ferlo), Mali (Gourma) and Niger (Fakara) within these periods could be very valuable.

This study was mostly concentrated in solving the technical requirement of implementing tools to assess fodder biomass available at the end of growing season in Senegalese Sahel. All variables used were computed within the growing season as well. However, fodder biomass is used by pastoral herd along the dry season. The questions that could be raised in the case of Senegalese rangelands are: i) how is changing the forage across rangelands during the dry season? and ii) how they could be monitored using existing remote sensing datasets? The conclusions made by Jacques *et al.* (2014) in the Gourma region of Mali could be used as pioneer information to answer these questions. In addition, only some efforts were made to understand the relation between satellite-derived variables (i.e., phenological metrics and status water indicators) and ecosystem properties (e.g., ecoregions and land cover classes). Further research is needed to improve our understanding of their spatial and temporal patterns across Sahelian ecosystems. Another aspect that we will investigate in the future is the assessment of forage quality using field measurements and satellite based products. This was not developed in the research presented here because of the absence of quantitative information about the forage value for Senegalese rangelands. Such information on forage quality combined with the dry matter production (and may be with many others factors to be defined) could allow to estimate approximately the production of milk and meat from pastoral herd at the end of the growing season.

During this PhD, we had the opportunity to work with other researchers on subjects more or less connected to our research theme. The collaboration

was fruitful for us (and the scientific community) as resulted on several article papers not included in this dissertation and published on high ranked journals (e.g., (Brandt *et al.*, 2014), (Brandt *et al.*, 2015), (Brandt *et al.*, 2016a), (Tian *et al.*, 2016) and (Brandt *et al.*, 2016b)). Beyond this PhD dissertation we will continue research to improve the proposed approaches, to develop or readapt deterministic tools, and to use their outputs for investigating the environmental change in the Sahel focusing specially on its relationship with livestock farming systems, climate variability and human induced effects (e.g., bushfires).

References

Acquah, H.D.-G. (2012). A bootstrap approach to evaluating the performance of akaike information criterion (aic) and bayesian information criterion (bic) in selection of an asymmetric price relationship. *Journal of Agricultural Sciences, 57*, 99-110

Afanador, N.L., Tran, T.N., & Buydens, L.M.C. (2013). Use of the bootstrap and permutation methods for a more robust variable importance in the projection metric for partial least squares regression. *Analytica Chimica Acta, 768*, 49-56

AGRHYMET (1996). Méthodologie de suivi des zones à risque. In: AGRHYMET FLASH, Bulletin de Suivi de la Campagne Agricole au Sahel. Centre Régional Agrhymet, B.P. 11011, Niamey, Niger, vol. 2, No. 0/96. p. 2

Akpo, L.E. (1990). Dynamique des systèmes écologiques sahéliens: Structure spécifique, productivité et qualité des herbages, le forage de widdu thiengoly. Mémoire : Diplome d'etudes approndies de biologie végétale. In, *Faculté des Sciences et Techniques* (p. 65): Université Cheikh Anta Diop

Ali, A. (2010). Variabilité et changements du climat au sahel : Ce que l'observation nous apprend sur la situation actuelle. *Le dossier : Aléas climatiques : quelles réalités, quelles évolutions ?, Grain de sel No 49*, 13-14

Ali, M., Saadou, M., & Jean, L. (2007). Phénologie de quelques espèces ligneuses du parc national du « w » (niger). *Science et changements planétaires / Sécheresse, 18*, 354-358

Alkemade, R., Reid, R.S., van den Berg, M., de Leeuw, J., & Jeuken, M. (2013). Assessing the impacts of livestock production on biodiversity in rangeland ecosystems. *Proceedings of the National Academy of Sciences, 110*, 20900-20905

Allen, R.G., Pereira, L.S., Raes, D., & Smith, M. (1998). Crop evapotranspiration. In. FAO irrigation and drainage paper no. 56. FAO, Rome

Altman, N.S. (1990). An introduction to kernel and nearest-neighbor nonparametric regression. *The American Statistician, 46*, 175-185

Anyamba, A., Small, J., Tucker, C., & Pak, E. (2014). Thirty-two years of sahelian zone growing season non-stationary ndvi3g patterns and trends. *Remote Sensing, 6*, 3101-3122

Anyamba, A., & Tucker, C.J. (2005). Analysis of sahelian vegetation dynamics using noaa-avhrr ndvi data from 1981–2003. *Journal of Arid Environments, 63*, 596-614

Asrar, G., Fuchs, M., Kanemasu, E.T., & Hatfield, J.L. (1984). Estimating absorbed photosynthetic radiation and leaf area index from spectral reflectance in wheat. *Agro., 1*, 300-306

Asrar, G., Kanemasu, E.T., Jackson, R.D., & Pinter Jr, P.J. (1985). Estimation of total above-ground phytomass production using remotely sensed data. *Remote Sensing of Environment, 17*, 211-220

Atkinson, P.M., Jeganathan, C., Dash, J., & Atzberger, C. (2012). Inter-comparison of four models for smoothing satellite sensor time-series data to estimate vegetation phenology. *Remote Sensing of Environment, 123*, 400-417

Atzberger, C. (2013). Advances in remote sensing of agriculture: Context description, existing operational monitoring systems and major information needs. *Remote Sensing, 5*, 949-981

Baret, F., & Guyot, G. (1991). Potentials and limits of vegetation indices for lai and apar assessment. *Remote Sensing of Environment, 35*, 161-173

Baret, F., Guyot, G., & Major, D.J. (1989). Tsavi: A vegetation index which minimizes soil brightness effects on lai and apar estimation. *Proceedings of the 12th Canadian Symposium on Remote Sensing, Vancouver, Canada, July 10 – 14*, 1355 – 1358

Baret, F., Hagolle, O., Geiger, B., Bicheron, P., Miras, B., Huc, M., Berthelot, B., Niño, F., Weiss, M., Samain, O., Roujean, J.L., & Leroy, M. (2007). Lai, fapar and fcover cyclopes global products derived from vegetation. *Remote Sensing of Environment, 110*, 275-286

Baret, F., Weiss, M., Lacaze, R., Camacho, F., Makhmara, H., Pacholcyzk, P., & Smets, B. (2013). Geov1: Lai and fapar essential climate variables and fcover global time series capitalizing over existing products. Part1: Principles of development and production. *Remote Sensing of Environment, 137*, 299-309

Bartholomé, E., & Belward, A.S. (2005). Glc2000: A new approach to global land cover mapping from earth observation data. *International Journal of Remote Sensing, 26*, 1959-1977

Baup, F. (2007). Apport des données envisat/asar pour le suivi des surfaces continentales : Application à la zone sahélienne. In, *Télédétection spatiale radar* (p. 202): UNIVERSITE TOULOUSE III - PAUL SABATIER

Bégué, A. (1993). Leaf area index, daily intercepted par and spectral vegetation indices: A sensitivity analysis for regular-clumped canopies. *Remote Sensing of Environment, 46*, 1-25

Bégué, A. (2002). Télédétection et production végétale. *Habilitation à Diriger des Recherches, Université Pierre et Marie Curie, France*, 112 pages

Bégué, A., Vintrou, E., Ruelland, D., Claden, M., & Dessay, N. (2011). Can a 25-year trend in soudano-sahelian vegetation dynamics be interpreted in terms of land use change? A remote sensing approach. *Global Environmental Change, 21*, 413-420

Begue, A., Vintrou, E., Saad, A., & Hiernaux, P. (2014). Differences between cropland and rangeland modis phenology (start-of-season) in mali. *International Journal of Applied Earth Observation and Geoinformation, 31*, 167-170

Belsley, D.A., Kuh, E., & Welsh, R.E. (1980). Regression diagnostics: Identifying influential data and sources of collinearity. *New York, Wiley*

Bénié, G.B., Kaboré, S.S., Goïta, K., & Courel, M.F. (2005). Remote sensing-based spatio-temporal modeling to predict biomass in sahelian grazing ecosystem. *Ecological Modelling, 184*, 341-354

Bille, J.C. (1975). Analyse mathématique des relevés de végétation en zone sahélienne. In, *Premier colloque sur l'inventaire et la cartographie des pâturages tropicaux africains* (pp. 333-334). Bamako, 3 au 8 mars 1975

Bille, J.C. (1977). Etude de la production primaire nette d'un écosystème sahélien. Travaux et documents de l'orstom. 85 pages.

Birth, G.S., & McVey, G. (1968). Measuring the color of growing turf with a reflectance spectroradiometer. *Agronomy Journal, 60*, 640-643

Bishop, C.M. (1995). Neural networks for pattern recognition. In: Oxford University Press, Inc., p482

Borra, S., & Di Ciaccio, A. (2010). Measuring the prediction error. A comparison of cross-validation, bootstrap and covariance penalty methods. *Computational Statistics & Data Analysis, 54*, 2976-2989

Boser, B.E., Guyon, I.M., & Vapnik, V.N. (1992). A training algorithm for optimal margin classifiers. In D. Haussler (Ed.), *Proceedings of the 5th Annual ACM Workshop on Computational Learning Theory* (pp. 144--152): ACM Press

Boudet, G. (1984). Recherche d'un équilibre entre production animale et ressources fourragères au sahel. *Bulletin de la société languedocienne de géographie, 18*, 167-178

Box, E.O., Holben, B.N., & Kalb, V. (1989). Accuracy of the avhrr vegetation index as a predictor of biomass, primary productivity and net co2 flux. *Vegetatio, 80*, 71-89

Brandt, M., Hiernaux, P., Tagesson, T., Verger, A., Rasmussen, K., Diouf, A.A., Mbow, C., Mougin, E., & Fensholt, R. (2016a). Woody plant cover estimation in drylands from earth observation based seasonal metrics. *Remote Sensing of Environment, 172*, 28-38

Brandt, M., Mbow, C., Diouf, A.A., Verger, A., Samimi, C., & Fensholt, R. (2015). Ground and satellite based evidence of the biophysical mechanisms behind the greening sahel. *Global Change Biology*, 1-11

Brandt, M., Tappan, G., Diouf, A.A., Beye, G., Mbow, C., & Fensholt, R. (2016b). Woody vegetation regeneration and die off in response to rainfall variability in the west african sahel (senegal). *Journal of Biogeography*, JBI-16-0269 (Under review)

Brandt, M., Verger, A., Diouf, A.A., Baret, F., & Samimi, C. (2014). Local vegetation trends in the sahel of mali and senegal using long time series fapar satellite products and field measurement (1982–2010). *Remote Sensing, 6*, 2408-2434

Breiman, L., Friedman, J., Ohlsen, R., & Stone, C. (1984). *Classification and regression trees*. Wadsworth, Monterey, CA, 1984

Breman, H., & Cissé, A.M. (1977). Dynamics of sahelian pastures in relation to drought and grazing. *Oecologia, 28*, 301-315

Breman, H., & de Ridder, N. (1991). Manuel sur les pâturages des pays sahéliens. In Karthala (Ed.) (p. 481). Paris, CTA Wageningen

Breshears, D., & Barnes, F. (1999). Interrelationships between plant functional types and soil moisture heterogeneity for semiarid landscapes within the grassland/forest continuum: A unified conceptual model. *Landscape Ecology, 14*, 465-478

Brooks, N. (2004). Drought in the african sahel: Long-term perspectives and future prospects. Tyndall centre working paper no. 61. Available from www.Tyndall.Ac.Uk.

Brosofske, K.D., Froese, R.E., Falkowski, M.J., & Banskota, A. (2014). A review of methods for mapping and prediction of inventory attributes for operational forest management. *Forest Science, 60*, 733-756

Bruegel, M., & Stanziani, A. (2004). Pour une histoire de la « sécurité alimentaire ». *Belin | Revue d'histoire moderne et contemporaine, 3 - no51-3*, 7 à 16

Carrière, M. (1996). Impact des systèmes d'élevage pastoraux sur l'environnement en afrique et en asie tropicale et subtropicale aride et subaride. In. Elevage et environnement à la recherche d'un équilibre, CIRAD-EMVT. p70

Cesaro, J.-D., Magrin, G., & Ninot, O. (2010). Atlas de l'élevage au sénégal : Commerce et territoires. In. Projet de recherche ATP Icare. CIRAD. p36

Chantarat, S., Mude, A.G., Barrett, C.B., & Carter, M.R. (2009). Designing index based livestock insurance for managing asset risk in nothern kenya In (p. 40. mimeo)

Chen, J., Jönsson, P., Tamura, M., Gu, Z., Matsushita, B., & Eklundh, L. (2004). A simple method for reconstructing a high-quality ndvi time-series data set based on the savitzky–golay filter. *Remote Sensing of Environment, 91*, 332–344

Chen, J.M., & Black, T.A. (1992). Defining leaf area index for non-flat leaves. *Plant, Cell & Environment, 15*, 421-429

Chipman, H.A., George, E.I., & McCulloch, R.E. (2010). Bart: Bayesian additive regression trees. *Annals of Applied Statistics, 4*, 266-298

Cissé, M.I. (1980). The browse production of some trees of the sahel: Relationships between maximum foliage biomass and various physical parameters. In H.N.L. Houerou (Ed.), *Browse in africa* (pp. 205-210). Addis Ababa, AA, Ethiopia: ILCA

Clevers, J.G.P.W. (1988). The derivation of a simplified reflectance model for the estimation of leaf area index. *Remote Sensing of Environment, 25*, 53-69

Climate Hazard Group (2015). Rainfall estimate and potential evapotranspiration data. Available online. In. ftp://ftp.chg.ucsb.edu/pub/org/chg/products/geowrsi/archives/ (accessed on 18 December 2015)

Cole, S.R. (1999). Simple bootstrap statistical inference using the sas system. *Computer Methods and Programs in Biomedicine, 60*, 79-82

Colombo, R., Busetto, L., Fava, F., Di Mauro, B., Migliavacca, M., Cremonese, E., Galvagno, M., Rossini, M., Meroni, M., Cogliati, S., Panigada, C., Siniscalco, C., & di Cella, U.M. (2011). Phenological monitoring of grassland and larch in the alps from terra and aqua modis images. *Italian Journal of Remote Sensing, 43*, 83-96

Confais, J., & Le Guen, M. (2006). Premiers pas en régression linéaire avec sas. *Modulad, no35*, 220-363

Cornet, A. (1984). Utilisation de modéles simples de bilan hydrique et de production de biomasse pour déterminer les potentialités de production de parcours en zone sahélienne sénégalaise. In W. Siderius (Ed.), *Proceedings of the workshop on land evaluation for extensive grazing* (pp. 207–228): ILRI, Wageningen

Cover, T., & Hart, P. (1967). Nearest neighbor pattern classification. *IEEE Transactions on Information Theory, 13*, 21-27

CSE (2010). Rapport sur l'état de l'environnement au sénégal. In, *Ministère de l'environnement et de la protection de la nature* (p. 266): Centre de Suivi Ecologique

CSE (2013). Suivi de la production végétale 2013 au sénégal. In. Rapport National, Centre de Suivi Ecologique de Dakar, Senegal

CSE (2014). Suivi de la production végétale 2014 au sénégal. In. Rapport National, Centre de Suivi Ecologique de Dakar, Senegal

Daget, P., & Godron, M. (1995). *Pastoralisme: Troupeaux, espaces et sociétés*: Universités francophones, Hatier, Evreux. 510 p

Dardel, C., Kergoat, L., Hiernaux, P., Mougin, E., Grippa, M., & Tucker, C.J. (2014). Re-greening sahel: 30years of remote sensing data and field observations (mali, niger). *Remote Sensing of Environment, 140*, 350-364

De Benedetto, D., Castrignanò, A., & Quarto, R. (2013). A geostatistical approach to estimate soil moisture as a function of geophysical data and soil attributes. *Procedia Environmental Sciences, 19*, 436-445

Deering, D.W., J.W. Rouse, R.H. Haas, & J.A. Schell (1975). Measuring "forage production" of grazing units from landsat mss data. In, *Proceedings of the 10th International Symposium on Remote Sensing of Environment* (pp. 1169-1178)

Demarty, J., Chevallier, F., Friend, A.D., Viovy, N., Piao, S., & Ciais, P. (2007). Assimilation of global modis leaf area index retrievals within a terrestrial biosphere model. *Geophysical Research Letters, 34*, n/a-n/a

DEPA (2012). Direction de l'elevage et des productions animales : Bilan provisoire de l'opération sauvegarde du bétail (osb) 2012. Available online. In: http://csa.sn/site/index.php?option=com_phocadownload&view=category&download=86:amr-2012-contribution-elevage-osb-bilan-provisoirepdf&id=3:rapports-comptes-rendus&Itemid=61&start=12. (Accessed on 28 March 2015)

Desbois, D. (1999). Introduction to partial least square regression with pls procedure in sas. *Modulad, no24*, 41-97

Devineau, J.-L. (1999). Seasonal rhythms and phenological plasticity of savanna woody species in a fallow farming system (south-west burkina faso). *Journal of Tropical Ecology, 15*, 497-513

Di Bella, C.M. (2002). Utilisation de données satellitaires couplées a un modèle de fonctionnement pour l'évaluation de la production prairiale à l'échelle nationale. In, *Unité Climat-Sol-Environnement* (p. 138): Institut National de Recherche Agronomique

Diallo, O., Diouf, A., Hanan, N.P., Ndiaye, A., & PrÉVost, Y. (1991). Avhrr monitoring of savanna primary production in senegal, west africa: 1987-1988. *International Journal of Remote Sensing, 12*, 1259-1279

Dicko, M.S., Djitèye, M.A., & Sangaré, M. (2006). Les systèmes de production animale au sahel. *Sécheresse, 17*, 83-97

Dinku, T., Ceccato, P., Grover-Kopec, E., Lemma, M., Connor, S.J., & Ropelewski, C.F. (2007). Validation of satellite rainfall products over east africa's complex topography. *International Journal of Remote Sensing, 28*, 1503-1526

Diouf, A., Brandt, M., Verger, A., Jarroudi, M., Djaby, B., Fensholt, R., Ndione, J., & Tychon, B. (2015). Fodder biomass monitoring in sahelian rangelands using phenological metrics from fapar time series. *Remote Sensing, 7*, 9122-9148

Diouf, A., & Lambin, E.F. (2001). Monitoring land-cover changes in semi-arid regions: Remote sensing data and field observations in the ferlo, senegal. *Journal of Arid Environments, 48*, 129-148

Diouf, A., Sall, M., Wélé, A., & Dramé, M. (1998). Méthode d'échantillonnage de la production primaire sur le terrain. *Document technique du Centre de suivi écologique de Dakar*, 9

Diouf, A.A., Djaby, B., Diop, M.B., Wele, A., Ndione, J.-A., & Tychon, B. (2014). Simple regression functions for estimating herbaceous fodder production in senegal rangelands from the s10 ndvi of spot-vegetation. In, *Proceedings of the 27th Symposium of the International Association of Climatology* (pp. 284-289). Dijon, France, 2-5 Juillet

Diouf, A.A., Hiernaux, P., Brandt, M., Faye, G., Djaby, B., Diop, M.B., Ndione, J.A., & Tychon, B. (2016). Do agrometeorological data improve optical satellite-based estimations of herbaceous yield in sahelian semi-arid ecosystems? . *Remote Sensing, 8 (Under Review)*

Djitèye, M.A., & Penning de Vries, F.W.T. (1991). *La productivité des pâturages sahéliens : Une étude des sols, des végétations et de l'exploitation de cette ressource naturelle*. Wageningen: Pudoc

Domingos, P. (2012). A few useful things to know about machine learning. *Commun. ACM, 55*, 78-87

Doorenbos, J., & Pruitt, W.O. (1977). Guidelines for predicting crop water requirements. In. FAO Irrigation and Drainage paper No. 24. FAO, Rome

du Plessis, W.P. (1999). Linear regression relationships between ndvi, vegetation and rainfall in etosha national park, namibia. *Journal of Arid Environments, 42*, 235-260

Efron, B. (2004). The estimation of prediction error: Covariance penalties and cross-validation. *Journal of American Statistical Association, 99*, 619-632

Eklundh, L., & Jönsson, P. (2011). Timesat 3.1 manuel software. In, *Lund University: Lund, Sweden* (p. 79)

Eklundh, L., & Jönsson, P. (2015). Timesat 3.2 manuel software. In, *Lund University: Lund, Sweden* (p. 88)

Fan, W., Liu, Y., Xu, X., Chen, G., & Zhang, B. (2014). A new fapar analytical model based on the law of energy conservation: A case study in china. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 7*, 3945-3955

FAO (1994). *Digital soil map of the world. Cdrom*. Food and Agriculture Organization of the United Nations: Rome, Italy

FAO (2000). Manuel de détermination et de mise en place d'un système d'information pour la sécurité alimentaire et l'alerte rapide (sisaar). Division de l' analyse du developpement agricole et economique. P 135

FAO (2009a). Global land cover network: Senegal land cover mapping. Available online. In. http://www.glcn.org/databases/se_landcover_en.jsp (Accessed on 02-11-2015)

FAO (2009b). Land cover changes: Senegal 1990-2005. Available online. In. http://www.glcn.org/downs/prj/senegal/Sen_lc_change_report_dec08.pdf (Accessed on 02-11-2015)

FAO (2012). Executive brief of the food and agriculture organization of the united nations: The sahel crisis 2012. In: Available online: http://www.fao.org/fileadmin/user_upload/sahel/docs/EXECUTIVE%20BRIEF%20TCE%206%20July.pdf (Accessed on 18 April 2015)

FAOSTAT (2016). Faostat: Http://faostat3.Fao.Org/ (accessed on 11/03/2016). In: Food and Agriculture Organization of the United Nations: Statistics Division

Faye, G. (2013). Complémentarité des capteurs radar pour le suivi de la végtation et de l'humidité du sol en zone semi-aride, application à la zone sahélienne : Cas du ferlo au sénégal. In, *Télédétection spatiale de la Biosphère* (p. 142): Université Cheikh Anta Diop

Fensholt, R., Anyamba, A., Stisen, S., Sandholt, I., Pak, E., & Small, J. (2007). Comparisons of compositing period length for vegetation index data from polar-orbiting and geostationary satellites for the cloud-prone region of west africa. *Photogrammetric Engineering & Remote Sensing, 73*, 297-309

Fensholt, R., Sandholt, I., & Rasmussen, M.S. (2004). Evaluation of modis lai, fapar and the relation between fapar and ndvi in a semi-arid environment using in situ measurements. *Remote Sensing of Environment, 91*, 490-507

Fensholt, R., Sandholt, I., Rasmussen, M.S., Stisen, S., & Diouf, A. (2006). Evaluation of satellite based primary production modelling in the semi-arid sahel. *Remote Sensing of Environment, 105*, 173–188

FEWS NET (2015). Fews data downloads. Available online. In. http://earlywarning.usgs.gov/fews/datadownloads (accessed on 18 December 2015)

Field, C.B., Randerson, J.T., & Malmström, C.M. (1995). Global net primary production: Combining ecology and remote sensing. *Remote Sensing of Environment, 51*, 74-88

Fitzpatrick, R.G.J., Bain, C.L., Knippertz, P., Marsham, J.H., & Parker, D.J. (2015). The west african monsoon onset: A concise comparison of definitions. *Journal of Climate, 28*, 8673-8694

Foody, G.M., Boyd, D.S., & Cutler, M.E.J. (2003). Predictive relations of tropical forest biomass from landsat tm data and their transferability between regions. *Remote Sensing of Environment, 85*, 463-474

Fraser, R.H., & Li, Z. (2002). Estimating fire-related parameters in boreal forest using spot vegetation. *Remote Sensing of Environment, 82*, 95-110

Frère, M., & Popov, G. (1979). Agrometeorological crop monitoring and forecasting. In: FAO plant production and protection paper No.17. FAO, Rome

Funk, C.C., & Brown, M.E. (2006). Intra-seasonal ndvi change projections in semi-arid africa. *Remote Sensing of Environment, 101*, 249-256

GCOS (2006). Systematic observation requirements for satellite-based products for climate supplemental details to the satellite-based component of the implementation plan for the global observing system for climate in support of the unfccc*, 107*, p 103

Glantz, M. (1976). The politics of natural disaster. Praeger publishers, new york

Gobron, N., Pinty, B., Mélin, F., Taberner, M., Verstraete, M.M., Robustelli, M., & Widlowski, J.-L. (2007). Evaluation of the meris/envisat fapar product. *Advances in Space Research, 39*, 105-115

Goward, S.N., & Huemmrich, K.F. (1992). Vegetation canopy par absorptance and the normalized difference vegetation index: An assessment using the sail model. *Remote Sensing of Environment, 39*, 119 – 140

Grainger, A. (2013). Controlling tropical deforestation. In: Taylor & Francis

Graves, S., Asner, G., Martin, R., Anderson, C., Colgan, M., Kalantari, L., & Bohlman, S. (2016). Tree species abundance predictions in a tropical agricultural landscape with a supervised classification model and imbalanced data. *Remote Sensing, 8*, 161

Günlü, A., Ercanli, I., Başkent, E.Z., & Çakır, G. (2014). *Estimating aboveground biomass using landsat tm imagery: A case study of anatolian crimean pine forests in turkey*

Guyon, I., Weston, J., Barnhill, S., & Vapnik, V. (2002). Gene selection for cancer classification using support vector machines. *Machine Learning, 46*, 389-422

Ham, F., & Fillol, E. (2010). Pastoral surveillance system and feed inventory in the sahel. In conducting national feed assessments. *FAO: Rome, Italy, 1998*, 83–114

Hanan, N.P., Prince, S.D., & Bégué, A. (1995). Estimation of absorbed photosynthetically active radiation and vegetation net production efficiency using satellite data. *Agricultural and Forest Meteorology, 76*, 259-276

Hanes, J.E. (2014). *Biophysical applications of satellite remote sensing*. Springer Remote Sensing/Photogrammetry: Springer

Hanes, J.M., Richardson, A.D., & Klostermann, S. (2013). Mesic temperate deciduous forest phenology In M.D. Schwartz (Ed.), *Phenology: An integrative environmental science* (pp. 211-224). New York, NY, USA: Springer

Hastie, T., Tibshirani, R., & J., F. (2009). *The elements of statistical learning: Data mining, inference, and prediction*. Springer-Verlag, NY, USA

Herman, A., Kumar, V.B., Arkin, P.A., & Ko usky, J.V. (1997). Objectively determined 10-day african rainfall estimates created for famine early warning systems. *Int. J. Remote Sensing, 18*, 2147-2159

Herrero, M., & Thornton, P.K. (2013). Livestock and global change: Emerging issues for sustainable food systems. *Proceedings of the National Academy of Sciences, 110*, 20878-20881

Herrmann, S., Wickhorst, A., & Marsh, S. (2013). Estimation of tree cover in an agricultural parkland of senegal using rule-based regression tree modeling. *Remote Sensing, 5*, 4900-4918

Hickler, T., Eklundh, L., Seaquist, J.W., Smith, B., Ardö, J., Olsson, L., Sykes, M.T., & Sjöström, M. (2005). Precipitation controls sahel greening trend. *Geophysical Research Letters, 32, L21415*

Hiernaux, P. (1983). Distribution des pluies et production herbacée au sahel:Une méthode empirique pour caractériser la distribution des précipitations journalières et ses effets sur la production herbacée. Premiers résultats acquis dans le sahel malien. In: Doc. Prog. N° AZ 98, CIPEA, Bamako.

Hiernaux, P., Ayantunde, A., Kalilou, A., Mougin, E., Gérard, B., Baup, F., Grippa, M., & Djaby, B. (2009a). Trends in productivity of crops, fallow and rangelands in southwest niger: Impact of land use, management and variable rainfall. *Journal of Hydrology, 375*, 65-77

Hiernaux, P., & Le Houérou, H.N. (2006). The rangelands of the sahel. *Sécheresse, 17*, 51-71

Hiernaux, P., Mougin, E., Diarra, L., Soumaguel, N., Lavenu, F., Tracol, Y., & Diawara, M. (2009b). Sahelian rangeland response to changes in rainfall over two decades in the gourma region, mali. *Journal of Hydrology, 375*, 114-127

Hiernaux, P.H. (1980). Inventory of the browse potential of bushes, trees and shrubs in an area of the sahel in mali: Method and initial results. In H.N.L. Houerou (Ed.), *Browse in africa* (pp. 197-210). Addis Ababa, AA, Ethiopia: ILCA

Hiernaux, P.H.Y., Cissé, M.I., Diarra L., & De Leeuw, P.N. (1994). Fluctuations saisonnières de la feuillaison des arbres et des buissons sahéliens. Conséquences pour la quantification des ressources fourragères. *Revue d'élevage et de médecine vétérinaire des pays tropicaux, 47*, 117-125

Hiernaux, P.H.Y., & Justice, C.O. (1986). Suivi du développement végétal au cours de l'été 1984 dans le sahel malien. *International Journal of Remote Sensing, 7*, 1515-1531

Hothorn, T. (2016). Cran task view: Machine learning & statistical learning. In. https://cran.r-project.org/web/views/MachineLearning.html (Accessed on 13/03/2016)

Huber, S., & Fensholt, R. (2011). Analysis of teleconnections between avhrr-based sea surface temperature and vegetation productivity in the semi-arid sahel. *Remote Sensing of Environment, 115*, 3276-3285

Huber, S., Fensholt, R., & Rasmussen, K. (2011). Water availability as the driver of vegetation dynamics in the african sahel from 1982 to 2007. *Global and Planetary Change, 76*, 186-195

Huete, A., Didan, K., Miura, T., Rodriguez, E.P., Gao, X., & Ferreira, L.G. (2002). Overview of the radiometric and biophysical performance of the modis vegetation indices. *Remote Sensing of Environment, 83*, 195-213

Huete, A.R. (1988). A soil-adjusted vegetation index (savi). *Remote Sensing of Environment, 25*, 295 – 309

Ibrahim, Y., Balzter, H., Kaduk, J., & Tucker, C. (2015). Land degradation assessment using residual trend analysis of gimms ndvi3g, soil moisture and rainfall in sub-saharan west africa from 1982 to 2012. *Remote Sensing, 7*, 5471-5494

Ikegami, M., Barrett, C.B., & Chantarat, S. (2012). Dynamic effects of index based livestock insurance on household intertemporal behavior and welfare. In, *Research Conference on Microinsurance* (p. 30). University of Twenty, Institute for Innovation and Governace Studies

Immunetrics (2016). Immunetrics-mechanistic modeling. Available online. In. http://immunetrics.com/technology/mechanistic-modeling.php (Accessed on 11/03/2016)

ISRA (2003). Rapport national sur l'état des ressources zoogénétiques au sénégal. In: Ministère de l'agriculture et de l'élevage: Institut Sénégalais de Recherches Agricoles. 54 pages.

Ivits, E., Cherlet, M., Horion, S., & Fensholt, R. (2013). Global biogeographical pattern of ecosystem functional types derived from earth observation data. *Remote Sensing, 5*, 3305-3330

Jackson, R.D., & Huete, A.R. (1991). Interpreting vegetation indices

Preventive Veterinary Medicine, 11, 185-200

Jacques, D.C., Kergoat, L., Hiernaux, P., Mougin, E., & Defourny, P. (2014). Monitoring dry vegetation masses in semi-arid areas with modis swir bands. *Remote Sensing of Environment, 153*, 40-49

Jarlan, L., Mangiarotti, S., Mougin, E., Mazzega, P., Hiernaux, P., & Ledantec, V. (2008). Assimilation of spot/vegetation ndvi data into a sahelian vegetation dynamics model. *Remote Sensing of Environment, 112*, 1381-1394

Jin, Y., Yang, X., Qiu, J., Li, J., Gao, T., Wu, Q., Zhao, F., Ma, H., Yu, H., & Xu, B. (2014). Remote sensing-based biomass estimation and its spatio-temporal variations in temperate grassland, northern china. *Remote Sensing, 6*, 1496-1513

Jobard, I., Chopin, F., Berges, J.C., & Roca, R. (2011). An intercomparison of 10-day satellite precipitation products during west african monsoon. *International Journal of Remote Sensing, 32*, 2353–2376

Johnson, C.J., Nielsen, S.E., Merrill, E.H., McDonald, T.E., & Boyce, M.S. (2006). Resource selection functions based on use-availability data: Theoretical motivation and evaluation methods. *J. Wildl. Manage, 70*, 347-357

Jönsson, P., & Eklundh, L. (2004). Timesat– a program for analyzing time-series of satellite sensor data. *Computers & Geosciences, 30*, 833-845

Justice, C.O., & Hiernaux, P.H.Y. (1986). Monitoring the grasslands of the sahel using noaa avhrr data: Niger 1983. *International Journal of Remote Sensing, 7*, 1475-1497

Kandasamy, S., Baret, F., Verger, A., Neveux, P., & Weiss, M. (2013). A comparison of methods for smoothing and gap filling time series of remote sensing observations: Application to modis lai products. . *Biogeosciences, 10*, 4055-4071

Ke, Y., Im, J., Park, S., & Gong, H. (2016). Downscaling of modis one kilometer evapotranspiration using landsat-8 data and machine learning approaches. *Remote Sensing, 8*, 215

Knyazikhin, Y., Martonchik, J.V., Myneni, R.B., Diner, D.J., & Running, S.W. (1998). Synergistic algorithm for estimating vegetation canopy leaf

area index and fraction of absorbed photosynthetically active radiation from modis and misr data. *Journal of Geophysical Research: Atmospheres (1984–2012), 103*, 32257-32275

Knyazikhin, Y.V., Marshak, A.L., & Myneni, R.B. (2004). *Three-dimensional radiative transfer in vegetation canopies*. in Three-Dimensional Radiative Transfer in the Cloudy Atmosphere: Springer-Verlag, New York

Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection *International Joint Conference on Artificial Intelligence, 14*, 1137-1145

Kouadio, L., Djaby, B., Duveiller, G., El Jarroudi, M., & Tychon, B. (2012). Decay kinetics of the green area and yield estimation of winter wheat. *Biotechnologie, Agronomie, Société et Environnement, 5*, 179-191

Kuhn, M., & Johnson, K. (2013). *Applied predictive modeling*. Springer: New York, NY, USA.

Kuhn, M., Weston, S., Keefer, C., & Coulter, N. (2012). Cubist models for regression. Available online. In (pp. http://cran.r-project.org/web/packages/Cubist/vignettes/cubist.pdf (Accessed on 25-06-2015))

Kuhn, M., Wing, J., Weston, S., Williams, A., Keefer, C., Engelhardt, A., Cooper, T., Mayer, Z., & the R Core Team (2014). Caret: Classification and regression training. R package version 6.0-24. Available online. In. http://CRAN.R-project.org/package=caret (Accessed on 25-06-2015)

Kumar, M., & Monteith, J.L. (1981). Remote sensing of crop growth. In S. H. (Ed.), *Plants and the daylight spectrum* (pp. 133-144)

Lacaze, R., Smets, B., Calvet, J.C., Camacho, F., Tansey, K., Baret , F., Ramon, D., Montersleet, B., Roujean, J.L., Wandrebek, L., & et al. (2015). Sentinel-3 for the copernicus global land service: Monitoring the continental ecosystems at global scale In, *Sentinel-3 for Science workshop*. Venice (Italy).

Le Houerou, H.N. (1989). *The grazing land ecosystems of the african sahel*: Heidelberg: Springer-Verlag

Le Roux, X., Gauthier, H., Bégué, A., & Sinoquet, H. (1997). Radiation absorption and use by humid savannah grassland, assessment using remote sensing and modelling. *Agricultural and Forest Meteorology, 85*, 117 – 132

Leroux, M. (1995). La dynamique de la grande sécheresse sahélienne / dynamics of the great sahelian drought. *Revue de géographie de Lyon*, 223-232

Li, L., Du, Y., Tang, Y., Xin, X., Zhang, H., Wen, J., & Liu, Q. (2015). A new algorithm of the fpar product in the heihe river basin considering the contributions of direct and diffuse solar radiation separately. *Remote Sensing, 7*, 6414

Liaw, A., & Wiener, M. (2002). Classification and regression by randomforest. *R News, 2*, 18--22.

Lind, M., & Fensholt, R. (1999). The spatio-temporal relationship between rainfall and vegetation development in burkina faso. *Geografisk Tidsskrift/Danish Journal of Geography*, 43-55

Liu, H.Q., & Huete, A.R. (1995). A feedback based modification of the ndvi to minimize canopy background and atmospheric noise. *IEEE Trans. Geosci. Remote Sens., 33*, 457-465

Lo Seen, D., Mougin, E., Rambal, S., Gaston, A., & Hiernaux, P. (1995). A regional sahelian grassland model to be coupled with multispectral satellite data. Ii: Toward the control of its simulations by remotely sensed indices. *Remote Sensing of Environment, 52*, 194-206

Lu, D. (2005). Aboveground biomass estimation using landsat tm data in the brazilian amazon. *International Journal of Remote Sensing, 26*, 2509-2525

Lumley, T., Diehr, P., Emerson, S., & Chen, L. (2002). The importance of the normality assumption in large public health datasets. *Annual Review of Public Health, 23*, 151-169

Magrin, G. (2008). Dynamiques territoriales et place de l'élevage au sénégal. In (p. 22). Documents de travail Icare: Cirad, Montpellier

Maignen, R. (1965). Notice explicative, carte pédologique du sénégal au 1/1.000.000. In (p. 63): Publication ORSTOM

Marill, T., & Green, D.M. (1963). On the effectiveness of receptors in recognition systems. *Information Theory, IEEE Transactions on, 9*, 11-17

Marteau, R., Moron, V., & Philippon, N. (2009). Spatial coherence of monsoon onset over western and central sahel (1950-2000). *J. Clim., 22*, 1313-1324

Mbow, C., Fensholt, R., Nielsen, T.T., & Rasmussen, K. (2014). Advances in monitoring vegetation and land use dynamics in the sahel. *Geografisk Tidsskrift-Danish Journal of Geography, 114*, 84-91

Mbow, C., Fensholt, R., Rasmussen, K., & Diop, D. (2013). Can vegetation productivity be derived from greenness in a semi-arid environment? Evidence from ground-based measurements. *Journal of Arid Environments, 97*, 56-65

McCallum, I., Wagner, W., Schmullius, C., Shvidenko, A., Obersteiner, M., Fritz, S., & Nilsson, S. (2010). Comparison of four global fapar datasets over northern eurasia for the year 2000. *Remote Sensing of Environment, 114*, 941-949

Mehmood, T., Liland, K.H., Snipen, L., & Sæbø, S. (2012). A review of variable selection methods in partial least squares regression. *Chemometrics and Intelligent Laboratory Systems, 118*, 62-69

Meroni, M., Fasbender, D., Kayitakire, F., Pini, G., Rembold, F., Urbano, F., & Verstraete, M.M. (2014a). Early detection of biomass production deficit hot-spots in semi-arid environment using fapar time series and a probabilistic approach. *Remote Sensing of Environment, 142*, 57-68

Meroni, M., Marinho, E., Sghaier, N., Verstrate, M., & Leo, O. (2013). Remote sensing based yield estimation in a stochastic framework — case study of durum wheat in tunisia. *Remote Sensing, 5*, 539-557

Meroni, M., Rembold, F., Verstraete, M., Gommes, R., Schucknecht, A., & Beye, G. (2014b). Investigating the relationship between the inter-annual variability of satellite-derived vegetation phenology and a proxy of biomass production in the sahel. *Remote Sensing, 6*, 5868-5884

Mitchell, T.M. (2006). The discipline of machine learning. In (p. 9): http://www.cs.cmu.edu/~tom/pubs/MachineLearning.pdf (Accessed on 13/03/2016)

Monteith, J. (1972). Solar radiation and productivity in tropical ecosystems. *J. Applied Ecology, 19*, 747-766

Mora, C., Frazier, A.G., Longman, R.J., Dacks, R.S., Walton, M.M., Tong, E.J., Sanchez, J.J., Kaiser, L.R., Stender, Y.O., Anderson, J.M., Ambrosino, C.M., Fernandez-Silva, I., Giuseffi, L.M., & Giambelluca, T.W. (2013). The projected timing of climate departure from recent variability. *Nature, 502*, 183-187

Moron, V., Robertson, A.W., & Ward, M.N. (2006). Seasonal predictability and spatial coherence of rainfall characteristics in the tropical setting of senegal. *Mon. Weather Rev., 134*, 3248-3262

Mougenot, B., Bégué, A., Chehbouni, G., Escadafal, R., Heilman, P., Qi, J., Royer, A., & Watts, C. (2000). Applications of vegetation data to resource management in arid and semi-arid rangelands. *In "VEGETATION-2000", Belgirate (IT)*, 433-434

Mougin, E., Lo Seena, D., Rambal, S., Gaston, A., & Hiernaux, P. (1995). A regional sahelian grassland model to be coupled with multispectral satellite data. I: Model description and validation. *Remote Sensing of Environment, 52*, 181-193

Mróz, M., & Sobieraj, A. (2004). Comparison of several vegetation indices calculated on the basis of a seasonal spot xs time series, and their suitability for land cover and agricultural crop identification. *Tech. Sci, 7*, 39-66

Myneni, R.B., Hall, F.B., Sellers, P.J., & Marshak, A.L. (1995). The interpretation of spectral vegetation indices. *IEEE Transactions on Geoscience and Remote Sensing, 33*, 481 – 486

Myneni, R.B., Hoffman, S., Knyazikhin, Y., Privette, J.L., Glassy, J., Tian, Y., Wang, Y., Song, X., Zhang, Y., Smith, G.R., Lotsch, A., Friedl, M., Morisette, J.T., Votava, P., Nemani, R.R., & Running, S.W. (2002). Global products of vegetation leaf area and absorbed par from year one of modis data. *Remote Sensing of Environment, 83*, 214-231

Neville, P.G. (2013). Controversy of variable importance in random forests. *Journal of Unified Statistical Techniques*, 15-20

Novella, N.S., & Thiaw, W.M. (2013). African rainfall climatology version 2 for famine early warning systems. *JOURNAL OF APPLIED METEOROLOGY AND CLIMATOLOGY, 52*, 588 - 606

O'Reilly, R.C. (2016). Computational cognitive neuroscience laboratory-mechanistic model. Available online. In. https://grey.colorado.edu/oreilly/index.php/Mechanistic_Model (Accessed on 11/03/2016)

Olofsson, P., Eklundh, L., Lagergren, F., Jönsson, P., & Lindroth, A. (2007). Estimating net primary production for scandinavian forests using data from terra/modis. *Advances in Space Research, 39*, 125-130

Olsson, L., Eklundh, L., & Ardö, J. (2005). A recent greening of the sahel – trends, patterns and potential causes. *Journal of Arid Environment, 63*, 556-566

Omotosho, J.B., Balogun, A.A., & Ogunjobi, K. (2000). Predicting monthly and seasonal rainfall, onset and cessation of the rainy season in west africa using only surface data. *International Journal of Climatology, 20*, 865-880

Penning de Vries, F.W.T., & Djitèye, M.A. (1982). La productivité des pâturages sahéliens: Une étude des sols, des végétations et de l'exploitation de cette ressource naturelle. In (p. 525): Centre for Agricultural Publishing and Documentation

Perry Jr, C.R., & Lautenschlager, L.F. (1984). Functional equivalence of spectral vegetation indices. *Remote Sensing of Environment, 14*, 169-182

Pettorelli, N. (2013). *The normalized difference vegetation index (oxford, 2013; pubd online may. 2015). Oxford scholarship online, accessed 08 jun. 2016*

Pettorelli, N., Gaillard, J.-M., Mysterud, A., Duncan, P., Chr. Stenseth, N., Delorme, D., Van Laere, G., Toïgo, C., & Klein, F. (2006). Using a proxy of plant productivity (ndvi) to find key periods for animal performance: The case of roe deer. *Oikos, 112*, 565-572

Pinzon, J., & Tucker, C. (2014). A non-stationary 1981–2012 avhrr ndvi3g time series. *Remote Sensing, 6*, 6929-6960

Potter, C.S., Randerson, J.T., Field, C.B., Matson, P.A., Vitousek, P.M., Mooney, H.A., & Klooster, S.A. (1993). Terrestrial ecosystem production: A process model based on global satellite and surface data. Available online. In: http://www.escholarship.org/uc/item/52z028wn

Powell, S.L., Cohen, W.B., Healey, S.P., Kennedy, R.E., Moisen, G.G., Pierce, K.B., & Ohmann, J.L. (2010). Quantification of live aboveground forest biomass dynamics with landsat time-series and field inventory data: A comparison of empirical modeling approaches. *Remote Sensing of Environment, 114*, 1053-1068

Prince, S.D. (1991). Satellite remote sensing of primary production: Comparison of results for sahelian grasslands 1981-1988. *International Journal of Remote Sensing, 12*, 1301-1311

Prince, S.D., & Goward, S.N. (1995). Global primary production: A remote sensing approach. *Journal of Biogeography, 22*, 815-835

Prince, S.D., Kerr, Y.H., Goutorbe, J.P., Lebel, T., Tinga, A., Bessemoulin, P., Brouwer, J., Dolman, A.J., Engman, E.T., Gash, J.H.C., Hoepffner, M., Kabat, P., Monteny, B., Said, F., Sellers, P., & Wallace, J. (1995). Geographical, biological and remote sensing aspects of the hydrologic atmospheric pilot experiment in the sahel (hapex-sahel). *Remote Sensing of Environment, 51*, 215-234

Proud, S.R., Zhang, Q., Schaaf, C., Fensholt, R., Rasmussen, M.O., Shisanya, C., Mutero, W., Mbow, C., Anyamba, A., Pak, E., & Sandholt, I. (2014). The normalization of surface anisotropy effects present in seviri reflectances by using the modis brdf method. *IEEE Transactions on Geoscience and Remote Sensing, 52*, 6026–6039

Qi, J., Chehbouni, A., Huete, A.R., Kerr, Y.H., & Sorooshian, S. (1994). A modified soil adjusted vegetation index. *Remote Sensing of Environment, 48*, 119-126

Rambal, S. (1980). Modélisation de l'utilisation de l'eau et de la production végétale d'une steppe à ranthérium suaveloens desf de la zone aride sud tunisienne. In (p. 188): Université du Languedoc, Montpellier

Rasmussen, M.S. (1992). Assessment of millet yields and production in northern burkina faso using integrated ndvi from the avhrr. *International Journal of Remote Sensing, 13*, 3431 – 3442

Redelsperger, J.-L., Thorncroft, C.D., Diedhiou, A., Lebel, T., Parker, D.J., & Polcher, J. (2006). African monsoon multidisciplinary analysis: An international research project and field campaign. *Bulletin of the American Meteorological Society, 87*, 1739-1746

Refaeilzadeh, P., Tang, L., & Liu, H. (2009). Cross-validation. In L. Liu & M.T. ÖZsu (Eds.), *Encyclopedia of database systems* (pp. 532-538): Springer US

Richardson, A.D., Anderson, R.S., Arain, M.A., Barr, A.G., Bohrer, G., Chen, G., Chen, J.M., Ciais, P., Davis, K.J., Desai, A.R., Dietze, M.C., Dragoni, D., Garrity, S.R., Gough, C.M., Grant, R., Hollinger, D.Y., Margolis, H.A., McCaughey, H., Migliavacca, M., Monson, R.K., Munger, J.W., Poulter, B., Raczka, B.M., Ricciuto, D.M., Sahoo, A.K., Schaefer, K., Tian, H., Vargas, R., Verbeeck, H., Xiao, J., & Xue, Y. (2012). Terrestrial biosphere models need better representation of vegetation phenology: Results from the north american carbon program site synthesis. *Global Change Biology, 18*, 566-584

Richardson, A.D., Black, T.A., Ciais, P., Delbart, N., Friedl, M.A., Gobron, N., Hollinger, D.Y., Kutsch, W.L., Longdoz, B., Luyssaert, S., Migliavacca, M., Montagnani, L., Munger, J.W., Moors, E., Piao, S., Rebmann, C., Reichstein, M., Saigusa, N., Tomelleri, E., Vargas, R.V., & rlagin, A. (2010). Influence of spring and autumn phenological transitions on forest ecosystem productivity. *Philos. Trans. R. Soc. B-Biol. Sci., 365*, 3227–3246

Richardson, A.J., & Wiegand, C.L. (1977). Distinguishing vegetation from soil background information. *Photogramnetric Engineering and Remote Sensing, 43*, 1541-1552

Rojas, O., Rembold, F., Royer, A., & Negre, T. (2005). Real-time agrometeorological crop yield monitoring in eastern africa. *Agronomy for Sustainable Development, Springer Verlag (Germany), 25*, 63-77

Rouse, J.W.J., Haas, R.H., Deering, D.W., Schell, J.A., & Harlan, J.C. (1974). Monitoring the vernal advancement and retrogradation (green wave effect) of natural vegetation. In (p. 371p): NASA/GSFC Type III Final Report, Greenbelt, MD.

RPCA (2010). L'élevage au sahel et en afrique de l'ouest. In, *26ème réunion annuelle du Réseau de Prévention des Crises Alimentaires (RPCA) Accra (Ghana), 14-16 décembre 2010*: Comité inter-État de lutte contre la sécheresse au Sahel. 10 pages.

Rudorff, B.F.T., & Batista, G.T. (1990). Yield estimation of sugarcane based on agrometeorological-spectral models. *Remote Sensing of Environment, 33*, 183-192

RuleQuest Research (2015). Rulequest research: Data mining with cubist. In. Available online: http://www.rulequest.com/cubist-info.html (accessed on 13 august 2015)

Running, S.W., Nemani, R.N., Glassy, J.M., & Thornton, P.E. (1999). Modis daily photosynthesis (psn) and annual net primary production (npp) product. Mod17 algorithm theoretical basis document. (atbd), version 3. In

Ruppert, D., Wand, M.P., & Carroll, R.J. (2009). Semiparametric regression during 2003–2007. *Electronic journal of statistics, 3*, 1193-1256

Santin-Janin, H., Garel, M., Chapuis, J.L., & Pontier, D. (2009). Assessing the performance of ndvi as a proxy for plant biomass using non-linear models: A case study on the kerguelen archipelago. *Polar Biology, 32*, 861-871

Seaquist, J.W., Olsson, L., & Ardö, J. (2003). A remote sensing-based primary production model for grassland biomes. *Ecological Modelling, 169*, 131-155

Senay, G. (2004). Crop water requirement satisfaction index (wrsi): Model description. In (pp. 7. http://iridl.ldeo.columbia.edu/documentation/usgs/adds/wrsi/WRSI_readme.pdf (Accessed on 17-07-2015))

Senay, G. (2008). Modeling landscape evapotranspiration by integrating land surface phenology and a water balance algorithm. *Algorithms, 1*, 52-68

Senay, G.B., & Verdin, J. (2001). Using a gis-based water balance model to assess regional crop performance. In, *Proceedings of the Fifth International Workshop on Application of Remote Sensing in Hydrology*. October 2-5, 2001, Montpellier, France

Senay, G.B., & Verdin, J. (2002). Evaluating the performance of a crop water balance model in estimating regional crop production. In, *Integrated Remote Sensing at the Global, Regional and Local Scale*. Denver, CO USA: ISPRS Comission I Mid-Term Symposium in conjunction with Pecora 15/Land Satellite Information IV Conference

Senay, G.B., & Verdin, J. (2003). Characterization of yield reduction in ethiopia using a gis-based crop water balance model. *Canadian Journal of Remote Sensing, 29*, 687-692

Senay, G.B., Verdin, J.P., & Rowland, J. (2011). Developing an operational rangeland water requirement satisfaction index. *International Journal of Remote Sensing, 32*, 6047-6053

Shuttleworth, J. (1992). *Evaporation*. In Handbook of Hydrology, Maidment D (ed.). McGraw-Hill: New York; 4.1–4.53.

Silleos, N.G., Alexandridis, T.K., Gitas, I.Z., & Perakis, K. (2006). Vegetation indices: Advances made in biomass estimation and vegetation monitoring in the last 30 years. *Geocarto International, 21*, 21-28

Sivakumar, M.V.K. (1988). Predicting rainy season potential from the onset of rains in southern sahelian and sudanian climatic zones of west africa *Agric. For. Meteorol., 42*, 295-305

Smets, B., Eerens, H., Jacobs, T., & Royer, A. (2010). Dry matter productivity (dmp). *BioPar Product User Manual. Available online:*

http://web.vgt.vito.be/documents/BioPar/g2-BP-RP-BP053-ProductUserManual-DMPV0-I1.00.pdf, (accessed on 27-07-2014)

Spiess, A.-N., & Neumeyer, N. (2010). An evaluation of r(2)as an inadequate measure for nonlinear models in pharmacological and biochemical research: A monte carlo approach. *BMC Pharmacology, 10*, 6-6

Stancioff, A., Staljanssens, M., & Tappan, G. (1986). Mapping and remote sensing of the resources of the republic of senegal: A study of the geology, hydrology, soils, vegetation and land use potential. In, *SDSU, Remote Sensing Institute, SDSU-RSI-86-01* (p. 655)

Statistics Solutions (2013). Statistics solutions. Normality [www document]. Retrieved from http://www.Statisticssolutions.Com/academic-solutions/resources/directory-of-statistical-analyses/normality/ (accessed on 6 october 2016)

Tagesson, T., Fensholt, R., Cropley, F., Guiro, I., Horion, S., Ehammer, A., & Ardö, J. (2015). Dynamics in carbon exchange fluxes for a grazed semi-arid savanna ecosystem in west africa. *Agriculture, Ecosystems & Environment, 205*, 15-24

Tanya, H. (2009). Parametric and nonparametric: Demystifying the terms In (p. 5): http://www.mayo.edu/mayo-edu-docs/center-for-translational-science-activities-documents/berd-5-6.pdf (Accessed on 13/03/2016)

Tappan, G.G., Sall, M., Wood, E.C., & Cushing, M. (2004). Ecoregions and land cover trends in senegal. *Journal of Arid Environments, 59*, 427-462

Tarnavsky, E., Grimes, D., Maidment, R., Black, E., Allan, R., Stringer, M., Chadwick, R., & Kayitakire, F. (2014). Extension of the tamsat satellite-based rainfall monitoring over africa and from 1983 to present. *J. Appl. Meteorol. Climatol., 53*, 2805–2822

Tian, F., Brandt, M., Liu, Y.Y., Verger, A., Tagesson, T., Diouf, A.A., Rasmussen, K., Mbow, C., Wang, Y., & Fensholt, R. (2016). Remote sensing of vegetation dynamics in drylands: Evaluating vegetation optical depth (vod) using avhrr ndvi and in situ green biomass data over west african sahel. *Remote Sensing of Environment, 177*, 265-276

Tian, F., Fensholt, R., Verbesselt, J., Grogan, K., Horion, S., & Wang, Y. (2015). Evaluating temporal consistency of long-term global ndvi datasets for trend analysis. *Remote Sensing of Environment, 163*, 326-340

Tian, Y., Dickinson, R., Zhou, L., Zeng, X., Dai, Y., Myneni, R., Knyazikhin, Y., Zhang, X., Friedl, M., & Yu, H. (2004). Comparison of seasonal and spatial variations of leaf area index and fraction of absorbed photosynthetically active radiation from moderate resolution imaging spectroradiometer (modis) and common land model. *Journal of Geophysical Research: Atmospheres, 109*

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society, Series B, 58*, 267-288

Toté, C., Patricio, D., Boogaard, H., van der Wijngaart, R., Tarnavsky, E., & Funk, C. (2015). Evaluation of satellite rainfall estimates for drought and flood monitoring in mozambique. *Remote Sensing, 7*, 1758-1776

Touré, I., Ickowicz, A., Wane, A., Garba, I., Gerber, P., Atté, I., Cesaro, J.D., Diop, A.T., Djibo, S., Ham, F., Hamadoun, M., Khamis, Y., Niang, I., Saleh, O.M., Métais, T., Saley, M., Sow, N.A., Toutain, B., & Yahaya, S. (2012). *Atlas of trends in pastoral systems in sahel*: FAO and CIRAD

Tracol, Y. (2004). Étude des variations interannuelles de la production herbacée des pâturages sahéliens : Exemple du gourma malien. In, *Ecologie des systèmes continentaux* (p. 270): UNIVERSITE TOULOUSE III - PAUL SABATIER

Tucker, C.J., Vanpraet, C., Boerwinkel, E., & Gaston, A. (1983). Satellite remote sensing of total dry matter production in the senegalese sahel. *Remote Sensing of Environment, 13*, 461-474

Tucker, C.J., Vanpraet, C.L., Sharman, M.J., & Itterstum, G.V. (1985). Satellite remote sensing of total herbaceous biomass production in the senegalese sahel: 1980–1984. *Remote Sensing of Environment, 17*, 233-249

Valenza, J. (1977). Dynamisme de quelques types de pâturages naturels sahélo-soudaniens en république du sénégal. In, *XIIIe International Grassland Congress LEJPZIG 18-27*

van Wijk, M.T., & Williams, M. (2005). Optical instruments for measuring leaf area index in low vegetation: Application in arctic ecosystems. *Ecological Applications, 15*, 1462-1470

Varlet-Grancher, C., Bonhomme, R., Chartier, M., & Artis, P. (1982). Efficience de la conversion de l'énergie solaire par un couvert végétal. *Acta Oecologica, Oecologia Plantarum, 3*, 3-26

Venables, W., & Ripley, B. (2002). Modern applied statistics with s. In. Springer-Verlag, New York. Available online: https://www.stats.ox.ac.uk/pub/MASS4/VR4stat.pdf

Verdin, J., & Klaver, R. (2002). Grid cell based crop water accounting for the famine early warning system. *Hydrological Processes, 16*, 617-1630

Verger, A., Baret , F., & Weiss, M. (2008). Performances of neural networks for deriving lai estimates from existing cyclopes and modis products. *Remote Sensing of Environment, 112*, 2789-2803

Verger, A., Baret, F., & Weiss, M. (2011). A multisensor fusion approach to improve lai time series. *Remote Sensing of Enviroment, 115*, 2460-2470

Verger, A., Baret, F., & Weiss, M. (2013a). Geov2/vgt: Near real time estimation of global biophysical variables from vegetation-p data. In, *Analysis of Multi-temporal Remote Sensing Images, MultiTemp 2013: 7th International Workshop on the* (pp. 1-4)

Verger, A., Baret, F., Weiss, M., Filella, I., & Peñuelas, J. (2015). Geoclim: A global climatology of lai, fapar, and fcover from vegetation observations for 1999-2010 *Remote Sensing of Environment*, http://dx.doi.org/10.1016/j.rse.2015.1005.1027

Verger, A., Baret , F., Weiss, M., Kandasamy, S., & Vermote, E. (2013b). The cacao method for smoothing, gap filling and characterizing seasonal anomalies in satellite time series. *IEEE Transactions on Geoscience and Remote Sensing, 51*, 1963-1972.

Vescovo, L., Wohlfahrt, G., Balzarolo, M., Pilloni, S., Sottocornola, M., Rodeghiero, M., & Gianelle, D. (2012). New spectral vegetation indices based on the near-infrared shoulder wavelengths for remote detection of grassland phytomass. *International Journal of Remote Sensing, 33*, 2178-2195

Vintrou, E., Bégué, A., Baron, C., Saad, A., Lo Seen, D., & Traoré, S. (2014). A comparative study on satellite- and model-based crop phenology in west africa. *Remote Sensing, 6*, 1367-1389

VITO (2005). Template product information package: S10 (spectral reflectance and ndvi). In: Available online: http://www.vgt4africa.org/PublicDocuments/S10-NDVI-Product-Sheet.pdf (Accessed on 27 March 2015)

Vrieling, A., De Beurs, K.M., & Brown, M.E. (2011). Variability of african farming systems from phenological analysis of ndvi time series. *Climatic Change, 109*, 455–477

Wei, H., Heilman, P., Qi, J., Nearing, M.A., Gu, Z., & Zhang, Y. (2012). Assessing phenological change in china from 1982 to 2006 using avhrr imagery. *Front. Earth Sci., 6*, 227–236

Weiss, M., Baret , F., Garrigues, S., Lacaze, R., & Bicheron, P. (2007). Lai, fapar and fcover cyclopes global products derived from vegetation. Part 2: Validation and comparison with modis collection 4 products. *Remote Sensing of Environment, 110*, 317-331

Weiss, M., Baret, F., Lacaze, R., Ramon, D., Wandrebek, L., Smets, B., & Verger, A. (2014). Near real time, global, lai, fapar and cover fraction products derived from proba-v, at 300 m and 1 km resolution. In, *Proceedings of the Global Vegetation Monitoring and Modeling* (pp. 284-289). Avignon, France, 3–7 February 2014

White, A.B., Kumar, P., & Tcheng, D. (2005). A data mining approach for understanding topographic control on climate-induced inter-annual vegetation variability over the united states. *Remote Sensing of Environment, 98*, 1-20

Wikistat (2013). Statistique & big data analytics de statisticien à data scientist. Available online. In. http://wikistat.fr/ (Accessed on 11/08/2013)

Willmott, C.J., & Matsuura, K. (2005). Advantages of the mean absolute error (mae) over the root mean square error (rmse) in assessing average model performance. *Clim. Res., 30*, 79–82

Wispelaere, G.d. (1980). Les photographies aériennes témoins de la dégradation du couvert ligneux dans un géosystème sahélien sénégalais : Influence de la proximité d'un forage. *L'arbre en Afrique tropicale : la fonction et le signe, 17*, 155-166

Wold, S. (2001). Personal memories of the early pls development. *Chemometrics and Intelligent Laboratory Systems, 58*, 83-84

Xiaoping, W., Ni, G., Kai, Z., & Jing, W. (2011). Hyperspectral remote sensing estimation models of aboveground biomass in gannan rangelands. *Procedia Environmental Sciences, 10*, 697-702

Xie, P., & Arkin, P.A. (1996). Analysis of global monthly precipitation using gauge observations, satellite estimates, and numerical model prediction. *J. Climate, 9*, 840-858

Xie, P., & Arkin, P.A. (1997). A 17-year monthly analysis based on gauge observations, satellite estimates, and numerical model outputs. *Bulletin of the American Meteorological Society, 78*, 2539-2558

Zhang, C., Kovacs, J., Liu, Y., Flores-Verdugo, F., & Flores-de-Santiago, F. (2014). Separating mangrove species and conditions using laboratory hyperspectral data: A case study of a degraded mangrove forest of the mexican pacific. *Remote Sensing, 6*, 11673-11688

Zhang, X., & Ni-meister, W. (2014). Remote sensing of forest biomass. In J.M. Hanes (Ed.), *Biophysical applications of satellite remote sensing* (pp. 63-98): Springer Remote Sensing/Photogrammetry

Zheng, D., Rademacher, J., Chen, J., Crow, T., Bresee, M., Le Moine, J., & Ryu, S.-R. (2004). Estimating aboveground biomass using landsat 7 etm+ data across a managed landscape in northern wisconsin, USA. *Remote Sensing of Environment, 93*, 402-411

Zhu, Y., Liu, K., Liu, L., Wang, S., & Liu, H. (2015). Retrieval of mangrove aboveground biomass at the individual species level with worldview-2 images. *Remote Sensing, 7*, 12192

Zine, S. (2004). Contribution de la télédétection satellitale radar pour le suivi des paramètres de surface d'une zone sahélienne agro-pastorale. In, *Sciences de l'information géographique* (p. 166): Université de Marne-la-Vallée

Appendices

Figure 5.1 – A1. Spatial distribution of percentage of missing data (MD) in yearly FAPAR time series from 1999 to 2013

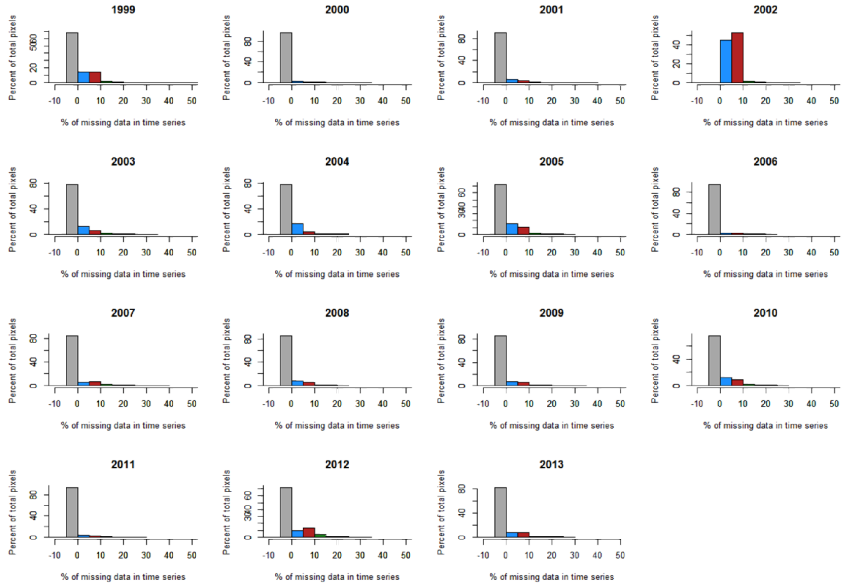

Figure 5.2 – A2. Histograms of percentage of missing data in yearly FAPAR time series from 1999 to 2013

Table 5.1 – A3. Description, unit and selection status after recursive feature elimination and variable inflation control of 17 agrometeorological variables provided by the GeoWRSI water balance model and used in the study. The 10 underlined variables correspond to those used for model development.

Variables	Description	Unit	Selection status
WRSI	Water requirement satisfaction index	%	No
SOSp	Start of the rainy season	dekad	Yes
EOSp	End of the rainy season	-	No
AETi	Actual evapotranspiration accumulated over the initial stage of the growing season	mm	No
AETv	Actual evapotranspiration accumulated over the vegetative stage of the growing season	-	Yes
AETf	Actual evapotranspiration accumulated over the flowering stage of the growing season	-	Yes
AETr	Actual evapotranspiration accumulated over the ripening stage of the growing season	-	No
WDEFi	Water deficit accumulated over the initial stage of the growing season	-	Yes
WDEFv	Water deficit accumulated over the vegetative stage of the growing season	-	Yes
WDEFf	Water deficit accumulated over the flowering stage of the growing season	-	Yes
WDEFr	Water deficit accumulated over the ripening stage of the growing season	-	Yes
WSURi	Surplus water accumulated over the initial stage of the growing season	-	No
WSURv	Surplus water accumulated over the vegetative stage of the growing season	-	Yes
WSURf	Surplus water accumulated over the flowering stage of the growing season	-	Yes
WSURr	Surplus water accumulated over the ripening stage of the growing season	-	Yes
PPTc	Cumulated rainfall during the rainy season	-	No
PPTm	Averaged rainfall during the rainy season	-	No

Table 5.2 – A4. Mean signed difference (in dekads) between onset metrics calculated from rainfall and FAPAR data across the agricultural and natural vegetation land cover classes in the study area.

Land cover classes	Mean signed difference in dekads	
	Start of season SOSp-SOS	End of season EOSp-EOS
Herbaceous	1.3	0.1
Agriculture	0.3	-1.4
Shrubs very open	0	-2.1
Shrubs open to closed	0	-1.7
Trees very open	0.3	-1.6
Trees open to closed	0.4	-2.2
Overall mean	0.4	-1.5

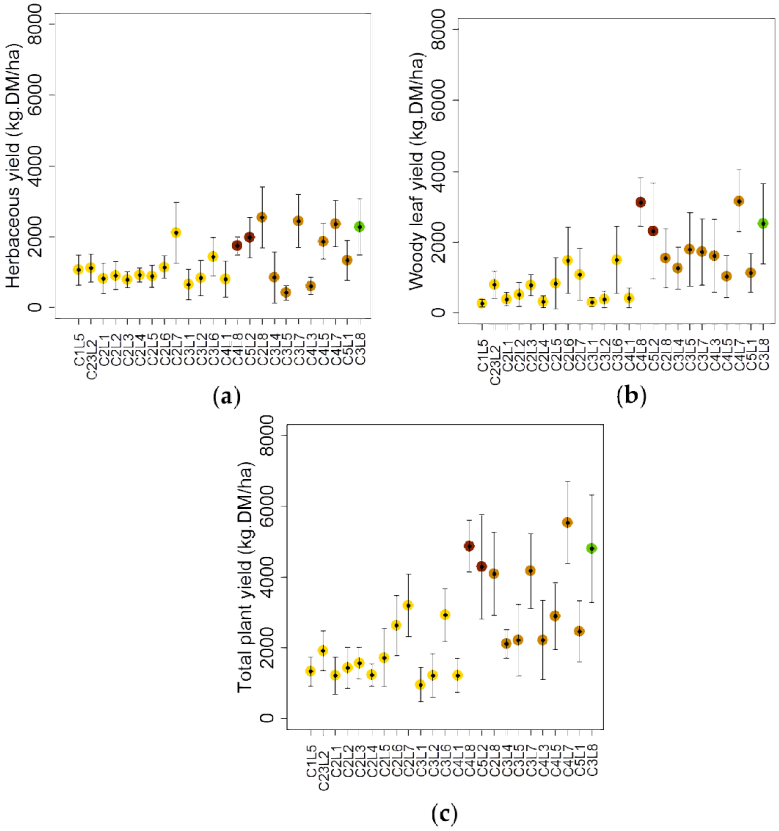

Figure 5.3 – A5. Averaged (a) herbaceous, (b) woody leaf and (c) total plant yield of the 24 sites used in this study and their corresponding standard deviation (error bar) for the 2000-2015 period.