Universidad de Valladolid

eSDUVa
Escuela de Doctorado Universidad de Valladolid

PROGRAMA DE DOCTORADO EN FISICA

TESIS DOCTORAL:

Identifying Desert locust breeding areas by means of Earth Observation in Mauritania

Presentada por Diego Gómez Aragón para optar
al grado de
Doctor por la Universidad de Valladolid

Dirigida por:

Dra. Julia Sanz Justo

Dr. Pablo Salvador González

Dr. Carlos Casanova Mateo

Valladolid, 2019

Identifying desert locust breeding areas by means of Earth Observation in Mauritania

Doctorate programme on Applied Physics "Earth observation, environmental modelling and remote sensing"

Diego Gómez Aragón

PhD Advisor: Dr. Jose Luis Casanova (UVA, Spain)
PhD Director/s: Dra. Julia Sanz Justo, Dr. Pablo Salvador and Dr. Carlos Casanova Mateo (UVA, Spain)

"Nothing would be done at all if a man waited till he could do it
so well that no one could find fault with it"

J. H. Newman

Acknowledgements

Firstly, I would like to show my sincere gratitude to José Luis Casanova to trust on me and provide a great support during my PhD. Without his recommendations, advices and knowledge, this thesis would not have been possible. I would also like to thank my PhD directors Julia Sanz, Pablo Salvador and Carlos Casanova for their continuous support during this time, for their patience, motivation, and immense knowledge. Their guidance helped me in all the time of research and writing of this thesis. I could not have imagined having better advisors and mentors for my PhD study.

I would like to thank to the rest of my fellow lab-mates at LATUV (Miguel and Dani) their support, the stimulating discussions about programming, encouragement and good advices during my research, what has permitted to enrich this work from various perspectives.

Last but not least; I would like to thank my family: my parents and sister for supporting me to get to the point where I am today, and to my partner and friends to encourage me to keep on until this desired goal.

TABLE OF CONTENTS

Abstract *1*

Resumen *3*

List of abbreviations *5*

Chapter 1.- INTRODUCTION *7*

1.1. Research hypotheses **7**

1.2. Antecedents and research interest **7**

1.2.1. Historical background of desert locust 7
1.2.2. Remote sensing 8
1.2.3. Artificial Intelligence applied on Remote sensing data 12
1.2.4. Species Distribution Models 15

1.3. Description of the study area **19**

1.3.1. Location 19
1.3.2. Topography and drainage 20
1.3.3. Climatic conditions 23
1.3.4. Geology and geomorphology 26
1.3.5. Land cover and soil type 27
1.3.6. Vegetation and animal life 28

1.4. The biology of desert locust **29**

1.4.1. Locust Phases 32
1.4.2. Life cycle 33
1.4.3. Environmental circumstances of the habitat 34
1.4.4. Migration and seasonal distribution 35
1.4.5. Important terms 36

1.5. Thesis structure, specific goals and objectives **37**

Chapter 2. Identification of potential wadis using SWAT hydrological model *39*

2.1 Introduction **39**

2.2. Materials and Methods **42**

2.2.1. Input data 42
2.2.2. Methods 47

2.3. Results and discussion **55**

2.3.1. Model set up 55
2.3.2. Sensitive parameters 58
2.3.3. Calibration and validation 60
2.3.4. Distances between desert locust and PRoW 64

2.4. Conclusion..66

Chapter 3. Soil moisture analysis to locate breeding areas using machine learning techniques 69

3.1. Introduction 69

3.1.1. Species occurrence data 70

3.1.2 Data cleaning..70

3.1.3. Sampling bias 70

3.1.4. Absence, pseudo-absence or background points 71

3.1.5. Extraction and preparation of environmental data 71

3.2. Materials and Methods..72

3.2.1 Survey data..72

3.2.2. Satellite data 73

3.2.3. Methods..74

3.3. Results 86

3.4. Discussion 92

3.5. Conclusions 94

Chapter 4. Multivariate ensemble model to detect desert locust 97

4.1. Introduction 97

4.2. Materials and methods..98

4.2.1. Field data and pseudo-absences..98

4.2.2. Environmental variables..100

4.2.3. Methods..103

4.3. Results 113

4.3.1. Model evaluation 113

4.3.2. Variable importance..115

4.3.3. Response curves and living thresholds 117

4.4 Discussion 120

4.5 Conclusion..123

Chapter 5. General discussion and conclusions 125

Chapter 6. References.. 129

Chapter 7. List of publications 161

Abstract

Desert locust (Schistocerca gregaria) has severely influenced crop production in northern Africa and Middle East since antiquity. To prevent or mitigate its effects on local communities, it is necessary to precisely locate those areas where they breed. Previous works have relied on precipitation and vegetation indices obtained by satellite **remote sensing**, however many authors (Browning et al., 1990; Gay et al., 2018) agree on the necessity to improve desert locust **prevention systems**.

In this PhD thesis, we have explored 3 different novel approaches to locate desert locust breeding areas in **Mauritania**. (1) Firstly, the SWAT hydrological model was used to locate wadis that may host desert locust given their **favourable ecological conditions**. (2) Secondly, the influence of **soil moisture** (SM) using the European Space Agency Climate Change Initiative Soil Moisture (ESA CCI SM) product was assessed over desert locust breeding sites using **Artificial Intelligence** (AI) and more specifically machine learning techniques. (3) Finally, we have generated a multivariate ensemble model using a combination of the frequently used Species Distribution Models (SDMs) in ecology.

The results in (2) showed a good correlation between general monthly soil moisture patterns and hopper presences. It was found that an area becomes suitable for breeding when the minimum SM values are over 0.07 m^3/m^3 during 6 days or more. On the other hand, the identified wadis by means of SWAT hydrological model (1) did not find significant influence on locust presences for the studied period. Many uncertainties in precipitation records, as well as poor river gauge data were encountered, what impeded adequate calibration and validation procedures. Longer and more accurate data records (precipitation and river gauge) may permit to further develop this approach in the near future. Furthermore, the third approach showed highly satisfactory model results (KAPPA & TSS = 0.901 and ROC = 0.986) to detect hopper desert locust in solitary phase, implying that our model can identify suitable environmental conditions for breeding. This study also confirms the potential of the SMAP satellite from NASA to retrieve critical temperatures due to its time pass, in addition to reinforcing the NDVI product from MODIS as a reliable environmental predictor (3).

These results demonstrate the validity of the methodologies exposed in this PhD thesis to identify favourable **breeding areas** in Mauritania. Earth observation techniques can retrieve periodically important environmental variables such as soil moisture, surface temperature or vegetation status to be managed by machine learning algorithms over remote and large areas. Thus, our work may be of interest to authorities of affected countries or international organizations to complement or improve current ongoing monitoring techniques and warning systems.

Keywords: *Artificial Intelligence; Breeding areas; Desert locust; Favourable ecological conditions; Mauritania; Prevention systems; Remote sensing; Soil moisture*

Resumen

La langosta del desierto (Schistocerca gregaria) ha provocado graves daños a la producción agrícola del norte de Africa y Oriente Medio desde tiempos antiguos. Con el objetivo de prevenir o mitigar sus efectos sobre las poblaciones locales, es necesaria una precisa localización de aquellas áreas donde crían y se reproducen. Tradicionalmente, los trabajos que han abordado esta problemática han utilizado variables como la precipitación o índices de vegetación obtenidos mediante **teledetección** con satélites; sin embargo, son varios los expertos que señalan la necesidad de mejorar los actuales **sistemas de prevención** de langosta del desierto (Browning et al., 1990; Gay et al., 2018).

En esta tesis doctoral, presentamos 3 novedosos procedimientos para detectar posibles zonas de reproducción o cría para la langosta en **Mauritania**. (1) El primer método se basa en el uso del modelo hidrológico SWAT, el cual se ha utilizado para localizar wadis. Según la bibliografía consultada, estos lugares presentan unas **condiciones ecológicas favorables** para la presencia de langosta. (2) El segundo método estudia la influencia del producto de humedad del suelo generado por la Agencia Europea del Espacio bajo su iniciativa de Cambio Climático (ESA CCI SM) en la zonas de cria de la langosta del desierto mediante técnicas de **Inteligencia Artificial** (IA). (3) Finalmente, se ha generado un modelo multi-variable a partir de la combinación de varios modelos de distribución de especies (SDMs).

Los resultados obtenidos en (2) muestran una buena correlación entre los valores mensuales de **humedad del suelo** y las presencias de langosta en estado juvenil “saltamontes”, más conocidos por el término inglés “hopper”. Se ha observado que aquellas zonas donde los valores mínimos de humedad sobrepasan los 0.07 m^3/m^3 durante 6 o más días, son áreas con mayor predisposición a ser zona de cría para la langosta. Por otro lado, se han identificado geográficamente wadis mediante el modelo SWAT y estudiado la posible relación con la presencia de langosta en las inmediaciones (1). Los resultados no confirman una directa relación que condicione la presencia de langostas, al menos para el periodo de tiempo estudiado. Se ha observado bastante incertidumbre en los datos de precipitación y caudal de agua, lo que ha provocado que la calibración y validación del modelo no sea la más óptima posible. Unas series temporales más largas y precisas de precipitación y caudales, permitiría una mejora de los resultados obtenido en este capítulo para estudios futuros. Además, el método desarrollado en el tercer capítulo ha mostrado unos buenos resultados en términos de capacidad predictiva del modelo (KAPPA & TSS = 0.901 and ROC = 0.986). De esa manera, el modelo permite localizar zonas de cría para la langosta del desierto en estado solitario. Los resultados de este tercer capítulo demuestran el potencial del satélite recientemente lanzado SMAP para registrar críticas temperaturas del suelo, además de confirmar el producto NDVI del Terra-MODIS como un fiable indicador ambiental (3).

Los resultados obtenidos en esta tesis doctoral demuestran la validez de las metodologías de machine learning propuestas para identificar **zonas de cría** para langosta en Mauritania. Los métodos de observación de la Tierra mediante satélite son capaces de recopilar información ambiental relevante sobre la humedad del suelo, estado de la vegetación o temperatura del suelo; y esa información es gestionada mediante algoritmos de inteligencia artificial. Por tanto, se concluye que esta tesis doctoral puede ser del interés de las autoridades de países afectados u organismos internacionales para complementar o mejorar las actuales técnicas de monitoreo y sistemas de alerta.

Palabras Clave: *Condiciones ecológicas favorables; Humedad del suelo; Inteligencia Artificial; Langosta del desierto; Mauritania; Sistemas de prevención; Teledetección; Zonas de cría.*

List of abbreviations

AI: Artificial Intelligence

ANN: Artificial Neural Network

CFSR: Climate Forecast System Reanalysis

CTA: Classification Tree Analysis

DLIS: Desert Locust Information Service

EMca: Ensemble Model Committee Averaging

EMciInf: Ensemble Model confidence interval lower

EMciSup: Ensemble Model confidence interval upper

EMmean: Ensemble Model mean

EMwmean: Ensemble Model weighted mean or sum of probabilities

FAO: Food and Agriculture Organization of the United Nations

FDA: Flexible Discriminant Analysis

GAM: Generalized Additive Model

GBM: Generalized Boosting Model

GIS: Geographic Information Systems

GLM: Generalized Linear Model

GRDC: Global Runoff Data Centre

HWSD: Harmonized World Soil Database

LAI: Leaf Area Index

LST: Land Surface Temperature

MARS: Multiple Adaptive Regression Splines

NDVI: Normalized Difference Vegetation Index

RF: Random Forest

ROC: Receiver Operating Characteristic

SDM: Species Distribution Model

SM: Soil Moisture

SRE: Surface Range Envelop or Bioclim

SRTM: Shuttle Radar Topographic Mission

SWAT: Soil and Water Assessment Tool

TSS: True Skill Statistic

Chapter 1.- INTRODUCTION

1.1. Research hypotheses

This study aims to improve the current early warning systems to detect desert locust breeding areas, and it is based on the latest remote sensing technology from Earth observation, hydrologic models and artificial intelligence algorithms. This work has been motivated to mitigate the social impact of locust pests in the affected countries, aiming to provide technological support to local authorities and decision makers.

1.2. Antecedents and research interest

1.2.1. Historical background of desert locust

Desert locust outbreaks have been a problem since antiquity, and periodically have caused devastation over local communities in northern Africa and Middle East countries. It is well documented by ancient literature: in the Har-ra list (Assyria - the Ashurbanipal Royal Library, 669-626 B.C.), in decorations found in Egyptian tombs (6th Dynasty, 2420-2270 B.C.), and in Biblical, Rabbinical, Greek and Roman literature, while control measures are also reported during Biblical, Grecian, Roman, Mishnaic, Talmudic, Byzantine as well as in modern times (Nevo, 1996). They affect local economies and living conditions, decreasing yield production in areas already affected by water scarcity and extreme weather conditions.

Desert locust is the earliest diverging species among the genus Schistocerca and the unique one settled in Africa, what indicates its high adaptability to the local conditions. Unlike other species of the same genus, it has kept some of its original traits such as the ability to change their behavior (Song et al., 2017).

Some of these areas have many political, economic and environmental constraints for food production, where still farming and herding are their main means of livelihood (FAO, 1994). Food scarcity and economic damages are direct consequences of locust pests, in addition to environmental impacts associated with the use of pesticides in control operations. According to (Brader et al., 2006), the desert locust invasion of 2004 reduced up to 80 % the expected cereal production in Burkina Faso, 90 % in Mali and between 90 to 100% in Mauritania. The food consumption over those areas needed to be minimized due to the food crisis, and external aid was supplied to mitigate the effects. In spite of the efforts to control the plague (2003 to 2005), whose campaign rose up to US$400 million, the socio-economic impacts lasted until 2006.

In order to avoid crop and pasture losses, as well as reducing control spending, Food and Agriculture Organization of the United Nations (FAO) highlights the necessity to implement preventive strategies over the countries with seasonal breeding areas (Cressman, 1999). And those efforts have been mainly located in the western region of Africa aiming to control desert locust outbreaks before they spread over larger areas. Preventive controls are recommended to handle migrant plagues in terms of cost-benefit analysis, reducing the current impact but also diminishing the likelihood of incidence in the following years. There are some analyses such as (Joffe, 1998) where real and modeled data were used to evaluate the economic effects of dessert locust. His simulation was based on crop losses with and without control measurements against desert locust plagues, and then comparing those benefits against the real cost of the control measurements. It turned out that the control measurements were effective to fight against locust but not economically efficient (massive use of pesticides over areas not needed) due to bad management, excessive spending and lack of information. Nevertheless, sudden and located desert locust outbreaks are real threats to local farmers, and they urge actions to save their crops.

Despite the long pest occurrence of desert locust, control efforts have been in vain at least until late 20th century. Large monitoring areas or lack of data are some of the reasons that account for this time lag (FAO, 2004).

1.2.2. Remote sensing

Remote sensing is the field that studies and models the phenomena occurring on the Earth surface and the iterations with the atmosphere (Lillesand et al.,

2014) by means of reflected or emitted electromagnetic energy that is collected by a sensor at certain distance (Campbell & Wynne, 2011).

There is a current necessity to monitor and control natural as well as anthropogenic dynamics such as environmental hazards, urban growth, meteorological and climatological prediction, natural resources to implement efficient management policies. It turns out that remote sensing is a good asset to solve those necessities (Camps-Valls, 2009). The physical principle of remote sensing techniques is that the target object reflects, absorbs, and emits electromagnetic radiation in a different way depending of their molecular composition and shape. Depending on the type of energy used by the sensor to acquire the data, we would differentiate two types: active and passive. Passive sensors measure energy that is naturally available. They can retrieve reflected energy when the sun illuminates the Earth, although they can detect energy which has been naturally emitted by the objects when is large enough such as in the thermal infrared spectrum (Fig. 1).

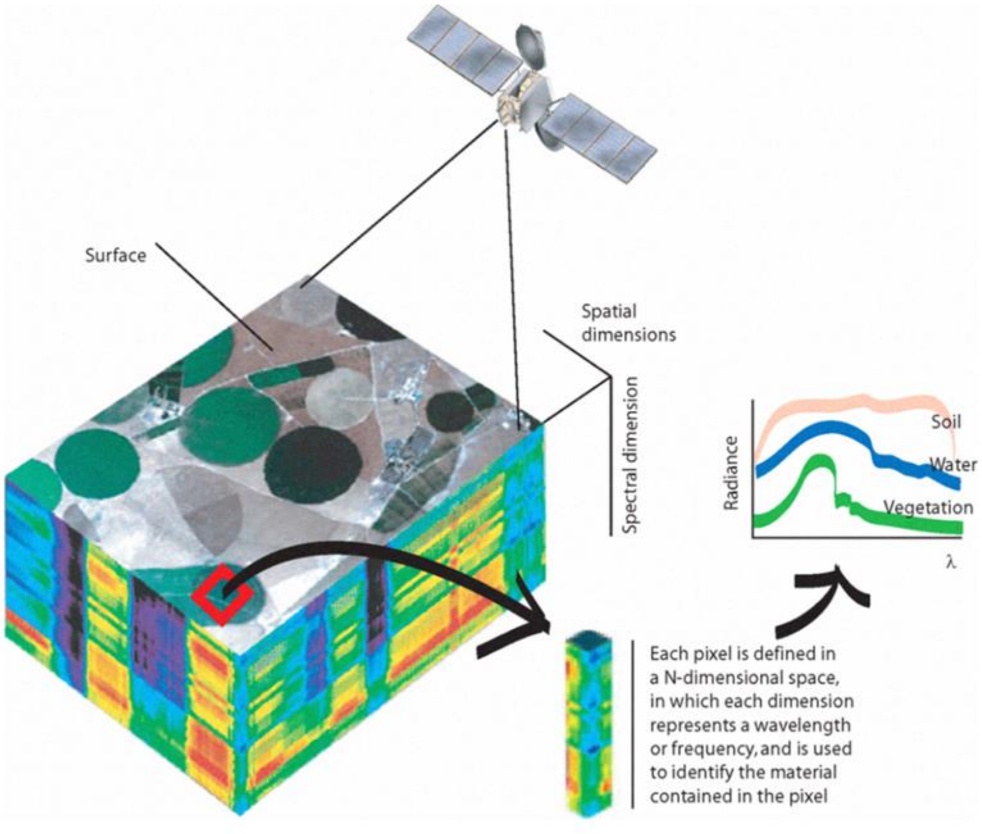

Surface
Spatial dimensions
Spectral dimension
Radiance
Soil
Water
Vegetation
λ
Each pixel is defined in a N-dimensional space, in which each dimension represents a wavelength or frequency, and is used to identify the material contained in the pixel

Figure 1. Principles of imaging spectroscopy. Source: (Camps-Valls., 2009)

On the other hand, active sensors such as radar emit their own energy to illuminate towards the target object. The active antenna measures the travelling time and the backscatter of the transmitted energy that has been reflected on the Earth's surface (NRCAN, 2015). These systems are independent of the weather conditions, and they can measure during day or night time. As a result, some characteristics of the image are determined by the object features: geometry (size, shape, roughness, orientation), dielectrics (water content, aggregate state, salt content, mineralogy), and motion (in azimuth direction and range direction). In addition, radar images are influenced by the characteristics of the radar sensor or the acquisition parameters (repeat frequency, pulse repetition frequency, bandwidth, polarization, incidence angle, imaging mode and orbit direction).

Satellite remote sensing is nowadays a great asset to study inaccessible or complicated regions, as well as a cost effective method to monitor a wide range of environmental parameters, with a good temporal and spatial resolution (Melesse et al., 2007).

Some authors have proposed the use of remote sensing platforms to monitor large and inaccessible locust breeding areas (Tucker et al., 1985; Tappan et al., 1991; Ceccato et al., 2007, Pekel et al., 2011, Waldner et al., 2015; Renier et al., 2015, Piou et al., 2017), which usually occur away from crops (Symmons, 1991). Remote sensed vegetation and precipitation are being used to derive potential grasshopper and locust habitats (Tappan et al., 1991) by means of satellite platforms as LANDSAT, NOAA, Meteosat, SPOT, TERRA or AQUA (Latchininsky et al., 2017). International organizations such as the Desert Locust Information Service (DLIS) from FAO have been using Earth observation methods since the 80's to assess favourable environmental conditions to desert locust (Latchininsky et al. 2017). Nevertheless, ongoing monitoring techniques present some limitations in arid environments. The vegetation is usually sparse and geomorphological features are not always well identified (Piou et al., 2013; Lazar et al., 2015). Moreover, one of the major problems is precipitation detection by satellite remote sensing. Detection probabilities may range from 70% to 20 % in arid and semiarid regions, with a high overestimation of rainfall occurrences (Dinku et al., 2010). Some studies have used precipitation datasets to determine breeding sites (Tucker et al., 1985; Cressman, 2013; Lazar et al., 2015), nevertheless remote sensing precipitation datasets do not seem to be precise enough over arid and semiarid environments, where precipitation is scarce, random, brief and intense. Surface soil moisture is a very important variable to understand hydrological processes in arid environments, playing a fundamental role in water and energy exchange between land and atmosphere (Wang & Qu, 2009). Furthermore, understanding its spatial and temporal evolution could be crucial for many environmental and ecological applications, such as to determine breeding areas for desert locust. A moist soil would keep the eggs hydrated, and such condition with high soil temperatures would enhance their quick hatching (Symmons & Cressman, 2001).

Recently, new satellite platforms such as Soil Moisture Ocean Salinity (SMOS) or Sentinel 1 are used to retrieve soil moisture information. There is a new ongoing project named Soil Moisture for dEsert Locust earLy Survey (SMELLS) funded by the European Space Agency (ESA) that aims to provide a 10 day soil moisture map of the north of Africa to identify suitable areas for egg laying. More information can be found at the site http://smells.isardsat.com.

Prior studies have used satellite information to identify green vegetation by means of vegetation indexes (Popov et al., 1991; Despland et al., 2004). Vegetated soils are good breeding habitats to maintain and breed hoppers and adults (Tucker et al., 1985); however it might not be a suitable area to lay eggs and identify the beginning of a locust generation, what is our aim. Furthermore, vegetation identification with remote sensing techniques in arid or semiarid regions can be sometimes complicated due to the high red soil reflectance and imagery resolution. In addition to that, some temperature patterns may have effects on locust development (Chappell, 1983; Zhang et al., 2009).

Nowadays, the main efforts are focused on establishing preventive measurements to retrieve meteorological variables and vegetation so as to locate breeding areas of desert locust (FAO, 2016). This rising concern to control the population number before they become a plague, led FAO to develop the desert Locust Information Service (DLIS). This project aims to assess and warn about potential outbreaks, and provide the necessary information to operate an early warning system based on Earth observation systems and field work, in addition to other projects as the already cited SMELLS from ESA.

Current technology such as satellite or drone remote sensing may improve data collection, overcoming temporal and spatial difficulties with cost effective methods, what makes these tools essential to tackle locust outbreaks (Ceccato et al., 2007).

1.2.3. Artificial Intelligence applied on Remote sensing data

Artificial intelligence (AI) dates back to the 1950s, and since then it has being developed keeping up with advances in computer science. AI could be described as the theory and development of computer systems able to perform tasks normally requiring human intelligence (Hansen et al., 2017).

These techniques are nowadays applied on many fields ranging from economics or medicine up to journalism. Whenever data management is needed, AI arises as an appropriate tool to solve complicated tasks. Computers allow us to do complex operations in short periods of time. This fact has led us to a research area that had not being explored yet: “teaching machines to predict a likely outcome by looking at patterns on datasets”. This technique is called Machine Learning, and it is a particular approach to AI (Michalski et al., 2013). We can distinguish between three types of learning (Brownlee, 2014): supervised learning, unsupervised learning and semi-supervised learning.

1.1.1.1. Supervised Machine Learning

Most of the practical machines learning techniques use supervised learning. In this method, you have independent input variables (x) and one output and dependent variable (Y). The algorithm will aim to explain the output (Y) based on the inputs (x) by means of a function f(x).

Thus, this function can predict the output variable (Y) when independent variables or predictors (x) are used as input data. It is referred as “Supervised Learning” because the algorithm will “learn” based on the training dataset since it has the correct output, and when predictions go wrong, they can be corrected to improve its accuracy. The learning process usually stops when the algorithm has reached an acceptable performance.

Furthermore, there are two types of supervised learning methods: Classification and Regression. We can say that there is a classification problem when the output variable has two or more classes (Kotsiantis, 2007). On the other hand, regression problems are characterized by having real numbers as variable outputs (Criminisi et al., 2014).

1.1.1.2. Unsupervised Machine Learning

Unsupervised learning methods take into account only independent variables (x) and it does not contain any output or observed dependent variables (Y). The aim of this method is to model the underlying structure of the data so as to obtain some patterns and learn from them (Hastie et al., 2009). Unlike supervised learning, the algorithms are not assisted by any output variable (Y) to find structures in the data. This type of problems is divided into clustering and association. The clustering problem aims to discover how the data groups due to its inherent characteristics. The association method aims to seek rules to describe large portions of the data. The most popular unsupervised learning algorithm for clustering problem is k-means, and for association problems "A-priori" (Brownlee, 2014).

1.1.1.3. Semi-Supervised Machine Learning

There is a third type of learning method named Semi-Supervised Machine Learning. It addresses problems where only part of the data is labelled with the output variable (Y), and it stands in the mid-way between supervised and unsupervised learning methods.

In this approach, supervised learning techniques can build a model to predict unlabelled data and feed that data back into the supervised learning algorithm as training data. Finally, that model can predict new unseen data (Brownlee, 2014).

1.1.1.4. General considerations in Machine Learning

In summary, there are so many machine learning algorithms to be applied on a wide variety of problems, and new algorithms are being generated on a daily basis. Depending on the nature of the problem and the type of the available data, we need to apply a certain bunch of machine learning algorithms. Decision Trees, K-Means, K-Nearest Neighbors, Neural Networks, Random Forests or Support Vector Machines are just a few examples of the available wide range of existing machine learning algorithms.

One of the first machine learning applications on remote sensing was done by (Huang & Jensen, 1997). They used a machine learning approach to automated building of knowledge bases for image analysis systems incorporating GIS data. It was a wetland classification which was compared against two conventional methods. The study concluded that the machine learning method based on decision trees was of good quality for image analysis. Since then, the remote sensing community have been using machine learning techniques as usual working tools (Melgani & Bruzzone, 2004; Ahmad et al., 2010; Rhee & Im, 2017).

Most of the machines learning projects share a similar workflow (Fig. 2).

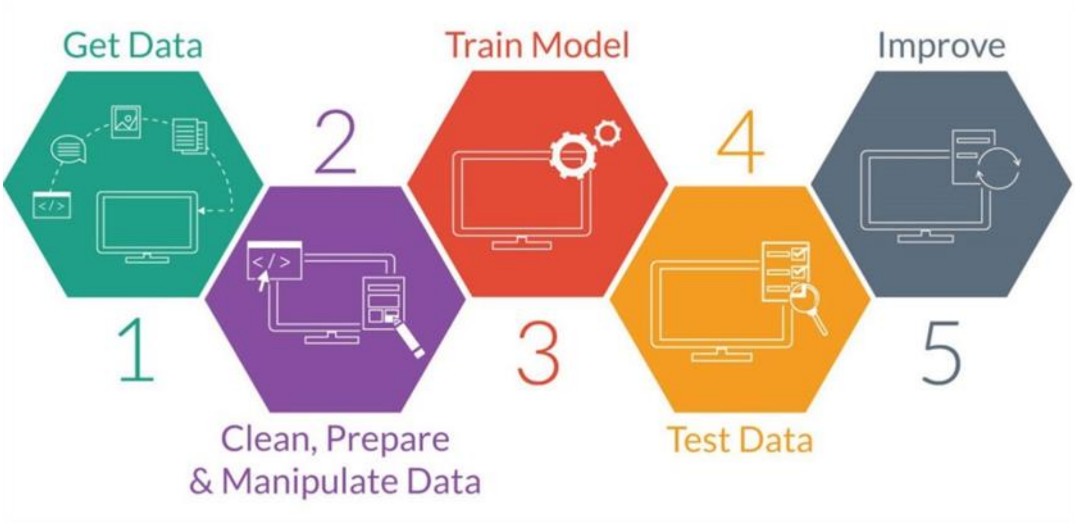


Figure 2. Usual workflow in machine learning projects. Source: https://machinelearning-blog.com/2017/11/19

Firstly, we need to gather the data. In remote sensing, data mostly comes from satellites or drones. Then, preprocess (clean, prepare and manipulate) the data in order to expect the best results from the machine learning algorithms. Some of these algorithms can perform better if the data is prepared in a specific way (standardized, scaled, centered or normalized). Afterwards, the dataset needs to be split into training and testing subsets, and an appropriate algorithm should be chosen to build our model. As previously seen, each algorithm may suit to certain circumstances so it is very important to identify which are the most suitable to our specific purposes. Later on, the model should be trained and in some instances tuned to aim for the best performance. Finally, the model ought to be validated and improved to achieve the highest accuracy to our samples. The predictor model would thus be constructed. This is a summarized version about what implies to construct a machine learning predictive model.

Two of the most recurrent problems in machine learning is called under-fitting and over-fitting. We say that one model is over-fitted when it has learned too well all the details in the training dataset, even the noise or random fluctuations within the data, so that it may affect the performance in the new data (Fig. 3). The model will not have the ability to generalize with new data in order to generate predictions (Brownlee, 2017).

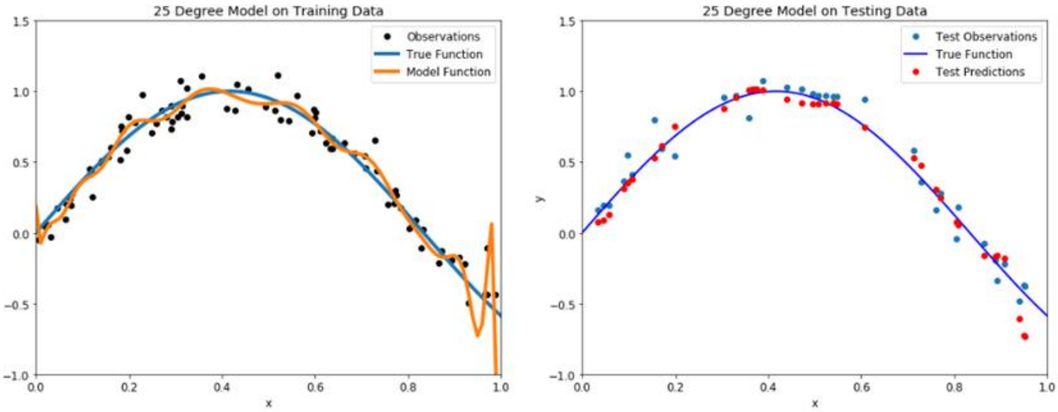


Figure 3. Over-fit 25 degree polynomial model on training (left) and testing (right) datasets. Source: https://towardsdatascience.com

Whereas a model is under-fitted when it does not describe the training data. It is easier to detect under-fitting in a model in comparison with overfitting, because it shows up during the calibration phase (Fig. 4).

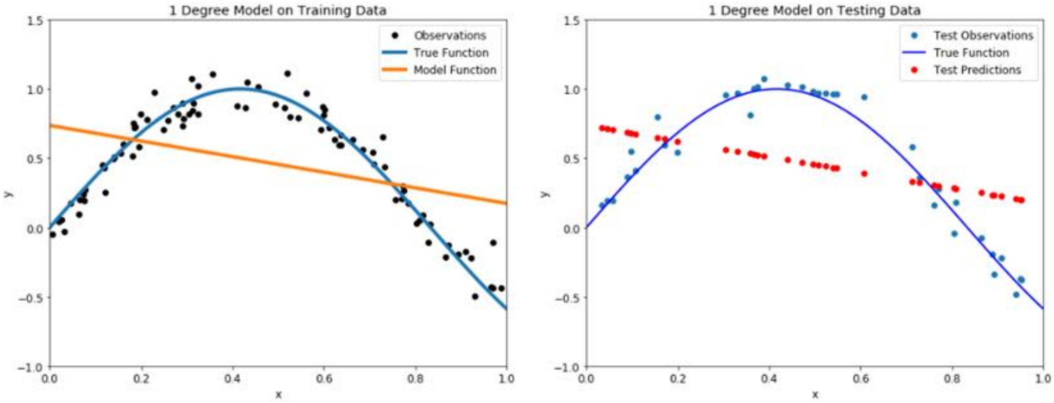


Figure 4. Under fit 1 degree polynomial model on training (left) and testing (right) datasets. Source: https://towardsdatascience.com

1.2.4. Species Distribution Models

Species Distribution Models (SDMs) are numerical tools that derive from Artificial Intelligence (AI) with ecological purposes. They analyze the link between species occurrences and environmental factors, providing an ecological insight to predict species distribution over space or time given certain environmental characteristics (Elith & Leathwick, 2009). Their machine learning methods increase traditional predictive performance and their capacity to incorporate complex interaction among variables (Anderson et al., 2006), being eligible to work with large ecological datasets (Robinson et al., 2014).

Species Distribution Modelling (SDM) or environmental niche modelling is a computer algorithm technique to predict the geographical distribution of species, based on the known distribution or true presence values. The environment of the species is usually characterized by variables such as climate, soil type, water depth or land cover among others (Elith & Leathwick, 2009). The accuracy of the predictive model will vary depending on the correct number of feature selecting, as well as sufficient data available for the modelling process. These models are applied on many fields such as conservation, ecology and evolution. SDM may facilitate land management (Hoffer, 1975; Kessell, 1976; Strahler, 1981), weed or pest species risk assessment (Sutherst & Maywald, 1985; Busby, 1991) and studies of climate impacts on the biota (Busby, 1986; Nix & Busby, 1986).

SDM also relies on geographic information science (GIScience) and remote sensing since it requires geospatial data for spatial prediction. Environmental and terrain modelling has been identified as one of the three major subdomains in GIScience. They have demonstrated to be very useful tools in conservation biogeography purposes. And they are especially interesting when there are geographical missing data on the species distribution since SDM can use interpolation techniques to fill the gaps of information. This is very valuable when information about the environmental variables is available but the observations of the species distribution are sparse. Tasks such as biodiversity inventory, biodiversity prospecting (designing biodiversity surveys – predicting new occurrences), gap analysis, prioritizing areas for conservation (reserve design) and environmental impact analysis (determining how human activities including resource management might affect critical habitat for species of conservation concern) become more feasible by means of interpolation (Franklin, 2013). There are several publications that effectively used SDM to fill in the geographical gaps in species distributions (Table 1).

Paper	**Topic**
Dark, **2004**	Application invasives: spatial autoregressive model contrasting invasive versus non-invasive non-native plant species
Thuiller *et al.*, **2006**	Application climate change: temperate areas of Europe predicted to lose tree functional diversity while boreal areas gain
Elith & Leathwick, **2007**	Methods: examine effect of background sample and multivariate response on SDM performance
Osborne *et al.*, **2007**	Methods: Local regression methods may perform better for interpolation but global methods for extrapolation
Guisan *et al.*, **2007**a	Methods: Effect of change in spatial grain on SDM performance
Tsoar *et al.*, **2007**	Methods: compared six presence-only SDM methods
Jiménez-Valverde *et al.*, **2008**	Concepts and methods: best models for realized versus potential distribution, performance as a function of prevalence, inaccuracy of SDMs
Wisz *et al.*, **2008**	Methods: Sample size effect on presence-only SDM methods
Marmion *et al.*, **2009**	Methods: compared five consensus methods for SDM
Puschendorf *et al.*, **2009**	Application pathogens: predict potential pathogen distribution from climate data
Beaumont *et al.*, **2009**	Application invasives: predict invasive plant species distributions from native versus entire distribution
Williams *et al.*, **2009**	Application new occurrences: compared SDM methods for predicting undiscovered populations of rare plant species
Franklin, **2010a**	Concepts and methods: SDMs have been linked to other models to forecast impacts of environmental change on biodiversity
Platts *et al.*, **2010**	Application conservation planning: predict distribution of forest plant taxa in biodiversity hotspot for conservation prioritization

Paper	Topic
Elith *et al.*, **2011**	Methods: describes MaxEnt in statistical terms; links species and data characteristics to implementation decisions
Dubuis *et al.*, **2011**	Application conservation: compared statistical models of species richness to estimates from stacked SDMs
Václavík & Meentemeyer, **2012**	Application invasives: SDMs of invasive pathogen from different stages of invasion
Hof *et al.*, **2012**	Application climate change: incorporate predator and prey distribution predictions into SDM forecast of climate change impacts
Junker *et al.*, **2012**	Application conservation: changes in human impacts variables predict changes in suitable habitat for great apes
Naujokaitis-Lewis *et al.*, **2013**	Application climate change: linked SDM-population models are sensitive to model uncertainty

Table 1. Review of publications that have used SDM to fill in gaps in species distribution. They are sorted in chronological order. Source: Franklin, 2013

(Platts et al., 2010) applied SDMs to carry out spatial predictions of plant species richness within a biodiversity hotspot and suggested that, due to models are more uncertain for endangered species; they should be generated iteratively with direct fieldwork. (Williams *et al.*, 2009) compared different SDMs in accordance with their ability to predict the distributions of rare plant species. In addition, they established on-field surveys based on those predictions to verify their assumptions. As a result, they discovered new populations of those rare species. (Puschendorf et al., 2009) used SDM techniques to predict the potential distribution of the amphibian chytrid fungus disease in Costa Rica. They were able to identify climatic and topographic areas with less likelihood of appearance for this pathogen.

Based on the success of previous literature, SDMs were used to analyse the role of remotely sensed environmental variables to identify desert locust breeding areas in solitary phase.

1.3. Description of the study area

1.3.1. Location

The Islamic Republic of Mauritania is a state located in the Maghreb region of western Africa. It is a vast country of 1,030,700 km^2 with large arid plains and only one continuous water flow, the Senegal River. It is the 11th largest country in Africa and it is bordered by Senegal (south), Mali (east), Algeria (north-east), Western Sahara (north) and the Atlantic Ocean (west). Nouakchott is the capital and its largest city, with approximately one third of the country's population. Due to the large extension of Mauritania, it is subdivided in 13 administrative regions or provinces (Table 2 & Fig. 5). This study area was selected to be one of the major breeding and recession regions for desert locust (Culmsee, 2002).

Province	**Capital**	**Area (km^2)**	**Population (2013)**
Adrar	Atar	235,000	62,658
Assaba	Kiffa	36,600	325,897
Brakna	Aleg	33,000	312,277
Dakhlet Nouadhibou	Nouadhibou	23,090	123,779
Gorgol	Kaédi	13,600	335,917
Guidimaka	Sélibaby	10,300	267,029
Hodh Ech Chargui	Néma	182,700	430,668
Hodh El Gharbi	Ayoun el Atrous	53,400	294,109
Inchiri	Akjoujt	46,800	19,639
Nouakchott-Nord	Dar-Naim	306	366,912
Nouakchott-Ouest	Tevragh-Zeina	146	165,814
Nouakchott-Sud	Arafat	252	425,673
Tagant	Tidjikja	98,340	80,962
Tiris Zemmour	Zouérat	252,900	53,261
Trarza	Rosso	67,800	272,773

Table 2. Principal regions or provinces in Mauritania with its capital, area and population. Source: http://www.africanbib.biz

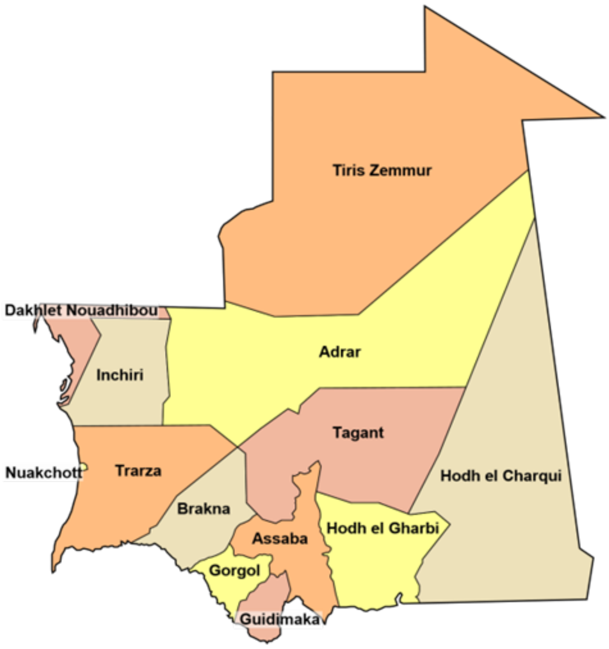


Figure 5. Map of Mauritania with its regions. Source: http://www.mapsopensource.com

1.3.2. Topography and drainage

Mauritania is a merely flat landscape with vast plains which are interrupted by a few rocky outcrops. The center of the country is formed by a series of sandstone plateaus giving with a few spring-fed oases. The most relevant feature of this area is the Guelb er Richat, or "the Eye of the Saraha". It is a greatly eroded dome consisting of a variety of intrusive and extrusive igneous rocks with different resistance to weathering. Due to climatic conditions and the low precipitation regime, there are no significant lakes or rivers. The highest spot of the country is Kediet Ijill with 915 m; while its lowest point is Te-n-Dghamcha with 5 m below the sea level (Gerteiny et al., 2018). In Fig. 6, the digital elevation model of the country can be seen. It has been generated using SRTM dataset with 30 m of spatial resolution. It should be noted that maximum and minimum height differs from the reported data in the bibliography. This fact can be accounted for the spatial resolution of the SRTM sensor, whose image pixels cover an area of 30 m x 30 m = 900 m^2. The relief and drainage of Mauritania are influenced by its arid conditions. The coastal plains are below 45 m, while the higher plains of the centre range from 180 to 230 m. In the interior plains, there are many tablelands with differences in height that are joined by long and smooth slopes of around 2 %. The slope map (Fig. 7) shows that most of the country has between 2 and 5 percentage of slope.

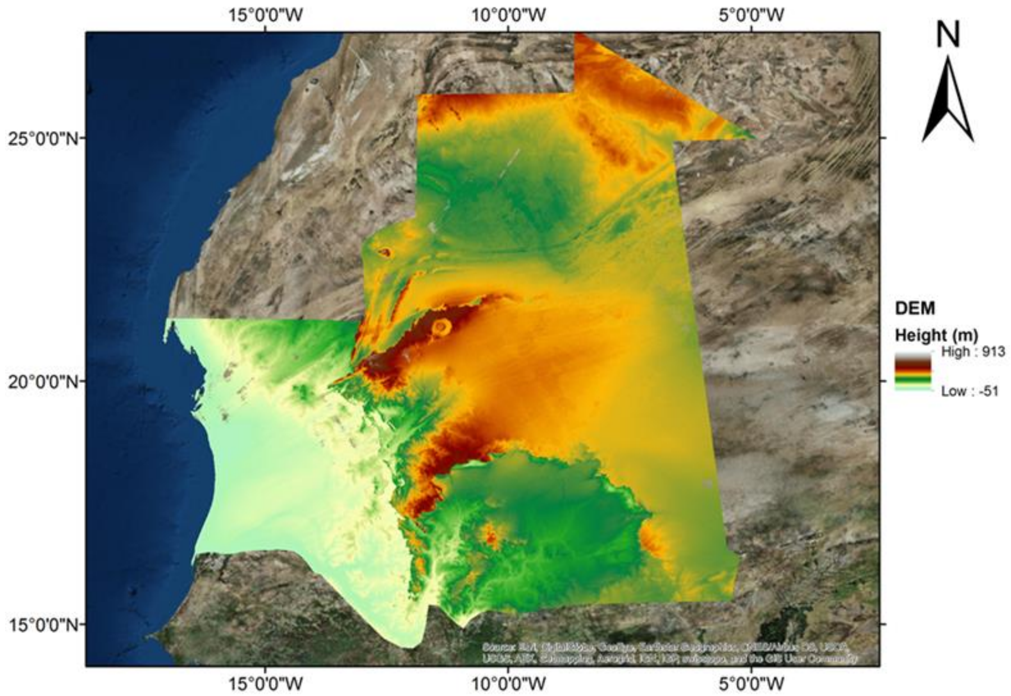

15°0'0"W
10°0'0"W
5°0'0"W
25°0'0"N
20°0'0"N
15°0'0"N
N
DEM
Height (m)
High : 913
Low : -51
Source: Esri, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AEX, Getmapping, Aerogrid, IGN, IGP, swisstopo, and the GIS User Community

Figure 6. Digital Elevation Model of Mauritania with 30 meters of spatial resolution from SRTM dataset.

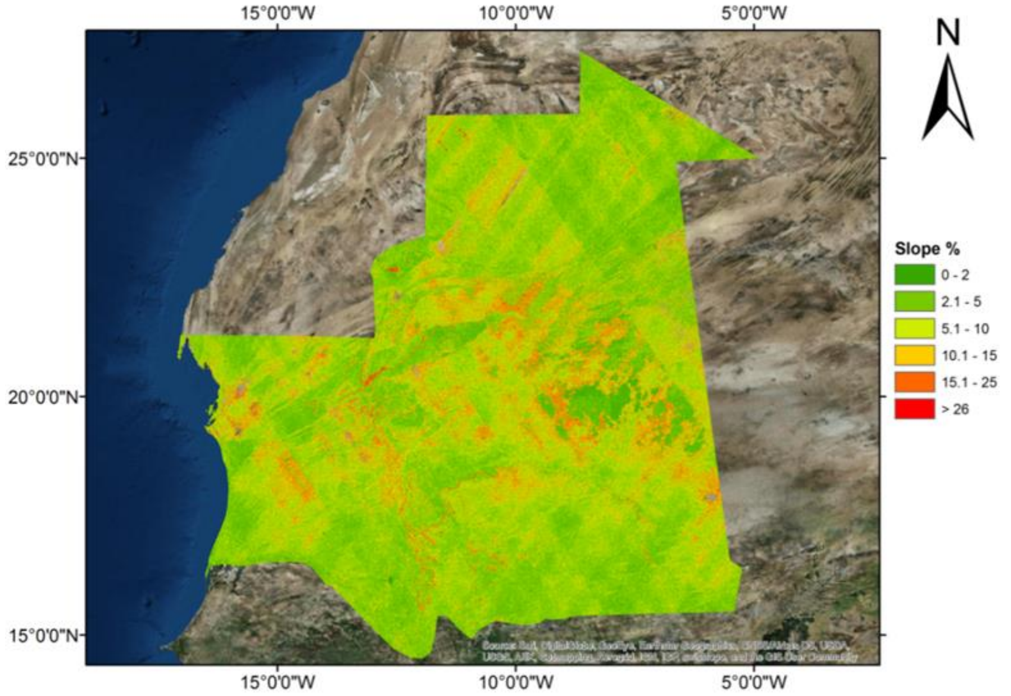

15°0'0"W
10°0'0"W
5°0'0"W
25°0'0"N
20°0'0"N
15°0'0"N
N
Slope %
0 - 2
2.1 - 5
5.1 - 10
10.1 - 15
15.1 - 25
> 26
Source: Esri, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AEX, Getmapping, Aerogrid, IGN, IGP, swisstopo, and the GIS User Community

Figure 7. Slope map of Mauritania generated from DEM-SRTM at 30 m. spatial resolution.

The topographical conditions of Mauritania allow high infiltration rates of the scarce precipitation that falls in the country. The most common slope ranges from 0 – 2 %, with just few areas over 10 %. Rocky outcrops are outlined by slopes greater than 15 % in the centre of the country.

In the south of Mauritania, there are a few seasonal flows which are tributaries of the Senegal River: Karakoro, Gorgol, Kolinbiné. They are subjected to have seasonal floods during the summer months. In the rest of the country, the plateaus are cut by dry river beds or also referred as “wadis”. When these rare floods occur, wadis lead the water and dissipate it over areas called “guelts”. Fig. 8 shows some visual examples of those geomorphological features in the Adrar province of Mauritania. In the north and eastern parts of the country, the precipitation is as rare and slight that hardly ever ends up in runoff (Trape, 2009).

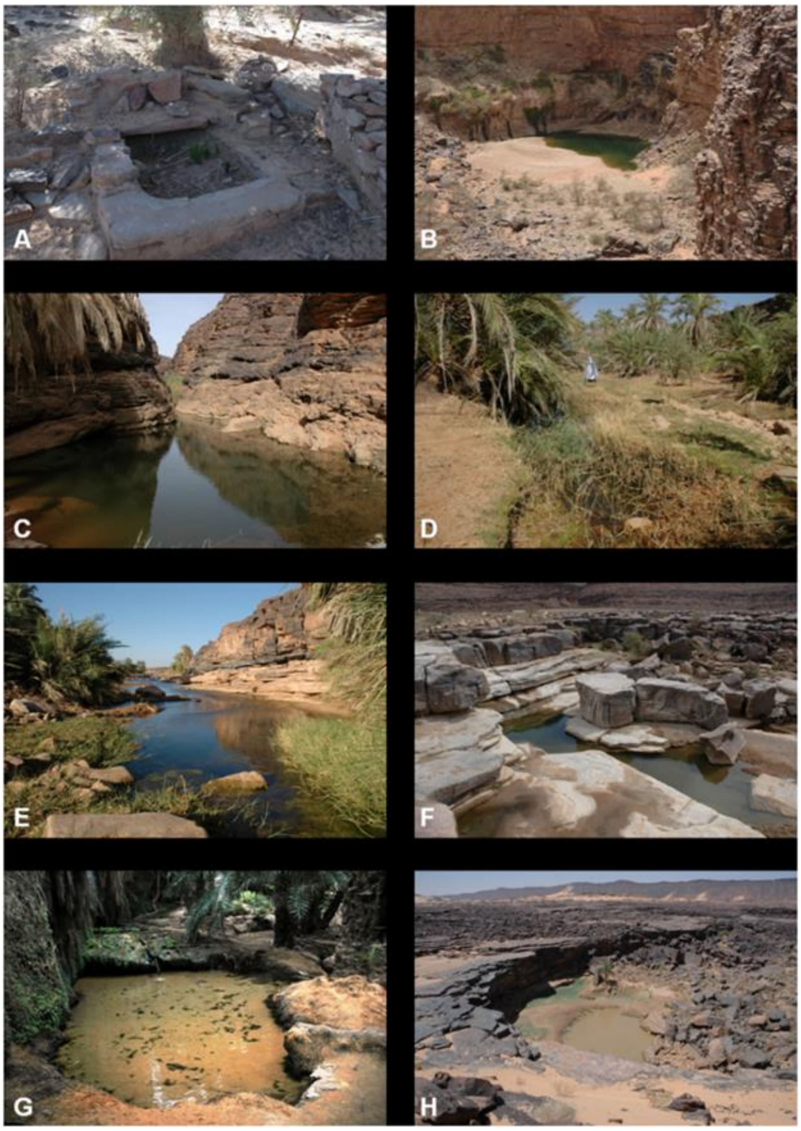


Figure 8. Gueltas and springs from the Seguellîl wadi basin. A: The former spring of Ted at Ksar Torchane; B: Ilîj guelta.; C: The first guelta of Molomhar; D: Pond at Agueni; E: Tachot guelta; F: Hamdoun guelta; G: Terjit springs; H: Toungad guelta. Source: Trape, 2009

1.3.3. Climatic conditions

According to (Gerteiny et al., 2018), the aridity of the Mauritanian climate is due to the north-eastern trade winds, which blow constantly in the north and throughout most of the year in the rest of the country. They have a very sheer drying effect and it is enhanced by the Harmattan season. This is a western African season that occurs between the end of November and the middle of March (Minka et al., 2014). It is featured by dry and dusty trade wind from the north-east or east. Then, it blows from the Sahara Desert over west Africa into the Gulf of Guinea. Depending on local circumstances, these winds affects differently concerning temperature.

Precipitation is generally scarce in Mauritania (Fig. 9). With the exception of few winter precipitation episodes that may occur due to mid-latitude atmospheric disturbances, precipitation is usually conveyed by south or south-west winds.

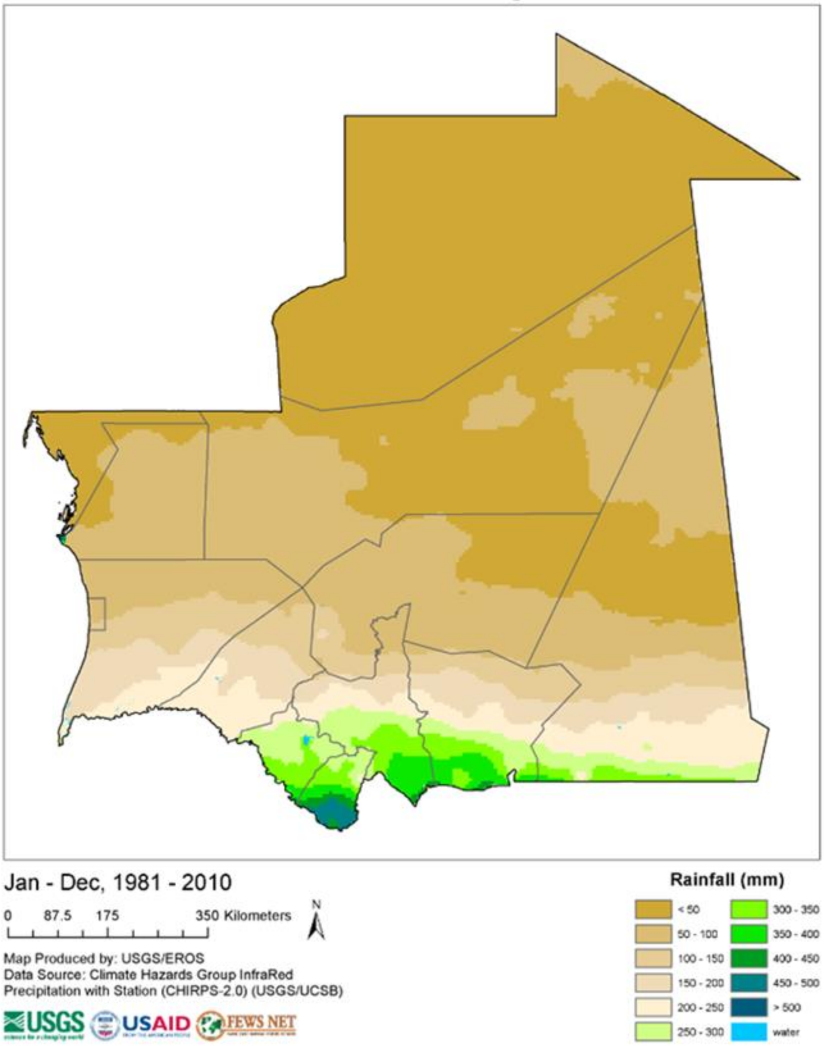


Figure 9. Mauritania historical average rainfall (1981 - 2010). Source: USGS/EROS

In average, Sélibabi (southernmost part) receives by 635 mm between June and October. Whereas Kiffa, which is situated more in the north, has round 355 mm between mid-June and mid-October. In the Tagant province, Tidjikdja has about 180 mm between July and September. In Adrar, Atar receives approximately by 177 mm between mid-July and September and Nouâdhibou, between 25 and 50 mm. during its rainy season: September, October and November, being common the stormy showers (Gerteiny et al., 2018).

Typically, summer month temperatures are rather high. In the afternoons, they can reach 30ºC across the country, while the highest temperatures throughout the day may reach up to 40ºC.

According to Koppen classification (Kottek et al., 2006), two climate types are present in Mauritania: Hot Desert Climate ("BWh") and Hot Semi-arid Climate ("BSh"). The BWh is predominant in most of the country covering round 70 % of the country, which spatially coincides with part of the Sahara Dessert and the Sahelian belt. BSh accounts for the southernmost strip, where the rainfall average is higher, in addition to cooler and less fluctuating "day-night" temperatures in comparison with BWh. The maximum average temperatures on a monthly basis are found from May to October, while precipitation is higher from June to October (Fig. 10). As already mentioned, rainfall is scarce and very difficult to monitor due to the lack of ground based stations, although it is well known that they might be intense and cause floods in some instances (see the link http://floodlist.com/tag/mauritania)

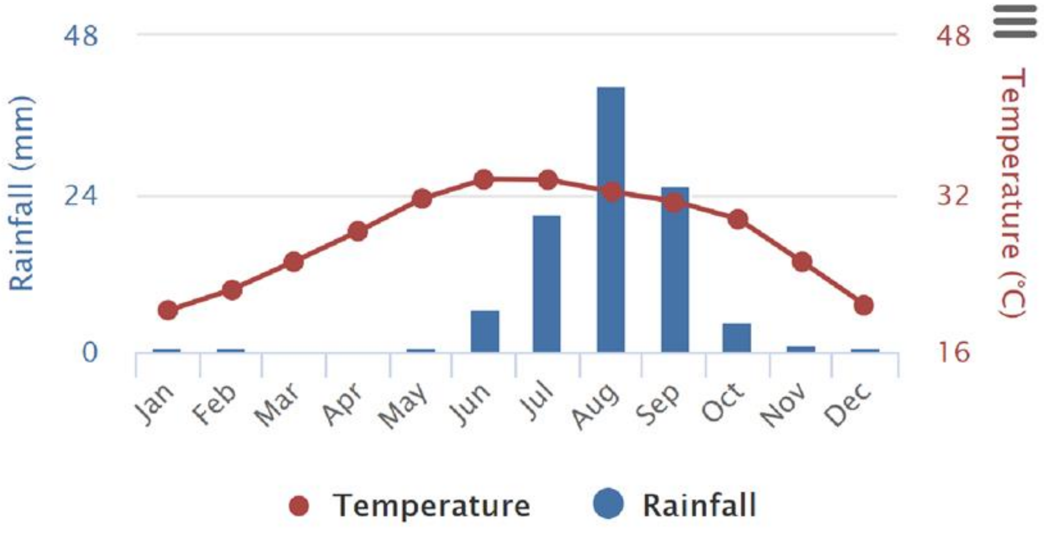


Figure 10. Average Monthly temperature and precipitation in Mauritania from 1901 to 2015. Source: World Bank

1.3.4. Geology and geomorphology

Based on litho-stratigraphical and structural geology, 5 formations are distinguished in Mauritania: Réguibat shield, Taoudeni basin, Tindouf basin, Mauritanide Belt and Coastal Basin (Fig.11).

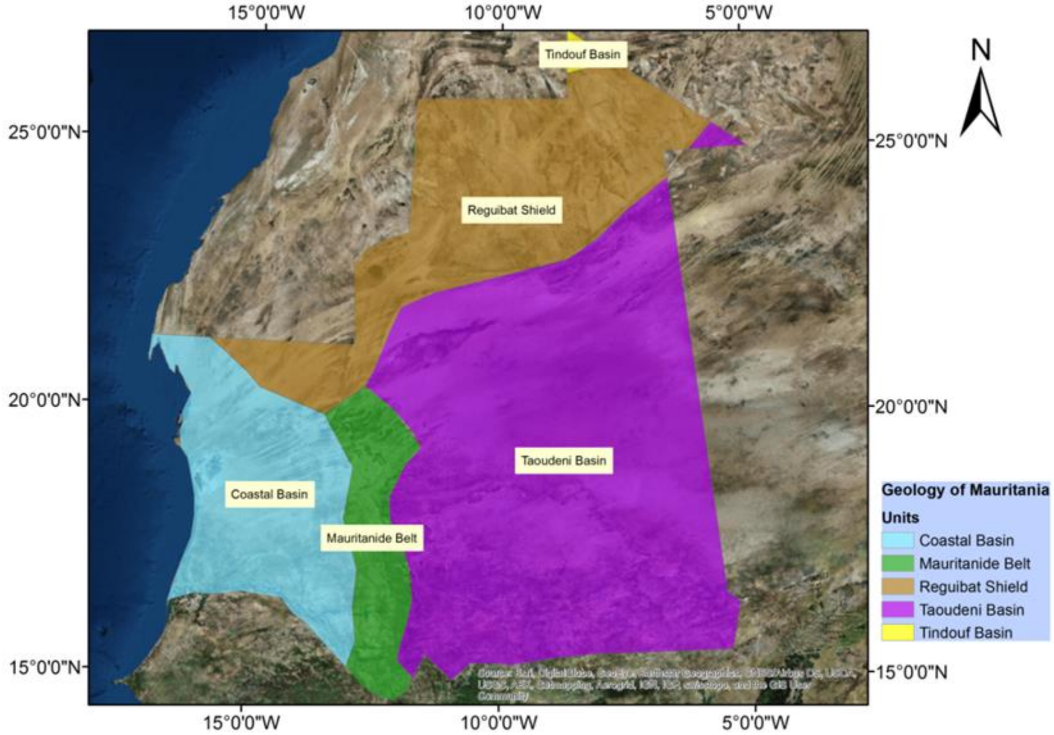


Figure 11. Geological units of Mauritania. Source United States Geological Survey (USGS)

(1) Réguibat shield is the northern part of exposure of the west African Craton that extends along the north of Mauritania and west of Argelia. It consists of Precambrian methamorphic and intrusive rocks (gneisses, greenstones, granites) probably older than 3,000 Ma that lay on the western part, while in the eastern part younger rocks (granite, granodiorite, orthogneiss and pegmatites) between 2,200 and 2,000 Ma in age overlay them. Limestones, dolomites and sand-stones arise flanking the whole shield. It has a complex history of magmatism and orogenesis.

(2) The Taoudeni basin is the largest sedimentary formation in NW Africa and covers greatly the west African Craton in Mauritania and Mali. Sedimentary fill can reach over 3000 meters thick (Wright, 1985) and their age vary from mid-late Proterozoic (Precambrian) to Cretaceous. Sedimentary lithologies such as

sandstones, siltstones, mudstones, turbidite facies, conglomerates, stromatolitic limestones and dolomites are present in the basin (Bradley et al., 2015) outcropping some of them mainly in the Adrar, Tagant and Assaba regions (Perez de Ayala, 2011).

(3) The Tindouf basin, whose depocenter is 8 km deep, is filled with sediments of Cambrian to Carboniferous age (Selley, 1997). Deposits such as sandstones, marlstones, fine sandstones, limestones, marly limestones and evaporitic series are displayed in stratigraphic sequence (Guerrak, 1989) lying discordant over the basement.

(4) The Mauritanide Belt is an orogen that follows a north-south axis between Tindouf and Taoudeni basin. This belt was formed between 320-270 Ma during the Hercynian orogeny and it includes Old Pan-African belt material and Palezoic allochthonous sediments (tectonised and metamorphosed during such orogeny) (Villeneuve, 2005).

(5) The Coastal Basin extends southwards from Nuadibú cape to the estuary of the Senegal River. The depth of the deposits increases westwards, being its depocenter offshore. Quaternary sand dunes mostly cover the inland part of the basin. In Nouakchott, a drilled borehole proves a depth of 5000 meters of basin sediments, although further offshore samples turned out being thicker deposits. Gypsum, anhydrite, salt, green-black clay deposits or pyrite are the oldest sediments, corresponding to the Permian-Triassic system (250-200 Ma). Limestones, sandstones, dolomites, reef carbonate deposits or sands with abundant shells are mainly the sedimentary layers that alternate in sequence (Perez de Ayala, 2011).

1.3.5. Land cover and soil type

The Sahel belt is a transitional eco-climatic and biogeographic region in Africa between the Sahara to the north and the Sudanian Savanna to the south (Huntington et al., 1834). Due to its less arid conditions and higher precipitation rates regarding Sahara, land use and soil types are more diverse. In Mauritania, by the parallel 17.5 ºN, two main zones are separated. Land uses under Saharan conditions (17.5 – 27 ºN) are mainly non-consolidated and consolidated bare areas, being the sand dunes their predominant soil type. However, other varieties are also observed: Arenosols, Leptosols, rock outcrops, Calcisols or Regosols. Under the influence of Sahel conditions (14.5-17.5 ºN), land uses widen (range grasslands, savannas, steppes, shrublands or diffuse agricultural fields), as precipitation regime increases. Soil types such as arenosols, regosols and leptosols are predominant. Lixiol, cambisols, Gleysols, Vertisols and Fuvisols are frequent in the southernmost part due to the influence of the Senegal River and its tributaries. A distribution map of the different land cover and soil map units can be found below (Fig. 12).

The land cover information was extracted from the Globcover regional (Africa) archive (Arino et al., 2007; FAO, 2009). We selected the land cover dataset covering December 2004 – December 2006, with 300 m spatial resolution. The GlobCover is an ESA initiative which began in 2005 in partnership with JRC, EEA, FAO, UNEP, GOFC-GOLD and IGBP. It develops a service capable to deliver global composites and land cover maps, using observations from the 300m MERIS sensor from ENVISAT satellite (Leroy, 2006). The soil information was derived from the Harmonized World Soil Database v 1.2 (FAO, 2012). This dataset provides valuable information with 30 arc-second resolution (1 km) about the basic properties of the soils (texture, drainage, depth, FAO's name, bulk density, available water storage capacity or organic carbon).

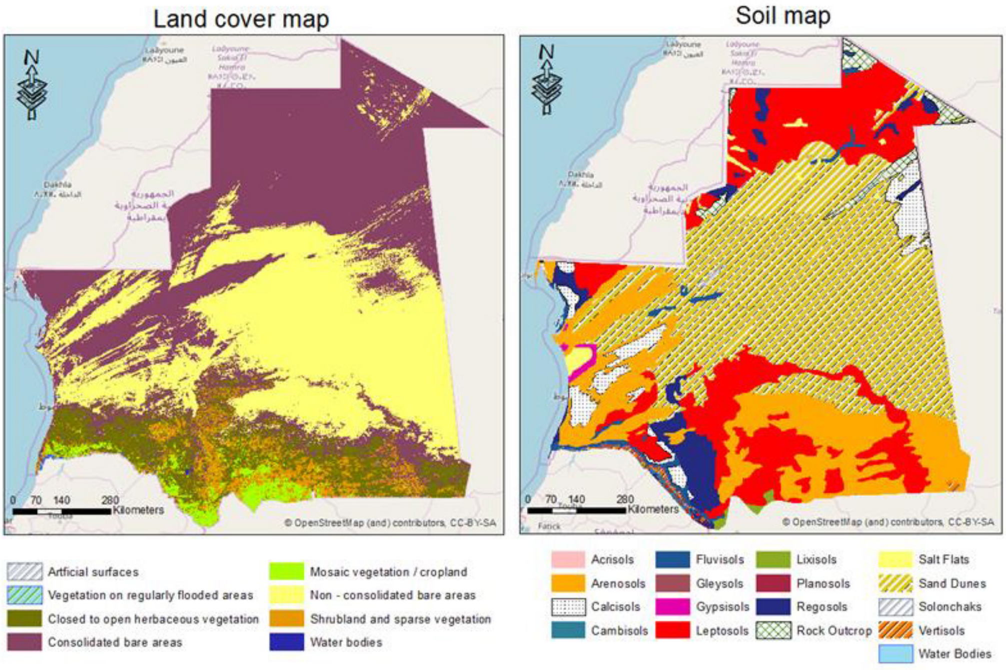


Figure 12. Land cover (ESA - GLOBCOVER) and soil map of Mauritania (FAO - HWSD).

1.3.6. Vegetation and animal life

Vegetation depends on the level of aridity, which increases from south to north. In the south, where the Sahel climatic conditions are prevalent, the area is characterized by a discontinuous belt of vegetation. Trees are rarely found, and the most prevalent individuals are acacias, euphorbia bushes, large tufts of morkba, fields of cram-cram, or Indian sandbur (Cenchrus biflorus, a prickly grass). By the center of the country, the steppe fades off and desert environment takes over. Vegetation is constrained to places such as wadis, in which the water continues flowing underneath or to oases.

In spite of the pressure of hunting, the southernmost part of the country that lies in the steppe is still frequented by gazelles, ostriches, warthogs, panthers, hyenas, and lynx (Gerteiny et al., 2018). Crocodiles are found in the guelta (pocket of water that forms in drainage canals or wadis in the Sahara (Lickens, 2010)). The only big mammal that goes into the desert northwards is the addax antelope.

Birdlife is quite rich in Mauritania, with over 500 different species recorded. The most remarkable ones are scissor-tailed kite, Nubian bustard, Arabian bustard, houbara bustard Egyptian plover, golden nightjar, chestnut-bellied starling, Kordofan lark and Sudan golden sparrow (Wheatley, 1995).

Despite the unfavourable climatic conditions, some insects such as desert locust have been able to adapt quite well to the limiting circumstances of arid or semi-arid environments. But what makes this insect so special is its ability to change its behaviour and cause plagues (Pener & Simpson, 2009).

1.4. The biology of desert locust

Schistocerca gregaria (Forskål, 1775) or desert locust is an insect that belongs to the Acrididae family, having three main stages throughout its life cycle: egg, hopper and adult. With breeding purposes, females lay their eggs when certain moist soil conditions are met from 5 to 10 cm deep (Uvarov, 1977). Depending on some environmental variables such as SM, temperature or wind, the egg development may last between 10 and 65 days (Pedgley, 1981; Symmons & Cressman, 2001).

The nymphs or young locusts are wingless and they moult between 5-6 times as its body grows to prepare the individual for flying and reproduction purposes. To better characterize them, after each shedding they are referred as “Instar” followed by an increasing number up to VI. After the last moult, the new individuals (fledgling) already possess immature wings. When locust wings harden, the individuals obtain fully capabilities to fly and this phase is called immature adults. They have the capacity to migrate from their original breeding area to more suitable environments. After several days, those immature adults become sexually mature and capable to copulate and lay eggs to complete their life cycle (Symmons & Cressman, 2001). During this phase, they are very mobile, with high capacity to migrate aiming for food (Bennet, 1976).

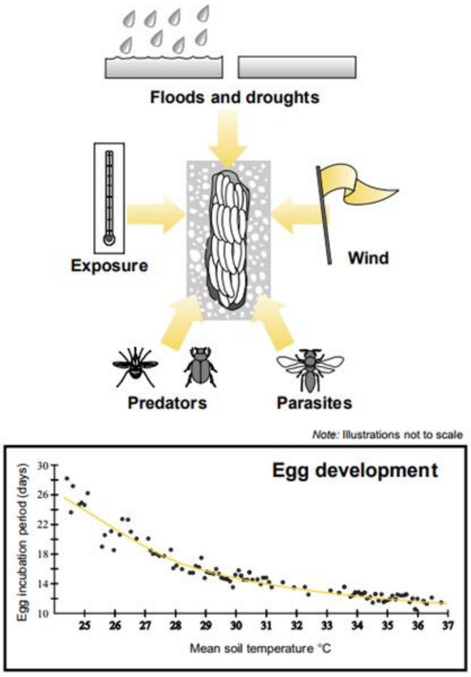


Figure 13. Top image shows some of the main influential variables in egg success. Below image expresses the rate of egg development as function of soil temperature. Egg mortality can be caused by several factors that vary from habitat to habitat. Source: Symmons & Cressman, 2001

Alike other species in the animal kingdom, desert locust has a phase polyphenism that implies drastic changes when population density increases, either in adult or nymph stage (Pener & Simpson, 2009; Simpson et al., 2011; Song et al., 2017). Even though behavioural gregarization may occur within hours (Ellis, 1962), it takes several generations to fully display gregarious characters (Ernst et al., 2015). The phase transition induces physiological changes in lifespan, metabolism, immune responses and reproductive physiology (Pener & Yerushalmi, 1998; Verlinden et al., 2009; Wang & Kang, 2014; Cullen et al., 2017). In their solitarious phase, locusts are generally bigger (Ernst et al., 2015) and they present higher fecundity and smaller eggs (Maeno & Tanaka, 2009).

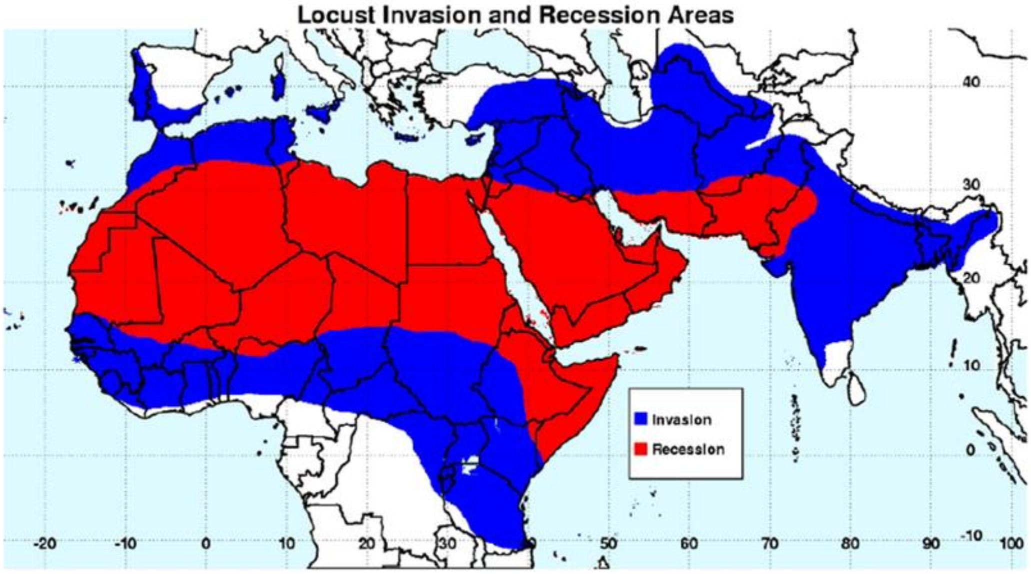


Figure 14. Distribution range of African migratory locust: recession area in red, invasion area in blue. Source: Waloff, 1966

Solitarious desert locust populations are usually constraint into the recession areas (Fig. 14), where annual rainfall is less than 200 mm (Tratalos & Cheke, 2006). However, they are able to increase rapidly their numbers when suitable conditions are met (Pedgley, 1981). These insects are very well adapted to arid environments with erratic but sometimes high intensity precipitation episodes (Uvarov, 1966). Some environmental events such as green vegetation blooms or rainfall are closely linked to the desert locust development, having triggering effects and enhancing outbreaks (Tucker et al., 1985; Hielkema et al., 1986). Temperature variability has also been demonstrated to have effects on some Schistocerca *species* as described by (Yu et al., 2009). This work indicated that the frequency of locust outbreaks may be altered by changes in climatic patterns. Among many environmental factors that may affect locusts, SM is the variable that mostly influences egg-laying location, egg-survival and egg-hatching rate (Liu et al., 2008), in addition to temperature (Nishide & Tanaka, 2016).

Generally, female locusts prefer open and warm sites of dry, soft and sandy soils in which, over 6 cm of depth have enough moist soil conditions (Popov, 1958; Uvarov, 1977). Successful breeding conditions are usually triggered by rainfall which provides enough moisture to the soil enhancing egg laying, development and hatching (Tucker et al., 1985), as well as an adequate vegetation for their hoppers to feed on (Bennett, 1976; Tratalos & Cheke, 2006). The success of preventive measures is subjected to the inaccessibility of some important breeding areas (Symmons & Cressman, 2001). Within the recession area, there are some seasonal breeding locations in which the lack of rain may cause that some are not infested for a particular year. So that, although breeding areas are constraint to the recession area, they may vary in accordance to suitable ecological conditions (Symmons & Cressman, 2001).

1.4.1. Locust Phases

Locusts exhibit two behavioural phases, solitarious and gregarious. Solitary individuals occur at low densities and present individual behaviours (Pener & Yerushalmi, 1998). They breed under favourable habitat circumstances. When vegetation and soil moisture are limiting, desert locust may migrate, regroup themselves in smaller areas with better conditions, or even die (Ceccato et al., 2007). Regrouping means to raise density numbers, and many publications relate this fact with behavioural changes and gregarization (Bouaichi et al., 1996; Despland et al., 2000; Sword et al., 2000). On the other hand, the gregarious phase makes them to fuse into bands or swarms. In adult phase, these groups of individuals move towards favourable environments, with high soil moisture conditions and vegetation. These environmental factors favour mating and breeding activities, while they have enough food supply. However, when the resources become scarce, they move to more favourable areas.

It has been documented that the gregarious behaviour in desert locust is evoked by touching their back legs (Simpson et al., 2001). In this study, desert locust in solitarious phase was subjected to physical contact in different parts of their body by mechanical stimulation. There was a significant change from solitarious to gregarious when the outer face of a hind femur had been stimulated (Fig. 15). Whereas 10 other body regions did not cause the same behavioural reaction.

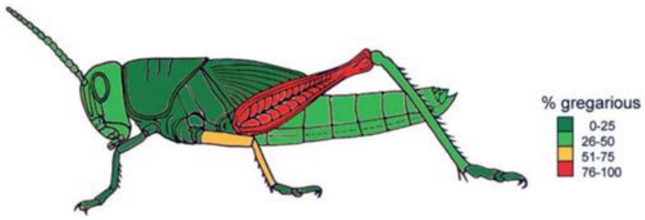


Figure 15. Body-colour image of a locust nymph in the solitarious phase that represents the effect after 4-h of mechanic-stimulation. The colours are based on median values for each treatment group. Source: Simpson et al., 2001.

This change raises the serotonin levels, enhancing colour changes and forming coherent group formation with greater activity (Anstey et al., 2009). To validate this assumption, studies such as (Guo et al., 2013) injected different doses of serotonin concentration into the head cavities of fourth-stadium gregarious nymphs. It was observed that lower doses of serotonin and 30 min of isolation from the group provoked a significant behavioural shift toward the solitarious phase state. The locust behavioural phase also affects the hatching time of the eggs (Nishide et al., 2015).

There is a transition phase between solitarious and gregarious named “transiens”. There have been numerous biological experiments at the individual level to show how this conversion occurs, but the effects of the environment and other stimuli urges to be further explored (Topaz et al., 2012).

1.4.2. Life cycle

Desert locust life cycle has three main stages: Egg, Hopper and Adult (Fig. 16), and each phase length varies according to the environmental conditions of the habitat.

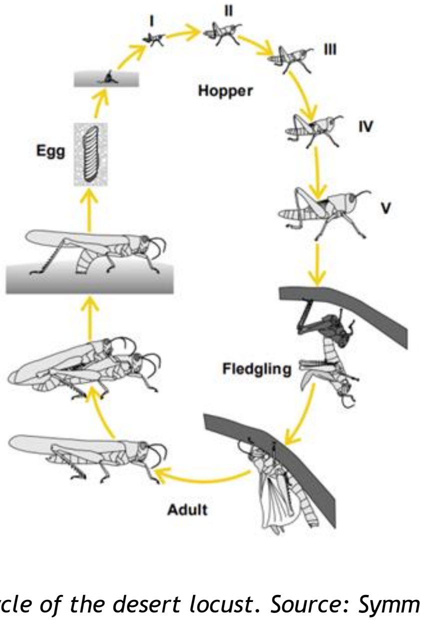
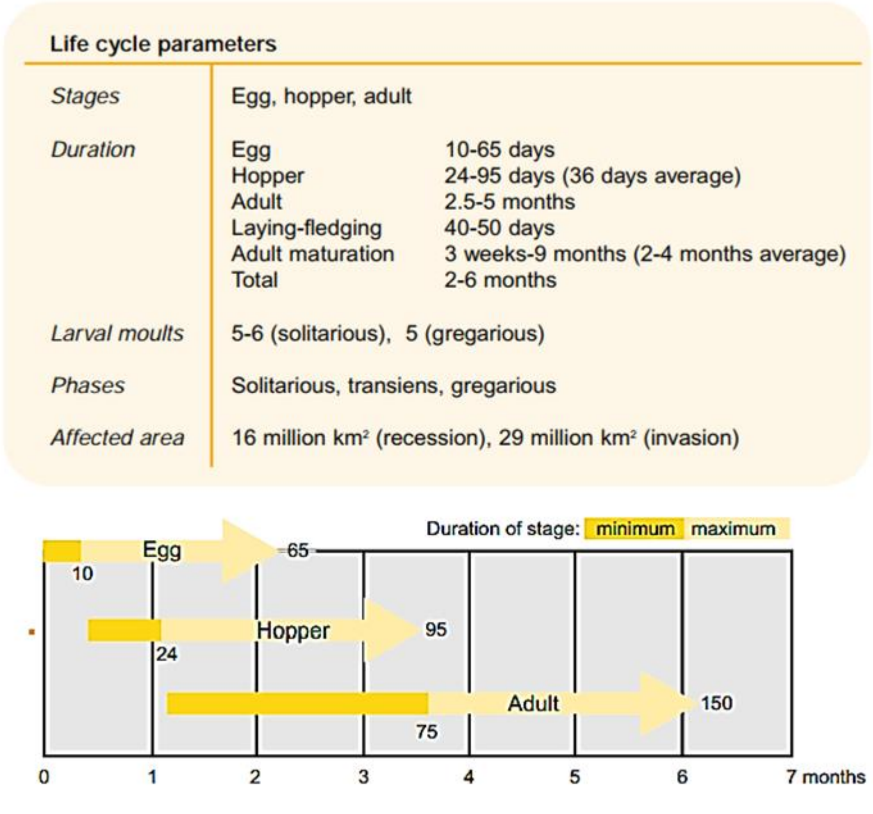


Figure 16. The life cycle of the desert locust. Source: Symmons & Cressman, 2001

As clearly detailed by (Symmons & Cressman, 2001), female individuals lay their eggs under moist soil conditions from 5 to 15 cm deep, with preference on bare soils.



Life cycle parameters

Stages	Egg, hopper, adult	
Duration	Egg Hopper Adult Laying-fledging Adult maturation Total	10-65 days 24-95 days (36 days average) 2.5-5 months 40-50 days 3 weeks-9 months (2-4 months average) 2-6 months
Larval moults	5-6 (solitarious), 5 (gregarious)	
Phases	Solitarious, transiens, gregarious	
Affected area	16 million km² (recession), 29 million km² (invasion)	

Figure 17. Life cycle parameters and Duration of each stage. Source: Symmons & Cressman, 2001

1.4.3. Environmental circumstances of the habitat

During non-limiting food conditions, desert locust are in solitarious phase scattered over zones known as “recession areas” (Tucker et al., 1985). Prior studies have shown the importance of vegetation density to account for phase changes: solitarious to gregarious (Cisse et al., 2013). A decrease in vegetation density seems to be an important factor to enhance the gregarious phase, where the individuals are more voracious and form groups or swarms (Duranton & Lecoq, 1990; Renier et al., 2015). This phenomenon is an adaptation to the extreme conditions of arid environments, where precipitation is scarce but sometimes intense (Uvarov, 1966). Other studies aimed to analyse precipitation (Cressman, 2013; Lazar et al., 2015) or soil moisture (Tucker et al., 1985) to describe good habitat conditions for breeding.

Laboratory studies have demonstrated the sensitivity of eggs from different hopper species to small changes in soil temperature (Nishide et al., 2015). Hence, outbreaks are led by an aggregation of variables and that usually occurs in areas smaller than 10,000 km2 (Van Huis et al., 2007). Outbreaks might not end up being a plague and it may take at least one year to be established as so (Cressman, 2008). According to (Song et al., 2017), desert locust is the earliest diverging species among the genus Schistocerca and the unique settled in Africa. This study suggests that the ancestral Schistocerca species was rather similar with respect to the current desert locust, indicating high adaptability to local conditions. On the contrary, other species of Schistocerca have lost and regained some traits throughout evolution (e.g. ability to change their behaviour).

1.4.4. Migration and seasonal distribution

Desert locust is a migratory species (Dingle, 2009) that moves when the habitat conditions are not favourable. In solitarious phase, the individuals do migrate within the limits of the recession zone (Fig. 14) using the dominant winds. Thus, the Sahel region and the Indo-Pakistan desert are usual summer breeding areas. While northwest Africa and by the Red sea are common winter breeding zones. Nevertheless, changes in precipitation patterns or other environmental circumstances may alter the location of the breeding sites (Symmons & Cressman, 2001). Fig. 18 shows a map of expected zones of migration and usual breeding zones according to the season, corresponding to the date "June 2013".

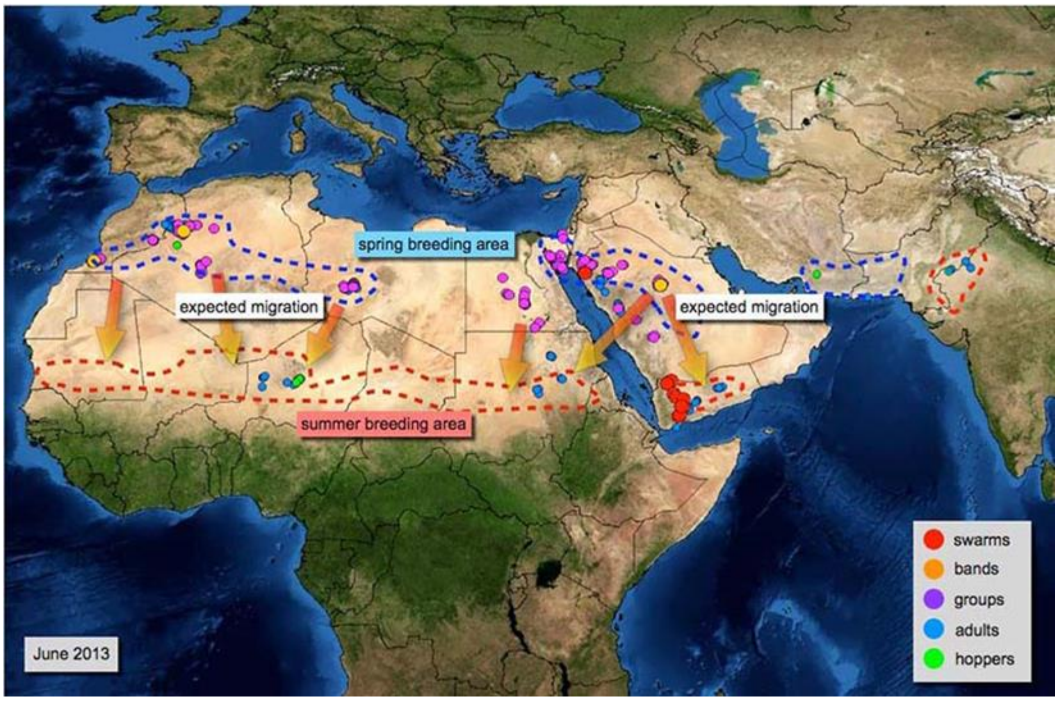


Figure 18. Desert locust shift from spring to summer breeding areas. Source: FAO

1.4.5. Important terms

In order to understand the desert locust problematic, there are a few terms that require further explanation and they are well detailed in (Symmons & Cressman, 2001). The plagues of desert locust occur when certain events make locust populations to grow in number. It is usually started with a calm period (recession) that may experience some punctual outbreaks and upsurges. This situation may lead to either develop a plague, or return to a recession period. The duration of a plague may vary from months to years. There are five important terms that is convenient to explain: Recessions, Outbreaks, Upsurges, Plagues and Declines.

The periods when desert locust are found at low densities and do not cause major damages are referred as “Recession periods”. Locust inhabits areas where agricultural fields are not threatened and hopper bands and swarms are very rare. It is estimated that the recession area that can be seen in Fig. 14 includes more than 30 countries, whose extension goes up to 16 million km^2.

Outbreaks and Upsurges are the link periods between recession and plagues. Outbreak is referred to the period of time, normally some months, in which there is a concentration of individuals that gregarize and multiplicate rapidly. They often go unnoticed by local authorities since locusts groups are disperse, and densities are sheltered in crops and vegetation. Some outbreaks can lead to upsurges when they occur at the same time, and each of them is able to breed successfully during two or three generations at medium to high density, what implies to have individuals in transient to gregarious behavioural phase. It is possible to have several upsurges in different regions at the same time, and they might either merge and generate a major plague or just fade off and disappear without great damages. It is considered that a band or swarm of locusts are plague when their existence overcomes one year, and the affected area is reasonable large.

One plague may mainly be established for two reasons. Firstly, favourable environmental conditions need to remain over long time to enhance breeding. Secondly, the control operations by local authorities should fail. Historical records show plague episodes that lasted by 13 years, and covered around 29 million km^2 (nearly twice in size the recession area) (FAO, 2009).

1.5. Thesis structure, specific goals and objectives

This doctoral thesis is structured into four different sections. The first one introduces the problematic, hypothesis and historical background. In addition, it describes the study area and details the biology of desert locust. The second chapter aims to identify potential wadis using SWAT hydrological model, and relate the results with desert locust presences. The third chapter studies the role of remotely sensed surface soil moisture to locate breeding areas of desert locust. The fourth chapter includes a wide range of environmental variables to build a high predictive model to detect breeding sites. A general discussion will address the key findings and relate them with prior studies and publications related to the topic. Finally, we conclude with a summary of this study and future scopes.

Chapter 2. Identification of potential wadis using SWAT hydrological model

2.1 Introduction

Mauritania, as many other countries within the Sahel region in Africa, suffers severe water scarcity. This resource is essential in any ecosystem, and its quantity imposes thresholds to living organisms. In addition, it is a key resource for humans, our well-being and economic activities such as agriculture; grazing or industrial production lay upon it.

In developing countries with arid and semi-arid environments, that is an extra drawback to achieve poverty mitigation and sustainable development, being likely to cause serious conflicts among neighbouring countries (Klemas & Pieterse, 2015). Furthermore, these areas are more vulnerable to Climate Change effects on water resources (Beuhler, 2003). Variations in the water cycle are expected; hence water availability and demand will vary accordingly. Globally, irrigation water may decrease (Haddeland et al., 2014) and precipitation patterns change (IPCC, 2014), affecting agriculture, ecosystems and fresh water supply. At this stage, it urges to reduce the vulnerability of these areas, and mitigate Climate Change impacts on water resources by means of proper assessments (Xia, 2016).

According to (Huang et al., 2010), one-third of the Earth surface is under arid or semi-arid climatic conditions, and water availability conditions forms of life. In Mauritania, agricultural activities are only found in the south of the country, where precipitation regimes are higher (Fig. 19).

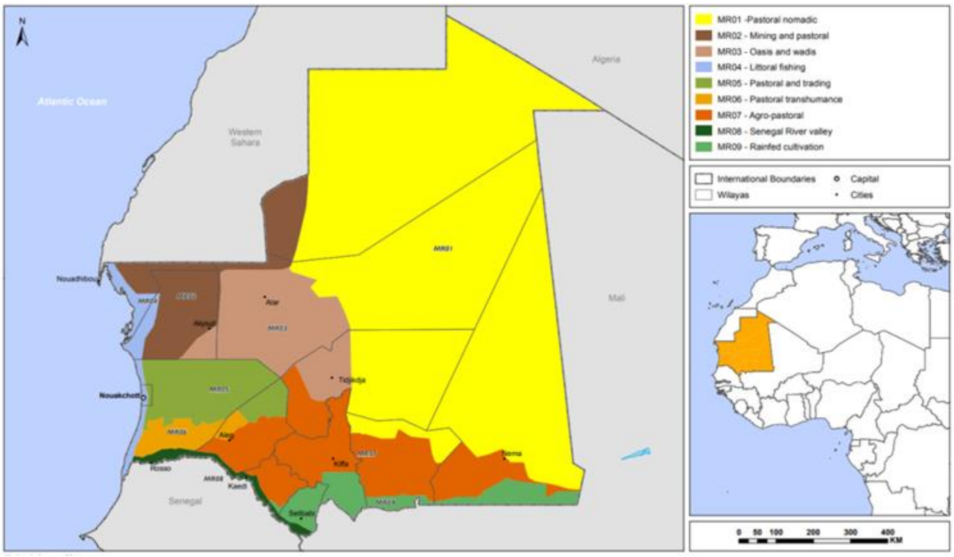


Figure 19. Livelihood areas within the study area (Mauritania), which are conditioned by water availability. Source: United States Agency International Development (USAID) formed by members of USGS, USDA, NASA, NOAA.

Under low precipitation regimes, groundwater is the major water supply for agricultural and domestic uses. Wadis are ephemeral dry rivers that temporarily drain arid or semi-arid areas after heavy rain episodes. Owing to their erratic occurrence and severe energy conditions, the river bed is poorly sorted from clay to gravel range, with high permeability rates. Many publications agree on its importance to alleviate water shortage, and their implications at local and regional scale, such as groundwater recharge (Subyani, 2004), agriculture and human settlements (Ward et al., 2001), richer biodiversity (Springuel et al., 1997; Ali et al., 2000), higher water table level (Edmunds, 2002), breeding areas for insects such as desert locust (Van Der Werf, 2005) or punctual flood hazards that may neglect crops, towns and roads.

Wadis constitute a principal breeding site for desert locust populations (https://earthobservatory.nasa.gov/IOTD/view.php?id=2799). A convenient soil moisture content provides a perfect environment to egg development and hatching, in addition to provide green and fresh vegetation that offer shelter and food to the new-born hoppers (Fig. 19). Seasonal breeding sites are also associated with sandy cultivated wadis (Popov, 1958; Stower et al., 1958). These first signs found in bibliography highlight the role of wadis and its intrinsic environmental characteristics to host desert locust, at least at early stages. In particular, the most vulnerable areas are at the wadi outflow deltas with usual presence of cultivated crops and earth dams to store the water from the floods (Maxwell-Darling, 1936; Kassas, 1957). Moreover, these areas present the finest and more fertile sediments to support lush vegetation growth (Woldewahid, 2004).

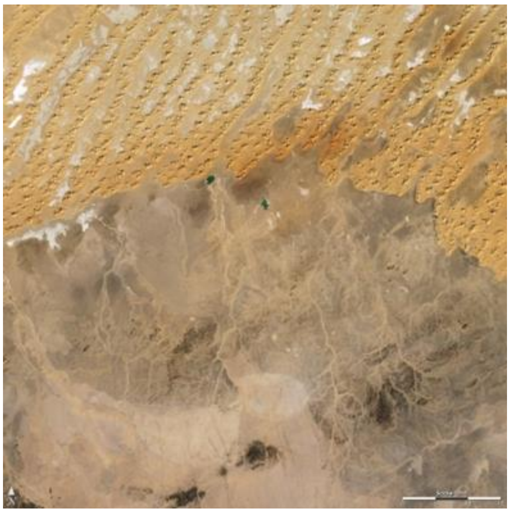


Fig 19. Wadi image acquired by Landsat 5-TM over the Sahel region in Africa. Source: Image by Robert Simmon, NASA GSFC, based on Landsat Thematic Mapper data archived by the Global Land Cover Facility.

Conventional remote sensing methods of wadi identification tend to fail for the following reasons (Liu et al., 2016): Physical shape methods are not effective facing numerous geomorphological land units and irregularities, weak spectral contrast with background, and there is a conflict between global and local accuracy. Wadi systems are complex, anisotropic and with numerous tiny tributaries. To better understand the water availability, as well as the natural dynamics in arid regions, (Klemas & Pieterse, 2015) suggests improving the water management and monitoring by means of remote sensing and conventional hydrologic measurements together. Remote sensing techniques on their own have shown to be insufficient in order to detect potential runoffs in arid or semi-arid environments. And ground mapping on the field would be effective but very costly and time consuming.

This chapter aims to identify potential runoffs or wadis in Mauritania, based on the hydrological model Soil and Water Assessment Tool (SWAT). Given the importance of these hydrological systems in arid and semi-arid environments, wadi identification has been addressed to assess its relationship with desert locust presences.

2.2. Materials and Methods

2.2.1. Input data

2.2.1.1. SWARMs database

Schistocerca WARning and Management System (SWARMS) is a database used by the Desert Locust Information Service (DLIS) at FAO for Desert Locust global monitoring and early warning. It compiles desert locust data since 1985 that has been collected by national survey and control teams of affected countries. It geo-locates field observations on a daily basis, although some uncertainties may be expected (Javaar Bacar, 2011; Renier et al., 2015). For this study, we selected hoppers on a solitarious phase as the target population because it has reduced mobility (lack of wings) so that hopper records have likely been born on the area, benefited from favourable environmental conditions. There were 12627 hopper sightings for the time span 1985-2017. Even though the database contemplates absence records, they were not considered for two reasons. During recession periods, individuals are mostly solitarious (solitarious phase) and many times go unnoticed for survey teams (Meynard et al., 2017). And second, the overall of absences is very low in comparison with presence records so that they are very unbalance.

2.2.1.2. Soil data (Harmonised World Soil Database)

Soil information was derived from the Harmonized World Soil Database (HWSD) v 1.2 (FAO, 2012). Soil Unit Composition of each pixel provides information about sixteen soil properties for topsoil (0-30cm) and subsoil (30-100cm) according to the FAO Revised Legend (FAO, 1990).

These are the properties: Organic Carbon, pH(H2O), Cation Exchange Capacity in soil and clay, Total Exchangeable Bases (TEB), Base saturation %, Sodicity, Calcium carbonate, Gypsum, Sand fraction, Silt fraction, Clay fraction, Salinity, USDA Texture, Reference Bulk Density, Soil Drainage, and Soil Phase information. The resolution of this dataset is 30 arc-second (~ 1 km for our study area). This comprehensive harmonized soil information is of great importance to understand local land and water limitations, land potential productivity, soil erosion or biodiversity (Nachtergaele et al., 2010). For the purpose of this chapter, it helped us to complete the soil information input required by the Soil and Water Assessment Tool (SWAT) hydrological model.

2.2.1.3. Digital Elevation Model (SRTM)

Topographic information was extracted from NASA's Shuttle Radar Topography Mission (SRTM) with a 1 arc-second, or about 30 metres resolution.

The NASA/NGA Shuttle Radar Topography Mission (SRTM) retrieved interferometric radar data to generate a near-global topography data product for latitudes within 60 degrees' north latitude and 54 degrees' south latitude (Rodriguez et al., 2006). It was an 11 - day mission accomplished in February 2000. In order to gather topographic elevation data of Earth's surface, SRTM used the technique of interferometry where two images are taken from slightly different viewpoints of the same area. The little difference between both images enables scientists to identify the surface's elevation of the terrain. This data can be downloaded free of charge at http://srtm.csi.cgiar.org/SELECTION/inputCoord.asp.

Most voids have been filled with elevation data from SRTM 3 arc-second or about 90 metres so as to obtain a continuous Digital Elevation Model (DEM).

2.2.1.4. Global cover – Land cover

The LULC (Land Use/Land Cover) of Mauritania is a derived product from the original raster based on Globcover regional (Africa) archive (Arino et al., 2007). Having two available products, we chose the land cover dataset covering December 2004 – December 2006, with 300 m spatial resolution. The GlobCover is an ESA initiative which began in 2005 in partnership with JRC, EEA, FAO, UNEP, GOFC-GOLD and IGBP. It develops a service capable to deliver global composites and land cover maps, using observations from the 300 m MERIS sensor from ENVISAT satellite (Defourny, 2006). The MERIS 300 m Full Resolution Full Swath (FRS) products are the unique data source of the GLOBCOVER project.

MERIS was a wide field-of-view imaging spectrometer on-board ENVISAT satellite, which was launched in 2002. The 15 spectral bands of the sensor covered from about 412.5nm to 900nm (Rast et al., 1999). The field of view angle of the instrument was 68.5 ° around nadir, covering a swath width of 1150 km at 800 km height, what made a global coverage of the Earth in 3 days (Bicheron et al., 2008). The processing chain that was used to generate and deliver land cover maps can be seen in Fig. 20.

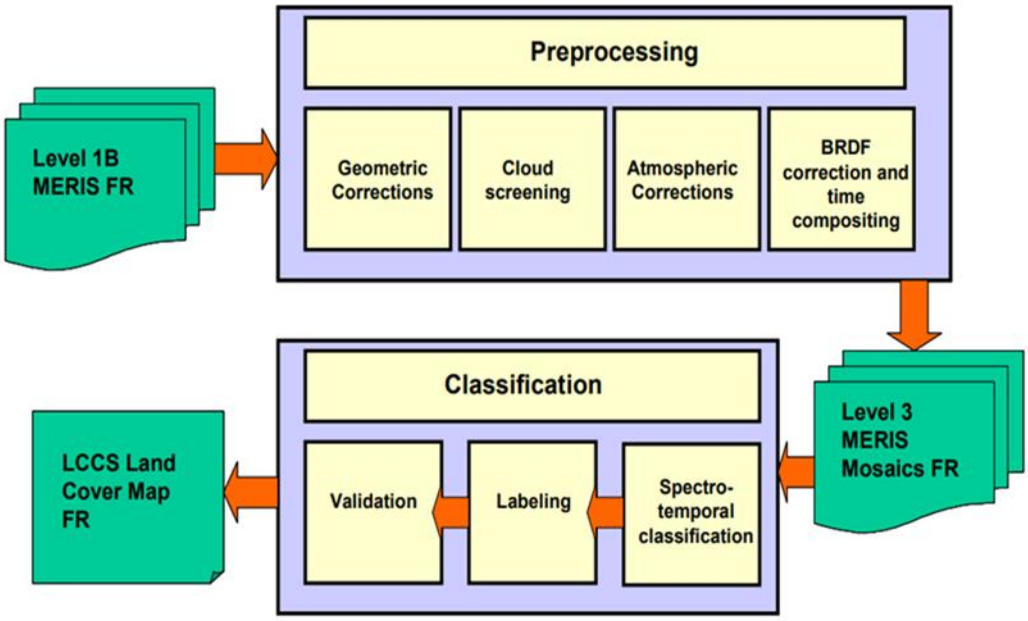
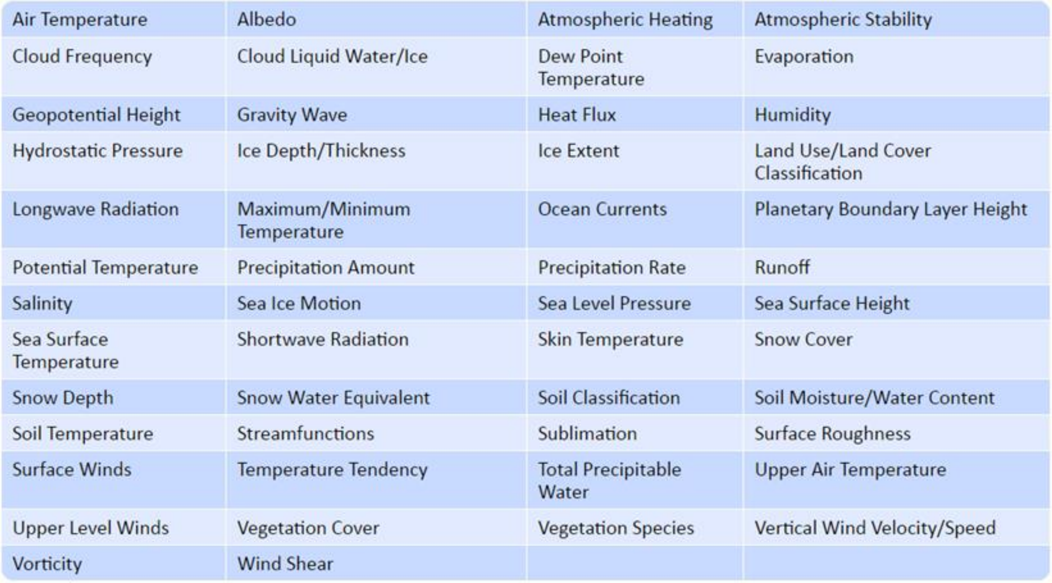


Figure 20. Algorithmic principle of the Globcover chain. Source: Bicheron et al., 2008

2.2.1.5. CFSR (Rainfall, Insolation, Air Temperature, Relative Humidity, Wind)

The Climate Forecast System Reanalysis (CFSR) is a third generation reanalysis product that covers from 1979 to 2017. It is a global, high resolution and coupled atmosphere-ocean-land surface-sea ice system designed to provide the best estimate of the state of these coupled domains over this period (National Center for Atmospheric Research Staff, 2017). The CFSR includes the following information: (1) coupling of atmosphere and ocean during the generation of the 6 hour guess field; (2) an interactive sea-ice model; and (3) assimilation of satellite radiances.

The CFSR global atmosphere data has an approximate resolution of 38 km. The global ocean resolution is 0.25 ° at the equator, extending to a global 0.5 ° beyond the tropics (Saha et al., 2014). This reanalysis is considered superior to previous National Centers for Environmental Prediction (NCEP) reanalysis for the following reasons (He & Zhao, 2018): improved model, finer resolution, advanced assimilation schemes, atmosphere-land-ocean-sea ice coupling, assimilates satellite radiances rather than retrievals. Nevertheless, some uncertainty is expected owing to differences between reanalysis data and observations. And this could be caused by errors in the numerical model, quality of the observed data or errors in the assimilation system (Bengtsson et al., 2004). The CFSR product consists of periodic weather forecasts (every hour) provided by the National Weather Service's NCEP Global Forecast System. Every hour, the CFSR analysis includes: forecast data, predicted from the previous analysis hour, and the data from the analysis utilized to reinitialize the forecast models every six hours. The CFSR dataset contains a wide range of historic expected variables (Fig. 21) for each hour for any land location in the world (Saha et al., 2010).



Air Temperature	Albedo	Atmospheric Heating	Atmospheric Stability
Cloud Frequency	Cloud Liquid Water/Ice	Dew Point Temperature	Evaporation
Geopotential Height	Gravity Wave	Heat Flux	Humidity
Hydrostatic Pressure	Ice Depth/Thickness	Ice Extent	Land Use/Land Cover Classification
Longwave Radiation	Maximum/Minimum Temperature	Ocean Currents	Planetary Boundary Layer Height
Potential Temperature	Precipitation Amount	Precipitation Rate	Runoff
Salinity	Sea Ice Motion	Sea Level Pressure	Sea Surface Height
Sea Surface Temperature	Shortwave Radiation	Skin Temperature	Snow Cover
Snow Depth	Snow Water Equivalent	Soil Classification	Soil Moisture/Water Content
Soil Temperature	Streamfunctions	Sublimation	Surface Roughness
Surface Winds	Temperature Tendency	Total Precipitable Water	Upper Air Temperature
Upper Level Winds	Vegetation Cover	Vegetation Species	Vertical Wind Velocity/Speed
Vorticity	Wind Shear		

Figure 21. Summary of the available variables contained in the CFSR dataset. Source: NCEP

CFSR dataset was used to overcome the lack of meteorological data provided by ground-based stations. Ground based meteorological stations are scarce in Africa what makes difficult to monitor reliably weather variables at large scale (Van den Berg & Feinstein, 2011). Products such as the CFSR dataset is a valuable option as reported in many publications (Dile & Srinivasan, 2014; Fuka et al., 2014; Monteiro et al., 2016; Worqlul et al., 2017). For the purpose of this study, we selected a daily product of the following variables: Temperature (°C), Wind speed (m/s), Air Relative Humidity (%) and Solar Radiation (MJ/m^2).

2.2.1.6. Precipitation data from ground based stations

The Global Historical Climatology Network (GHCN) is an integrated database of climate summaries from land surface stations across the globe. The data are obtained from more than 20 sources. Some datasets are more than 175 years old, and their frequency may be down to 1 hour. GHCN is the official archived dataset, and it serves as a replacement product for older National Centers for Environmental Information (NCEI) maintained datasets that are designated for daily temporal resolution (https://www.ncdc.noaa.gov/data-access/land-based-station-data/land-based-datasets/global-historical-climatology-network-ghcn).

There are only 13 ground based stations in Mauritania with some missing data in their records. Nevertheless, they can provide valuable information to calibrate and validate the SWAT model in smaller areas. Given the availability of other data sources, Kaedi station was the only one that was used to calibrate and validate the model in a smaller scale. Located at latitude: 16.16 °, longitude: -13.508 ° and 22.9 m. of elevation, Kaedi station provides an important approximation of the rainfall over the Ghorfa basin.

2.2.1.7. GRDC river gauge data

The Global Runoff Data Centre (GRDC) is a repository for the world's river discharge data and associated metadata that compiles records from more than 9,300 stations in 160 countries. This international archive covers up to 200 years of short and long term hydrological studies. The main aim of the GRDC organization is to facilitate data to researches, universities and other organizations to carry out environmental and global climate trends. Discharge data and data records have been recorded either on a daily or monthly basis for non-commercial applications.

The GRDC operates with the support of the World Meteorological Organisation (WMO) and the research on climate variability and global change. The German Federal Institute of Hydrology (Bundesanstalt für Gewässerkunde or BfG) hosts the GRDC in Koblenz. Further information about this institution or its data can be found at the following site: http://www.bafg.de/GRDC/EN/Home/homepage_node.html.

Only one suitable dataset was available within the study area, the Ghorfa Aval station. It is located at 15.53 °N and 12.7 °W and has measured river gauge data of the Ghorfa river from 1979 to 1985, covering an approximate drainage area of 5050 km^2.

2.2.2. Methods

2.2.2.1. Model description and model components

In arid and semiarid environments, the identification of zones that are susceptible to water drainage is challenging. Owing to the scarcity of rainfall and its intensity, the water tends to runoff by channels that throughout the year are dried. In some instances, these wadis end up in pools with no outlet point. Seemingly, they create a favourable environment as breeding location for desert locust. In order to identify and locate wadis in Mauritania, SWAT (Soil and Water Assessment Tool) hydrological model was used. It has been developed by the US Department of Agriculture - Agricultural Research Service (USDA-ARS) and Texas A&M AgriLife Research.

SWAT simulates quantity and quality of surface and ground water, as well as it estimates environmental impacts on land use, management, nutrient cycle or Climate Change within a watershed over long periods of time. This model is continuous, semi-distributed and physically based and aims to provide a better insight about runoff, transmission losses, groundwater recharge, evapotranspiration, erosion rate, crop growth, biomass estimation, soil moisture content, irrigation management, groundwater flow, reach routing, nutrient and pesticide loading (Neitsch et al., 2011).

SWAT models an entire river watershed on a topographic basis and divides it into smaller and linked catchment areas, “sub-basins”. These sub-basins are dissociated in Hydrological Response Units (HRUs) that result from Land Use – Land Cover (LULC), soil type and slope combination so as to identify areas with homogeneous response, thus increasing the spatial accuracy of the model on a daily time step. SWAT model comprises two phases: land phase and channel/routing phase (Neitsch et al., 2011). The model requires certain input datasets such as elevation, land use, soil type and meteorological data.

In order to analyse the prediction uncertainty of the SWAT model, SWAT-CUP (SWAT Calibration Uncertainty Program) was used (Abbaspour et al., 2013). This software integrates several uncertainty/calibration analysis approaches. In this study, the algorithm SUFI-2 (Sequential Uncertainty Fitting Version 2) was used to carry out sensitive analysis, calibration and validation for the unique basin where data was available. This algorithm is very efficient in localizing optimum parameter ranges and in terms of number of simulations (Schuol et al., 2008).

2.2.2.2. Hydrological Processes

SWAT allows a number of different physical processes (Fig. 22) to be simulated in a watershed such as surface runoff, infiltration, evapotranspiration (ET), lateral flow, percolation to shallow and deep aquifers and channel routing (Arnold et al., 1998). The tool has also a weather simulation model that is capable to generate daily data for rainfall, solar radiation, relative humidity, wind speed or temperature from the average monthly variables of these data, in case there is some missing data in the observed meteorological variables.

The CFSR climate dataset was used to incorporate the following meteorological variables: wind speed, solar radiation, maximum and minimum temperatures and relative humidity. Precipitation is the driving force in hydrological processes in arid and semi-arid environments, so that we aimed to achieve the most accurate available dataset over the region, the Kaedi ground based station. The coverage time of our model ranges from 1979 to 1985, seeking for a temporal overlap with the available river gauge data to calibrate and validate the model.

The Land phase controls the amount of water, sediment, nutrient and pesticide loadings to the main channel in each sub-basin. The hydrological cycle simulated in SWAT is based on the water balance (Equation 1):

$$SWt = SW_{\circ} + \sum_{i=1}^{t}(Rday - Qsurf - Ea - Wseep - Qgw) \qquad \textit{(Eq. 1)}$$

Where:

- SWt and SW_o are the final and initial soil water content on day i (mm)
- t is the time (days)
- Rday (mm H_2O) is the overall precipitation on day i
- Qsurf (mm) the surface runoff on day i
- Ea the evapotranspiration on day i (mm)
- Wseep the water that enters into the soil unsaturated zone on day i (mm)
- Qgw is the return flow on day i (mm) (Neitsch et al., 2011).

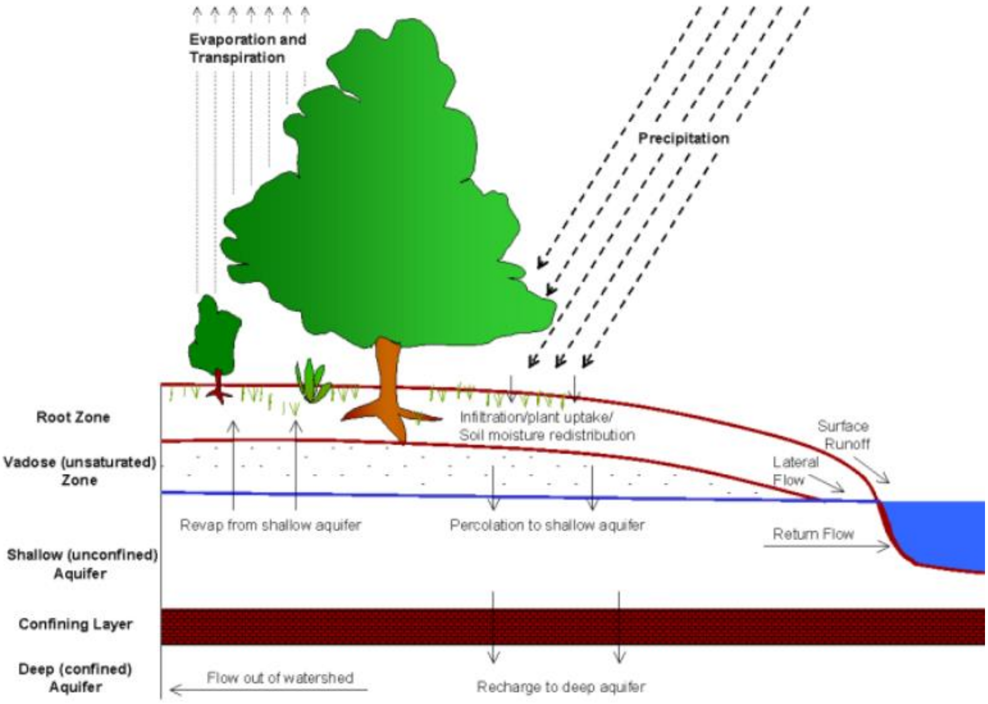

Evaporation and Transpiration
Precipitation
Root Zone
Vadose (unsaturated) Zone
Shallow (unconfined) Aquifer
Confining Layer
Deep (confined) Aquifer
Infiltration/plant uptake/ Soil moisture redistribution
Surface Runoff
Lateral Flow
Revap from shallow aquifer
Percolation to shallow aquifer
Return Flow
Flow out of watershed
Recharge to deep aquifer

Figure 22. Schematic representation of the hydrologic cycle in SWAT (Neitsch et al., 2011)

Surface runoff occurs when the water provided to the ground surface is larger than the rate of infiltration (Fig. 23). When a soil is dry, the rate of infiltration is very high, and then the rate decreases as the soil moisture increases. If more water is provided than it is possible to infiltrate, the runoff would start off (Neitsch et al., 2011). SWAT model has two approaches to estimate surface runoff: SCS curve number (SCS, 1972) and the Green & Ampt infiltration method (Green & Ampt, 1911). The Soil Conservation Service (SCS) curve number (CN) is a function of the soil's permeability, land use and antecedent moisture conditions (Soil Conservation Service, 1972) whereas the Green and Ampt infiltration method calculates infiltration as a function of the wetting front metric potential and effective hydraulic conductivity (Green and Ampt, 1911).

SWAT was set to use a derived SCS-Curve Number method (Soil Conservation Service, 1972). It was chosen because in areas where precipitation is scarce, random and brief, it seems more convenient to use daily overall amounts rather than intensities, which would be inaccurate. In addition to that, this approach has been selected to have been widely used in prior works under similar arid conditions (Mohammad & Adamowski, 2015; Adam et al., 2017).

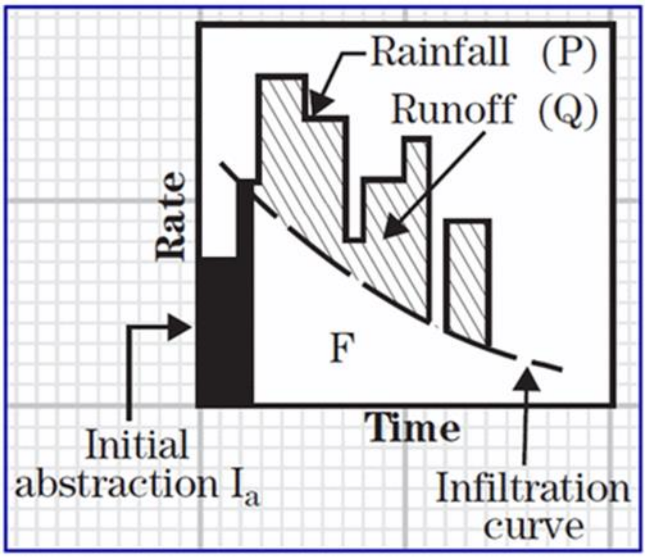


Figure 23. Components of SCS Runoff equation. Source: Patel, 2016

The SCS Curve Number method is an efficient and widely used method to estimate the runoff that rainfall may provoke in a particular area. It was originally developed to measure singular storm events, although nowadays it is also used to infer average runoff values across time. It requires only a few parameters: the amount of rainfall and the Curve Number which is based on land use, hydrologic soil group, hydrologic condition and treatment of the area. The general equation for the SCS Curve Number approach is explained following TR-55 (USDA-NRCS, 1986):

$$Q = (P - Ia)^2 / (P - Ia) + S \qquad \textit{(Eq. 2)}$$

Where:

- *Q = runoff (in)*
- *P = rainfall (in)*
- *S = potential maximum retention after runoff begins*
- *I_a = initial abstractions (in)*

Initial abstraction (mm) consist of the losses before the runoff begins, and it includes water retained in surface depressions, water intercepted by vegetation, evaporation, and infiltration. It is highly variable but generally is correlated with soil and cover parameters, and it is approximated to the following equation 3:

$$Ia = 0.2 * S \qquad \textit{(Eq. 3)}$$

If we substitute I_a in equation 2, we obtain:

$$Q = (P - 0.2 * S)^2 / (P + 0.8 * S) \qquad \textit{(Eq. 4)}$$

Where S is related to the soil and cover conditions of the watershed through the CN. CN has a range of 0 to 100, and S is related to CN by:

$$S = 1000 / CN - 10 \qquad \textit{(Eq. 5)}$$

The following figures (Fig. 24-25) solve equations 3-4 and 3-5 for a range of CN's and rainfall.

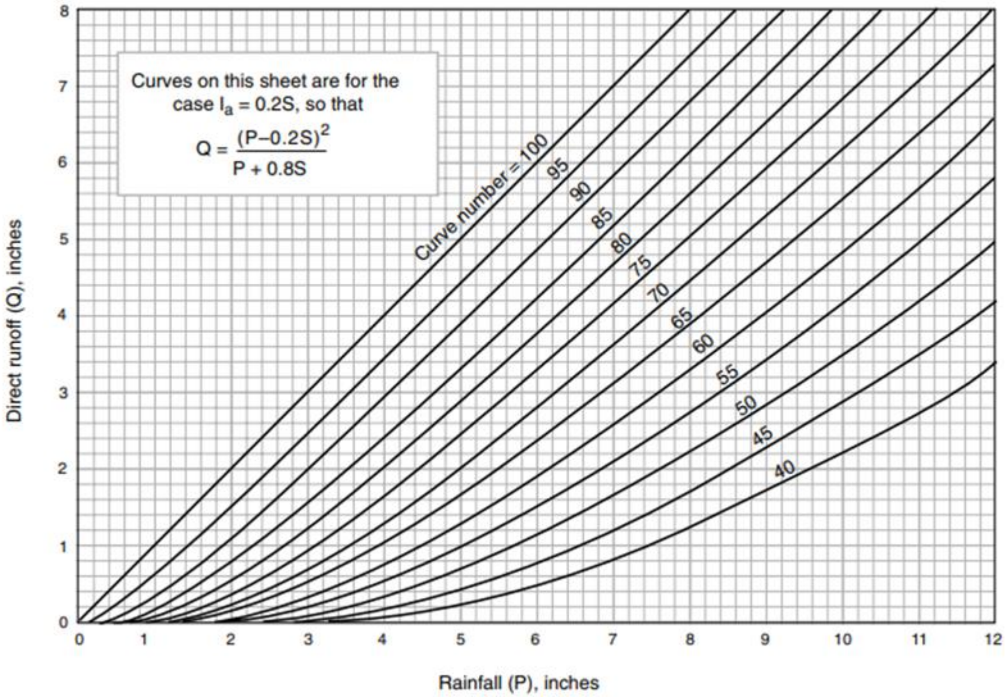
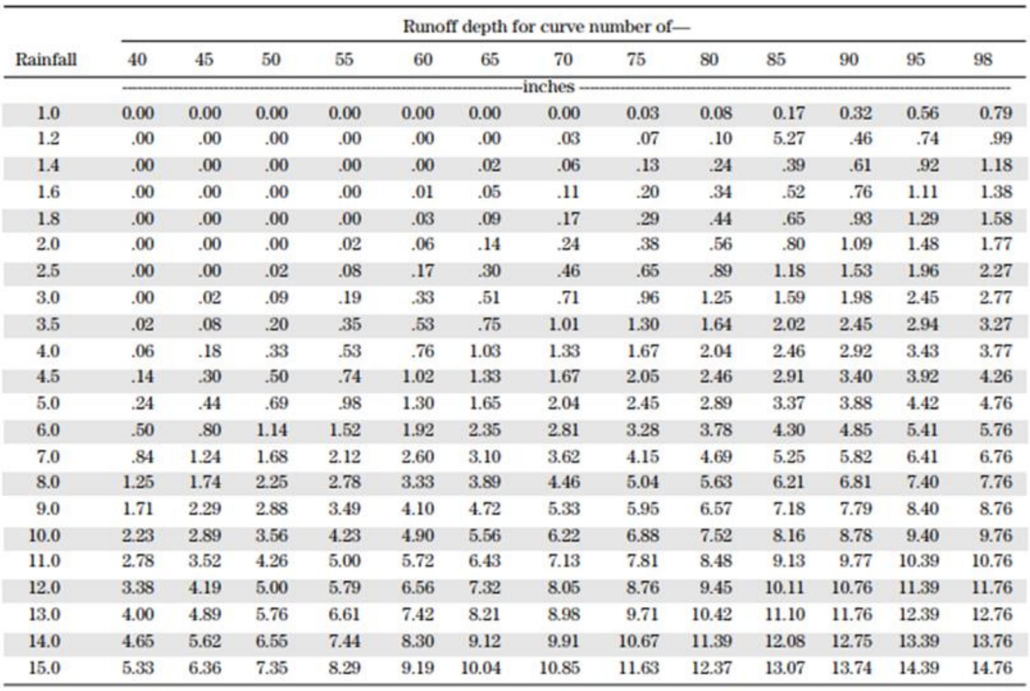


Figure 24. SCS runoff. Source: Profesor Pattel, 2016; (http://www.professorpatel.com/curve-number-introduction.html)



	Runoff depth for curve number of—												
Rainfall	40	45	50	55	60	65	70	75	80	85	90	95	98
	inches												
1.0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.08	0.17	0.32	0.56	0.79
1.2	.00	.00	.00	.00	.00	.00	.03	.07	.10	5.27	.46	.74	.99
1.4	.00	.00	.00	.00	.00	.02	.06	.13	.24	.39	.61	.92	1.18
1.6	.00	.00	.00	.00	.01	.05	.11	.20	.34	.52	.76	1.11	1.38
1.8	.00	.00	.00	.00	.03	.09	.17	.29	.44	.65	.93	1.29	1.58
2.0	.00	.00	.00	.02	.06	.14	.24	.38	.56	.80	1.09	1.48	1.77
2.5	.00	.00	.02	.08	.17	.30	.46	.65	.89	1.18	1.53	1.96	2.27
3.0	.00	.02	.09	.19	.33	.51	.71	.96	1.25	1.59	1.98	2.45	2.77
3.5	.02	.08	.20	.35	.53	.75	1.01	1.30	1.64	2.02	2.45	2.94	3.27
4.0	.06	.18	.33	.53	.76	1.03	1.33	1.67	2.04	2.46	2.92	3.43	3.77
4.5	.14	.30	.50	.74	1.02	1.33	1.67	2.05	2.46	2.91	3.40	3.92	4.26
5.0	.24	.44	.69	.98	1.30	1.65	2.04	2.45	2.89	3.37	3.88	4.42	4.76
6.0	.50	.80	1.14	1.52	1.92	2.35	2.81	3.28	3.78	4.30	4.85	5.41	5.76
7.0	.84	1.24	1.68	2.12	2.60	3.10	3.62	4.15	4.69	5.25	5.82	6.41	6.76
8.0	1.25	1.74	2.25	2.78	3.33	3.89	4.46	5.04	5.63	6.21	6.81	7.40	7.76
9.0	1.71	2.29	2.88	3.49	4.10	4.72	5.33	5.95	6.57	7.18	7.79	8.40	8.76
10.0	2.23	2.89	3.56	4.23	4.90	5.56	6.22	6.88	7.52	8.16	8.78	9.40	9.76
11.0	2.78	3.52	4.26	5.00	5.72	6.43	7.13	7.81	8.48	9.13	9.77	10.39	10.76
12.0	3.38	4.19	5.00	5.79	6.56	7.32	8.05	8.76	9.45	10.11	10.76	11.39	11.76
13.0	4.00	4.89	5.76	6.61	7.42	8.21	8.98	9.71	10.42	11.10	11.76	12.39	12.76
14.0	4.65	5.62	6.55	7.44	8.30	9.12	9.91	10.67	11.39	12.08	12.75	13.39	13.76
15.0	5.33	6.36	7.35	8.29	9.19	10.04	10.85	11.63	12.37	13.07	13.74	14.39	14.76

Figure 25. Runoff depth for selected CN's and rainfall amounts. Values that are not shown in the table may be interpolated. Source: www.njscdea.ncdea.org

The routing phase is related to the movement of water, sediments, nutrient and pesticide in the channel network of the watershed to the outlet point.

SWAT assumes that the main channels have a trapezoidal shape. So that water speed and direction were defined by Manning's equation for this study. Between the two available water routing methods: Variable Storage Coefficient "VSC" (Williams et al., 1969) and Muskingum (Chow et al., 1988), only the first one "VSC" was chosen to build the model. It performs better in those cases when the flow is combined (floodplain flow + channel flow) for long channels (Williams et al., 2011).

Penman-Monteith equation (Monteith, 1965) was selected to calculate the evapotranspiration because it is more robust in arid areas and involves more meteorological variables (Kingston et al., 2009). We used two years as "warm up" period to prepare the model and avoid initial state conditions such as antecedent soil moisture (Schuol et al., 2008).

As a consequence of the vast extension of Mauritania (1.070.00 km^2), and due to SWAT software requirements, the minimum catchment extension to generate runoff as default was 350,000 ha. The priority has been given to those main wadis or streams that may collect and drain water country-wide, enhancing the connectivity among them and providing a better insight about its hydrological behaviour on a large scale. To run the SWAT model, it implies river gauge data to finally calibrate and validate the model approach.

Within the model, the water can be stored in different reservoirs at Hydrologic Response Unit (HRU) level: soil, shallow aquifer, deep aquifer and snow. Further details about SWAT description and functions can be found in (Neitsch et al., 2011; Arnold et al., 2012). The software used for this study is Arc SWAT (Extension and graphical user input interface that embedded Swat model into ESRI GIS software). SWAT-CUP software (Abbaspour et al., 2013) was used to do sensitive analysis, calibration and validation for a small basin in the south of Mauritania.

2.2.2.3. Model building

The process to build up the SWAT model has been done as follows. Firstly, SWAT delineates a first draft of potential runoff over the study area with tributaries and outlet points, which urges to visually be corrected whenever possible. From now onwards, this model outcome is what we consider as "Potential Runoffs or Wadis" PRoW, being aware that may be coincident with real rivers such as the Senegal River. The output obtained by the model is a feature dataset with the likely streams, and they have been visually corrected based on Google earth imagery. SWAT stream delineation is a good approach for those areas where information is scarce and the area very remote (Luo et al., 2011). The second step may be more troublesome since the study area lacks of long and appropriate river gauge data to perform uncertainty analysis, calibrate and validate the SWAT model. Quantitative analysis is very often needed to clearly identify the water balance of a certain region, although some alternatives are proposed such as extrapolation of response information from gauged to ungauged basins, remote sensing data, the application of process-based hydrological models in which climate inputs are specified or measured, or the use of combined meteorological - hydrological models that do not require specific precipitation inputs (Sivapalan et al., 2003).

Other studies have examined alternative approaches including a priori parameter estimation from physical watershed characteristics (Atkinson et al., 2008); regionalization of model parameters (Vandewiele & Elias, 1995) or hydrologic indices (Yadav et al., 2007; Zhang et al., 2008), application of satellite remote sensing (Lakshmi, 2006), and the use of process-based distributed hydrologic models (Moretti & Montanari, 2008). (Srinivasan et al., 2010) highlights the possibility to use hydrological models that use physically based inputs both spatially and temporally, as far as there is a comprehension of the model interrelationships to accomplish reasonable predictions over ungauged river basins. SWAT model was originally developed to operate in large-scale ungauged basins with little or no calibration efforts (Arnold et al., 1998). Most of its parameters can be estimated automatically using the GIS interface and meteorological information combined with internal model databases (Srinivasan et al., 1998).

Studies such as (Srinivasan et al., 2010) proved that given appropriate input data, SWAT is able to provide satisfactory simulations for the water budget of a basin. Obviously, optimal solutions would address uncertainty analysis, calibration and validation against river gauge data, but alternative methods are also possible and have been tested to offer approximations for basins without river gauge data.

In this study, we could only use one river gauge station to carry out the uncertainty analysis, calibration and validation steps (Fig. 26).

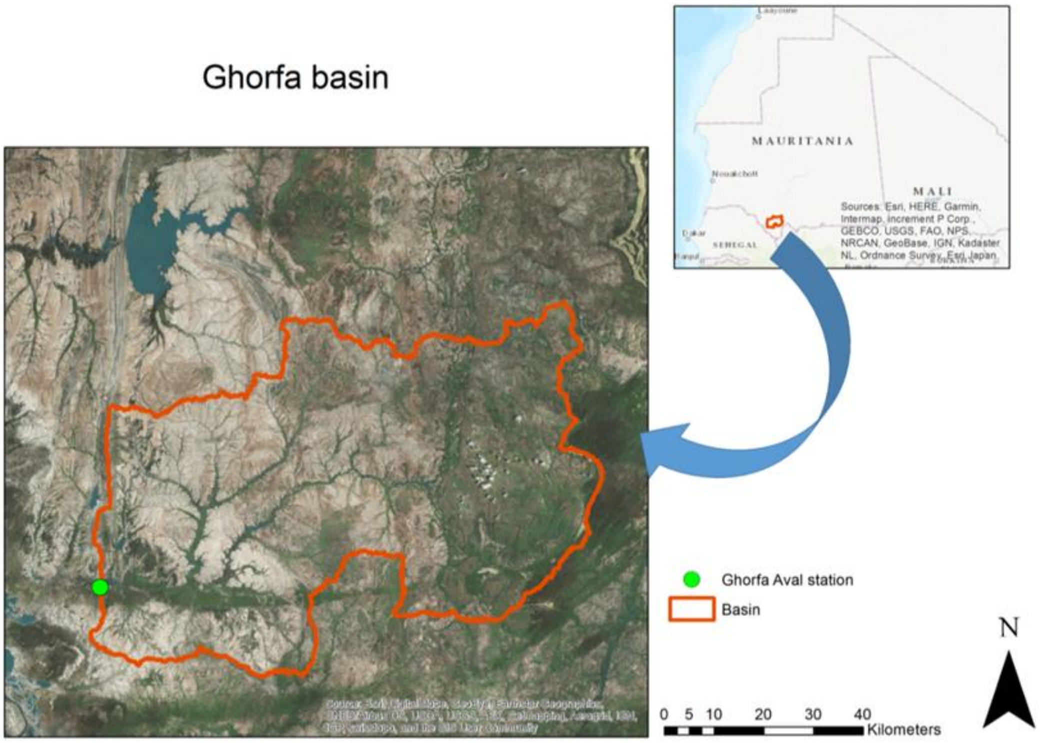

Ghorfa basin
MAURITANIA
Nouakchott
MALI
Dakar
SENEGAL
Sources: Esri, HERE, Garmin, Intermap, increment P Corp., GEBCO, USGS, FAO, NPS, NRCAN, GeoBase, IGN, Kadaster NL, Ordnance Survey, Esri Japan
Ghorfa Aval station
Basin
N
0 5 10 20 30 40
Kilometers

Figure 26. Ghorfa basin location within the study area.

Finally, the desert locust presences from the SWARMS database were taken into account to compute distances with respect to the PRoW generated by SWAT. Records of hoppers in solitarious phase were used from 1985 to 2017. Data extraction and analysis was done by ArcGis 10.3. Despite the fact that the calibration and validation period of the SWAT model do not overlap with the sighting records, it does not affect the geographical location of the wadis but only the quantitative analysis of the flow.

2.3. Results and discussion

2.3.1. Model set up

The elevation of the watershed delineated was varying from -41 to 795 m, where 78 % of the delineated watershed surface is under 350 m. The land use map (Fig. 27) indicates that the major land cover of the area is arid consolidated and unconsolidated grounds (SWRN) with 70.84 %, followed by Range Grasses (19.09 %) and far behind are Agricultural Land Generic (8.87 %), Pasture (1.19 %) and Range Brush (0.01 %). Leptosols (34.78 %), Sand Dunes (27.23 %) and Arenosols (31.73 %) are the predominant soil types, while Regosols (3.36 %), Calcisols (2.76 %) and Gleysols (0.15 %) are less relevant. The slope of the terrain is mainly between 0 - 2 % for most of the watershed.

The model was run from 1st January 1979 to 31st December 1985 with a 1 year warm up period and default parameters. The graphical result of the model simulation is shown in Fig. 28, where potential runoff or Wadis (PRoW) are in blue. The total watershed area is 505,474 km^2 and it has been divided into 78 sub-basins. The area uncovered by any sub-basin would not meet the criteria to generate runoff owing to land use, soil type or slope conditions of the terrain, and the rainfall would be infiltrated.

Due to the lack of river gauge data, the SWAT model was only calibrated for the Ghorfa Basin (Fig. 26).

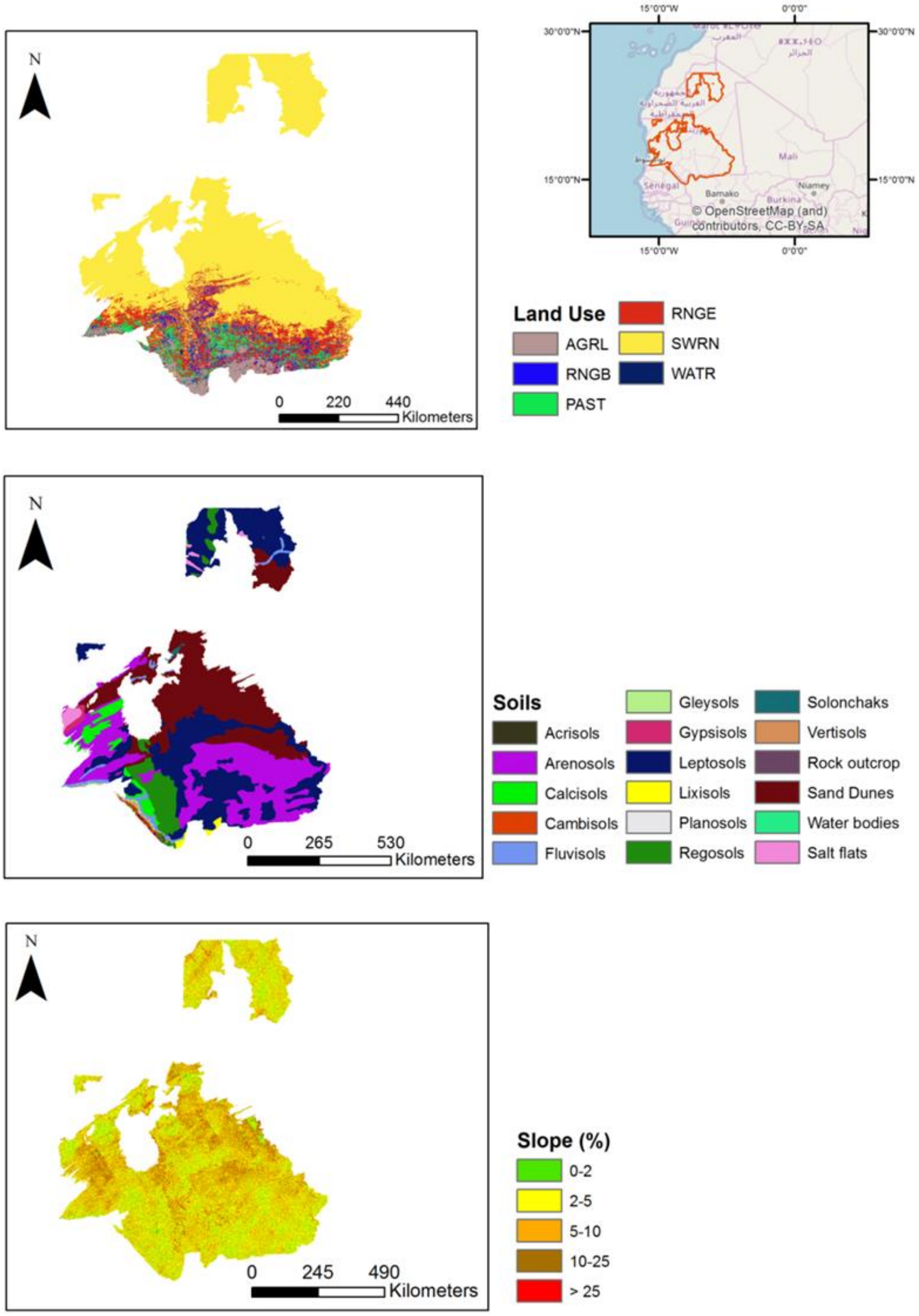
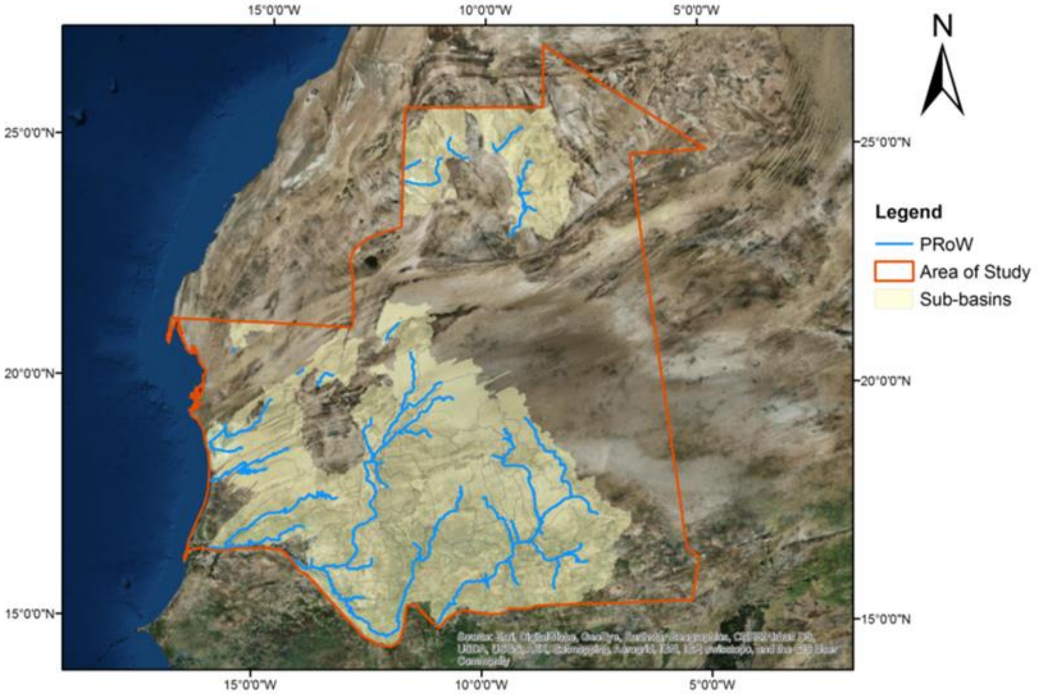


Figure 27. Land Use, Soil and Slope maps of the delineated watershed within our study area.



Source: Esri, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AEX, Getmapping, Aerogrid, IGN, IGP, swisstopo, and the GIS User Community

Figure 28. Potential Runoff or Wadi "PRoW" delineation as a graphical result of SWAT model in Mauritania. Areas with no potential runoff within the study area have not been assigned with sub-basin feature by SWAT.

To quantitatively assess the fitness of the model, the pre-calibrated model has been compared to observe monthly flows for the period 1980 -1985 (Fig. 29) with SWAT-CUP software and SUFI-2 algorithm.

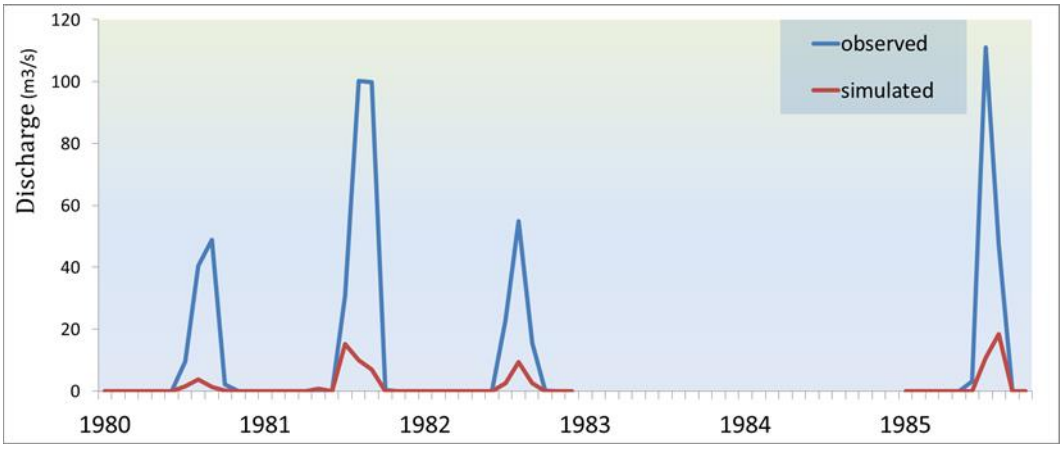


Figure 29. Comparison between modelled and observed total monthly discharge at Ghorfa Aval station for the time period 1980 – 1985.

A priori, there is a clear underestimation of the SWAT model in comparison with the observed discharge data. There is a time gap by 1984 due to river gauge missing data. In spite of the marked deviation, the beginning and the end of the peaks are well identified by the model so that there are signs of improvement if we calibrate the model against certain sensitive parameters. The final goal is to obtain satisfactory prediction accuracy for our model. The NSE (Nash-Sutcliffe statistic) and R^2 values for the simulation were 0.05 and 0.56 respectively. Sensitivity, calibration and validation were carried out over the Ghorfa basin (unique dataset available with river gauge data in the study area).

2.3.2. Sensitive parameters

Sensitive analysis was done for the river flow in order to identify which of the parameters have more influential effects on the model. This task permits us to save time during calibration and validation stages. Only the most sensitive parameters will be adjusted during calibration.

Table 3 compiles the list of parameters used in the sensitivity analysis, as well as the rank of each parameter according to the p-value and t-stat. The t-stat value provides sensitivity information, so that the larger the absolute value is, the more influential will be the parameter in the model. Whereas the smaller the p-value is, the more sensitive.

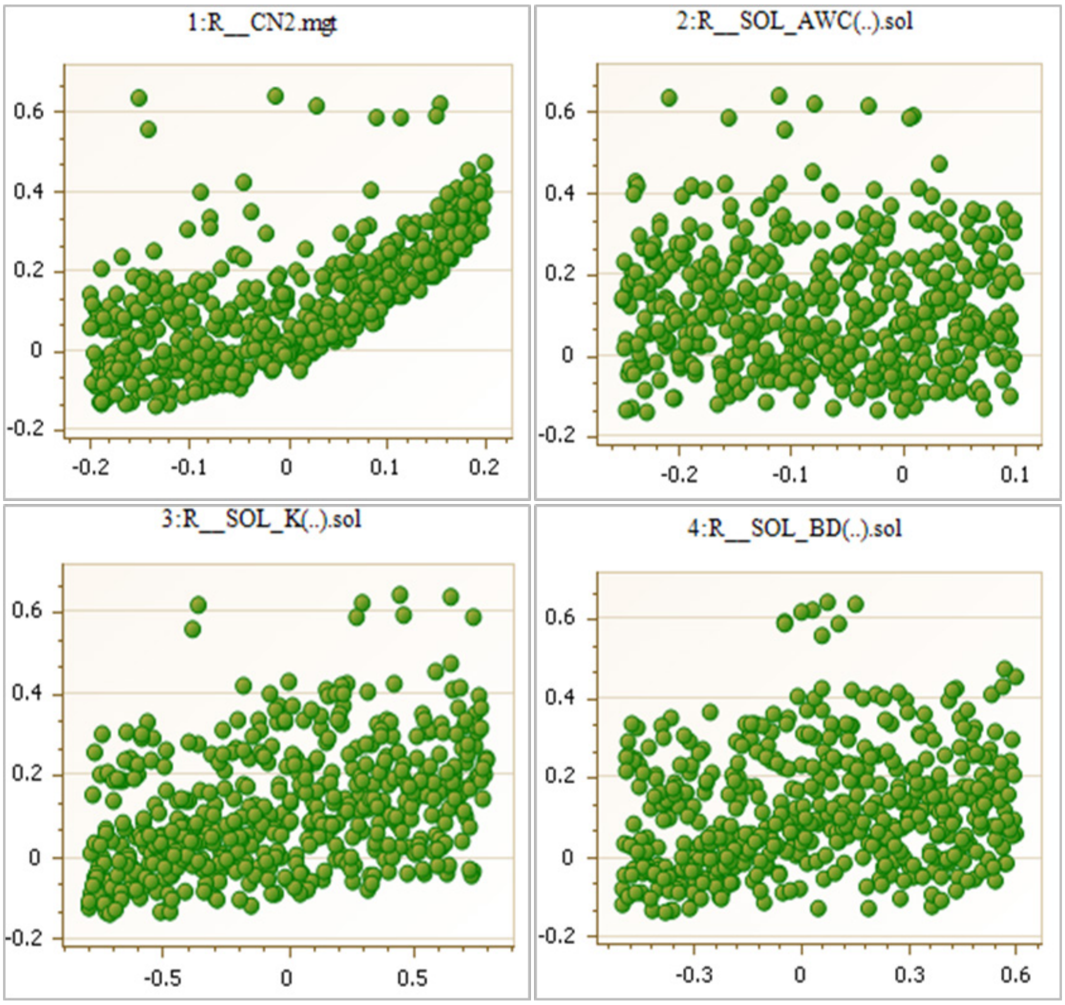
The Curve Number CN2 parameter showed to be the most influential or sensitive for the study conditions. Then saturated hydraulic conductivity and moist bulk density are ranked 2nd and 3rd respectively. Plant uptake compensation factor and available water capacity still have an acceptable p-value so that they will also be included in the model calibration.

Sensitivity ranking	Parameter	Description	t-stat	p-value
1	R_CN2.mgt	Curve number	27.163	0.000
2	R_SOL_K (1).sol	Saturated hydraulic conductivity	15.37	0.000
3	R_SOL_BD (1).sol	Moist bulk density	10.38	0.000
4	R_EPCO.bsn	Plant uptake compensation factor	3.450	0.000

Sensitivity ranking	Parameter	Description	t-stat	p-value
5	R_SOL_AWC (1).sol	Available water capacity	-2.217	0.002
6	R_ESCO.bsn	Soil evaporation compensation factor	-1.909	0.006
7	R_SURLAG.bsn	Surface runoff lag coefficient	1.503	0.133

Table 3. Sensitive parameters and their ranking according to p-value for Ghorfa basin.

Dot plots are the representation of parameter values or relative changes versus the objective function. They depict the distribution of sampling points as well as parameter sensitivity. Fig. 30 shows the dot plots of each parameter used during the sensitive analysis. These plots agree with the t-stat values obtained in Table 3, where CN2 is the most sensitive parameter.



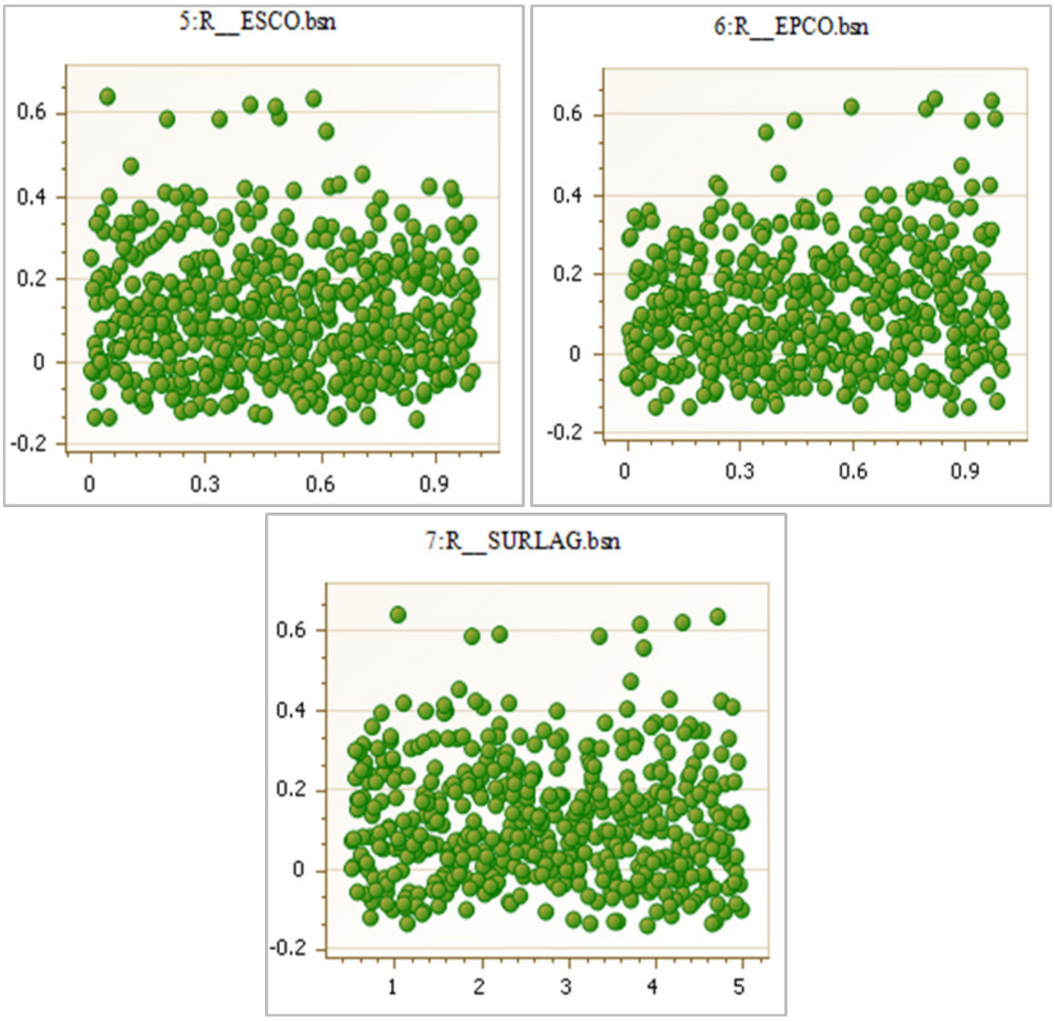


Figure 30. Dot plots of the parameters under the sensitivity analysis. The Y-axis indicates the NSE values, and the X -axis indicates the value of parameters.

2.3.3. Calibration and validation

Based on river gauge availability, the hydrologic analysis goes from 1979 – 1985. Due to the short available data, we only use the year 1979 as a warm up period. Calibration was done from 1st January 1980 to 31st December 1983, and validation from 1984 to 1985. Based on the sensitivity analysis, the first 5 parameters (CN2, SOL_K, SOL_BD, EPCO, SOL_AWC) were used to fit the model. The parameter values have been adjusted within the range suggested by the SWAT-CUP software after 4 iterations. The sensitive parameters and their fitted range values after calibration are shown in Table 4.

Sensitive parameter	Default parameter range	Parameter range derived from calibration
R_CN2.mgt	-0.2 to 0.2	-0.12 to 0.02
R_SOL_K (1).sol	-0.8 to 0.8	-0.29 to 0.15
R_SOL_BD (1).sol	-0.5 to 0.6	0.09 to 0.30
R_EPCO.bsn	0 to 1	0.52 to 0.74
R_SOL_AWC (1).sol	-0.2 to 0.4	-0.32 to -0.14

Table 4. Sensitive parameters with their default and fitted range of values

To assess model performance against observed discharge values, the coefficient of determination (R^2) and Nash-Sutcliffe (NSE), P-factor and R-factor statistics were used (Table 5).

R^2 or coefficient of determination is a statistic that indicates the proportion of the variance in the dependent variable that is predictable from the independent variable(s). It ranges from 0 to 1, being values closer to 1 the optimal solutions.

NSE or Nash-Sutcliffe coefficient is an indicator of the model's ability to predict about the 1:1 line between observed and simulated data. Values equal to 1 show perfect fit, while values under 0 would indicate that the model is predicting no better than using the average of the observed data (MathWorks, 2018).

P-factor and R-factor:

P-factor is the percent of observations enveloped by the model results, the 95% Prediction Uncertainty or 95PPU. The 95PPU contains the 95% of predictive uncertainty or measured data corresponding to the behavioural parameters, and the corresponding uncertainty bounds. R-factor is the thicknesses of the 95PPU envelop (Abbaspour, 2007).

The most optimal solution would have a P-factor equal to 1, and an R-factor equal to zero. It would mean that the model would perfectly represent the study basin, and no measurement errors or other uncertainty sources would be involved in the process. Nevertheless, these instances are never found in reality so that efforts must aim to obtain a P-factor as high as possible (close to 1), while R-factor should be as narrow as possible (close to 0) (Abbaspour, 2015).

Before calibration, the NSE and R^2 were 0.05 and 0.56 respectively, which is a relatively low agreement between simulated and observed values. These results show the low ability of the model without calibration to estimate quantitatively the flow over the basin. Nevertheless, Fig. 31 shows the ability of the model to coincide temporally with runoff episodes. Given that the Ghorfa basin is dry throughout the year, this is a good approximation to estimate runoff arrival to the river gauge station. After calibration, the NSE and R^2 values greatly improved those metrics with 0.83 and 0.84 respectively, with 69 % of the observed data covered by the 95PPU.

Fitness statistic	Before Calibration	After Calibration	Validation
R^2	0.56	0.84	0.54
NSE	0.05	0.83	0.16
P-factor	0.65	0.69	0.50
R-factor	0.00	0.40	0.07

Table 5. Evaluation metrics to assess the model performance before and after calibration, as well as for validation stage.

Thus, the model was successfully calibrated for the Ghorfa basin with satisfactory fitness statistics for the time span 1980 to 1983. Validation statistics show a low predictive skill to quantitative estimate the flow discharge in the basin. As well as it was described in the pre-calibrated phase, the model is able to identify temporally when the runoffs occurred as seen in Fig. 32-33. However, given the low performance of the model, quantitative estimation is somewhat no possible with the current available data for the basin. The differences may be due to inaccurate or missing meteorological data, errors in other input data sets such as land cover and soil information where the resolution may be too coarse or even some errors during the pre-processing of the data. The missing information of river gauge data is described in Fig. 33.

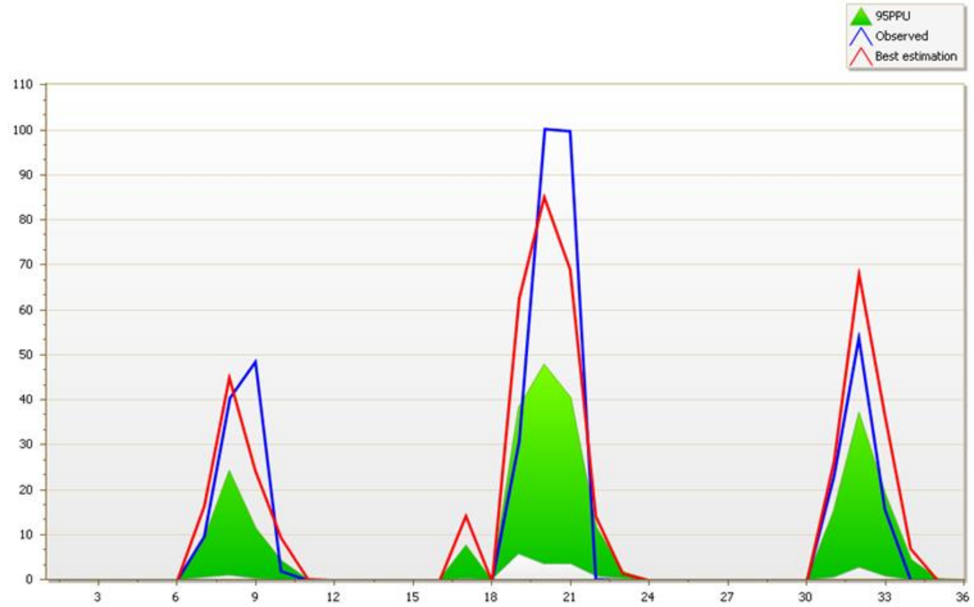
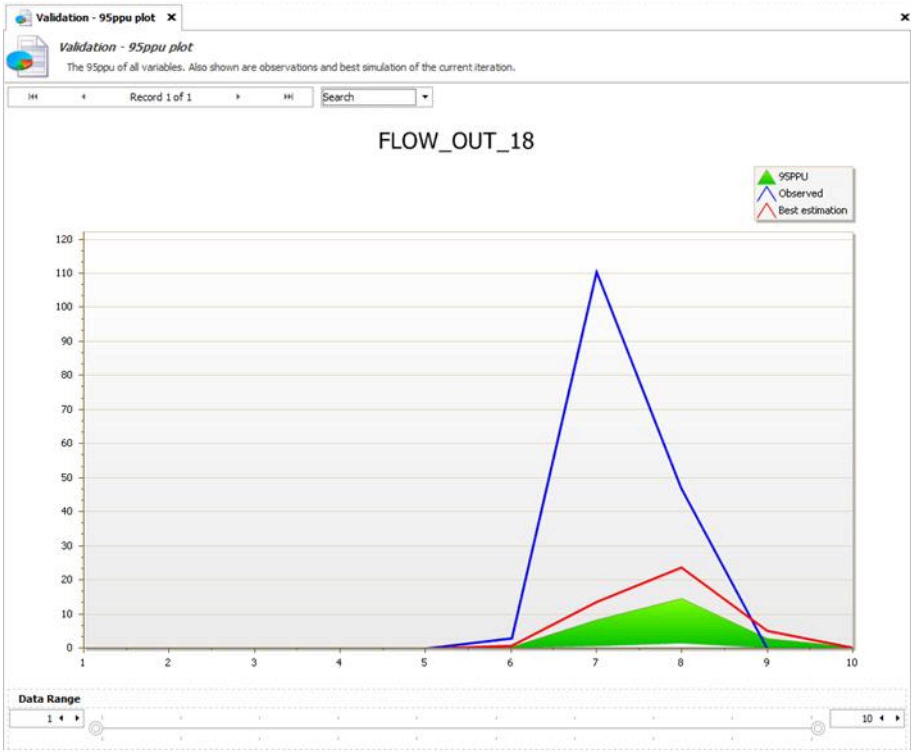


Figure 31. Best simulated runoff with 95ppu for the calibration phase, where the x-axis represents the months and y-axis the runoff in m^3/s.


Validation - 95ppu plot

Validation - 95ppu plot

The 95ppu of all variables. Also shown are observations and best simulation of the current iteration.

Figure 32. Best simulated runoff with 95ppu for the validation phase, where the x-axis represents the months and y-axis the runoff in m^3/s.

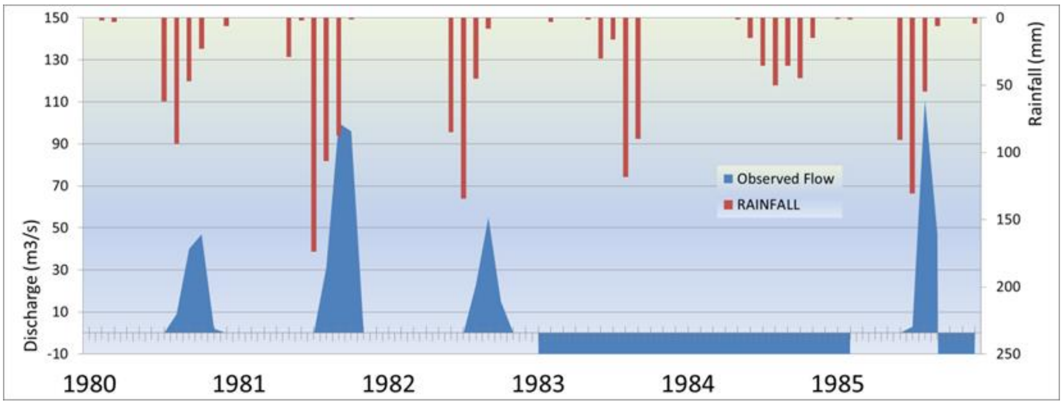


Figure 33. Comparison between observed river gauged data and precipitation recorded at the closest rain gauged station during the survey time at Ghorfa basin. X-axis represents the time and y-axis the amounts for both variables: precipitation (red, at the top) and river gauge data (blue, at the bottom). Flow values below 0 correspond to missing data

Model uncertainties can be accounted for the great variations in topography and rainfall, both spatially and temporally. Wadis show a high dependence on precipitation, where no groundwater supply is observed to maintain a continuous flow. On the contrary, these areas usually have high infiltration rates providing a good mechanism of groundwater recharge due to their geomorphological characteristics (Al-Adamat, 2003). The discharge at Ghorfa Aval station reaches its maximum slightly later than precipitation reaches its maximum. It occurs for the 4 peaks observed between 1980 and 1985. The absence of data during 1983 and 1984 impedes to analyse the effects of precipitation over such period. There is a good correlation between the observed precipitation and the discharged observed, nevertheless our model have not been able to be trained appropriately to capture the uncertainties.

2.3.4. Distances between desert locust and PRoW

The density map of absolute presences (Fig. 34) indicates the most usual areas to find solitarious hoppers. It gives an insight about the potential areas to breed without any further environmental analysis of the habitat or meteorological conditions. Occurrences are mainly found in the centre-west of the country, with lower rates in the south. The centre-east and north-east of Mauritania reports almost no records of sightings. The computed distance between the potential runoffs or wadis and hopper presences (1985 - 2017) is shown in Fig. 35.

These figures need to be taken cautiously. The lack of occurrence records does not mean absence data, but just not been recorded. The statistical figures of the computed distances are shown in Fig. 35. The minimum distance to the closest PRoW was 1.4 m; the mean and median were 38280 and 31160 m respectively; the maximum distance was 275000 m and the standard deviation was 3,680.

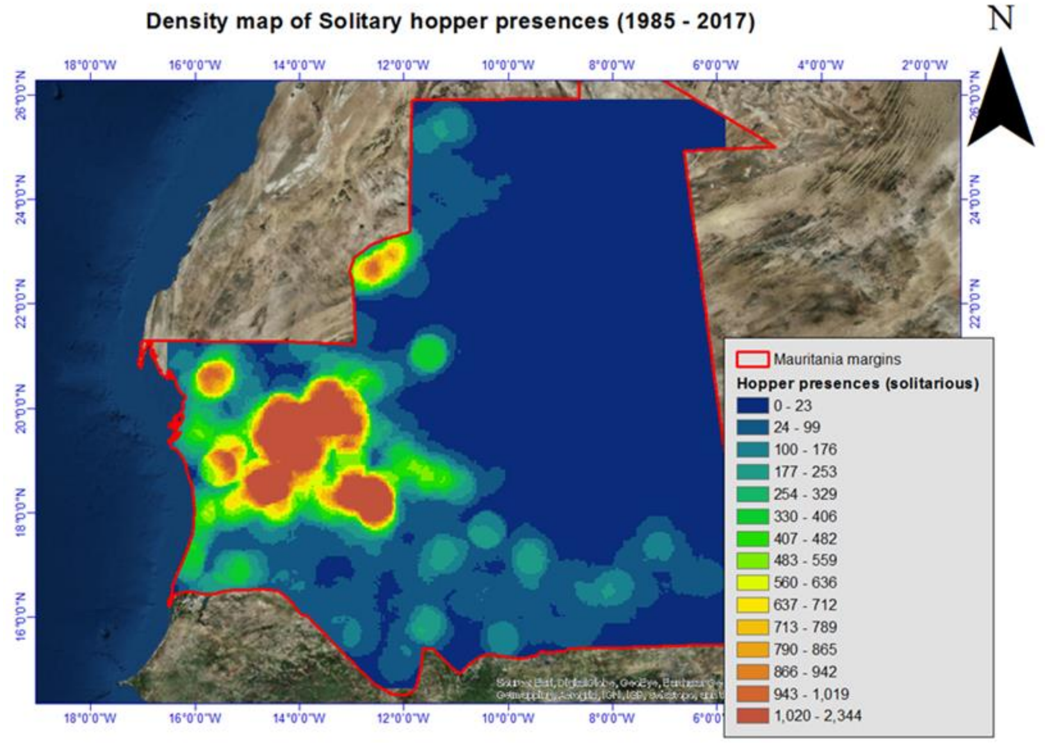


Figure 34. Density map of solitarious hopper by absolute number of records within the study area from 1985 to 2017

It is has been observed that the distance between hopper presences and PRoW is very variable. It means that not just wadis might influence hopper occurrences, as already known by the scientific community (Popov, 1958; Simpson et al., 1999; Despland & Simpson, 2000). Nevertheless the distances in some particular areas fit very well, this highlights the role of some wadis in locust development. As far as these insects are highly influenced by different environmental conditions such as wind (Culmsee, 2002), soil moisture (Popov, 1958) or surface temperature (Haskell, 1962), the morphology of the terrain is also important as detailed in (Culmsee, 2002). Desert locust breeding areas are very dependent on rainfall and flooding (Voss & Dreiser, 1997), and the second is channelled by wadis in arid environments such as in our study area. In absence of superficial or rainfall water, humans and vegetation lay upon wadis alluvium for water supply. (Ahmed et al., 2007) details that accessible water is mainly concentrated at the wadi beds.

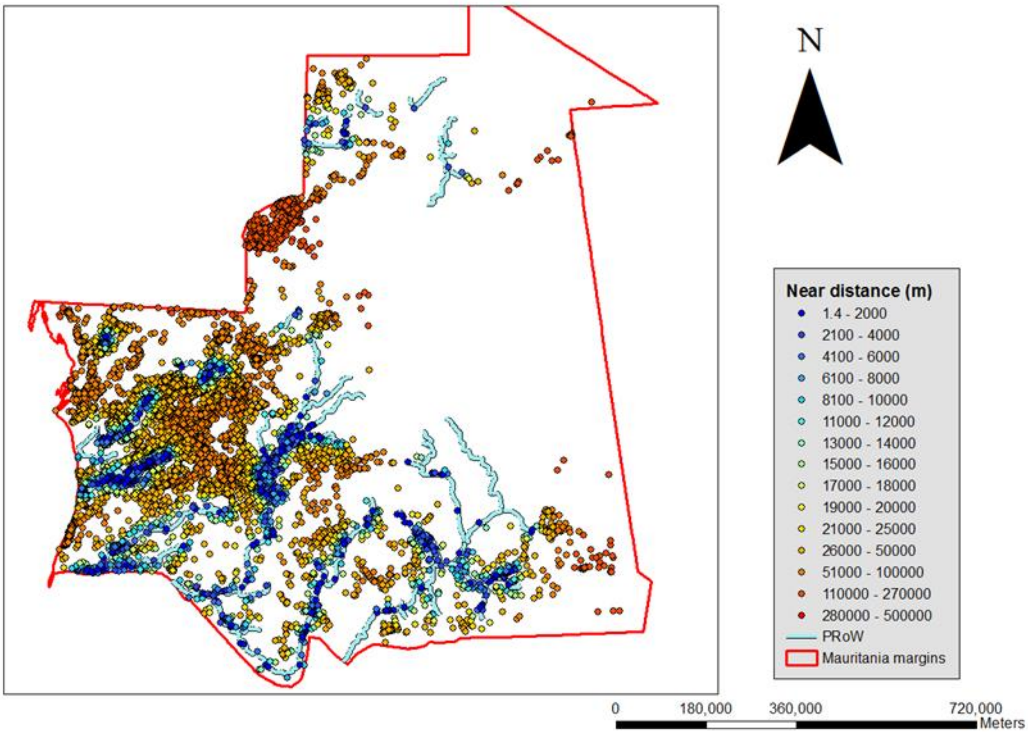


Figure 35. Map of the distances between the hopper occurences and the SWAT PRoW.Red colours are associated with the occurences found at maximum distances, while blue colours correspond to hoppers recorded at close sites.

2.4. Conclusion

This chapter addresses a hydrological model approach to locate wadis taking into account geo-physical variables such as elevation of the terrain, land type, slope or land cover. SWAT model has delineated potential runoff networks or wadis over the entire surface of Mauritania. As ground data was scarce and not well distributed over the entire area, different data sources were needed during the process. Due to the aridity of the region, it was very difficult to carry out calibration and validation procedures. The only available flow database was located at Ghorfa basin, where there was an attempt to calibrate/validate the model at local extent and then expand or regionalised those results as detailed in (Srinivasan et al., 2010). Results present high uncertainty due to the short length of data records, with abundant missing data. In addition, most of the rivers run dry during most of the year. Moreover, the quantity and quality of the river gauge data was not enough to carry out quantitative analysis. On the other hand, and based on the morphological features of the terrain, SWAT identified potential drainage networks in case of high rainfall episodes. SWAT stream delineation may be a good approach for those areas where information is scarce and the area very remote (Luo et al., 2011). These channels have been used to compute the distance of historical occurrence records of desert locust. Given the role that wadis may play as breeding locations, the aim of this chapter was to study the suitability of SWAT hydrological model to, a priori, locate suitable areas for breeding.

Despite the promising results to detect wadis as favourable areas for breeding, the main problems of this methodology are data availability. It should be noted the impossibility to calibrate and validate the SWAT model in an adequate manner given the lack of rainfall and river gauge data for the study area. This methodology can be used as a first approach to locate breeding sites; but it requires improvements in terms of data quality. Nevertheless, remote sensing arises as an alternative to detect and monitor, on near real time, suitable breeding areas.

Chapter 3. Soil moisture analysis to locate breeding areas using machine learning techniques

3.1. Introduction

The aim of this chapter is to identify suitable SM conditions for desert locust eggs as well as to hopper desert locust in solitarious phase. It is based on SM estimations from ESA CCI SM product and ground based observations of hopper desert locust. Species Distribution Models (SDMs) were used to better understand the link between SM and desert locusts to predict their likely distribution across landscapes and breeding areas.

Traditionally, remotely sensed precipitation and vegetation estimates have been the main techniques to predict desert locust events. However, precipitation data presents high uncertainty in arid and semi-arid areas in Africa (Schmidt & Karnieli, 2000), so desert locust detection is to be improved (Bolten et al., 2009). On the contrary, soil moisture (SM) estimates are very promising despite very few studies have addressed the link between SM remote sensing and desert locust (Liu et al., 2008). This measures can be retrieved either by active or passive sensors.

SM measures have conventionally been ground based; therefore survey areas are usually limited for being an expensive and time consuming activity (Sun et al., 2005; Huang et al., 2006). As presented early in this work, SM is a limiting variable to the desert locust development, influencing locust egg laying site location and the survival of locust eggs (Liu et al., 2008).

Recently launched satellite platforms such as SMOS, SMAP and Sentinel 1 may greatly contribute to detect breeding sites for desert locust.

To analyse the link between species occurrences and environmental factors such as SM. SDMs require extensive data preparation prior to model building

(Franklin, 2010) in order to acquire good success during the process. Data preparation ought to compile the following steps (Hijmans & Elith, 2016):

3.1.1. Species occurrence data

It is one of the most limiting and difficult tasks. Depending on the nature of the species under analysis, they can be immobile or mobile, what might hinder collection works. Data collection is usually a field work that implies high cost and long survey periods to generate a reliable database. These records need to be geo-located and then cross-checked in a GIS software to avoid any error. Some discussions address which is the best method to model species, nevertheless this is a critical step that must provide reliable records. When there is some uncertainty in the records, efforts should focus firstly on improving the quality of the data (Lobo, 2008). Model performance improves, independently of the used method, when occurrence data is unbiased (Graham et al., 2007) and the number of records is large enough (Wisz et al., 2008).

3.1.2 Data cleaning

This is a very important step that cannot be omitted (Hijmans & Elith, 2016). Any error in the measurements would mislead the model so that efforts must focus on removing any data that makes no sense. In order to do that, we need to have a deep knowledge of the data and inspect the anomalies thoroughly. For instance, some plant records may be duplicated when they are split in different herbariums, or when by error the same sample is recorded twice. These kinds of errors are common and must be removed from the dataset alike in geolocation. Data cleaning is tedious but necessary if we want to build a robust predicting model.

3.1.3. Sampling bias

Frequently, sampling bias is found in many of the occurrence datasets (Hijmans et al., 2012). It can be critical to the accuracy of SDMs generated from presence-only datasets (Phillips et al., 2009), but options to correct for sampling bias are not always applied (Yackulic et al., 2012).

Sample collection often occurs over relatively accessible locations close to roads, urban settlements and rivers, and it is as a consequence of not doing it systematically or randomly. It implies that samples may not be representative of the true range of environmental conditions in which the species occurs

(Reddy & Davalos, 2003; Kadmon et al., 2004), and it usually happens with specimen data from open access data portals (Hortal et al., 2008).

(Phillips et al., 2009) developed a method to mitigate the geographical bias in SDMs with only-presence data. This methodology consists on generating background data or pseudo-absences with similar geographical bias as the presence records, so that accessible areas would be eligible to be chosen when no presence data is recorded. Background or pseudo-absence points are used during the training phase of the model building to highlight unfavourable environmental conditions for the species under survey. This approach has demonstrated to improve prediction accuracy of the models (Syfert et al., 2013).

3.1.4. Absence, pseudo-absence or background points

Some SDM algorithms use only presence data, while others in addition use pseudo-absence or back-ground data. Background data (Phillips et al., 2009) aim to characterize environments in the study area, and not aiming to find absence locations of the species under research. Thus, background points depict the area where the species has been found. Background data indicates the environmental domain of the species, whereas presence data establish under which conditions a species is more likely to be present than on average. Pseudo-absences are used to generate non-presence locations to feed logistic models. This method requires few assumptions that lay upon randomness of absence or presence records.

Nevertheless, these methods should not prevail unless there is a lack of survey-absence data. True absence data provides valuable information concerning prevalence of the species as well as geo-locations where the survey was carried out. Nevertheless, it can also be biased or incomplete (Kéry et al., 2010).

3.1.5. Extraction and preparation of environmental data

Environmental or predictor variables can be found in different formats such as geo-located data points or raster images. Previous literature and researcher experience may help to find predicting variable to include in the model. A careful selection of predictor variables may improve model performance, and it is particularly important when the objective is merely explanatory (Mellert et al., 2011). Finally, the environmental variables are used to fit a model to infer similarities to the occurrence locations as well as other measures such as species abundance.

When there is already a robust database, a wide range of machine learning algorithms are available to predict the occurrence of the species, or the variable of interest within the region under analysis. Some of these SDMs have also been designed to infer past or future distributions as far as environmental variables are provided to the model.

3.2. Materials and Methods

3.2.1 Survey data

In this chapter, we have selected 12027 solitarious hopper sightings from the SWARMs database for the time span 1985-2015, and they were spatially distributed as seen in Fig. 36. The generated pseudo absence distribution is presented in Fig. 37.

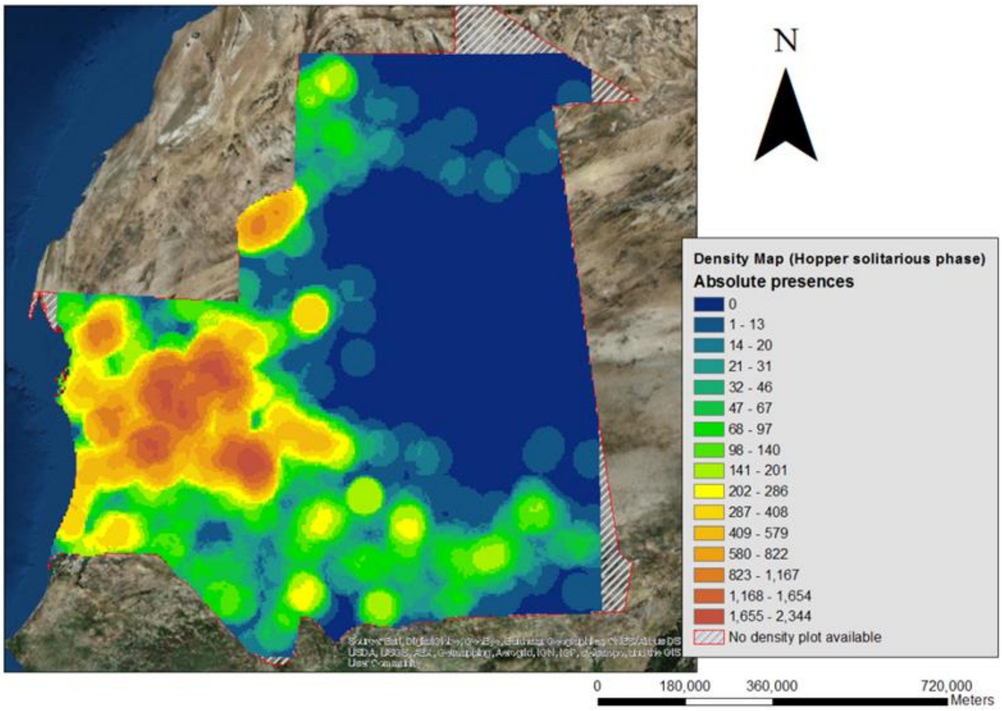


Figure 36. Density plot of solitarious hoppers between 1985 and 2015. Data presences comes from SWARMS database.

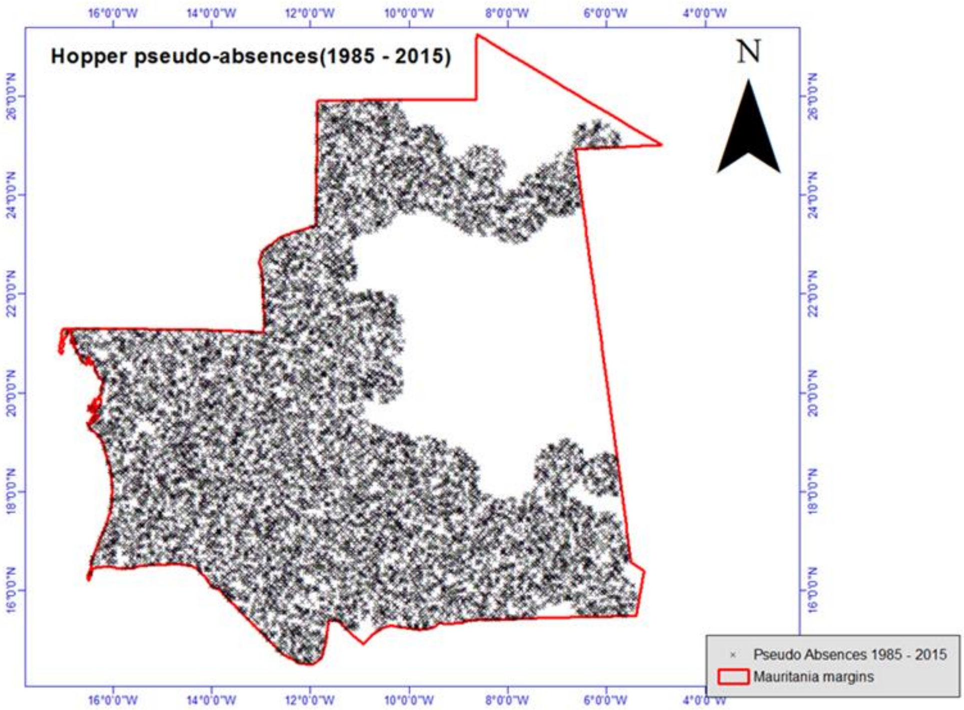


Figure 37. Computed pseudo absences in Mauritania with random dates from 1985 to 2015 within the "ever recorded locust sighting".

3.2.2. Satellite data

The SM dataset, generated via the Climate Change Initiative (CCI) of the European Space Agency (ESA) (ESA CCI SM v03.2), is a merged product from radar and radiometer sensors of the volumetric surface SM (up to 5 cm depth) expressed in m^3/m^3 units. Its spatial resolution is 0.25 ° and offers daily coverage worldwide from 1978 up to 2015 (Liu et al., 2012; Gruber et al., 2017; Dorigo et al., 2017). Aiming to provide the most complete and long-term consistent SM dataset, it comprises active data retrieved from C-band scatterometers on board of ERS-1, ERS-2, MetOp-A and MetOp-B (generated by the "TU Wien") and passive data obtained from microwave observations by the following sensors: Nimbus 7 SMMR, DMSP SSM/I, TRMM TMI, Aqua AMSR-E, Coriolis WindSat, GCOM-W1 AMSR2, and SMOS (generated by VU University Amsterdam in collaboration with NASA). This product has been validated against ground based reference measures or alternate estimates from other projects and sensors (Liu et al., 2012; Wagner et al., 2012).

There are three available harmonized products: merged passive, merged active and a combined active and passive SM product. For the purpose of this study, we have used the combined product for being the most complete one. It uses the pixel from either the active or passive source, or the average value of both depending on the performance of the vegetation optical depth (VOD) from the Advanced Microwave Scanning Radiometer for EOS (AMSR-E) C-band observations (Liu et al., 2012).

3.2.3. Methods

The applied methodology is based on SM analysis to estimate favourable patterns to desert locust breeding areas.

The ESA CCI SM v03.2 product was used to geographically compare the seasonal presence of solitarious hoppers of desert locust by months, with SM values from 1985 to 2015. Breeding areas in Mauritania vary widely throughout the year according to the National Centre for Prevention and Control of Desert Locust in Mauritania (CNLA). During summer months, desert locust usually breeds in southern parts of the country. From September to December, breeding occurs in the centre and the north-western part; and from December to May in the northern areas of Mauritania (Babah Ebbe, 2012). It is widely accepted that these insects have regional migrations following suitable environmental conditions (Van Huis et al., 2007).

The coordinates of each hopper in solitarious phase were extracted and its corresponding date from SWARMS database. In addition to that, those records can be also considered as “pseudo-absences” owing to hoppers in solitarious phase may go unnoticed at low densities (Meynard et al., 2017). Thus, we found convenient to randomly generate a grid of “pseudo-absences” as reported in other studies using SDMs (Zaniewski et al., 2002; Engler et al., 2004).

Pseudo-absence samples were computed based on two principles. Firstly, they were located within a maximum of 50 km radius mask created of ever recorded locust sighting (1985-2015), aiming to select areas with environmental and geophysical potentialities and to reduce geographical bias. This distance was selected to visually coincide with desert locust presences in the density map (Fig. 37), where most of the areas with no presences are masked out. Otherwise, it could misguide SDM predictions (Barnes et al., 2014).

Secondly, date allocation was done using a uniform random arrangement with R-software. Each pseudo-absence location was assigned a date within the first and the last hopper presence date of the SWARMS database (1985-2015). These pseudo-absence points were generated randomly and equally weighted to the presences (pseudo-absence and presence weighted sums are equal) for predicting species occurrences or distribution (Barbet-Massin et al., 2012). It may occur that some presences and pseudo-absences coincide geographically within the same pixel; however it is very unlikely that they have the same assigned date. Each pseudo-absence date has been randomly allocated from 1985 to 2015, what implies that they will likely not have the same SM values.

The duration of locust life cycles are variable, depending on the environmental conditions of the habitat (Showler, 2018), nevertheless we rely on the following premises to create the variables in our study. Eggs are laid at 5-10 cm depth, and the egg incubation period may ranges from 10 to 65 days (Pedgley, 1981). After hatching, nymph phase may last between 24 and 95 days since the egg was laid. Thus, under the most severe environmental circumstances, the maximum expected egg-hopper development time would be 95 days (Symmons & Cressman, 2001). SWARMS database registers the sighting date and phase, but not the age of each individual, so that we have established up to 95 days prior the sighting record as the time analysis. Fig. 38 shows the sequence of the proposed method as a flow chart.

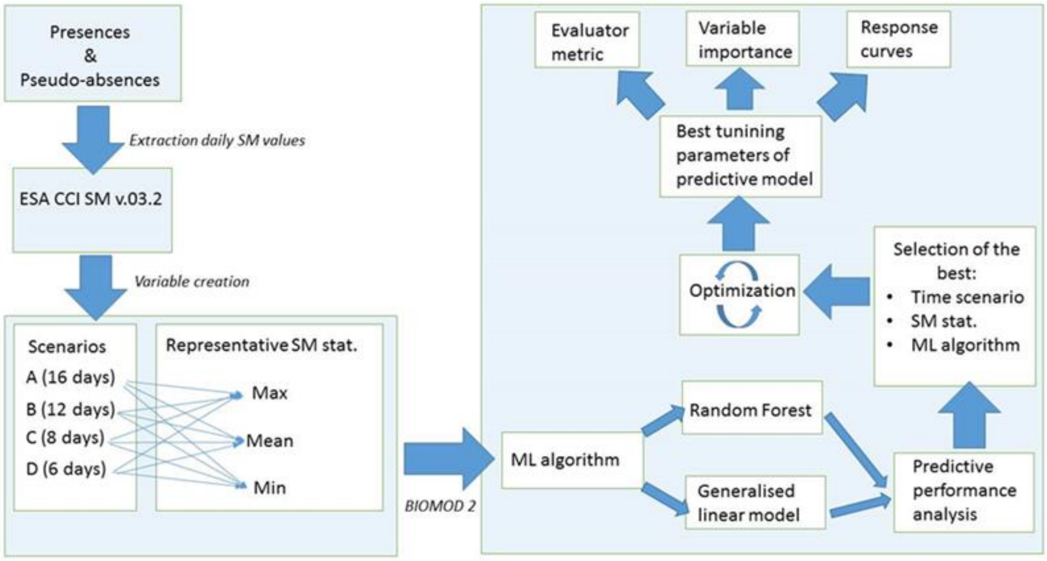


Figure. 38. Flow chart of the proposed methodology to study the link of ESA CCI SM with desert locust, using machine learning approach.

Given the coordinates of each presence and pseudo-absence record, the corresponding daily SM value was extracted based upon the sighting or assigned date, up to 95 days backwards. Based on these antecedent SM conditions, we generated variables dividing the analysis time into different time intervals (16, 12, 8 and 6 days) and assess the performance of the model with each of them. By this method, we aim to cover and differentiate critical events in the locust lifecycle such as egg-laying, egg-hatching and early stages of the nymph phase individuals as well as to deal with punctual missing data (Fig. 39).

Some areas of SM imagery had missing data due to the satellite revisit times used to generate ESA CCI SM v03.2. It was computed the minimum, mean, and maximum SM values within each time interval to be a representative value of such period. Then, we assess which descriptive statistic provides better information to the model in terms of performance. If no value was found for a particular time interval, the presence or absence record is not included in the model. In this way, we mitigate the effect that the missing information could provoke on the model results. Even though SM may vary greatly on a daily basis (Wang et al., 2014b), the biological evolution for egg and hopper development need some days to be altered (Symmons & Cressman, 2001), so that we found convenient this approach to generate the model variables.

Therefore, four different scenarios were studied: A, B, C and D. As previously mentioned, SM values were obtained, on a daily basis, up to 95 days before the presence or pseudo-absence date record. Each of the proposed scenarios contemplates a different division in terms of days: A = 16 days, B = 12 days, C = 8 days and D = 6 days. Hence, we aimed to obtain one representative SM value per each subdivision of time, within each scenario. In order to acquire this representative SM value, we have computed the minimum, mean and maximum out of the daily SM values contained in every time interval.

Fig. 39 shows variable creation for each scenario (A, B, C, and D) based on SM and presence and pseudo-absence dates. For instance, scenario (A) contemplates equal time intervals of 16 days so that (SM1) indicates the SM value on the local pixel between -95 to -80 days (both included) prior the presence or pseudo-absence date. (SM2) SM value on the local pixel between -79 to -64 days prior the presence or pseudo-absence date and the rest accordingly as detailed below.

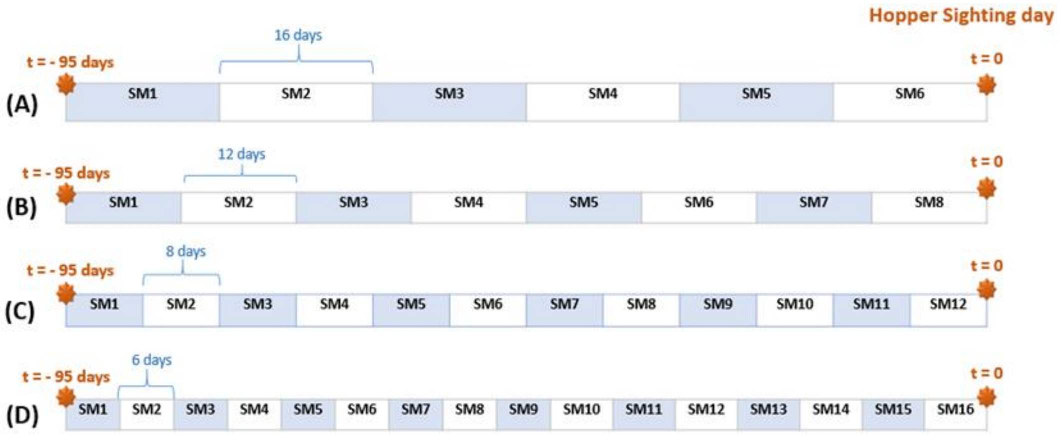


Figure 39. Variable names and their distribution back in time for four different scenarios: A, B, C and D. Time interval for scenario (A) is 16 days generating 6 variables, 12 days for (B) with 8 variables, 8 days for (C) with 12 variables and 6 days for (D) with 16 variables. Time equals to 0 (t = 0) corresponds to the presence or pseudo-absence sighting date. Within each scenario, 3 different alternatives are independently tested (minimum, mean and maximum SM value within the given time interval).

Some publications suggest the suitability of machine-learning (ML) approaches to model species distributions, since they may perform better than the traditional regression-based algorithms (Elith et al., 2006). BIOMOD2 tool (Thuiller et al., 2009) implemented for R software (R Development Core Team, 2012) was again used to build the models. Two different ML modelling techniques were tested to describe and model the link between desert locust and SM: Generalized Linear Model “GLM” (McCullagh & Nelder, 1989) and Random Forest “RF” (Breiman, 2001).

Generalized Linear Model

GLMs are flexible generalization of ordinary linear regression approaches that can be used also in classification problems. Owing to they do not constrain the data into unnatural scales, they allow non-linearity and non-constant variance structures in the data (Hastie and Tibshirani, 1990). These models assume the relationship between the mean of the response variable and the linear combination of the predictor variables through a link function. In linear regression models, it is assumed that residuals are normally distributed with constant variance; however, GLMs generalize this assumption permitting other types of error distributions. Data can be from several probability distributions such as normal, binomial, Poisson, negative binomial, or gamma distribution, many of which better fit the non-normal error structures of most ecological data (Guisan et al., 2002). Therefore, GLMs are more eligible to analyse ecological relationships than others which are poorly represented by classical Gaussian distributions (Austin, 1987).

The GLM function embedded in BIOMOD permits to use linear, quadratic or polynomial functions, and it can be selected either by the Akaike information criterion (AIC) developed by Hirotugu Akaike in 1971 (Akaike, 1973) or by Bayesian information criterion (BIC) developed by Gideon E. Schwarz in 1978 (Schwarz, 1978). Both are criterion for model selection among a certain number of models. During the model building phase, it is possible to raise the likelihood by including more parameters, though it may cause overfitting. Both BIC and AIC aim to fix this issue by adding a penalty term for the number of parameters included in the model.

This method provides a less restrictive form than classic multiple regressions by providing error distributions for the dependent variable other than normal and non-constant variance functions. When there is no linear relationship with the predicting variable, the algorithm can include more complex function such as polynomial in order to simulate skewed and bimodal responses. Some of the shortcomings of the GLM approach are the necessity to know prior hand which is the relationship between predicting variables and species occurrence, in addition to lack of flexibility in some instances to approximate to the true regression surface (McCullagh, 1989; Team Biomod, 2012).

GLM Simple: Used only linear terms:

$$Y_1 = X_1 + X_2 + X_3 + (X_1 * X_2) + (X_2 * X_3) \qquad \textit{(Eq. 6)}$$

GLM Quad: Used linear, 2nd and 3rd order:

$$Y_1 = X_1 + X_1{}^2 + X_1{}^3 + X_2{}^2 + X_3{}^3 \qquad \textit{(Eq. 7)}$$

GLM Poly: Use ordinary polynomial terms:

$$Y_1 = f(X_1 + X_1{}^2 + X_1{}^3) + f(X_2 + X_2{}^2 + X_2{}^3) + f(X_3 + X_3{}^2 + X_3{}^3) \qquad \textit{(Eq. 8)}$$

GLM is a very popular modelling approach that has been widely used to model and predict habitats and species distribution (Guisan & Zimmermann, 2000; Sanchez-Zapata et al., 2007). The formula object was set to be "quadratic" (default) and the information criteria for the stepwise selection procedure was the Akaike Information Criteria (AIC). GLM approach implemented in BIOMOD2 only runs on presence-absence data, so binomial distribution family was used.

Random Forest "RF"

The Random Forest algorithm was developed by (Breiman, 2001) as an extension of bagging classification Trees (Archer & Kimes, 2008). This method has been implemented in the "randomForest" library programmed by Andy Liaw and Matthew Wiener, and has the additional capability to apply random feature selection at each node and no stopping rule. This procedure diminish the correlation among trees so that the predicting error of the forest.

Random Forests grows plenty of classification trees, and each tree uses a bootstrapped sample from the original learning sample. For every forest tree, the model provides a predicted class for each observation. Thus, each tree gives a classification in which is counted the number of votes among all the trees in the forest.

Each tree grows as much as possible and there is no pruning. It is important to mention that the correlation between trees in the forest has consequences in the prediction error. When there is a high tree correlation, the forest error rate would be high too. Whereas having strong individual trees provide lower error rates, and makes them strong classifier. When the number of predicting variables is reduced, the correlation and the strength tend to decrease accordingly. Hence, there is an optimal range to the number of variables to find optimal solutions to the problems. Fig. 40 represents the architecture of random forest algorithm for 2 trees. Each internal node is a "test" on an attribute, every branch indicates the outcome of the test, and the leaf node represents a class label.

Random forest trees grow as follows (Breiman, 2001) (https://www.stat.berkeley.edu/~breiman/RandomForests/cc_home.htm#papers):

1. If the number of cases in the training set is N, sample N cases at random - but with replacement, from the original data. This sample will be the training set for growing the tree.

2. If there are M input variables, a number m<<M is specified such that at each node, m variables are selected at random out of the M and the best split on these m is used to split the node. The value of m is held constant during the forest growing.
3. Each tree is grown to the largest extent possible. There is no pruning.

Random Forest is widely used in the scientific community to regression or classification problems and it has demonstrated its efficiency working with large databases. It can deal with thousands of independent variables and analyse which of those are more relevant in the classification problem. In addition to that, it also offers an experimental technique to identify variable interactions.

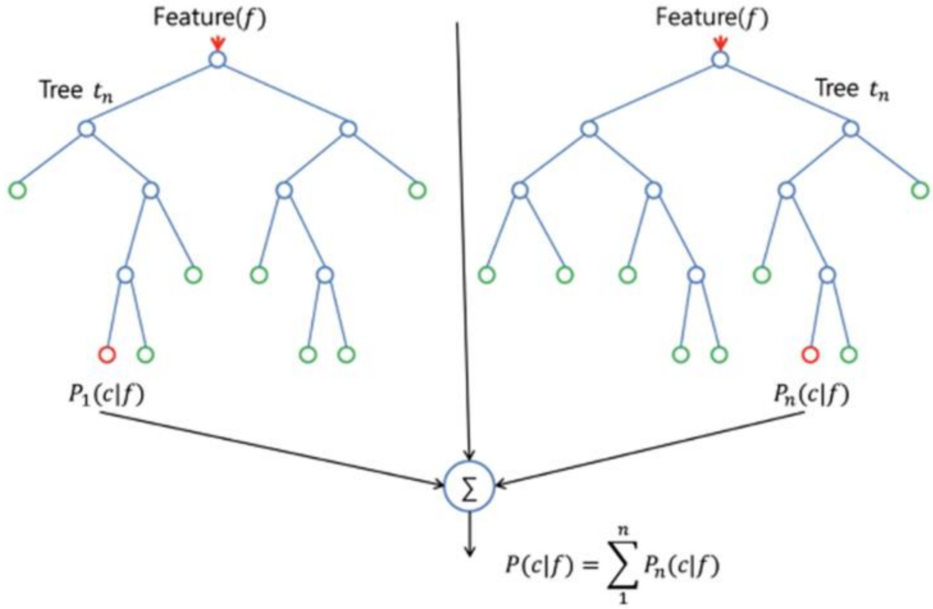


Figure 40. Architecture of Random Forest algorithm with 2 trees. Source: Niklas Donges: https://towardsdatascience.com

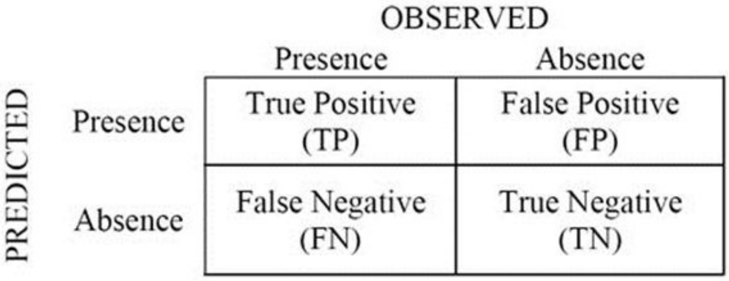
A RF classifier will have most of the hyper-parameters that decision tree classifiers have, in addition to the hyper-parameters of bagging classifiers. Most of the time, RF is very robust against overfitting since it creates random subsets that will use to build smaller trees. And then it combines the subtrees. The main disadvantage of RF is the processing time, which is closely linked with the number of generated trees. It turns out that accurate predictions require higher number of trees, so that slower processing times for the model. Biomod uses for RF a default value of 500 trees, although can be tuned through the “Biomod.Models” function, and it retrieves the variable importance within the model.

Random Forest algorithm is a flexible and easy to use ML approach that has been demonstrated to have good predictive performances in ecology and species distribution (Mi et al. 2017). It can be used both for classification and regression problems. The most important tuning parameters are the “mtry” (number of variables randomly selected at each split of the tree as it grows) and “ntree” (number of trees). We have set these two parameters with their default values: “ntree” = 500 (Elith & Graham, 2009; Benito et al., 2011) and “mtry” (in classification) = the squared of the number of variables (Genuer et al., 2010). The minimum size of terminal nodes “NodeSize” and the maximum number of terminal nodes “MaxNodes” were also left with their defaults values, which are 5 and null respectively (Thuiller et al., 2016).

Different predictive methods result in different accuracy measures, so that they need to be carefully evaluated. This procedure is referred as Model Evaluation.

Different measures can be used to evaluate the accuracy of the model, and each case would require different techniques according to the nature of the study (Fielding & Bell, 1997; Liu et al., 2011). Many of these statistical metrics to evaluate models are based on threshold values. Predicted values above that specific threshold would indicate a prediction of ‘presence’, whereas values below the threshold would indicate ‘absences’. Depending on the intrinsic characteristics of the study, it may be convenient to emphasize the weight of false absences; others give more weight to false presences, according to their prevalence in the dataset. Nevertheless, there are some statistic metrics to evaluate model performances that do not rely on thresholds such as the Area Under the Receiver Operating characteristic Curve (ROC-AUC) and the correlation coefficient (r).

Before starting to describe the model metrics, it is convenient to introduce a few basic terms in the subject. Sensitivity and specificity are products derived from a confusion matrix, which is widely used to evaluate binomial distribution models (Fig. 41). In the ecological field, a confusion matrix compares the capability of an ecological-habitat model to accurately predict observed presences and absences by tabulating true positives (TP), false positives (FP), false negatives (FN), and true negatives (TN) predictions. Sensitivity is a measure of commission error (TP/(TP+FN)) and specificity is a measure of omission error (TN/(TN+FP)). Sensitivity and specificity are derived independently of each other and also independent of prevalence, the proportion of presence locations. Sensitivity and specificity values range from 0, indicating a high error, to 1 which indicates very high agreement between observed and predicted values (Torres et al., 2008).



		OBSERVED	
		Presence	Absence
PREDICTED	Presence	True Positive (TP)	False Positive (FP)
	Absence	False Negative (FN)	True Negative (TN)

Figure 41. A confusion matrix that describes the predictive capacity of SDMs. TP = presence observed and predicted by model; FP = absence observed but predicted as a presence location; FN = presence observed but location predicted as absence; TN = absence observed and predicted by model.

There are many other model performance metrics that have been generated from Sensitivity and specificity (Fielding & Bell, 1997; Pearce & Ferrier, 2000) such as ROC-AUC, Cohen's kappa statistic or TSS.

In spite of the generalized use of some statistics to assess model performances in ecology, there is still an ongoing debate about their use (Allouche et al., 2006; Ruete & Leynaud, 2015). We decided to select 4 broadly used evaluation methods for cross-comparisons that are included in Biomod2: Relative Operating Characteristics "ROC" (Hanley & McNeil, 1982), Cohen's Kappa "KAPPA" (Monserud & Leemans, 1992), True Skill Statistic "TSS" (Allouche et al., 2006) and Accuracy.

1. The ROC evaluation method uses the area under the curve (AUC) to discriminate between events and non-events. It is a very common measure of model accuracy in order to evaluate classification machine learning models. The area under the receiver operating characteristic curve has several interpretations:

    a. The expectation that a uniformly drawn random positive is ranked before a uniformly drawn random negative.
    b. The expected proportion of positives ranked before a uniformly drawn random negative.

c. The expected true positive rate if the ranking is split just before a uniformly drawn random negative.
d. The expected proportion of negatives ranked after a uniformly drawn random positive.
e. The expected false positive rate if the ranking is split just after a uniformly drawn random positive.

In a ROC curve, the Y-axis represents the True Positives while the True Negatives are on the X-axis (Fig. 42). All models with points below the diagonal have worse performance than a model which makes predictions randomly. Its score ranges from 0 (worst score) to 1 (perfect score), and values under 0.5 are considered to indicate random chance (Fawcett, 2006). High ROC-AUC values indicate that the sites are suitable to species presences, while in contrast lower values indicate presence uncertainty.

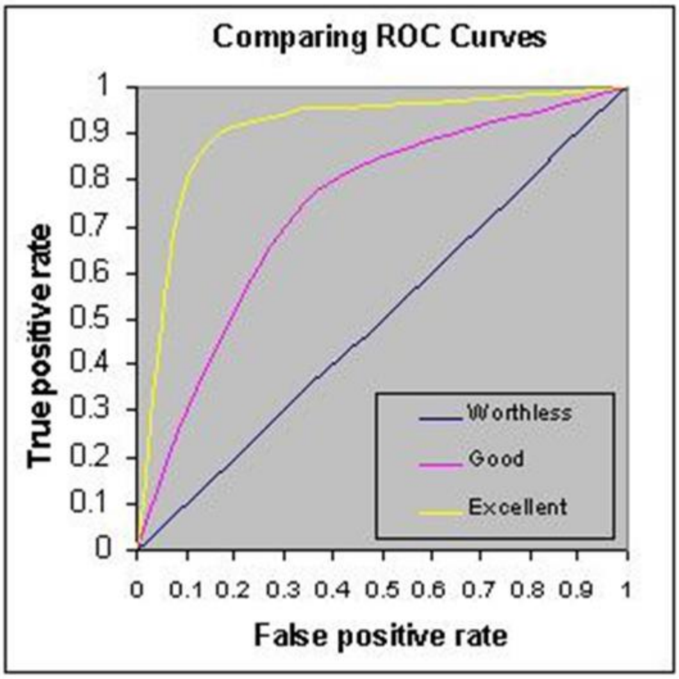


Figure 42. Roc curve comparison. Source: Thomas G. Tape, http://gim.unmc.edu

2. KAPPA statistic is one of the most used methods to measure model performance on presence-absence predictions and it indicates the relative accuracy of the forecast comparing with the random chance. In other words, it calculates the difference between how much agreement

is actually present (“observed” agreement) compared to how much agreement would be expected to be present by chance alone (“expected” agreement) (Viera & Garret, 2005). It ranges between -1 (the worst score) to 1 (perfect score), where values under 0 indicates no predictive skill. It corrects the overall accuracy in the prediction of a model by the accuracy expected to happen by random chance. Other advantages are its tolerance concerning zero values in the confusion matrix or the simplicity of the procedure (Manel et al., 2001).

Despite the wide use of Kappa among the scientific community, some authors have highlighted the limitations of this measurement to correctly assess model accuracies. Kappa is inherently dependent on prevalence, and this introduces bias and errors in the estimation of performance (Lantz & Nebenzahl, 1996; Allouche et al., 2006)

3. Accuracy expresses the fraction of the predictions that are correct, and ranges between 0 (the poorest) to 1 (the best). Typically, this metric does not provide enough information to ensure that a machine learning model is robust to make predictions on unseen data (Brownlee, 2014).

   Classification accuracy is essentially the number of correct predictions divided by the total number of predictions, multiplied by 100 to show it in percentage. But this measure can sometimes be misleading. In some instances, it is more recommended to choose a model with lower accuracy performance because it predicts better to that specific problem. It usually occurs in those machine learning problems where there is a great class imbalance. The model will predict with very high accuracy all those cases that tend to be majority in the dataset, whereas small occurrences will go unnoticed. This phenomenon is called Accuracy Paradox. To get further information about these metrics, access to the Collaboration for Australian Weather and Climate Research site “CAWCR” (CAWCR, 2015).

4. TSS statistic. It a metric developed by (Allouche et al., 2006) that keeps all the advantages of the Kappa statistic, but corrects its dependency to prevalence. It is widely used in ecology to be a simple and intuitive metric to evaluate species distribution models for presences and absences of species (Liu et al., 2009; Barbet-Massin et al., 2012).

Although these evaluation procedures could be used independently, it is recommended to use several of them to assess the accuracy of statistical models. Table 6 shows an index for classifying model prediction accuracy:

ACCURACY	AUC	KAPPA/TSS
Excellent or high	0.9 – 1	0.8 – 1
Good	0.8 – 0.9	0.6 – 0.8
Fair	0.7 – 0.8	0.4 – 0.6
Poor	0.6 - 0.7	0.2 – 0.4
Fail or null	0.5 – 0.6	0 – 0.2

Table 6. *Index for classifying model prediction accuracy (Thuiller et al. 2009).*

The Biomod2 package allows the user to randomly subset the original dataset into 2 subsets, calibration and validation. The dataset was divided into 70% of the data to calibrate the models and 30 % to validate the predictions. When found the best scenario and variables to choose, we repeated the process 5 times to the best performing algorithm to obtain a robust test of the model, where each replicate uses a unique random split 70% - 30% of the data (Thuiller et al., 2009). Presence and pseudo-absences were set to have the same importance in the calibration process, with a prevalence value of 0.5.

Based on model results, the best performing algorithm with the best scenario and representative statistic of SM values is selected. Then, an optimization process was applied to ensure that the settled algorithm is presenting the best possible performance (Brownlee, 2014). We tuned the algorithm hyper-parameters to find their best combination in terms of predictive performance, and finally an objective comparison of the results. The best tuning parameters were chosen to run the final model.

We used the response curves to assess the prediction of the model, which are independent of the used SDM algorithm. The response curves allow comparing the probability of presence based on ROC, TSS and Kappa metrics with the variables used in the model. It facilitates the interpretation of relationships between environmental variables and predicted responses of species, even though they may not be apparent from the outputs of the model (Elith et al., 2005). The contribution of each variable to the final model is analysed. So the higher the value is, the more influential the variable will be in the model; where a 0 value means no influence at all.

The aim is to evaluate desert locust presence probabilities to locate potential breeding areas, based on remotely sensed SM conditions.

3.3. Results

SM monthly averages (Fig. 43-44) suggest a spatial correlation with usual breeding areas, indicating high SM values in the south for the months: July, August, September and October; whereas higher values are found in the north and north-eastern part of Mauritania during December, January and February. In general, autumn breeding sites (blue dots in Fig. 44) do not show visual correlation with the monthly mean SM values.

GLM and RF algorithms were used with SM variables that relied upon various time intervals (16, 12, 8 and 6 days) and their maximum, minimum or mean (Table 7 and Table 8) SM values. Based on ROC, TSS and KAPPA statistics, we obtained performance scores with 1 iteration from an independent test dataset. The results showed that Random Forest “RF” obtained the best performance for our study, whereas GLM performed far behind.

The highest scores were obtained when the time interval was 6 days (scenario D) and the representative SM value was the minimum acquired within the time interval. According to Table 7, the RF algorithm obtained a high o very good performance with respect to ROC-AUC with 0.95 and good performance for Kappa and TSS statistics with 0.75. The sensitivity and specificity was over 87%. Slightly lower values are found when using the maximum or mean SM values across the scenario D, demonstrating the suitability of 6 days coverage time to build the SM variables of the model. Scenario A (16 days) obtained the worst model performance when using mean SM values as representative of the given interval. Nevertheless, this scenario still obtained a fair performance of 0.6 for TSS and kappa statistics, and ROC-AUC = 0.90 when using the minimum SM value across their time length.

Model performance increases when the time interval of the variables gets smaller and the representative SM value is the minimum for such period. Therefore, we suggest regarding at minimum SM values over 6 days period to link solitarious hopper presences and SM values of the ground.

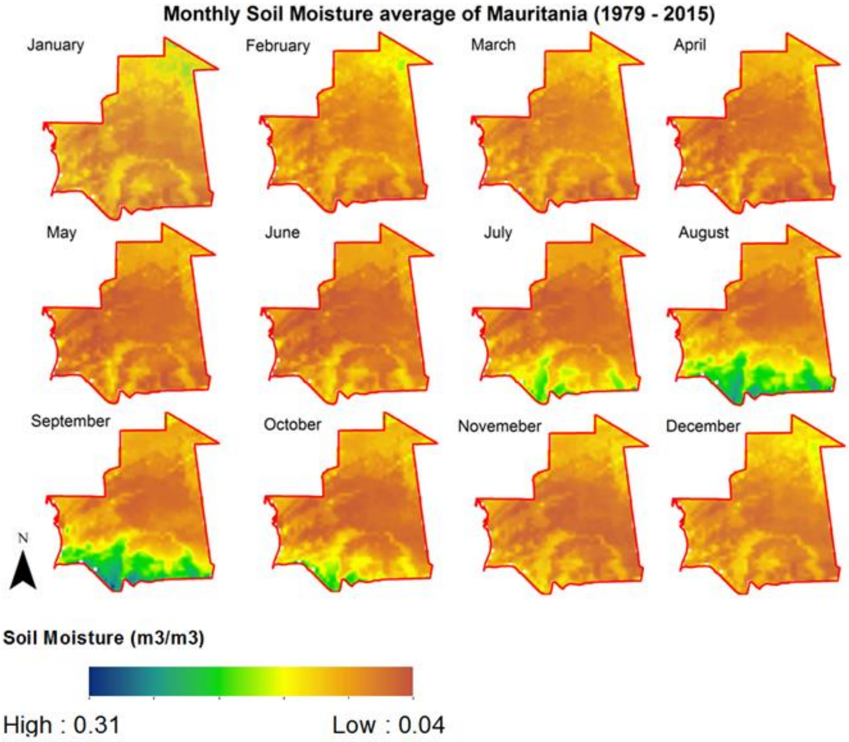


Figure 43. SM average per month for the time span 1985 – 2015, units is in m^3/m^3.

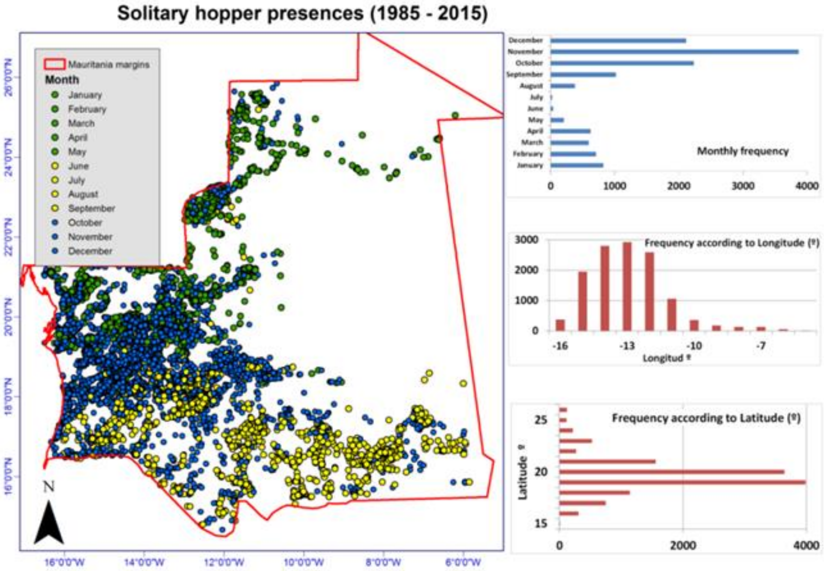
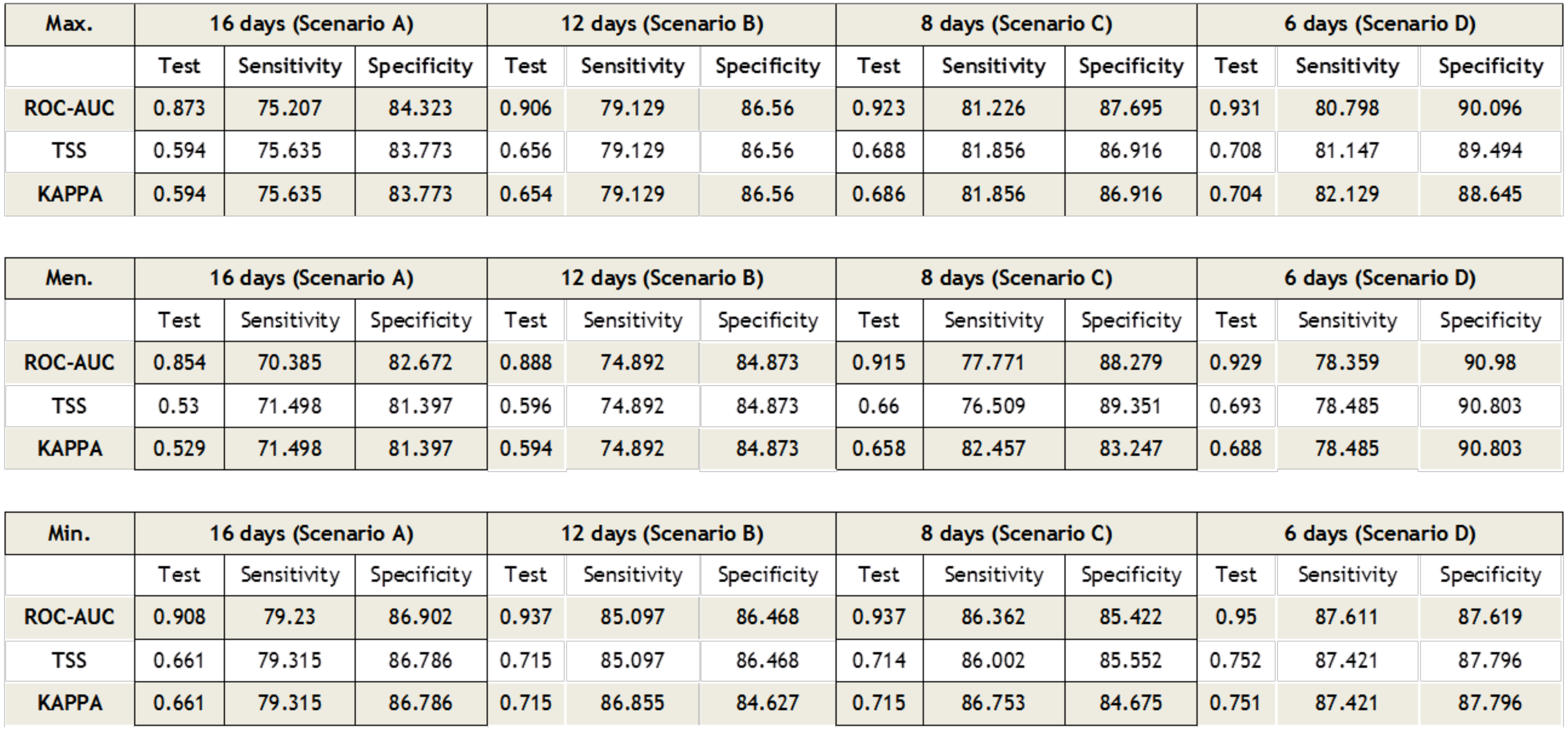


Figure 44. Location map of solitarious hopper presences reported from 1985 to 2015, grouped per months (on the left). Frequency histograms of presences based on months, latitude and longitude are found on the right.

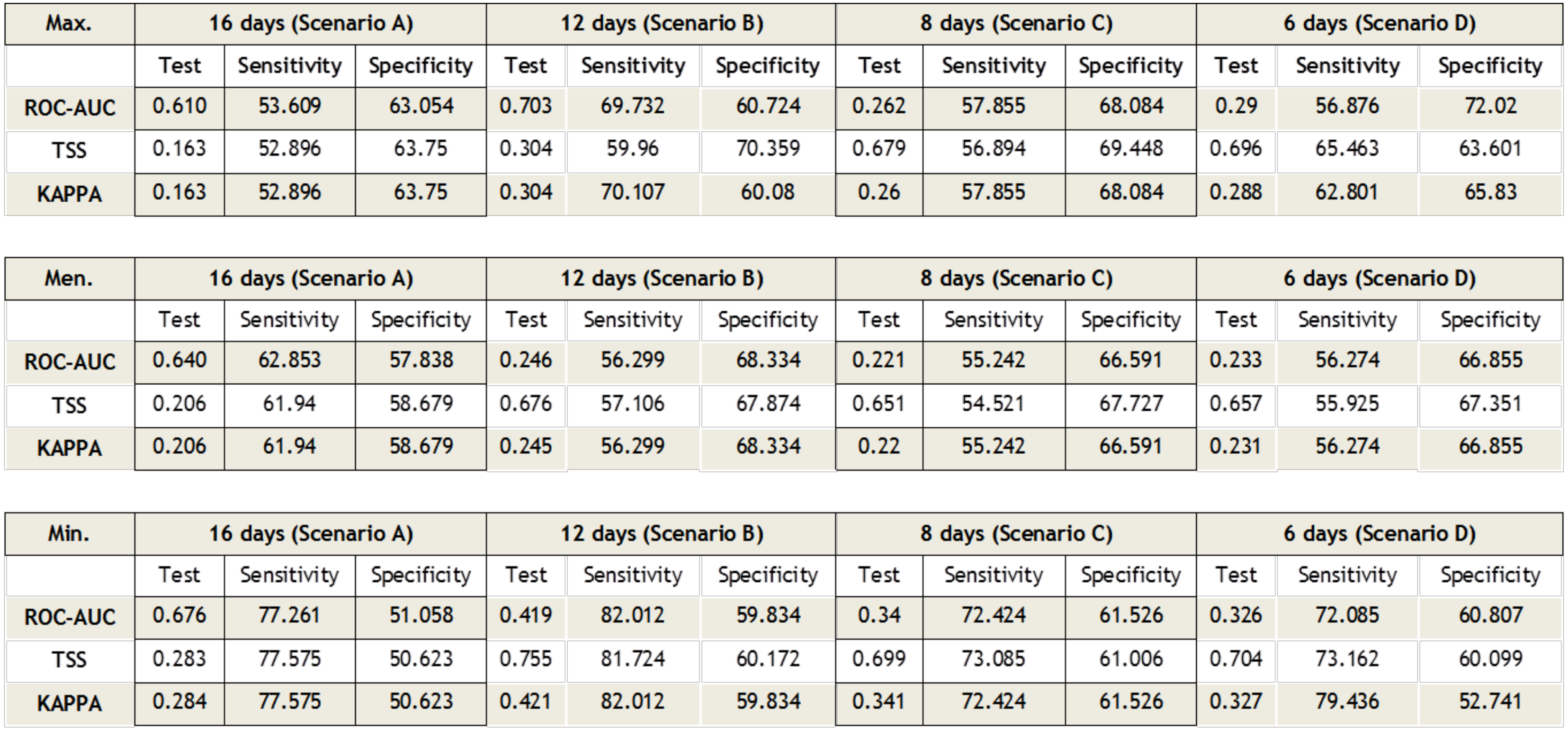


Max.	16 days (Scenario A)			12 days (Scenario B)			8 days (Scenario C)			6 days (Scenario D)		
	Test	Sensitivity	Specificity	Test	Sensitivity	Specificity	Test	Sensitivity	Specificity	Test	Sensitivity	Specificity
ROC-AUC	0.873	75.207	84.323	0.906	79.129	86.56	0.923	81.226	87.695	0.931	80.798	90.096
TSS	0.594	75.635	83.773	0.656	79.129	86.56	0.688	81.856	86.916	0.708	81.147	89.494
KAPPA	0.594	75.635	83.773	0.654	79.129	86.56	0.686	81.856	86.916	0.704	82.129	88.645

Men.	16 days (Scenario A)			12 days (Scenario B)			8 days (Scenario C)			6 days (Scenario D)		
	Test	Sensitivity	Specificity	Test	Sensitivity	Specificity	Test	Sensitivity	Specificity	Test	Sensitivity	Specificity
ROC-AUC	0.854	70.385	82.672	0.888	74.892	84.873	0.915	77.771	88.279	0.929	78.359	90.98
TSS	0.53	71.498	81.397	0.596	74.892	84.873	0.66	76.509	89.351	0.693	78.485	90.803
KAPPA	0.529	71.498	81.397	0.594	74.892	84.873	0.658	82.457	83.247	0.688	78.485	90.803

Min.	16 days (Scenario A)			12 days (Scenario B)			8 days (Scenario C)			6 days (Scenario D)		
	Test	Sensitivity	Specificity	Test	Sensitivity	Specificity	Test	Sensitivity	Specificity	Test	Sensitivity	Specificity
ROC-AUC	0.908	79.23	86.902	0.937	85.097	86.468	0.937	86.362	85.422	0.95	87.611	87.619
TSS	0.661	79.315	86.786	0.715	85.097	86.468	0.714	86.002	85.552	0.752	87.421	87.796
KAPPA	0.661	79.315	86.786	0.715	86.855	84.627	0.715	86.753	84.675	0.751	87.421	87.796

Table 7. Random Forest results per time-scenario, representative statistic to generate the SM variables (maximum, mean or minimum per each interval) and the model performance per statistical metric. Sensitivity and specificity are expressed in %.



Max.	16 days (Scenario A)			12 days (Scenario B)			8 days (Scenario C)			6 days (Scenario D)		
	Test	Sensitivity	Specificity	Test	Sensitivity	Specificity	Test	Sensitivity	Specificity	Test	Sensitivity	Specificity
ROC-AUC	0.610	53.609	63.054	0.703	69.732	60.724	0.262	57.855	68.084	0.29	56.876	72.02
TSS	0.163	52.896	63.75	0.304	59.96	70.359	0.679	56.894	69.448	0.696	65.463	63.601
KAPPA	0.163	52.896	63.75	0.304	70.107	60.08	0.26	57.855	68.084	0.288	62.801	65.83

Men.	16 days (Scenario A)			12 days (Scenario B)			8 days (Scenario C)			6 days (Scenario D)		
	Test	Sensitivity	Specificity	Test	Sensitivity	Specificity	Test	Sensitivity	Specificity	Test	Sensitivity	Specificity
ROC-AUC	0.640	62.853	57.838	0.246	56.299	68.334	0.221	55.242	66.591	0.233	56.274	66.855
TSS	0.206	61.94	58.679	0.676	57.106	67.874	0.651	54.521	67.727	0.657	55.925	67.351
KAPPA	0.206	61.94	58.679	0.245	56.299	68.334	0.22	55.242	66.591	0.231	56.274	66.855

Min.	16 days (Scenario A)			12 days (Scenario B)			8 days (Scenario C)			6 days (Scenario D)		
	Test	Sensitivity	Specificity	Test	Sensitivity	Specificity	Test	Sensitivity	Specificity	Test	Sensitivity	Specificity
ROC-AUC	0.676	77.261	51.058	0.419	82.012	59.834	0.34	72.424	61.526	0.326	72.085	60.807
TSS	0.283	77.575	50.623	0.755	81.724	60.172	0.699	73.085	61.006	0.704	73.162	60.099
KAPPA	0.284	77.575	50.623	0.421	82.012	59.834	0.341	72.424	61.526	0.327	79.436	52.741

Table 8. GLM results per time-scenario, representative statistic to generate the SM variables (maximum, mean or minimum per each interval) and the model performance per statistical metric. Sensitivity and specificity are expressed in %.

RF was the best performing algorithm, using scenario D and the minimum SM values obtained in each time interval. RF algorithm was tuned for the two most important hyper-parameters: the number of trees “ntree” (50, 500, 1000, 2000 and 4000) and the number of variables randomly sampled as candidates at each split “mtry” (2, 4, 6, 8 and 10). Firstly, we optimized the number of trees and secondly the mtry. As observed in Fig. 45, the default parameters established by Biomod2 for Random Forest (ntree = 500 and mtry = 4) obtained the best model performance, whose evaluator metrics did not greatly differed from other tuning options. The poorest performance was obtained with ntrees = 50 and mtry = 2 (lower value parameters than the default proposed by BIOMOD2). The increase of ntrees or mtry has not improved model results, with relatively very small changes in model performance. It is also noticeable how the ROC-AUC evaluator remains more or less constant across the different attempts, while the changes of TSS and KAPPA are slightly larger.

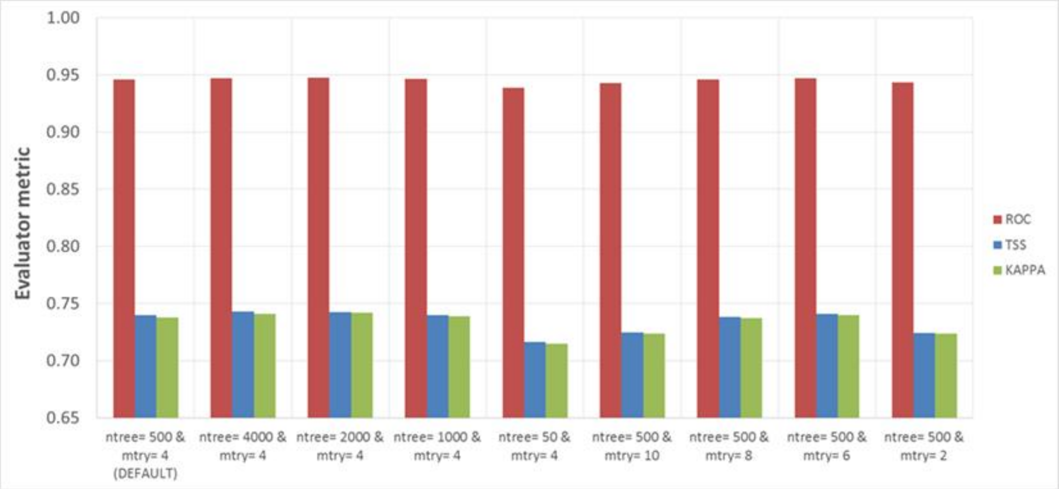


Figure 45. Comparison of different RF results using different tuning parameters, with scenario D and the minimum SM value per interval (best performances in the previous step). X-axis represents the parameter changes and Y-axis the model performance of each tuning combination according to ROC, KAPPA and TSS statistics.

Therefore, the best algorithm (RF) was optimized after the tuning phase with ntree = 500 and mtry = 4. And the best model results were obtained using the variables created with scenario D and the minimum SM reached at each time interval. Finally, we ran RF for 5 iterations to aim for robust results. Model performance scores are compiled in (Table 9).

RF 5 iterations	Test	Sensitivity	Specificity
ROC-AUC	0.946	84.911	89.105
TSS	0.740	85.468	88.461
KAPPA	0.738	87.325	86.508

Table 9. RF results after 5 iterations using the best scenario (6 days) with the minimum SM values obtained in each interval. Sensitivity and specificity are expressed in %.

The metric scores are in accordance with the ones obtained in Table 7 for the same scenario (D) and chosen variables (minimum SM). In general, testing values and sensitivity are slightly lower, while ROC-AUC and TSS specificity are somewhat higher. In essence, score values do not differ considerably when running more iterations and averaging their metrics. The impact of SM variables in the final model results (RF, scenario D, minimum SM) are summarized in Fig. 46.

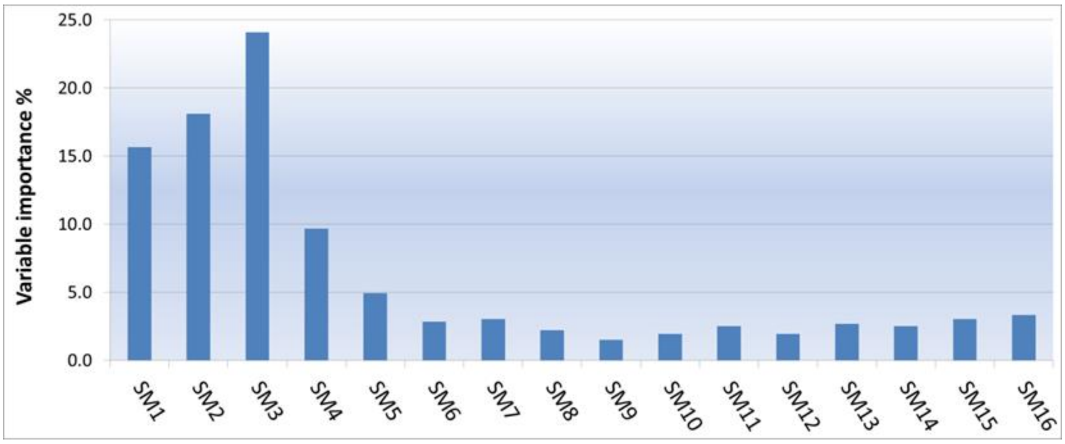


Figure 46. Variable importance in % of each variable from Scenario D (6 days), using the minimum SM value obtained in each time interval for RF.

The most relevant variables for the outcome model were SM1, SM2, SM3 and SM4 which stand for the minimum SM values obtained between 95 and 90, 89 and 84, 83 and 78, 77 and 72 days before the sighting record respectively. Fig. 46 indicates the greater impact of these mentioned variables (mostly over 10 %) in comparison with the rest, which do not overcome the 5 % per each. Fig. 47 shows the response curves of these four more relevant variables that are over 5

% of importance. The plots suggest some potential thresholds of SM content to increase the probability of presence. The minimum SM values acquired during SM1, SM2, SM3 and SM4 denote a positive influence in hopper occurrences. It is observed that the range of SM values in which the probability of presence is over 0.5 varies. Presence probabilities tend to keep steady by 0.5 when SM values reaches 0.15 for SM1, SM2 and SM4. SM3 keeps a high probability over that value. Nevertheless there is a common trend by the 0.07 (m^3/m^3) to increase the probability of presence within 72 and 95 days afterwards.

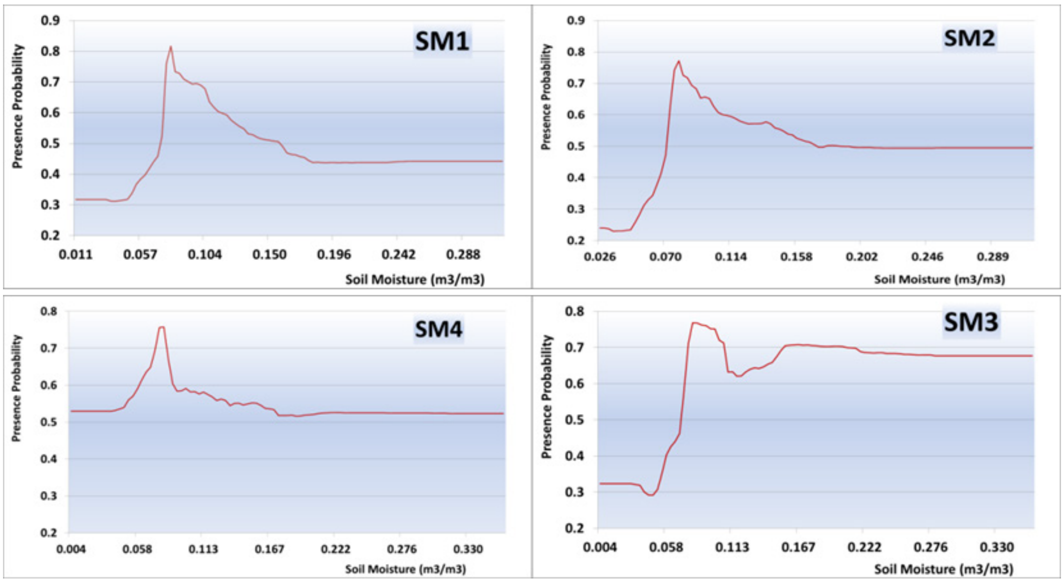


Figure 47. Response curves for hopper's desert locust for SM1, SM2, SM3 and SM4 variables for RF. The Y-axis represents the presence probability of the prediction, while X-axis stands for SM values.

3.4. Discussion

It is widely assumed that rainfall over 25 mm in two consecutive months is generally enough to locust breeding and development (FAO & WMO, 2016). Nevertheless, remotely sensed precipitation in arid environments has some limitations such as high rainfall overestimation due to sub-cloud evaporation (Dinku et al., 2011). Aiming to solve the problems associated with remote sensing precipitation, we have analysed the link from ESA CCI SM remote sensing product with field surveys of hopper desert locust from SWARMS – FAO. In addition, we assess the suitability of this SM product to derive desert locust breeding sites.

It has been long known the importance of SM in egg laying and development, as well as the role of fresh vegetation which is greatly determined by water availability in the soil (Pedgley, 1981). SM monthly averages suggest a spatial correlation with summer and winter breeding areas. It coincides with the regional climatic conditions of Mauritania as reported in other works (Van Huis et al., 2007; Babah Ebbe, 2012).

Winter rainfall is usual in the north while summer rain in the south of the country. Nevertheless, typical autumn breeding areas do not seem to be accounted for the monthly SM patterns. In arid environments, there is a direct relationship between rainfall and SM (Nicholson & Farrar, 1994; Brocca et al., 2013), so that problems such as sub-cloud evaporation (Dinku et al., 2011) may be avoided with the applied methodology. Despite ESA CCI SM only senses the first 5 cm of the top soil, and desert locust lay eggs usually at depth down to 10 cm; this system seems appropriate due to the strong relationship of the top SM with deeper layers (Albergel et al., 2008).

Our analysis reveals the importance of variable creation as a previous step to modelling. We have tested different time intervals for the variable creation. In addition, we have chosen different representative SM values for the given time-span (maximum, mean and minimum) and presence and pseudo-absence sites. Perhaps, the use of pseudo-absences may be controversial in certain fields because bring some sort of uncertainty into the results (Hastie & Fithian, 2013). However, their use is generally justified for providing a set of conditions available in the region that need to be included in the SDM (Phillips et al., 2009).

The highest performance was acquired by the RF algorithm when dividing the whole survey time into ranges of 6 days, and selecting the minimum SM as the variable value. Even though previous literature (Sanchez-Zapata et al., 2007) have used the GLM model with a binomial distribution to identify potential factors that determine species presences or absences, GLM approach did not perform well in our study. According to (Thuiller et al., 2009), our RF model has had an excellent performance based on ROC-AUC metric with 0.946, and a good performance for TSS and Kappa statistics with 0.740 and 0.738 respectively. The probability of hopper detection (sensitivity) is over 85 %, being able to correctly identify (specificity) over 86 % of the pseudo-absence records. The variables with more weight in the model results were SM1, SM2, SM3 and SM4, whose cover time range from 95 to 72 days before the sighting record. Locust eggs develop and hatch successfully when there is enough moisture in the soil (Shulov & Pener, 1963), whereas insufficient moisture may stop egg development or dry them out (Pedgley, 1981). Our results indicate that the minimum SM conditions over at least 6 days should remain higher than 0.07 m^3/m^3. This value is in accordance, although slightly lower, with the SM range proposed by (Escorihuela et al., 2018) which are between 0.10 and 0.20 m^3/m^3. Hopper mortality is closely linked to food shortage (Pedgley 1981), which in arid environments is closely linked with inadequate precipitation (Bennett, 1976; Teklu, 2003). Thus, remotely sensed SM may also be a good indicator of suitable conditions to infer hopper presences and locate breeding areas. A good understanding of the geographical relationship between desert locust populations and their potential breeding habitats can improve desert locust survey and control operations (Teklu, 2003).

The applied methodology offers very promising results to correctly identify breeding areas based on 30 years of SM values. The ESA CCI SM dataset is the most complete and consistent global SM data record available (Wagner et al., 2012). To the best knowledge of the authors, there has not been any previous desert locust analysis using this SM dataset. Given the acknowledged importance of SM for desert locust and the length of ESA CCI SM dataset, our results may signify a breakthrough to complement the ongoing locust monitoring techniques used until today.

3.5. Conclusions

This chapter has evaluated the importance of satellite SM products to locate breeding areas for desert locust in solitarious phase. The survey has used the most complete and consistent available SM dataset, the ESA CCI SM product.

A machine learning approach was used to assess the relationship between Desert locust presences and antecedent SM conditions and estimate the accuracy of our model. This chapter confirms the robustness of the applied methodology, where 30 years of locust records and SM values were used to feed the model. Some uncertainty is expected due to the use of pseudo-absence data, nevertheless the creation of pseudo-absences is an accepted method in ecology when there is a lack of true-absence data.

The monthly SM values suggest a spatial correlation with usual breeding areas in Mauritania. So far, desert locust suitable sites have been mainly delimited based on rainfall estimates from satellite remote sensing. However, some literature marks the high overestimation of these products over dry regions. Therefore, we suggest the use of ESA CCI SM product to overcome that problem either to complement other rainfall products or to substitute them in certain instances of high uncertainty.

Furthermore, we have modelled quantitatively the relationship between hopper presences and SM under different scenarios and variables. The best model performance was obtained by Random forest, when using the minimum SM value within 6 days interval, for a maximum survey time of 95 days before the sighting date. The validation phase acknowledged the suitability of this methodology to identify hopper presences with a ROC-AUC of 0.94 and TSS & Kappa of 0.74. The importance of SM thresholds and survey time has also been addressed: when the minimum SM value of a certain location overcomes 0.07 m^3/m^3 during 6 days or more, the area becomes favourable as breeding zone. However, these values should be taken carefully. Variable importance showed that the most relevant variables of the model would cover between 95 and 72 days before the sighting record. It implies, as highlighted in other works, that certain SM levels need to be maintained over time not just for egg laying, but egg development and hatching. So that, monitoring periods should be longer than 6 days to those favourable areas for a successful egg development and hatching.

According to these results, the observed SM during certain periods stands as a very reliable contributor to accurately predict hopper presences in Mauritania; and consequently its monitoring may reduce the locust impact on local communities. The next chapter aims to ensemble other environmental variables along with SM datasets to improve model performance. This innovative approach may correct some shortcomings of current desert locust early warning systems.

Chapter 4. Multivariate ensemble model to detect desert locust

4.1. Introduction

Previous studies have demonstrated the importance of vegetation density to account for desert locust phase changes from solitarious to gregarious (Cisse et al., 2013), what even affects the hatching time of the eggs (Nishide et al., 2015). Meanwhile, other authors aimed to explain variables such as precipitation (Cressman, 2013; Lazar et al., 2015) or moist status (Tucker et al., 1985) to describe good habitat conditions for breeding. In addition, some laboratory studies have demonstrated the egg sensitivity of different hopper species to small changes in soil temperature (Nishide et al., 2017). So then, favourable breeding conditions are to be a combination of many environmental circumstances.

The main concern of local authorities is to control desert locust populations before a plague is established. Outbreaks are led by an aggregation of variables that usually occurs in areas smaller than 10,000 km^2 (Van Huis et al., 2007). The outbreaks might not end up being a plague due to human intervention or natural limits, and it needs one year at least to be established as plague (Cressman, 2008).

This rising concern to control the population number before they become plague, led FAO to develop the Desert Locust Information Service (DLIS). This project aims to assess and warn about potential outbreaks, and provide the necessary information to operate an early warning system based on Earth Observation Systems and field work. Remote sensing satellite data is nowadays a great asset to study inaccessible or complicated regions, as well as a cost effective method to monitor a wide range of environmental parameters, with a good temporal and spatial resolution (Melesse et al., 2007). As previously described in prior chapters, a wide range of satellite platforms have been using to monitor habitat conditions in an individual basis: Aqua, Terra, NOAA, SPOT, Meteosat, etc.

In order to derive which are the most favourable conditions for desert locust to breed, a wide range of variables need to be explored. It was used a combination of the best performing species distribution models (SDM) to analyse the link between species occurrences and the habitat conditions.

Artificial intelligence algorithms have been introduced in ecological studies for modelling complex systems (Recknagel, 2003), hence to obtain reliable environmental assessments and deeper insight of ecology. Accordingly, reliable assessments will permit local authorities and decision makers to achieve adaptive management with appropriate vision and efficiency (Fukuda & Hiramatsu, 2008). Increasing its popularity, the process of running two or more related but different predictive models and then summarize the results into a single score (termed as "ensemble modelling") arises as a solution to intermodel variations, with clear advantages over single model forecasts (Araujo & New, 2007).

This chapter aims to explore the combination of different environmental variables retrieved by remote sensing techniques to derive optimal breeding conditions for solitarious desert locust. In order to obtain the best predictive results, it was used an ensemble model approach.

4.2. Materials and methods

4.2.1. Field data and pseudo-absences

Schistocerca WARning and Management System (SWARMS) is again the dataset of locust presence records used as "ground truth". As described in previous chapters, it has been used to analyse the link between environmental data and desert locust sightings. We chose solitarious hopper stage as our population target for two reasons: the solitarious phase accounts for non-restricting conditions (Simpson et al., 1999) and hoppers (wingless nymphs) have less mobility than adults due to their lack of wings (Showler, 2008). In order to prevent plagues, it is very important to locate suitable breeding areas, with egg lying being the first stage of their life cycle. Optimal egg development conditions will lead to new hatched hoppers, which cannot travel far distances owing to their lack of wings.

Figure 48. Keith Cressman (FAO-Senior Locust Forecasting Officer) , reviews with the desert locust survey team the use of maps in combination with GPS and compass to accurately determine the locations of survey stops and routes in the field (6 Mar 2000). Source: Locust Watch- FAO

We selected 750 solitarious hopper presences from 01/07/2015 to 01/07/2017. The reason to choose this time interval was to obtain remote sensing data from the new NASA sensor SMAP. A random grid of pseudo-absence points (Zaniewski et al., 2002) was generated using R software (R Development Core Team, 2016). Following the same procedure as detailed in **chapter 3**, we first selected the most suitable areas for hoppers (pink area on the right hand side map on Fig. 49), aiming to avoid geographical bias that might misguide model predictions (Barnes et al., 2014). Hopper records from 1985 to 2017 were used to generate a density map (Fig. 49). It was found that using a 50 km radius buffering mask from each recorded sighting enabled us to mask out those zones with no reported occurrences.

Each random point was then assigned with a random date from mid-2015 to mid-2017. Locust habitats are ephemeral and limited, thus making impossible long-term population studies at a single site (Greathead, 1966). A total of 750 points were created so as to be equal with presence records as observed in other approaches such as (Mateo et al., 2010). Using this approach, we sought to reduce geographical bias, selecting areas with high environmental or geophysical potentialities to observe hoppers.

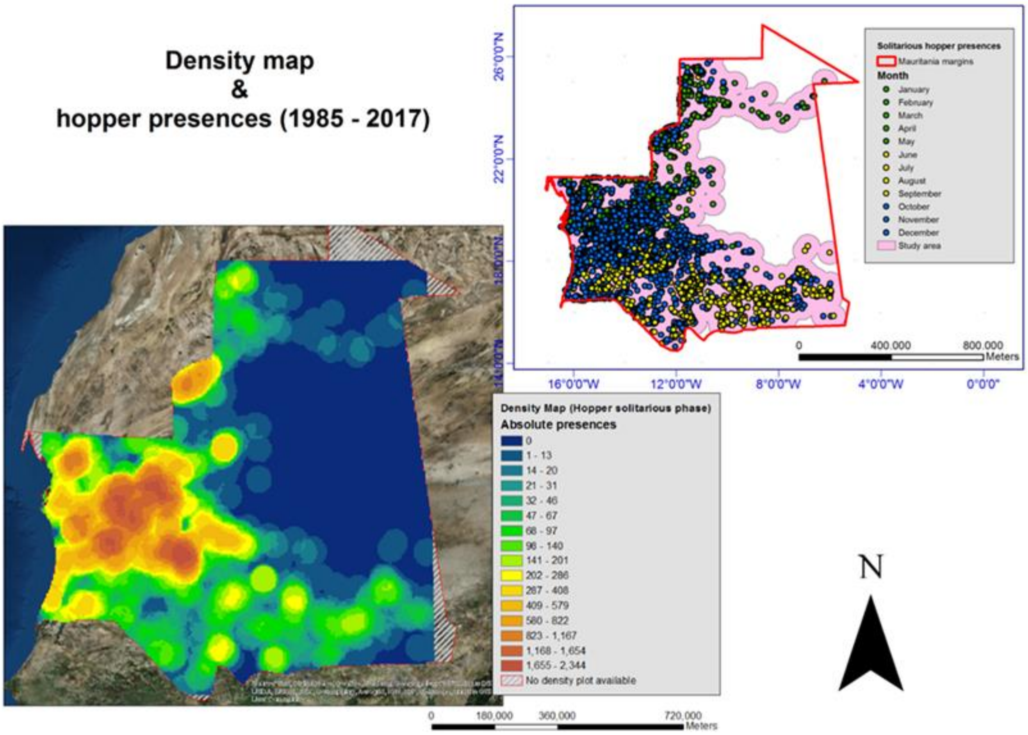


Figure 49. Density map of solitarious hoppers, between 1985 and 2017 (left). Hopper presence is depicted by months (right). The pink area corresponds with the buffering zone taken to create pseudo-absence points. Survey data: SWARMS database.

4.2.2. Environmental variables

4.2.2.1. SMAP L4 9 km EASE-Grid Surface and Root Zone Soil Moisture v3

Soil Moisture Active Passive (SMAP) L4 9 km EASE-Grid Surface and Root Zone Soil Moisture v3 Geophysical Data (SPL4SMGP) is a NASA product that provides global information about surface soil moisture (0-5 cm vertical average) and root zone soil moisture (0-100 cm vertical average), with 9 km spatial resolution, three hour temporal resolution and about 2.5 day latency. Both surface and root zone soil moisture are under 0.04 m^3/m^3 of uncertainty, measured by unbiased root-mean-square-error (Reichle et al., 2017). The SPL4SMGP product was validated against ground observations and is a cloud free product. It covers from 31 March 2015 up to today. SMAP estimates surface soil moisture (0-5 cm) by assimilating the brightness temperature into the NASA catchment land-surface model (Reichle et al., 2014). This model describes the vertical transfer of SM between surface and root zone, assuming unsaturated conditions (Pablos et al., 2018). It was decided to use both variables to cover different soil depths, and thus assess their importance within the model. In addition, we used other research products (not validated) from SMAP satellite (Fig. 50) such as Leaf Area Index (LAI) and surface temperature product, to test their influence on locust presence. LAI accounts for the proportion of the upper leaf area compared to the ground area, and is dimensionless. The LAI associated to each pixel may range from 0 (bare ground) to values greater than 1 (indicating a canopy with multiple layers of leaves per unit of soil surface) (Carlson & Ripley, 1997). Some studies relate LAI products with biomass (Fensholt et al., 2004), such that it was included in the model as a biophysical parameter that can indicate the presence of vegetation (Zheng & Moskal, 2009).

In Mauritania, vegetation is usually sparse and is not always well identified by NDVI at coarse resolutions (Piou et al., 2013). Surface temperature represents the mean soil temperature of the first 5 cm retrieved at 6:00 a.m. and 6:00 p.m. local solar time. We sought to include surface temperature, given the influence that soil temperature has on egg development as reported in (Nishide et al., 2017). Temperatures were converted to°C. To the best of our knowledge, no previous studies have used remotely sensed surface temperature for desert locust purposes. Further information about these products may be found at (https://smap.jpl.nasa.gov/mission/description).

Figure 50. SMAP satellite from NASA during testing phase Source: NASA/Robert Rasmison.

4.2.2.2. Land Surface Temperature and Emissivity 8-Day L3 Global 1km (MOD11A2)

Terra-MODIS Version 7, 8 days composite (MOD11A2) Land Surface Temperature (LST) was employed to analyse the temperature of the surface (https://modis.gsfc.nasa.gov/). The original dataset covers from 5th of March 2000 to ongoing, and the data is stored on a 1 kilometre Sinusoidal grid as the calculated average values of 8 days to obtain cloud free LST images. Terra satellite is a sun synchronous, near polar circular orbit that crosses the equator in descending mode by 10:30 a.m. local time (Fig. 51). This product is retrieved by the generalized split-window method (Wan & Li, 1997), using emissivity of the Thermal Infrared Channels (TIR) 31 and 32. It has been validated against ground based observations (Wan et al., 2002). Temperatures have been converted from Kelvin to Celsius degrees.

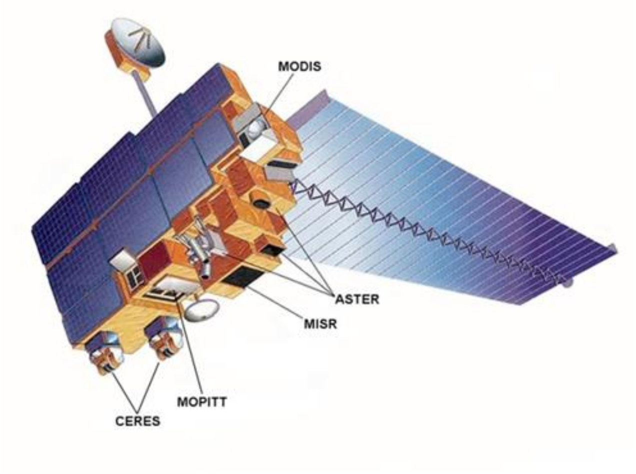


Figure 51. Terra spacecraft with on-board sensors. Source: NASA

4.2.2.3. Vegetation Indices 16-Day L3 Global 250m (MOD13Q1)

In order to characterize the presence of vegetation in our study area, Terra-MODIS Version 6, 16 days composite (MOD13Q1) Normalize Difference Vegetation Index (NDVI) was used for such purposes (Huete et al., 1999). Aiming to obtain a cloud free product, MOD13Q1 provides coherent temporal and spatial comparisons of the vegetation on ground. MODIS sensor retrieve data in 36 spectral bands which cover the wavelength range from 0.4 μm to 14.4 μm (https://modis.gsfc.nasa.gov/).

Even though some authors have exposed the problems to use NDVI in arid areas (Ceccato, 2005), others have shown the goodness of NDVI to reflect sensitively vegetation growth and vegetation cover, in addition to reduce negative impacts caused by clouds/shadows, atmospheric conditions or changes in solar angles (Gao et al., 2000; Vermote et al., 1997). Then, it was decided to include this product in our model to assess vegetation importance to shelter and feed desert locust (Popov, 1985; Uvarov, 1957), hence condition its presence.

4.2.3. Methods

BIOMOD2 platform (Thuiller et al., 2009) implemented for R software was the selected tool to build the models. This computer platform for species distribution modelling (SDM) contains 10 machine learning algorithms to model the relationship between given species and its environment. Two of these algorithms Generalized Linear Model “GLM” and Random Forest “RF” have already been explained in **Chapter 3**. The others are Generalized Additive Model “GAM”, Generalized Boosting Model “GBM”, Classification Tree Analysis “CTA”, Artificial Neural Network “ANN”, Surface Range Envelop or Bioclim “SRE”, Flexible Discriminant Analysis “FDA”, Multiple Adaptive Regression Splines “MARS” and Low Memory Multinomial Logistic Regresion “Maxent.Tsuroka”. These algorithms are described hereunder:

a) Generalized Additive Model

Generalized additive models (GAMs) (Hastie & Tibshirani, 1990) were developed to blend the benefits of GLMs with additive models. In essence, they are GLMs in which the functions are additive and the components smooth. Likewise, a link function is used in the GAM approach to set a relationship between the mean of the response variable and a ‘smoothed’ function of the explanatory variables. One of the main advantages is the GAM ability to tackle highly non-linear and non-monotonic relations between the response variable and the predictor variables.

This approach is also very used to model non-linear relations in ecology, as well as to acquire a better insight of the natural systems (Guisan et al., 2002). GAM algorithms are especially useful when the relations between the predicting variables and the response variable is expected to be rather complex or there is no sign to use a specific model for the data. They tend to generalise quite well the data using a class of equations named “smooothers”. These algorithms fit a smooth curve to each variable and then add the results.

The BIOMOD tool uses a cubic spline smoother, which is a set of polynomials up to degree 3. Identically to GLM, BIOMOD has an automated stepwise procedure to set the most significant variables for each species (Team Biomod, 2012).

$$Y = s\,(X_1, 4) + s\,(X_2, 4) + s\,(X_3, 4) \qquad \text{(Eq. 9)}$$

The user needs to select the number of degree of freedom. The value is 4 by default, what is similar to a polynomial of degree 3.

b) Generalized Boosting Model

This model is a combination of two methods: decision tree algorithms and boosting methods. It fits many decision trees to improve the accuracy of the model. A random subset of the data is done by means of the boosting approach per each new tree. In this new generated tree, the input data shall be weighted to enhance that poorly modelled data by previous trees has more chances to be selected in the new tree. When the first tree is fitted, the model will consider the prediction error of that tree to fit the next tree, and this process repeatedly. Considering previous fitted trees, the model tends to raise the accuracy obtained. This stepwise approach is unique to boosting (Elith et al., 2008).

Generalized Boosting Models have two important parameters that the user needs to specify: Interaction depth and Shrinkage. The interaction depth handles the number of splits in each tree. When such value is equal to 1, the model will not consider any iteration between the variable response and the environmental variables. The shrinkage parameter indicates the contribution of each tree to enlarge the model in a way that small values will permit the generation of many trees.

In order to obtain an optimal prediction, these two parameters need to be adjusted correctly to determine the most suitable number of trees. And the size of the dataset also plays an important role. Datasets with less than 500 presences usually require simple tree models (interaction depth = 2 or 3) with small shrinkage rates to permit the model to grow at least to 1000 trees (Ridgeway, 1999). Some of the most relevant advantages of these models are their capacity to work with large datasets with quite well performances, when the number of environmental variables is rather large in comparison to the number of observations, or their ability to solve problems related to missing values and outliers. In addition to that, they can be used with different type of response variables such as binomial, Gaussian or following Poisson distributions. Due to its stochastic nature, in general it has a good predictive performance that automatically finds the best fit of the model, although it needs at least of 2 predicting variables and the input of absence, pseudo-absence or background points to be run (Team Biomod, 2012). R-BIOMOD uses the gbm library programmed by Greg Ridgeway. This package implements the generalized boosted modelling framework following Friedman's Gradient Boosting Machine (Friedman, 2001). For more details: http://www.salford-systems.com/friedmankdd.php ; www.i-pensieri.com/gregr/ ModernPrediction/ L9boosting.pdf

c) Classification Tree Analysis

This is a good alternative to regression approaches. These types of algorithms do not have any prior assumption about the link between response and predicting variables. Classification Tree Analysis use recursive partitions of the dimensional space defined by the predictors into groups with relatively similar response. Tree building is done by splitting the data repeatedly, and it is defined by a simple rule based on a single independent variable. The data are separated into two exclusive and homogeneous groups at each split.

The algorithm aims to reduce the variance within the subset as much as it is possible. The heterogeneity of a node can be interpreted as a deviance of a Gaussian model (regression tree) or of a multinomial model (classification tree). As a result, there is a graph representing the deviance function of the cost-complexity parameter. The best tree is a trade-off between a high decrease of deviance and the smallest number of leaves. In the tree structures, leaves indicate class labels and branches represent conjunctions of features that lead to those class labels (Team Biomod, 2012).

Classification tree analysis is included in BIOMOD to be a good alternative to regression approaches. The tree length is controlled by a nested sequence of sub-trees by recursively cutting the less important splits in terms of explained deviance. BIOMOD selects the best trade-off between the number of leaves and the explained deviance through X-fold cross-validations (where X is the number of cross-validation that can be set by the user. Seemingly, there is no optimal number of cross-validation but trying and testing until find the best solution for the problem (Team Biomod, 2012).

d) Artificial Neural Network

Inspired by human nervous system, the artificial neural network (ANN) technique has demonstrated to be a very powerful tool to deal with multivariate time series analysis and huge amount of information (Zhang, 2018). The human brain holds hundreds of billions of interconnected neurons that perform parallel processing of information (Wang, 2003), so that ANN was originally thought to solve problems alike human brain would do. Nowadays, it has evolved greatly towards many other fields and applications such as computer vision, biology or medical diagnosis. One of the greatest contributors to this technique was the British statistician Brian Ripley (Ripley, 1996). An artificial neural network (ANN) is a flexible mathematical structure that is able to identify complex nonlinear relationships between input and output datasets (Hsu et al., 2005). It consists of one layer of neurons, nodes or units that stand as input, and one, two or three hidden layers of neurons. The output is a final layer of neurons (Fig. 52). The connexion between neurons of different layers are called weight.

It has been proved to be an efficient and useful technique to solve problems where the characteristics of the interactions are very unclear and difficult to describe. Results may vary with different runs, and the most optimal weight decay and the number of units in the hidden layer (with 3 as default) is chosen by means of N-fold cross-validation. The number of cross-validations can be selected by the user. Feed forward neural networks offer a flexible method to generalize linear regression functions. Even though they are non-linear regression models, the fact to have so many parameters makes them flexible enough to approximate any smooth function. ANN accuracy is merely handled by two parameters: the amount of weight decay and the number of hidden unit. NNET is the library used by BIOMOD to work with these types of algorithms (Team Biomod, 2012). As stated before, each run may give different results, so that N-fold cross-validation finds the best weight decay and the number of units in the hidden layer. In order to set this last parameter, there are different approaches such as (Wierenga & Kluytmans, 1994) where the number of units should be equal to the number of variables, or 75% of the number of variables (Venugopal & Baets, 1994). This is a time-consuming approach if the number of cross-validations is high. The output, h_i, of neuron i in the hidden layer can be seen in equation 10:

$$h_i = \sigma\left(\sum_{j=1}^{N} V_{ij}\, x_j + T_i^{hid}\right) \qquad \text{(Eq. 10)}$$

where σ () is called activation function, N corresponds to the number of input neurons, V_{ij} the weights, x_j inputs to the input neurons, and T_i^{hid} the threshold terms of the hidden neurons. The main purposes of the activation function are to introduce nonlinearity into the neural network, as well as to link the value of the neuron in the way that the neural network is not stopped by divergent neurons.

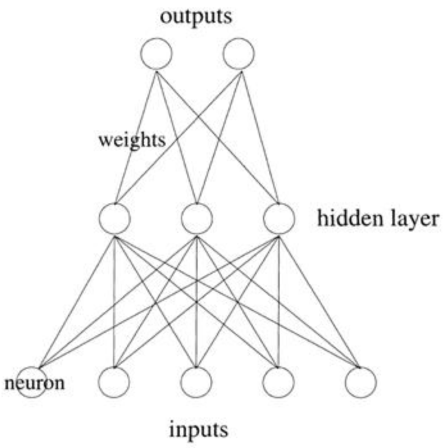


Figure 52. Architecture of a neural network. Source: Wang, 2003

The data provided to the input neurons are independent variables while the returned data from the output neurons are the response variables to the function being approximated by the neural network (Wang, 2003). Either inputs or outputs can be numeric, binary or even symbols if appropriately encoded, what enables neural networks to have a wide range of uses.

e) Surface Range Envelop or Bioclim “SRE”

This is a simple surface range envelop, similar to BioClim. It is an approach that only uses presence data to identify environmental conditions that best suits to the species under research (Busby, 1991). The “envelop” is defined between the maximum and minimum values of the environmental variables of every presence recorded in the data. Every location with all variables ranging within these maximum and minimum thresholds is included within the range. In order to avoid over-prediction due to outliers, the envelope can be shrunk at specified standard deviations or percentiles. This is one of the simplest methods to model the distribution of species and widely used in many works (Carpenter et al., 1993; Booth et al., 2014) that removes those presences which are close to be outside the envelop for being considered as outliers.This method does not provide probability of occurrence but directly the presence or absence of the species (Team Biomod, 2012). It is very simple and intuitive, with no need to provide absence data to the model that offers a ranking of the most important environmental variables; nevertheless it has some limitations too. SRE cannot use categorical variables, it is susceptible to over-predict, and it does not explain interaction between predictors or does not provide confidence levels (Araujo & Peterson, 2012).

f) Flexible Discriminant Analysis "FDA"

Flexible Discriminant Analysis (FDA) is a supervised classification model based on a mixture of linear regression models (Hastie, 1994). It uses optimal scoring to transform the dependent variable so that the data are in a better form for linear separation, and multiple adaptive regression splines to generate the discriminant surface. This method is an extension of the linear discriminant analysis. Biomod uses the library mda to implement this algorithm, and Multiple Adaptive Regression Splines MARS (see below) to improve model prediction (Team Biomod, 2012).

Multiple Adaptive Regression Splines "MARS"

Multiple Adaptive Regression Splines (MARS) is an implementation of methods to solve regression-type problems aiming to predict response variables from a set of independent variables (Friedman, 1991). It is non-parametric procedure that avoids any assumption about the nature of the relationships between independent and dependent variables. On the contrary, it builds the relationship from a set of coefficients that are driven by the regression data. MARS is especially proper to problems with high number of independent variables, performing better than other methods under similar circumstances of high dimensionality on the dataset or low order interaction effects between variables.

The major assumption in any linear process is that the coefficients are stable across all levels of the predictor variables and/or time. MARS is a very appropriate method to analyse data when the coefficients of the model have different optimal values across different levels of the explanatory variables. Essentially, it identifies and estimates a model whose coefficients differ based on the levels of the explanatory variables. A spline knot is a threshold value that pinpoints a change in the model coefficients, and it can be done automatically by the algorithm itself. Furthermore, complex nonlinear relationships can be set too. Biomod uses the mars function from the mda library programmed by Trevor Hastie and Robert Tibshirani. The MARS method automatically selects the necessary amount of smoothing for each independent variable and their order of relationship. However, it urges to determine the maximum level of interaction (Team Biomod, 2012). There are only two level of interactions implemented in Biomod to tune, with no further parameterisation to modify. If required, parameters may be changed at the private function but only recommended to experience users.

g) Maximum Entropy “Maxent.Phillips”

Maximum Entropy is a technique to study the problem of modelling species geographic distributions (Phillips et al., 2004). It is based on sequential-update algorithms that can deal with a very large number of predicting variables. Maxent is a general-purpose machine learning method with a simple and precise mathematical formulation, and it has a number of aspects that make it well-suited for species distribution modelling.

The core idea of the Maxent approach is to estimate a target probability distribution by finding the probability distribution of maximum entropy with a set of constraints. These constraints would indicate that the information about the target distribution is not complete. The available information is regarded as independent variables or features in the model

Some of the advantages of these algorithms are (Phillips et al., 2006): (1) it needs only presence data, in addition to environmental information. (2) It can use both continuous and categorical data, as well as the possibility to add interactions between different variables. (3) Efficient deterministic algorithms were developed to converge with the optimal (maximum entropy) probability distribution. (4) The Maxent probability distribution has a concise mathematical definition. More information about this approach can be found at (Phillips, 2017).

h) Low Memory Multinomial Logistic Regresion “Maxent.Tsuroka”

The Maxent.Tsuroka approach is data classification approach developed by Dr. Yoshimasa Tsuruoka that uses multinomial logistic regression, also known as maximum entropy (Jurka, 2012). The main aim of this classifier is to minimize memory consumption on very large datasets. Biomod uses the maxent package (Jurka & Tsuruoka, 2013) which provides a fast, low-memory maximum entropy classifier for a variety of classification tasks including ecology. Furthermore, there are some available hyper-parameters to prevent model overfitting and provide more accurate results.

The Biomod2 package allows the user to randomly subset the original dataset into 2 subsets with calibration-validation purposes. The 80% of the data was selected to calibrate the models and 20 % to validate the predictions. Then, we repeated the process 5 times to obtain a robust test of the models, where each replicate uses a unique random split 80% - 20% of the data (Thuiller et al., 2009). Presence and pseudo-absences were set to have the same importance in the calibration process, with a prevalence value of 0.5. Each parameter specification can be found at (https://www.rdocumentation.org/packages/biomod2/versions/3.3-7/topics/BIOMOD_Modeling).

BIOMOD2 package also offers the possibility to incorporate the best performing algorithms into ensemble models, which in many instances improve model prediction. These models combine the probabilities of some individual model predictions using their mean, coefficient of variation, median, confidence intervals, committee averaging or probability mean weight decay (Thuiller et al., 2016). In this chapter, this last technique (ensemble models) was selected to identify the potential distributions of hopper desert locust using ensemble species distribution models. To evaluate the performance of the model, 4 different metrics were taken into account: the Receiver Operating Curve *"ROC" (Hanley & McNeil, 1982), Cohen'S Kappa "KAPPA" (Monserud & Leemans, 1992), True Skill Statistic "TSS" (Allouche et al., 2006) and Accuracy.*

We chose six environmental variables from two different sensors to include in our model: MODIS (NDVI and LST) and SMAP (Soil Moisture Root Zone, Surface Soil Moisture, LAI and Surface Temperature). Soil temperature information was retrieved for both sensors in order to add complementary information since the time pass is different. In order to overcome the difficulties involved in locating desert locust breeding zones in arid environments, desert locust biology must be understood and the environmental predictors adapted accordingly.

We extracted the environmental variables from satellite imagery that correspond to presence and pseudo-absence sites. Each point is associated with 95 values that correspond to the evolution of each environmental variable back in time on a daily basis. As time resolution differs from satellite products, we created sub-variables based on a 16-day period to obtain continuous time series. The variable description is detailed in Table 10.

Variable	Explanation	Units
LST_1	Average value of Land Surface Temperature from Terra-MODIS between 95 and 81 days before survey date	ºC
LST_2	Idem as LST_1 between 80 and 65 days before the survey date	ºC
LST_3	Idem as LST_1 between 64 and 49 days before the survey date	ºC
LST_4	Idem as LST_1 between 48 and 33 days before the survey date	ºC
LST_5	Idem as LST_1 between 32 and 16 days before the survey date	ºC

Variable	Explanation	Units
LST_6	Idem as LST_1 between 15 days before the survey date and the survey date itself	ºC
NDVI_1	Average value of NDVI from Terra-MODIS between 95 and 81 days before survey date	-
NDVI_2	Idem as NDVI_1 between 80 and 65 days before the survey date	-
NDVI_3	Idem as NDVI_1 between 64 and 49 days before the survey date	-
NDVI_4	Idem as NDVI_1 between 48 and 33 days before the survey date	-
NDVI_5	Idem as NDVI_1 between 32 and 16 days before the survey date	-
NDVI_6	Idem as NDVI_1 between 15 days before the survey date and the survey date itself	-
LAI_1	Average value of Leaf Area Index from SMAP between 95 and 81 days before survey date	-
LAI_2	Idem as LAI_1 between 80 and 65 days before the survey date	-
LAI_3	Idem as LAI_1 between 64 and 49 days before the survey date	-
LAI_4	Idem as LAI_1 between 48 and 33 days before the survey date	-
LAI_5	Idem as LAI_1 between 32 and 16 days before the survey date	-
LAI_6	Idem as LAI_1 between 15 days before the survey date and the survey date itself	-
SMRZ_1	Average value of Soil Moisture Root Zone from SMAP between 95 and 81 days before survey date	m^3/m^3
SMRZ_2	Idem as SMRZ_1 between 80 and 65 days before the survey date	m^3/m^3
SMRZ_3	Idem as SMRZ_1 between 64 and 49 days before the survey date	m^3/m^3
SMRZ_4	Idem as SMRZ_1 between 48 and 33 days before the survey date	m^3/m^3
SMRZ_5	Idem as SMRZ_1 between 32 and 16 days before the survey date	m^3/m^3
SMRZ_6	Idem as SMRZ_1 between 15 days before the survey date and the survey date itself	m^3/m^3
SSM_1	Average value of Surface Soil Moisture from SMAP between 95 and 81 days before survey date	m^3/m^3
SSM_2	Idem as SSM_1 between 80 and 65 days before the survey date	m^3/m^3
SSM_3	Idem as SSM_1 between 64 and 49 days before the survey date	m^3/m^3
SSM_4	Idem as SSM_1 between 48 and 33 days before the survey date	m^3/m^3
SSM_5	Idem as SSM_1 between 32 and 16 days before the survey date	m^3/m^3

Variable	Explanation	Units
SSM_6	Idem as SSM_1 between 15 days before the survey date and the survey date itself	m^3/m^3
ST_1	Average value of Surface Temperature from SMAP between 95 and 81 days before survey date	ºC
ST_2	Idem as ST_1 between 80 and 65 days before the survey date	ºC
ST_3	Idem as ST_1 between 64 and 49 days before the survey date	ºC
ST_4	Idem as ST_1 between 48 and 33 days before the survey date	ºC
ST_5	Idem as ST_1 between 32 and 16 days before the survey date	ºC
ST_6	Idem as ST_1 between 15 days before the survey date and the survey date itself	ºC

Table 10. Environmental data used to derive hopper presence in our model. It contains the explanation of the variables, measurement units, mean values and their standard deviation.

For each point of presence or pseudo-absence, we extracted each of the variables explained in Table 10, aiming to obtain representative values for each time span before the sighting, so that we may observe some trends or patterns. NDVI stands for the longest temporal resolution, 16 days and then we established such length for the rest of the variables too. As mentioned by (Symmons & Cressman, 2001), hoppers may stay up to 95 days after the egg was laid before changing phase, under the longest case scenario.

Model fitting and prediction

Data split for calibration and testing was set to be random, with 80% of the original dataset to calibrate the model and 20% to evaluate it by ROC, TSS, Kappa and Accuracy metrics. This process to subset the original dataset into calibration and validation subsets was repeated randomly 5 times in order to obtain a robust test of the model (Thuiller et al., 2016). Presences and pseudo-absences were set to have the same weight in the model.

In this chapter, it was applied an ensemble technique excluding those individual models with TSS and kappa < 0.8 and ROC < 0.9. According to (*Thuiller et al., 2009) in* Table 11, those metric values would range from “fail” to “fair” model accuracies.

ACCURACY	ROC	KAPPA/TSS
Excellent or high	0.9 – 1	0.8 – 1
Good	0.8 – 0.9	0.6 – 0.8
Fair	0.7 – 0.8	0.4 – 0.6
Poor	0.6 - 0.7	0.2 – 0.4
Fail or null	0.5 – 0.6	0 – 0.2

Table 11. Index for classifying model prediction accuracy (Thuiller et al., 2009).

Model evaluation

The ensemble model algorithms in BIOMOD2 generate combined predictions based on different techniques: mean of probabilities (EMmean), confidence interval upper and lower (EMciSup and EMciInf respectively), being set 0.05 as the significance level for estimating it, committee averaging (EMca) or the weighted sum of probabilities (EMwmean). Those model predictions were evaluated against ROC, TSS, accuracy and Kappa metrics and then, we average the scores of the 5 runs to obtain a representative value of each model and metric. These results were compared against individual model performances. Biomod2 also provides information about the relative variable importance of each predictor to build the ensemble model, ranging from 0 to 1 (the higher the value, the higher the importance of the predictor). In addition, response curves offer the possibility to observe the sensitivity of the model for each predictor variable.

4.3. Results

4.3.1. Model evaluation

In overall, the best results have been obtained by the ensemble models in comparison with the individual algorithms of BIOMOD2 (Fig. 53-54).The Committee Averaging Ensemble Model (EMca) was the best approach to predict hopper desert locust with the highest metric scores (Kappa and TSS = 0.901, ROC =0.986). The proportion of presences that were correctly identified (sensitivity) was 95.18 %, while the true negative rate (specificity) was 94.96 %. The rest of the models have rather similar performance, although showed slightly lower scores (Table 12).

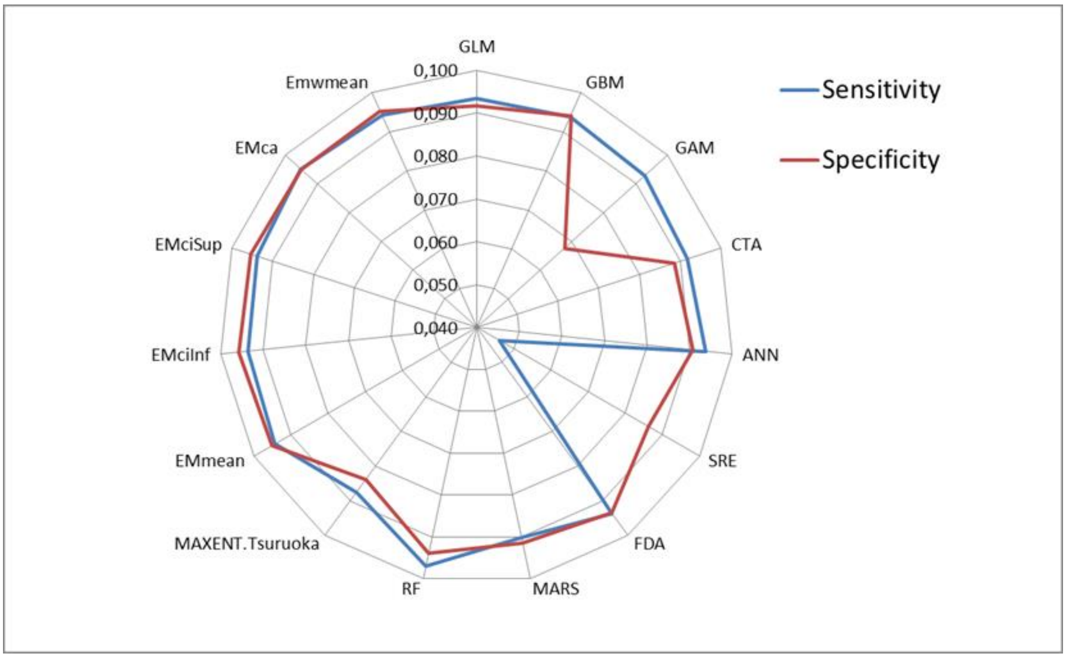


Figure 53. Comparison between individual algorithms and ensemble models in terms of sensitivity and specificity.

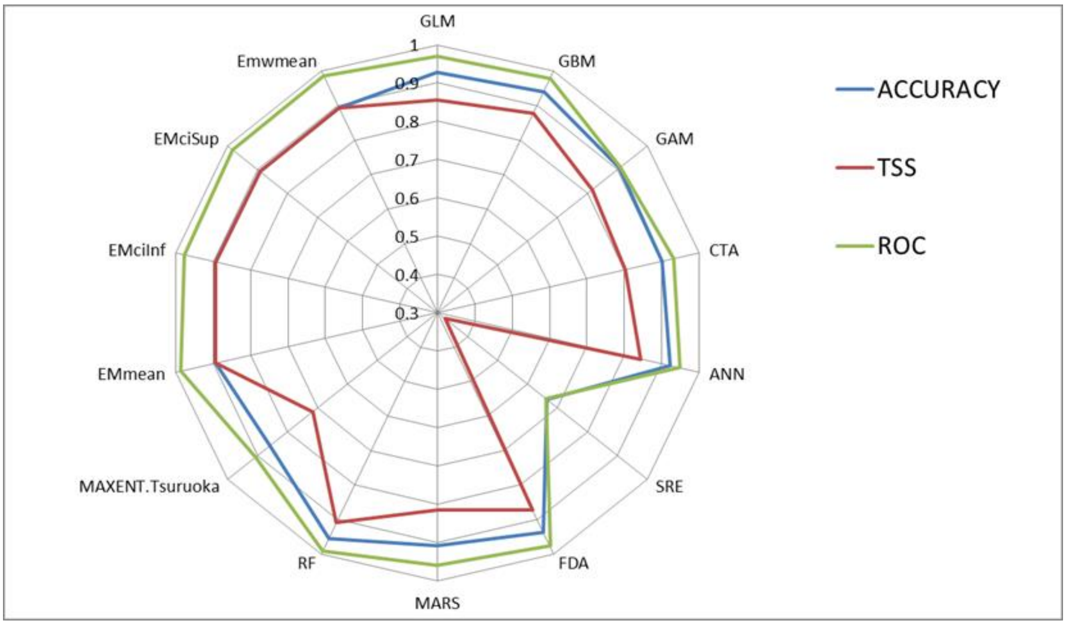


Figure 54. Comparison between individual algorithms and ensemble models in terms of Accuracy, TSS and ROC metrics.

	ACCURACY	TSS	ROC	KAPPA	Sensitivity %	Specificity %
GLM	0.928	0.856	0.969	0.855	93.525	91.748
GBM	0.939	0.878	0.978	0.878	93.669	94.126
GAM	0.908	0.816	0.911	0.816	92.986	67.657
CTA	0.901	0.802	0.932	0.801	91.654	88.532
ANN	0.922	0.845	0.948	0.844	93.669	90.769
SRE	0.665	0.325	0.662	0.327	46.187	86.294
FDA	0.937	0.874	0.977	0.874	93.813	93.566
MARS	0.908	0.817	0.961	0.817	90.216	91.469
RF	0.955	0.909	0.991	0.909	96.978	93.846
MAXENT.Tsuruoka	0.858	0.715	0.906	0.715	87.626	83.916
EMmean	0.895	0.895	0.986	0.947	94.240	95.198
EMciInf	0.895	0.895	0.978	0.948	93.695	95.723
EMciSup	0.892	0.892	0.982	0.946	93.810	95.323
EMca	0.951	0.901	0.986	0.901	95.180	94.965
Emwmean	0.895	0.895	0.986	0.947	94.188	95.260

Table 12. Predictive performance scores for the individual and ensemble models of hopper desert locust in Mauritania (2015-2017).

4.3.2. Variable importance

Results of the most influential environmental predictors for the EMca model are shown in Table 13 and Fig. 55. The soil temperature (from SMAP) between the previous 95 to 80 days and the previous 16 days of the hopper records, along with the NDVI values (from MODIS) obtained from the previous 16 days of the records are the most relevant environmental predictors in our model. Their normalized importance scores emphasize the association of each environmental variable with the probability of hopper presence.

Normalized Importance	Variable
0.154	ST_6
0.133	ST_1
0.106	NDVI_6
0.071	ST_4
0.054	NDVI_1
0.041	LST_6
0.038	SMRZ_3
0.035	SMRZ_6
Between 0.035 - 0.030	SMRZ_5, SMRZ_2, ST_5
Between 0.029 - 0.020	ST_3, SMRZ_4, SSM_2, SMRZ_1, LAI_4, SSM_1
Between 0.019 - 0.010	ST_2, NDVI_3, LST_5, SSM_5
Between 0.009 - 0.002	LAI_3, SSM_4, LST_3, NDVI_5, LST_1, SSM_6, LST_2, LAI_5, SSM_3, LST_4, LAI_2, LAI_6, LAI_1, NDVI_4, NDVI_2

Table 13. Ranking of the normalized variable importance for the 36 environmental predictors of the model. Variables sorted within the same range have been ordered by importance from left to right.

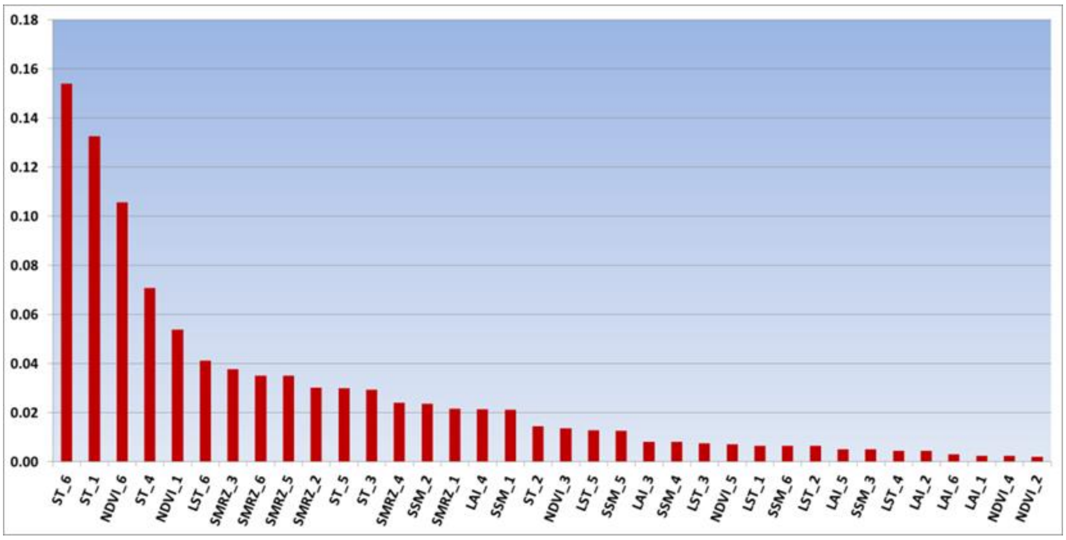


Figure 55. Normalized variable importance displayed graphically.

4.3.3. Response curves and living thresholds

Fig. 56 depicts the response curves for the eight most important variables in the model (Table 13). These response curves were calculated using the statistic "mean" to fix as constant the rest of variables when the predicted response is calculated for one of them (Thuiller et al., 2016). They show changes in the probability of solitarious hoppers across variable values. In other words, all the predictor variables are set to their mean value whereas the interested variable varies across its whole range of values, thus we assess the sensibility of the model to this variable not taking into account the relationship with other predictor variables (Elith et al., 2005).

Surface temperature curves indicate a great influence on hopper presence. The average value of surface temperature from SMAP between 95 and 81 days before the survey date (ST_1) and the average value of surface temperature from SMAP between 15 days before the survey date and the survey date itself (ST_6) differed greatly in SMAP acquisitions, and in general showed less probability of occurrence for lower temperatures compared to MODIS. Nevertheless, SMAP seemed to cover more critical temperatures during the day (minimum and close to the maximum throughout the day). ST_1 curve presented a probability increase from 25°C to 40°C, with a temperature optimum between 32.5 – 37.5°C. The NDVI_6 response showed that hopper desert locust were more likely to be found when NDVI ranges between 0.12 – 0.60 for the time covered up to 16 prior the sighting. LST_6 from MODIS showed an increase of probability when temperature was over 30°C. MODIS and SMAP surface temperature products vary their values due to their time pass, representing MODIS acquisitions the highest in general, and SMAP would indicate the average between close to the lowest at 6:00 am and still rather high temperature at 6:00 pm. SMRZ_3 and SMRZ_6 response curves increased in probability when SM values reached 0.21 m^3/ m^3.

The surface temperature retrieved by SMAP for the period (ST_6, ST_1, ST_4) and by MODIS (LST_6) proved to be highly influential with regard to increasing the probability of hoppers in our model. SM at root zone (SMRZ_3 and SMRZ_6) had less influence on the probability of presence, with a decrease in probability when values reach 0.125 m^3/ m^3 from 15 days before up to the date of occurrence.

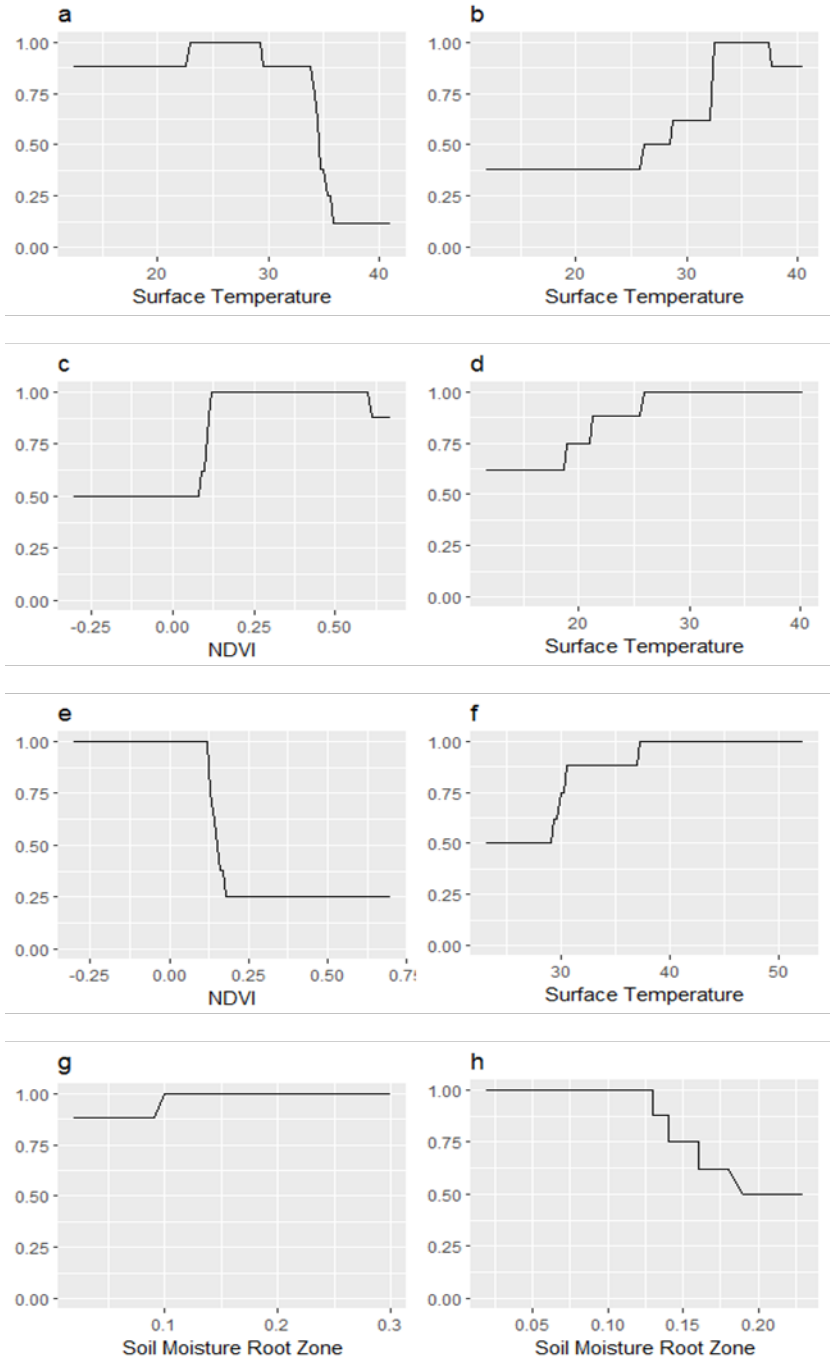


Figure 56. Response curves of the Ensemble Model committee averaging for the 8th most influential variables: (a) ST_6, (b) ST_1, (c) NDVI_6, (d) ST_4, (e) NDVI_1, (f) LST_6, (g) SMRZ_3, (h) SMRZ_6. X-axis represents variable values and Y-axis the probability of occurrence according to our committee average ensemble model. Variable units for Surface Temperature is degrees Celsius (ºC), Soil Moisture Root Zone is m^3/ m^3, NDVI and LAI are dimensionless.

Figure 57 depicts the comparison of surface temperature derived variables from SMAP. Some differences can be seen in terms of data distribution and interquartile ranges. For instance, ST_5 and ST_6 have lower median values in comparison with ST_1, ST_2 and ST_3, which present higher median temperatures and their interquartile range are narrower with higher temperature values. Furthermore, NDVI derived variables tend to increase their median values on presences from NDVI_1 to NDVI_6 as their interquartile range increases in terms of NDVI. This figure would explain how low NDVI values benefit desert locust at very early stages (egg phase) as seen in Fig. 56e for NDVI_1, whereas NDVI values need to increase so as to ensure the survival of new born hoppers as seen in NDVI_5 and NDVI_6 Fig. 56c. In both figures (Fig. 57-58), pseudo absence interquartile ranges remain equal across the time derived variables as it should be expected, demonstrating the role of pseudo absences to retrieve background values of the studied area.

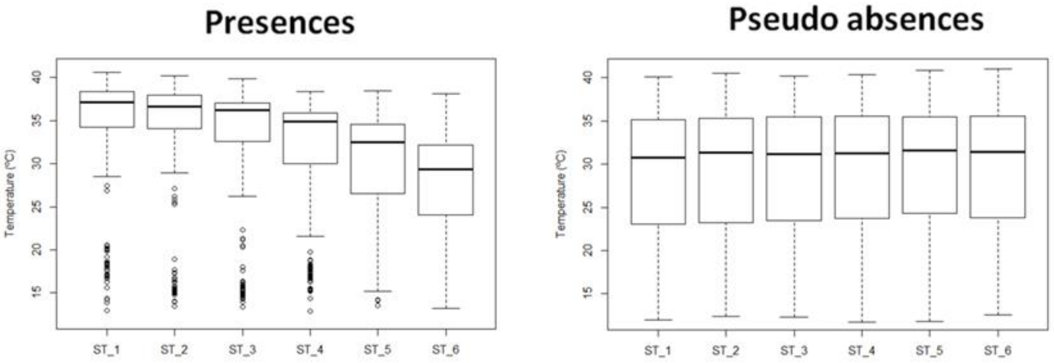


Figure 57. Distribution values of surface temperature for presence and pseudo absence records across the different time-based variables of temperature (see table 10).

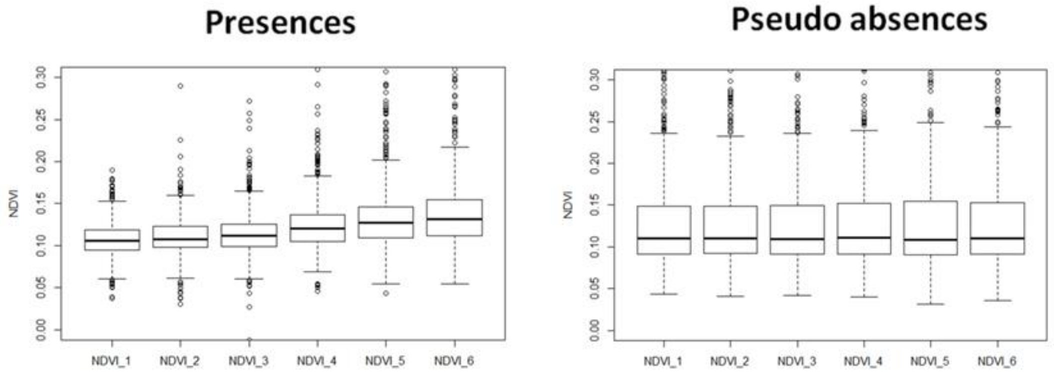


Figure 58. Distribution values of NDVI for presence and pseudo absence records across the different time-based variables of NDVI (see table 10).

4.4 Discussion

The aim of this chapter was to identify favourable environmental conditions for hoppers in the solitarious phase. Our approach is based on satellite remote sensing imagery and ground observations of hopper presence from 2015 to 2017, in Mauritania. We use Species Distribution Models (SDMs) based on machine learning techniques to predict presence/absence based on survey data to calibrate and validate our model.

In this chapter, it was studied the relationship between some environmental variables and hopper desert locust. NDVI, surface temperature, leaf area index, SM root zone and surface SM have been analysed for different time spans before the hopper sighting date. The target was to assess which are the most critical variables and time span to predict desert locust presences based on Artificial Intelligence (AI). Our results have proved the suitability of this methodology to apply and develop an early warning system to prevent and mitigate locust effects. In addition to the traditional environmental predictors such as precipitation or NDVI, it has been demonstrated the influence of land surface temperature on desert locust presence by means of two different products from the satellite sensors MODIS and SMAP.

The Committee Averaging Ensemble Model obtained a very good performance for KAPPA & TSS = 0.901, and ROC = 0.986 (Table 12). Although some studies accurately predict habitat suitability for desert locust based on individual variable analysis: rainfall (Dinku et al., 2010), vegetation (Ceccato, 2005, Lazar et al., 2015; Renier et al., 2015) or SM (Liu et al., 2008; Escorihuela et al., 2018), the present model improves the accuracy of previous approaches obtained through the goodness of SDM and machine learning algorithms (Elith & Leathwick, 2009).

We incorporate a new group of variables based on surface temperature from the SMAP satellite, and combine them with other factors to include complex interaction among the explanatory variables. To the best of our knowledge, no similar approaches currently exist for locating areas favourable to desert locust. Our ensemble approach obtained high model accuracy (Thuiller et al., 2009), being able to identify correctly 95.18 % of hopper presences. As variables were split by days, we distinguished different influences according to the survey time (Table 10). Variables with sub index 1 or 6 merely would account for strict egg and hopper stages respectively, whereas the rest of sub-indexes could not clearly be interpreted for any of those stages due to the time flexibility in locust development.

The most influential environmental variables were the surface temperature ST_6 and ST_1 retrieved by SMAP. This finding is consistent with previous studies that highlighted the importance of soil temperature conditions for egg viability and hatching phase (Hunter-Jones, 1970; Nishide et al., 2017), although this has not yet been applied to monitor suitable areas by remote sensing. Due to SMAP acquisition time, this sensor retrieves representative daily soil temperatures for our study zone. In spite of being the same physical variable, LST from Modis did not prove to be as relevant, with LST_6 standing as the 6th in importance (see Table 13). Then, time retrieval is a very important factor in arid or semi-arid areas owing to the high contrast day/night temperatures (Gunnigle et al., 2017) that may affect greatly the viability of eggs and new hatched hoppers (Hunter-Jones, 1970; Parker, 1930).

Field and laboratory experiments showed that heat controls the daily activities of some grasshopper species, with effects on food consumption, population density and egg production (Parker, 1930). In addition, (Nishide et al., 2017a) considered soil temperature as the leading factor for egg hatching activity because it occurs during low temperature hours, with minimum temperatures ranging from 20 and 24ºC (Hunter-Jones, 1964).

The findings of this study suggested that the optimal soil temperature range for egg development was between 32.5 and 37.4ºC for average temperatures of 16 days (Fig. 56). In accordance to that, (Hunter-Jones, 1964) laboratory experiments demonstrated a minimum temperature for embryonic development of 15.1ºC, nevertheless upper lethal limits showed to be dependent on time exposure, with constant temperatures over 38ºC (rare on the field). In the range 15.1 - 35ºC, the higher soil temperature, the faster egg development would be, whereas the egg development speed did not show to be only dependent on temperature over 35ºC. These statements coincide with our findings for ST_1, since prediction probability responds to temperature influence over 25ºC (which is the average of 16 days as explained in Table 10), proving egg viability under such circumstances. ST_6 showed an optimum in predicting hoppers between 23ºC and 29ºC, observing a drastic drop in occurrence probability over 34ºC. Then, hoppers would show more tolerance to values under 35ºC, whereas eggs seem to have a more constraint optimum range of soil temperatures. ST_4 response curve obtained its optimum for temperatures over 26ºC for the days covered (between 48 and 33 days before the survey date). In this variable, it is difficult to infer whether locust phase corresponds with egg or hopper stage because it varies according to environmental conditions of each record.

NDVI has been traditionally used to study habitat suitability for desert locust (Renier et al., 2015), since vegetation may provide food and shelter to hopper and adult individuals. In arid areas, vegetation conditions greatly the loss of soil water by evapotranspiration causing rapid SM depletion and, NDVI can be a good indicator along with other factors (Glenn et al., 2011). The most sensitive periods in which NDVI have any influence on our model are NDVI_1 (95 to 80 days before sighting) and NDVI_6 (16 up to the sighting day). As stated before, NDVI_1 would account for vegetation on the local pixel where the egg was laid. We could observe a drop in its response curve between 0.12 and 0.19 of NDVI, and this would explain unsuitable circumstances for egg stages. Whereas NDVI_6 response curve suggests favourable NDVI values from 0.12 to 0.60 during hopper stages. Our findings agree with previous studies that have stated similar minimum NDVI thresholds to discriminate breeding areas 0.13 (Despland et al., 2004) and 0.14 (Cherlet et al., 2017).

The SM in the root zone stood as the last of the 8th most important variables within our model. Previous works have related low SM content and large size grasshopper individuals with longer wings, while higher SM values would enhance mortality due to bacterial and fungous diseases (Parker et al., 1930).

Other studies pointed out the impact of SM on soil temperature to stop egg hatching when precipitation occurs (Nishide et al., 2017), the effects of prolonged flooding (>14 days) along with the increase of egg mortality (Woodman, 2015) or direct significance in egg lying, development and hatching (Liu et al., 2008). The SMRZ_3 response curve indicated a slight drop in probability from 0.12 to 0.19 m^3/m^3 of SM while to SMRZ_6, hopper's probability decreased from 0.13 to 0.19 m^3/m^3. So that, SM values over 0.19 m^3/m^3 indicate the lowest probability of occurrence for the time covered by SMRZ_3 and SMRZ_6 (see Fig. 56). These results are in accordance with prior publications to confirm SM influence, although they were less influent than the surface temperature. This fact can be explained by the effect that moisture has on surface temperature, enhancing the influence of this last one in the model. LAI or surface SM environmental variables do not seem to exert direct effect on desert locust distribution, although further improvements in spatial or temporal resolution of the remote sensing datasets might enhance their influence in the model.

These results concur with previous studies (Cherlet et al., 2000; Cherlet et al., 2017; Despland et al., 2004; Escorihuela et al., 2018; Lazar et al., 2016; Liu et al., 2008; Noy-Meir, 1973; Piou et al., 2017; Tratalos & Cheke, 2006) that individually survey certain environmental variables such as NDVI, SM or rainfall to identify habitat suitability for desert locust. Nevertheless, the use of pseudo-absences may bring some uncertainties into our model (Wisz & Guisan, 2009). They are often used in training data when there is a lack of information about the absence of species (Lobo et al., 2010). We therefore decided to incorporate them, given the lack of reliable and representative true absences in our study area. In order to build an effective SDM, pseudo-absences or background data need to be included together with presence data (Barbet-Massin et al., 2012), although there is still no consensus on how and where to sample these pseudo-absences nor how many are required.

Increased desert locust population is the result of an accumulation of several variables (Van Huis et al., 2007). By including new ecological variables such as root zone SM, LAI or surface temperature from more developed sensors (SMAP), our model obtains a very high predictive performance. As described in (Anderson et al., 2006), the machine learning algorithms embedded in SDMs are able to incorporate complex interactions among variables that traditional methods cannot, and they may outperform traditional methods. We obtain threshold values that feature the typical ecological niches for Mauritania.

In this study, we provide distribution constraints of the chosen environmental predictors using a machine learning species distribution model. These findings are the first step towards developing or improving operational early warning systems that may reduce survey and management operations. These tasks are essential vis-à-vis diminishing crop losses in areas already sensitive to food security issues (Ceccato et al., 2007) and where problems of social instability may emerge (Lecoq, 2003). Owing to the key role played by desert locust in agricultural production (Magor et al., 2008), it is essential to understand the environmental circumstances linked to locust damage. The methodology proposed in this chapter aims to improve or consolidate ongoing monitoring systems in order to keep major agricultural areas free of locusts. Early intervention has reduced the size of upsurges and plagues since the 1960s, helping to protect recent agricultural developments, crops and grazing of poor subsistence farmers in recession areas (Magor et al., 2008).

4.5 Conclusion

It has been verified the potentiality of Earth observation methods to identify potential habitats for solitarious desert locust in Mauritania. The methodology was based on machine learning algorithms for species distribution modelling with satellite remote sensing datasets. We obtained very high model accuracy with 0.901 Kappa and TSS metrics, 0.986 ROC and 0.951 accuracy for the Committee Averaging Ensemble Model. This chapter confirms the importance of previously exposed environmental variables such as NDVI or SM to detect desert locust presence, but note the greatly contribution of new variables such as Surface Temperature or Root Zone Soil Moisture from SMAP. We observed that for the purpose of this study, SMAP retrieves a more representative soil temperature than MODIS, and it may be as a consequence of the time acquisition. The most sensitive variables were the Surface Temperature and NDVI.

In addition, it was especially important to structure the environmental variables by time back from the sighting record, observing differences on model influence and curve responses within the same environmental variable. These results confirmed the hypothesis that several environmental conditions interfere in desert locust presence, and their combination may constrain its ecological niche. Even though data availability is limited, and temporal and spatial resolution of satellite images are still coarse, Sentinel 1 and Sentinel 2 satellites (ESA) may raise the robustness of the model. To further improve these results, true absence information as well as more complex algorithms may refine the results of this study. Future works will be oriented to develop and operational early warning system based on this methodology and results to prevent desert locust impacts.

Chapter 5. General discussion and conclusions

In this PhD research, various methodologies were explored to address the problematic to locate desert locust presences. The existing literature highlights the necessity to early identify locust breeding areas to avoid high population density (Gianessi, 2013), hence phase change from solitarious to gregarious and the formation of plagues (Skaf et al., 1990). This is a difficult task due to the dimensions and remoteness of the monitoring areas, in addition to the political instability and insecurity of those territories (FAO, 1994). An appropriate detection methodology facilitates control strategies, which aim to reduce populations to prevent plagues that damage crops and grazing (Van Huis et al., 2007).

Firstly, the role that wadis may play as breeding sites for desert locust was asessed. Secondly, we analyzed the link from ESA CCI SM remote sensing product with field surveys of hopper presences. And thirdly, we have narrowed the time span from 2015 to 2017 to use the latest satellite technology SMAP from NASA in order to monitor surface and root zone soil moisture.

It has been observed that potential drainage streams or wadis do not show special relationship with respect to solitarious hoppers in the time covered. The observed distance between records and potential wadis is very variable, suggesting the existence of more influential variables as pointed out by many other authors (Popov, 1958; Simpson et al., 1999; Despland & Simpson, 2000) to facilitate desert locust breeding. SWAT hydrological model was used to identify the runoff and wadis network in Mauritania.

Although the low precipitation regime in arid or semiarid environments, wadis may store great amount of groundwater due to its geo-stratigraphic features over long periods of time (Subyani, 2004). Some of the precipitation water percolates into the wadi soil to form local groundwater reservoirs, explaining the potential richness of vegetation and the relative higher SM content in comparison with no-wadi areas (Kassas & Imam, 1954). In general terms, this first approach could not offer significant results to link hopper records and wadi areas. The lack of river gauge data has been one the SWAT model limitations, which did not permit us to derive quantitative hydrological values of the study area. It should also be noted that locust records of SWARMS database are recorded from direct observation by locust survey teams, what implies that some areas cannot be monitored as frequently, or locusts may go unnoticed at low densities. Nevertheless, some results suggest a visual spatial correlation between hopper presences and wadi sites, what goes in accordance with studies such as (Tucker et al., 1985; Hielkema et al., 1986). Longer and more precise data records need to be taken to properly assess the validity of this approach. Visual analysis of desert locust presences suggests a geographical bias of the SWARMS database, which may be accounted for road networks and harsh conditions of the Sahara desert.

Despite ongoing desert locust monitoring techniques use rainfall to determine potential breeding sites (FAO & WMO, 2016), this variable presents some limitations in arid and semi-arid environments (Dinku et al., 2011). To improve management as well as forecasting techniques, the ESA CCI SM product and its suitability to locate breeding sites in Mauritania were analyzed. These results indicate a spatial correlation with traditional breeding areas according to the season of the year (Van Huis et al., 2007; Babah Ebbe, 2012). Despite ESA CCI SM only senses the first 5 cm. of the top soil, and desert locust lay eggs usually at depth down to 10 cm; this system seems appropriate due to the strong relationship of the top SM with deeper layers (Albergel et al. 2008). Our results provide a methodology using BIOMOD2 tool with highly predicting capacity (ROC-AUC = 0.95, TSS = 0.75 and Kappa = 0.75).

It was observed that the model performance improved when using narrower time intervals (6 days) and selecting the minimum SM value for that given time span. These findings suggest that an area becomes suitable for breeding when the minimum SM value remains higher than 0.07 m^3/m^3 over at least 6 days. Nevertheless, the area should be monitored for longer, since the success of egg-development is closely linked with the temporal evolution of SM as well as temperature (Shulov and Pener, 1963; Pedgley 1981). The applied methodology offers very promising results to correctly identify breeding areas based on 30 years of SM values. The ESA CCI SM dataset is the most complete and consistent global SM data record available (Wagner et al. 2012). To the best of our knowledge, there has not been any previous desert locust analysis using this SM dataset. Given the acknowledged importance of SM for desert locust and the length of ESA CCI SM dataset, our results may signify a breakthrough to complement the ongoing locust monitoring techniques used until today.

In the third chapter, we explore other variables derived from satellite remote sensing that seem to have influence on desert locust presence. Earth Science models tend to be stochastic (Christakos, 2012), with multiple iterations among variables that may enhance or diminish their impact over desert locust development in solitarious phase. For this reason, other environmental variables were included and assessed in the third chapter. Unfortunately, at the time this chapter was developed, the ESA CCI SM product was available only until 31/12/2015. Therefore, we could not use such SM product, but SMAP's soil moisture mission is expected to complement in the near future the ESA CCI SM initiative (Entekhabi et al., 2010, Dorigo et al., 2015).

We have evaluated the relationship between Surface and Root Zone Soil Mositure, LAI, NDVI and Land Surface Temperature variables with hopper desert locust. The target was to assess which are the most critical variables and time span to predict desert locust presences based on Artificial Intelligence. Our results indicate the suitability of this methodology to apply or develop an early warning system to prevent and mitigate locust effects. In addition to the traditional environmental predictors such as precipitation or NDVI, it was demonstrated the influence of land surface temperature on desert locust presence by means of MODIS and SMAP sensors. Note that the time interval was 16 days given the temporal resolution of the NDVI-MODIS product.

The Committee Averaging Ensemble Model obtained a very good performance for Kappa and TSS = 0.901, and ROC = 0.986 (Table 12). Even though there have been some studies predicting habitat suitability for desert locust based on individual variable analysis: rainfall (Dinku et al., 2010), vegetation (Ceccato, 2005, Lazar et al., 2015; Renier et al., 2015) or SM (Escorihuela et al. 2018), we aimed to develop a model combining different variables. Our ensemble approach obtained high model accuracy (Thuiller et al., 2009), being able to identify correctly 95.18 % of hopper presences. Variables were split by days, and we distinguished different influences according to the time interval (Table 10). Variables with sub index 1 or 6 merely would account for strict egg and hopper stages respectively, whereas the rest of sub-indexes could not clearly be interpreted for either stage due to time variation in locust development.

The most influential environmental variables were the surface temperature ST_6 and ST_1 retrieved by SMAP. This finding is consistent with previous studies that highlighted the importance of soil temperature conditions for egg viability and hatching phase (Hunter-Jones, 1970; Nishide et al., 2017), although have not yet been applied to monitor suitable areas by remote sensing. These findings suggest egg optimal range (soil temperature) between 32.5 and 37.4ºC for average temperatures of 16 days.

Traditionally, NDVI has been used to study habitat suitability for desert locust (Renier et al., 2015), since vegetation may provide food and shelter to hopper and adult individuals. In arid or semi-arid environments, vegetation influences the loss of soil water by evapotranspiration causing rapid SM depletion; thus NDVI can be a good proxy in combination with other factors (Glenn et al., 2011). NDVI_6 response curve suggests favourable NDVI values from 0.12 to 0.60 during hopper stages. Our findings agree with previous studies that have stated similar minimum NDVI thresholds to discriminate breeding areas 0.13 (Despland et al., 2004) and 0.14 (Cherlet et al., 2017).

Some studies have pointed out the impact of SM on soil temperature to stop egg hatching when precipitation occurs (Nishide et al., 2017), the effects of prolonged flooding (>14 days) along with the increase of egg mortality (Woodman, 2015) or direct significance in egg lying, development and hatching (Liu et al., 2008). In this model, SM variables seem to have less influence than others variables, although sensitive between 64 and 49 days, and 16 days prior the sighting record. Many publications relate brightness temperature and SM content (Rao et al., 1987), so that soil temperature and SM may be offering similar information to the model as observed in (Jin et al., 2014). Hence, temperature may have more predicting capabilities than SM in terms of predicting, but this statement needs further and future efforts to be confirmed.

This model provides an approach to incorporate new environmental variables such as Soil Moisture Root Zone, LAI or surface temperature from more developed satellite platforms to improve desert locust monitoring and prediction. The results obtained in this third chapter indicate a good performance to identify breeding sites for desert locust in solitarious phase.

In this PhD thesis, 3 different approaches were presented to locate breeding sites of desert locust. The results suggest that this study may help to improve ongoing operational early warning systems to detect desert locust breeding sites, or assist to create new predictive systems based on some of the presented methodologies. Future studies can expand these methods to other affected countries in order to implement a large-scale surveillance tool, using more data to improve the robustness of the presented models so that they may enhance their accuracy to locate potential breeding areas for desert locust.

Chapter 6. References

Abbaspour, C., Karim (2007), User manual for SWAT-CUP, SWAT calibration and uncertainty analysis programs, 93pp, Eawag: Swiss Fed. Inst. of Aquat. Sci. and Technol. Dubendorf, Switzerland. http://www.eawag.ch/organisation/abteilungen/siam/software/swat/index_EN (Accessed at 25/08/2016).

Abbaspour, K. C. (2013). SWAT-CUP 2012: SWAT calibration and uncertainty programs–a user manual. Eawag: Dübendorf, Switzerland, 103.

Abbaspour, K. C., Rouholahnejad, E., Vaghefi, S., Srinivasan, R., Yang, H., & Kløve, B. (2015). A continental-scale hydrology and water quality model for Europe: Calibration and uncertainty of a high-resolution large-scale SWAT model. Journal of Hydrology, 524, 733-752.

Adam, E. O., Elbasit, M. A., Solomon, T., & Ahmed, F. (2017). Integration of satellite rainfall data and curve number method for runoff estimation under semi-arid wadi system. International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences, 42.

Ahmad, S., Kalra, A., & Stephen, H. (2010). Estimating soil moisture using remote sensing data: A machine learning approach. Advances in Water Resources, 33(1), 69-80.

Ahmed, A. O. C., Nagasawa, R., Hattori, K., Chongo, D., & Perveen, M. F. (2007). Analytical hierarchic process in conjunction with GIS for identification of suitable sites for water harvesting in the oasis areas: Case study of the oasis zone of Adrar, northern Mauritania.

Akaike, H. (1973). Maximum likelihood identification of Gaussian autoregressive moving average models. Biometrika, 60(2), 255-265.

Al-Adamat, R. A., Foster, I. D., & Baban, S. M. (2003). Groundwater vulnerability and risk mapping for the Basaltic aquifer of the Azraq basin of Jordan using GIS, Remote sensing and DRASTIC. Applied Geography, 23(4), 303-324.

Albergel, C., Rüdiger, C., Pellarin, T., Calvet, J. C., Fritz, N., Froissard, F., ... & Martin, E. (2008). From near-surface to root-zone soil moisture using an exponential filter: an assessment of the method based on in-situ observations and model simulations. Hydrology and Earth System Sciences Discussions, 12, 1323-1337.

Ali, M. M., Dickinson, G., & Murphy, K. J. (2000). Predictors of plant diversity in a hyperarid desert wadi ecosystem. Journal of Arid Environments, 45(3), 215-230.

Allouche, O., Tsoar, A., & Kadmon, R. (2006). Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). Journal of applied ecology, 43(6), 1223-1232.

Anderson, R., Dudík, M., Ferrier, S., Guisan, A., J Hijmans, R., Huettmann, F & A Loiselle, B. (2006). Novel methods improve prediction of species' distributions from occurrence data. Ecography, 29(2), 129-151.

Anstey, M. L., Rogers, S. M., Ott, S. R., Burrows, M., & Simpson, S. J. (2009). Serotonin mediates behavioral gregarization underlying swarm formation in desert locusts. science, 323(5914), 627-630.

Araujo, M. B., & New, M. (2007). Ensemble forecasting of species distributions. Trends in ecology & evolution, 22(1), 42-47.

Araujo, M.B., Peterson AT (2012) Uses and misuses of bioclimatic envelope modeling. Ecology 93(7): 1527-1539.

Archer, K. J., & Kimes, R. V. (2008). Empirical characterization of random forest variable importance measures. Computational Statistics & Data Analysis, 52(4), 2249-2260.

Arino, O., Gross, D., Ranera, F., Leroy, M., Bicheron, P., Brockman, C., ... & Bourg, L. (2007). GlobCover: ESA service for global land cover from MERIS. In Geoscience and Remote Sensing Symposium, 2007. IGARSS 2007. IEEE International (pp. 2412-2415). IEEE

Arnold, J. G., Srinivasan, R., Muttiah, R. S., & Williams, J. R. (1998). Large area hydrologic modeling and assessment part I: model development 1. JAWRA Journal of the American Water Resources Association, 34(1), 73-89.

Arnold, J. G., Moriasi, D. N., Gassman, P. W., Abbaspour, K. C., White, M. J., Srinivasan, R., ... & Kannan, N. (2012). SWAT: Model use, calibration, and validation. Transactions of the ASABE, 55(4), 1491-1508.

Atkinson, P. M. (2008). Super-resolution mapping using the two-point histogram and multi-source imagery. In geoENV VI–Geostatistics for Environmental Applications (pp. 307-321). Springer, Dordrecht.

Austin, M.P. (1987). Models for the analysis of species response to environmental gradients. Vegetation 69, 35/45.

Babah Ebbe, M. A. O. (2003). Biogéographie du criquet pèlerin en Mauritanie: Fonctionnement d'une aire grégarigène et conséquences sur l'organisation de la surveillance et de la lutte anti-acridienne (No. AGP/DL/TS/31), Stations de recherche acridienne sur le terrain, séries techniques. FAO, Rome.

Babah Ebbe, M. A. O. (2010). Biogéographie du Criquet pèlerin en Mauritanie. Hermann, Paris, 1-286.

Babah Ebbe, M. A. O. (2012). Preventative control for desert locust pest in africa: experiences of mauritania. https://www.jircas.go.jp/sites/default/files/publication/proceedings/2012-session-41_0.pdf (Accesed at 21/11/2017).

Bacar Javar, M. E. H. (2011). Contribution à l'étude descriptive et causale de la chorologie du Criquet pèlerin (# Schistocerca gregaria# Forskål, 1775) en Mauritanie. http://agritrop.cirad.fr/572004/ (Accessed at 18/06/2017).

Barbet-Massin, M., Jiguet, F., Albert, C. H., & Thuiller, W. (2012). Selecting pseudo-absences for species distribution models: how, where and how many?. Methods in Ecology and Evolution, 3(2), 327-338.

Barnes, M. A., Jerde, C. L., Wittmann, M. E., Chadderton, W. L., Ding, J., Zhang, J., ... & Lodge, D. M. (2014). Geographic selection bias of occurrence data influences transferability of invasive Hydrilla verticillata distribution models. Ecology and evolution, 4(12), 2584-2593.

Barrett, E. C., & Hamilton, M. G. (1986). Potentialities and problems of satellite remote sensing with special reference to arid and semiarid regions. Climatic change, 9(1-2), 167-186.

Beaumont, L.J., Gallagher, R.V., Thuiller, W., Downey, P.O., Leishman, M.R. & Hughes, L. (2009) Different climaticenvelopes among invasive populations may lead to under-estimations of current and future biological invasions.Diversity and Distributions, 15, 409–420.

Benito Garzón, M., Alía, R., Robson, T. M., & Zavala, M. A. (2011). Intra-specific variability and plasticity influence potential tree species distributions under climate change. Global Ecology and Biogeography, 20(5), 766-778.

Bengtsson, L., Hagemann, S., & Hodges, K. I. (2004). Can climate trends be calculated from reanalysis data?. Journal of Geophysical Research: Atmospheres, 109(D11).

Bennett, L. V. (1976). The development and termination of the 1968 plague of the Desert locust, Schistocerca gregaria (Forskål)(Orthoptera, Acrididae). Bulletin of Entomological Research, 66(3), 511-552.

Beuhler, M. (2003) Potential Impacts of Global Warming on water resources in Southern California. Water Sci Technol 47: 165-168

Bicheron, P., Amberg, V., Bourg, L., Petit, D., Huc, M., Miras, B., ... & Leroy, M. (2008). Geolocation assessment of 300 m resolution MERIS Globcover ortho-rectified products. In Proceedings of the‘ 2nd MERIS/(A) ATSR User Workshop’, Frascati, Italy, 22– 26 September 2008 (ESA SP-666, November 2008).

Biomod team, Thuiller, W., Georges, D., Engler, R., & Lafourcade, B. (2012). BIOMOD: Tutorial.

Bolten, J. D., Brown, M., & Ceccato, P. N. (2009). Improving Desert Locust Decision Support in Africa and Asia using SMAP Soil Moisture Estimates.

Booth, T. H., Nix, H. A., Busby, J. R., & Hutchinson, M. F. (2014). BIOCLIM: the first species distribution modelling package, its early applications and relevance to most current MAXENT studies. Diversity & Distributions, 20(1),1-9.

Bouaichi, A., Simpson, S. J., & Roessingh, P. (1996). The influence of environmental microstructure on the behavioural phase state and distribution of the desert locust Schistocerca gregaria. Physiological Entomology, 21(4), 247-256.

Brader, L., Djibo, H., Faye, F. G., Ghaout, S., Lazar, M., Luzietoso, P. N., & Babah, M. O. (2006). Towards a more effective response to desert locusts and their impacts on food security, livelihoods and poverty. Multilateral evaluation of the 2003–05 Desert locust campaign. Food and Agriculture Organisation, Rome.

Bradley, D. C., Motts, H., Horton, J. D., Giles, S. A., & Taylor, C. D. (2015). Geologic map of Mauritania (phase V, deliverables 51a, 51b, and 51c): Chapter A1 in Second projet de renforcement institutionnel du secteur minier de la République Islamique de Mauritanie (PRISM-II) (No. 2013-1280-A1). US Geological Survey.

Breiman, L., Friedman, J., Stone, C. J., & Olshen, R. A. (1984). Classification and regression trees. CRC press.

Breiman, L. (2001). Random forests. Mach. Learn. 45: 5–32.

Brocca, L., Moramarco, T., Melone, F., & Wagner, W. (2013). A new method for rainfall estimation through soil moisture observations. Geophysical Research Letters, 40(5), 853-858.

Browning, K. A., Boulahya, M. S., Pedgley, D. E., Lyne, W. H., Rijks, D., & Symmons, P. M. (1990). Algerian Case Study and the Need for Permanent Desert Locust Monitoring: Discussion. Philosophical Transactions of the Royal Society of London Series B, 328, 581-583.

Brownlee, J. (2014). Machine learning mastery. URL: http://machinelearningmastery. com/discover-feature-engineering-howtoengineer-features-and-how-to-getgood-at-it. (Accessed at 07/03/2018).

Brownlee, J. (2017). Overfitting and underfitting with machine learning algorithms. https://machinelearningmastery.com/overfitting-and-underfitting-with-machine-learning-algorithms/ (Accessed at 17/09/2018).

Buerki, S., Callmander, M. W., Bachman, S., Moat, J., Labat, J. N., & Forest, F. (2015). Incorporating evolutionary history into conservation planning in biodiversity hotspots. Phil. Trans. R. Soc. B, 370(1662), 20140014.

Busby, J. R. (1986). A biogeoclimatic analysis of Nothofagus cunninghamii (Hook.) Oerst. in southeastern Australia. *Austral Ecology, 11*(1), 1-7.

Busby, J. R. (1991). BIOCLIM – a bioclimate analysis and prediction system. – In: Margules, C. R. and Austin, M. P. (eds), Nature conservation: cost effective biological surveys and data analysis. CSIRO, pp. 64–68.

Campbell, J. B., & Wynne, R. H. (2011). Introduction to remote sensing. Guilford Press.

Camps-Valls, G. (2009). Machine learning in remote sensing data processing. In Machine Learning for Signal Processing, 2009. MLSP 2009. IEEE International Workshop on (pp. 1-6). IEEE.

Carlson, T. N. & Ripley D. A. (1997). On the relation between NDVI, fractional cover, and leaf area index. Remote Sensing of Environment 62: 241-252.

Carpenter, G., Gillison, A. N., & Winter, J. (1993). DOMAIN: a flexible modelling procedure for mapping potential distributions of plants and animals. Biodiversity & Conservation, 2(6), 667-680.

Ceccato, P. N. (2005). Operational early warning system using spot-vegetation and terra-modis to predict desert locust outbreaks.

Ceccato, P., Cressman, K., Giannini, A., & Trzaska, S. (2007). The desert locust upsurge in West Africa (2003–2005): Information on the desert locust early warning system and the prospects for seasonal climate forecasting. International Journal of Pest Management, 53(1), 7-13.

Chappell, M. A. (1983). Metabolism and thermoregulation in desert and montane grasshoppers. Oecologia, 56(1), 126-131.

Cherlet, M., Mathoux, P., Bartholomé, E., & Defourny, P. (2000). Spot vegetation contribution to desert locust habitat monitoring. In Proceedings of the VEGETATION 2000 Workshop, Lake Maggiore, Italy (pp. 3-6).

Cherlet, M., Mathoux, P., Bartholomé, E., & Defourny, P. (2017). Spot vegetation contribution to desert locust habitat monitoring. http://www.vgt.vito.be/pages/vgtprep/vgt2000/cherlet.pdf (Accessed at 06/1/2018).

Christakos, G. (2012). Random field models in earth sciences. Courier Corporation.

Chow V.T., Maidment, D.R., Mays, L.W. (1988). Applied Hydrology. McGraw-Hill, Inc., New York, NY. http://agris.fao.org/agris-search/search.do?recordID=US201300485114 (Accessed at 12/10/2016).

Cisse, S., Ghaout, S., Mazih, A., Ebbe, B., Ould, M. A., Benahi, A. S., & Piou, C. (2013). Effect of vegetation on density thresholds of adult desert locust gregarization from survey data in Mauritania. Entomologia Experimentalis Et Applicata, 149(2), 159-165.

Collaboration for Australian Weather and Climate Research "CAWCR" (2015). http://www.cawcr.gov.au/projects/verification/#Methods_for_dichotomous_forecasts (accesed 14 November 2017)

Cressman, K. (1999). Monitoring desert locusts in the Middle East: An overview. Transformations of Middle Eastern Natural Environnements: Legacies ans Lessons; Coppock, J., Miller, JA, Albert, J., Bernhardsson, M., Kenna, R., Eds, 492.

Cressman, K. (2008). The use of new technologies in Desert Locust early warning. Outlooks on Pest Management, 19(2), 55-59.

Cressman, K., (2013). Role of remote sensing in desert locust early warning. Journal of Applied Remote Sensing, 7(1), 075098-075098.

Criminisi, A., Shotton, J., & Konukoglu, E. (2012). Decision forests: A unified framework for classification, regression, density estimation, manifold learning and semi-supervised learning. Foundations and Trends® in Computer Graphics and Vision, 7(2–3), 81-227.

Cullen, D. A., Cease, A. J., Latchininsky, A. V., Ayali, A., Berry, K., Buhl, J., ... & Ott, S. R. (2017). From molecules to management: mechanisms and consequences of locust phase polyphenism. In Advances in Insect Physiology (Vol. 53, pp. 167-285). Academic Press.

Culmsee, H. (2002). The habitat functions of vegetation in relation to the behaviour of the desert locust Schistocerca gregaria (Forskål)(Acrididae: Orthoptera)-a study in Mauritania (West Africa). Phytocoenologia, 32(4), 645-664.

Dark, S.J. (2004). The biogeography of invasive alien plantsin California: an application of GIS and spatial regressionanalysis. Diversity and Distributions, 10,1–9.

Defourny, P., Vancutsem, C., Bicheron, P., Brockmann, C., Nino, F., Schouten, L., & Leroy, M. (2006). GLOBCOVER: a 300 m global land cover product for 2005 using Envisat MERIS time series. In Proceedings of the ISPRS Commission VII mid-term symposium, Remote sensing: from pixels to processes (pp. 8-11). Enschede, the Netherlands.

Despland, E., & Simpson, S. J. (2000). Small-scale vegetation patterns in the parental environment influence the phase state of hatchlings of the desert locust. Physiological Entomology, 25(1), 74-81.

Despland, E., Collett, M., & Simpson, S. J. (2000). Small-scale processes in desert locust swarm formation: how vegetation patterns influence gregarization. Oikos, 88(3), 652-662.

Despland, E., Rosenberg, J., & Simpson, S. J. (2004). Landscape structure and locust swarming: a satellite's eye view. Ecography, 27(3), 381-391.

Dile, Y. T., & Srinivasan, R. (2014). Evaluation of CFSR climate data for hydrologic prediction in data-scarce watersheds: an application in the Blue Nile River Basin. JAWRA Journal of the American Water Resources Association, 50(5), 1226-1241.

Dingle, H. (2009). Migration. In Encyclopedia of Insects (Second Edition) (pp. 628-633).

Dingman, S. L. (2002). Water in soils: infiltration and redistribution. Physical hydrology.

Dinku, T., Ceccato, P., Cressman, K., & Connor, S. J. (2010). Evaluating detection skills of satellite rainfall estimates over desert locust recession regions. Journal of Applied Meteorology and Climatology, 49(6), 1322-1332.

Dinku, T., Ceccato, P., & Connor, S. J. (2011). Challenges of satellite rainfall estimation over mountainous and arid parts of east Africa. International journal of remote sensing, 32(21), 5965-5979.

Dorigo, W. A., Gruber, A., De Jeu, R. A. M., Wagner, W., Stacke, T., Loew, A., ... & Kidd, R. (2015). Evaluation of the ESA CCI soil moisture product using ground-based observations. Remote Sensing of Environment, 162, 380-395.

Dorigo, W.A., Wagner, W., Albergel, C., Albrecht, F., Balsamo, G., Brocca, L., Chung, D., Ertl, M., Forkel, M., Gruber, A., Haas, E., Hamer, D. P. Hirschi, M., Ikonen, J., De Jeu, R. Kidd, R. Lahoz, W., Liu, Y.Y., Miralles, D., Lecomte, P. (2017). ESA CCI Soil Moisture for improved Earth system understanding: State-of-the art and future directions. Remote Sensing of Environment, 2017, ISSN 0034-4257, https://doi.org/10.1016/j.rse.2017.07.001 (Accesed 14 November 2017).

Dubuis, A., Pottier, J., Rion, V., Pellissier, L., Theurillat, J.P.& Guisan, A. (2011) Predicting spatial patterns of plantspecies richness: a comparison of direct macroecologicaland species stacking modelling approaches. Diversity andDistributions, 17, 1122–1131.

Duranton, J.F.; Lecoq, M., (1990). Le Criquet Pèlerin au Sahel; Comité permanent inter-etats de lutte contre la sécheresse au Sahel: Ouagadougou, Brukina Faso.

Earth Observatory, NASA. https://earthobservatory.nasa.gov/IOTD/view.php?id=2799 (Accesed at 29 March 2018)

Edmunds, W. M. (2002). Wadi hydrology applications of geochemical and isotopic methods: a case study of wadi hawad, sudan. Hydrology of wadi systems, 23.

Elith, J., Ferrier, S., Huettmann, F., & Leathwick, J. (2005). The evaluation strip: a new and robust method for plotting predicted responses from species distribution models. Ecological modelling, 186(3), 280-289.

Elith, J., Graham, C. H., Anderson, R. P., Dudík, M., Ferrier, S., Guisan, A., ... & Li, J. (2006). Novel methods improve prediction of species' distributions from occurrence data. Ecography, 129-151.

Elith, J. & Leathwick, J. (2007) Predicting species distribu-tions from museum and herbarium records using multire-sponse models fitted with multivariate adaptive regressionsplines. Diversity and Distributions, 13, 265–275.

Elith J, Leathwick JR, Hastie T (2008) A working guide to boosted regression trees. Journal of Animal Ecology, 77(4): 802-813.

Elith, J., & Leathwick, J. R. (2009). Species distribution models: ecological explanation and prediction across space and time. Annual review of ecology, evolution, and systematics, 40, 677-697.

Elith, J., & Graham, C. H. (2009). Do they? How do they? WHY do they differ? On finding reasons for differing performances of species distribution models. Ecography, 32(1), 66-77.

Elith, J., Phillips, S.J., Hastie, T., Dudik, M., Chee, Y.E. &Yates, C.J. (2011) A statistical explanation of MaxEnt forecologists. Diversity and Distributions, 17,43–57.

Ellis, P. E. (1962). The behaviour of locusts in relation to phases and species. In Colloq. int. Cent. nat. Rech. sci. (Vol. 114, pp. 123-143).

Engler, R., Guisan, A., & Rechsteiner, L. (2004). An improved approach for predicting the distribution of rare and endangered species from occurrence and pseudo-absence data. Journal of applied ecology, 41(2), 263-274.

Entekhabi, D., Njoku, E. G., O'Neill, P. E., Kellogg, K. H., Crow, W. T., Edelstein, W. N., ... & Kimball, J. (2010). The soil moisture active passive (SMAP) mission. Proceedings of the IEEE, 98(5), 704-716.

Ernst, U. R., Van Hiel, M. B., Depuydt, G., Boerjan, B., De Loof, A., & Schoofs, L. (2015). Epigenetics and locust life phase transitions. Journal of Experimental Biology, 218(1), 88-99.

Escorihuela, M. J., Merlin, O., Stefan, V., Moyano, G., Eweys, O. A., Zribi, M., ... & Chihrane, J. (2018). SMOS based high resolution soil moisture estimates for Desert locust preventive management. Remote Sensing Applications. Society and Environment, 11, 140-150.

FAO - Agriculture Organization of the United Nations (1994). Desert Locust Guidelines (five volumes). Rome: FAO.

FAO, 2004. http://www.fao.org/ag/locusts/oldsite/LOCFAQ.htm (Accessed at 15/10/2018)

FAO - Agriculture Organization of the United Nations (2009). http://www.fao.org/ag/locusts/en/archives/2331/index.html (Accessed at 20/05/2018).

FAO/IIASA/ISRIC/ISSCAS/JRC (2012). Harmonized World Soil Database (version 1.2). FAO, Rome, Italy and IIASA, Laxenburg, Austria.

FAO & WMO; Agriculture Organization of the United Nations, (2016). Weather and Desert Locusts. http://www.fao.org/ag/locusts/common/ecg/2350/en/2016_WMOFAO_WeatherDLe.pdf (Accessed at 19/04/2018).

Fawcett, T. (2006). An introduction to ROC analysis. Pattern recognition letters, 27(8), 861-874.

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(3), 490-507.

Fielding, A. H., & Bell, J. F. (1997). A review of methods for the assessment of prediction errors in conservation presence/absence models. Environmental conservation, 24(1), 38-49.

Franklin, J. (2010) Moving beyond static species distribu-tion models in support of conservation biogeography.Diversity and Distributions, 16, 321–330.

Franklin, J. (2013). Species distribution models in conservation biogeography: developments and challenges. *Diversity and Distributions*, *19*(10), 1217-1223.

Friedman, J. (1991). Multivariate adaptive regression splines. Ann. Stat. 19: 1–141.

Fuka, D. R., Walter, M. T., MacAlister, C., Degaetano, A. T., Steenhuis, T. S., & Easton, Z. M. (2014). Using the Climate Forecast System Reanalysis as weather input data for watershed models. Hydrological Processes, 28(22), 5613-5623.

Fukuda, S., & Hiramatsu, K. (2008). Prediction ability and sensitivity of artificial intelligence-based habitat preference models for predicting spatial distribution of Japanese medaka (Oryzias latipes). Ecological Modelling, 215(4), 301-313.

Gao, X., Huete, A. R., Ni, W., & Miura, T. (2000). Optical–biophysical relationships of vegetation spectra without background contamination. Remote Sensing of Environment, 74(3), 609-620.

Gay, P. E., Lecoq, M., & Piou, C. (2018). Improving preventive locust management: insights from a multi-agent model. Pest management science, 74(1), 46-58.

Genuer, R., Poggi, J. M., & Tuleau-Malot, C. (2010). Variable selection using random forests. Pattern Recognition Letters, 31(14), 2225-2236.

Gerteiny, A., Deschamps, H., Stewart, C. & Toupet, C. (2018). Mauritania https://www.britannica.com/place/Mauritania (Access Date: 09/03/2018)

Gianessi, L. (2013). Desert Locust Plagues Managed with Insecticides. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.659.844&rep=rep1&type=pdf (Accessed at 30/08/2018).

Glenn, E. P., Neale, C. M., Hunsaker, D. J., & Nagler, P. L. (2011). Vegetation index-based crop coefficients to estimate evapotranspiration by remote sensing in agricultural and natural ecosystems. Hydrological Processes, 25(26), 4050-4062.

Graham, C.H., J. Elith, R.J. Hijmans, A. Guisan, A.T. Peterson, B.A. Loiselle and the NCEAS Predicting Species Distributions Working Group, (2007). The influence of spatial errors in species occurrence data used in distribution models. Journal of Applied Ecology 45: 239-247

Greathead, D. J. (1966). A brief survey of the effects of biotic factors on populations of the desert locust. Journal of Applied Ecology, 239-250.

Green, W. H., & Ampt, G. A. (1911). Studies on Soil Phyics. The Journal of Agricultural Science, 4(1), 1-24.

Gruber, A., Dorigo, W. A., Crow, W., Wagner W. (2017). Triple Collocation-Based Merging of Satellite Soil Moisture Retrievals. IEEE Transactions on Geoscience and Remote Sensing. PP. 1-13. 10.1109/TGRS.2017.2734070.

Guerrak, S., 1989. Time and space distribution of Palaeozoic oolitic ironstones in the Tindouf Basin, Algerian Sahara. Geological Society, London, Special Publications, 46(1), 197-212.

Guisan, A., & Zimmermann, N. E. (2000). Predictive habitat distribution models in ecology. Ecological modelling, 135(2-3), 147-186.

Guisan, A., Edwards Jr, T. C., & Hastie, T. (2002). Generalized linear and generalized additive models in studies of species distributions: setting the scene. Ecological modelling, 157(2-3), 89-100.

Guisan, A., Graham, C.H., Elith, J. & Huettmann, F. & theNCEAS Species Distribution Modelling Group (2007). Sen-sitivity of predictive species distribution models to changein grain size. Diversity and Distributions, 13, 332–340.

Guo, X., Ma, Z., & Kang, L. (2013). Serotonin enhances solitariness in phase transition of the migratory locust. Frontiers in behavioral neuroscience, 7, 129.

Haddeland, I. (2014). Global water resources affected by human interventions and climate change. PNAS vol. 111 no. 9 3251–3256

Hanley, J.A. & McNeil, B.J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology, 143, 29–36.

Hansen, M., Roca-Sales, M., Keegan, J. M., & King, G. (2017). Artificial Intelligence: Practice and Implications for Journalism.

Haskell, P. T., Paskin, M. W. J., & Moorhouse, J. E. (1962). Laboratory observations on factors affecting the movements of hoppers of the desert locust. Journal of Insect Physiology, 8(1), 53-78.

Hastie, T. J. and Tibshirani, R. (1990). Generalized additive models. Chapman and Hall.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). Unsupervised learning. In The elements of statistical learning (pp. 485-585). Springer, New York, NY.

Hastie, T., & Fithian, W. (2013). Inference from presence-only data; the ongoing controversy. Ecography, 36(8), 864-867.

He, W. P., & Zhao, S. S. (2018). Assessment of the quality of NCEP-2 and CFSR reanalysis daily temperature in China based on long-range correlation. Climate Dynamics, 50(1-2), 493-505.

Hielkema, J. U., Roffey, J., & Tucker, C. J. (1986). Assessment of ecological conditions associated with the 1980/81 desert locust plague upsurge in West Africa using environmental satellite data. International Journal of Remote Sensing, 7(11), 1609-1622.

Hijmans, R.J., 2012. Cross-validation of species distribution models: removing spatial sorting bias and calibration with a null-model. Ecology 93: 679-688.

Hijmans, R. J. and Elith, J. (2016). Species distribution modelling. http://rspatial.org/sdm/index.html (Access Date: 24/03/2018)

Hof, A.R., Jansson, R. & Nilsson, C. (2012). How biotic inter-actions may alter future predictions of species distribu-tions: future threats to the persistence of the arctic fox inFennoscandia. Diversity and Distributions, 18, 554– 562.

Hoffer, J. A., & Severance, D. G. (1975). The use of cluster analysis in physical data base design. In *Proceedings of the 1st International Conference on Very Large Data Bases* (pp. 69-86). ACM.

Hortal J, Jimenez-Valverde A, Gomez JF, Lobo JM, Baselga A (2008). Historical bias in biodiversity inventories affects the observed environmental niche of the species. Oikos 117: 847–858.

Hsu, K. L., Gupta, H. V., & Sorooshian, S. (1995). Artificial neural network modeling of the rainfall-runoff process. Water resources research, 31(10), 2517-2530.

Huang, X., & Jensen, J. R. (1997). A machine-learning approach to automated knowledge-base building for remote sensing image analysis with GIS data. Photogrammetric engineering and remote sensing, 63(10), 1185-1193.

Huang, G. B., Guo, Q. Y., Zhang, R. Z., Pang, L., Li, G., Chan, K. Y., & Yu, A. Z. (2006). Effects of conservation tillage on soil moisture and crop yield in a phased rotation system with spring wheat and field pea in dryland. Acta Ecologica Sinica, 4, 026.

Huang, J., Minnis, P., Yan, H., Yi, Y., Chen, B., Zhang, L., & Ayers, J. K. (2010). Dust aerosol effect on semi-arid climate over Northwest China detected from A-Train satellite measurements. Atmospheric Chemistry and Physics, 10(14), 6863-6872.

Huete, A., Justice, C., & Van Leeuwen, W. (1999). MODIS vegetation index (MOD13). Algorithm theoretical basis document, 3, 213.

Hunter-Jones, P. (1964). Egg development in the desert locust (Schistocerca gregaria Forsk.) in relation to the availability of water. Physiological Entomology, 39(1-3), 25-33.

Huntington, E. et al. (1834). Co A System of Modern Geography. p. 287

IPCC (2014). Climate Change 2014, synthesis Report Page: 56-60 http://ipcc.ch/pdf/assessment-report/ar5/syr/SYR_AR5_FINAL_full_wcover.pdf (Accessed at 05/11/2017).

Jimenez-Valverde, A., Lobo, J.M. & Hortal, J. (2008) Not asgood as they seem: the importance of concepts in speciesdistribution modeling. Diversity and Distributions, 14, 885–890.

Jin, M. S., & Mullens, T. (2014). A study of the relations between soil moisture, soil temperatures and surface temperatures using ARM observations and offline CLM4 simulations. Climate, 2(4), 279-295.

Joffe, S. R. (1998). Economic and policy issues in Desert Locust management: a preliminary analysis. FAO, Rome.

Junker, J., Blake, S., Boesch, C. et al. (2012) Recent declinein suitable environmental conditions for African great apes.Diversity and Distributions, 18, 1077–1091.

Jurka, T. P. (2012). Maxent: an R package for low-memory multinomial logistic regression with support for semi-automated text classification. The R Journal, 4(1), 56-59.

Jurka, T.P., & Tsuruoka, Y. (2013). Maxent: Low-memory multinomial logistic regression with support for text classification. r package version 1.3.3.1. https://CRAN.R-project.org/package=maxent.

Kadmon R, Farber O, Danin A (2004) Effect of roadside bias on the accuracy of predictive maps produced by bioclimatic models. Ecological Applications 14: 401–413.

Kassas, M., & Imam, M. (1954). Habitat and plant communities in the Egyptian Desert: III. The wadi bed ecosystem. Journal of Ecology, 42(2), 424-441.

Kassas, M. (1957). On the ecology of the Red Sea coastal land. Journal of Ecology, 45(1), 187-203.

Kéry M., B. Gardner, and C. Monnerat. (2010). Predicting species distributions from checklist data using site-occupancy models. J. Biogeogr. 37: 1851–1862

Kessell, S. R., & Whittaker, R. H. (1976). Comparisons of three ordination techniques. *Vegetatio*, *32*(1), 21-29.

Kingston, D. G., Todd, M. C., Taylor, R. G., Thompson, J. R., & Arnell, N. W. (2009). Uncertainty in the estimation of potential evapotranspiration under climate change. Geophysical Research Letters, 36(20).

Klemas, V., & Pieterse, A. (2015). Using Remote Sensing to Map and Monitor Water Resources in Arid and Semiarid Regions. The Handbook of Environmental Chemistry, 33-60. doi:10.1007/978-3-319-14212-8_2

Kottek, M., Grieser, J., Beck, C., Rudolf, B., & Rubel, F. (2006). World map of the Köppen-Geiger climate classification updated. Meteorologische Zeitschrift, 15(3), 259-263.

Kotsiantis, S. B., Zaharakis, I., & Pintelas, P. (2007). Supervised machine learning: A review of classification techniques. Emerging artificial intelligence applications in computer engineering, 160, 3-24.

Lantz, C.A. & Nebenzahl, E. (1996) Behavior and interpretation of the k statistic: resolution of two paradoxes. Journal of Clinical Epidemiology, 49, 431–434.

Lakshmi, G., Sudheer, K. P., & Chaubey, I. (2006). Auto calibration of complex watershed models using simulation-optimization framework. American Society of Agricultural and Biological Engineers, St. Joseph, MI, ASABE paper, (062126).

Latchininsky, A., Piou, C., Franc, A., & Soti, V. (2017). Applications of remote sensing to locust management. In Land Surface Remote Sensing (pp. 263-293).

Lawson, C. R., Hodgson, J. A., Wilson, R. J., & Richards, S. A. (2014). Prevalence, thresholds and the performance of presence–absence models. Methods in Ecology and Evolution, 5(1), 54-64.

Lazar, M., Diongue, A., Yang, J., Doumandji-Mitiche, B. and Lecoq, M. (2015). Location and Characterization of Breeding Sites of Solitary Desert Locust Using Satellite Images Landsat 7 ETM+ and Terra MODIS. Advances in Entomology, 3, 6-15. http://dx.doi.org/10.4236/ae.2015.31002

Lecoq, M. (2003). Desert locust threat to agricultural development and food security and FAO/international role in its control. In : Eighth Arab Congress of Plant Protection, El-Beida, Libya, October 12-16, 2003. (Eds) B. Bayaa, K.M. Makkouk, S.G. Kumari, I. El-Ghariani. Beyrouth : Arab Society for Plant Protection, 6 p.

Leroy M., P. Bicheron, C. Brockmann, U. Krämer, B. Miras, M. Huc, F. Ninô, P. Defourny, C. Vancutsem, D. Petit, V. Amberg, B. Berthelt, O. Arino and F. Ranera (2006). 'GlobCover: a 300 m global land cover product for 2005 using ENVISAT MERIS time series'ISPRS Commision VII Mid-Term Symposium: Remote Sensing: from Pixels to Processes, Enschede (NL).

Likens, G. E. (2010). Lake ecosystem ecology: A global perspective. Academic Press.

Lillesand, T., Kiefer, R. W., & Chipman, J. (2014). Remote sensing and image interpretation. John Wiley & Sons.

Liu, Z., Shi, X., Warner, E., Ge, Y., Yu, D., Ni, S., & Wang, H. (2008). Relationship between oriental migratory locust plague and soil moisture extracted from MODIS data. International Journal of Applied Earth Observation and Geoinformation, 10(1), 84-91.

Liu, C., White, M., & Newell, G. (2009). Measuring the accuracy of species distribution models: a review. In Proceedings 18th World IMACs/MODSIM Congress. Cairns, Australia (pp. 4241-4247).

Liu, Y.Y., Dorigo, W.A., Parinussa, R.M., de Jeu, R.A.M., Wagner, W., McCabe, M.F., Evans, J.P., van Dijk, A.I.J.M. (2012). Trend-preserving blending of passive and active microwave soil moisture retrievals, Remote Sensing of Environment, 123, 280-297, doi: 10.1016/j.rse.2012.03.014. (accesed 21 November 2017)

Liu, Y., Chen, X., Yang, Y., Sun, C., & Zhang, S. (2016). Automated extraction and mapping for desert wadis from Landsat imagery in arid West Asia. Remote Sensing, 8(3), 246.

Lobo, J. M. (2008). More complex distribution models or more representative data?. Biodiversity informatics, 5.

Lobo, J. M., Jiménez‐Valverde, A., & Hortal, J. (2010). The uncertain nature of absences and their importance in species distribution modelling. Ecography, 33(1), 103-114.

Luo, Y., Su, B., Yuan, J., Li, H., & Zhang, Q. (2011). GIS techniques for watershed delineation of SWAT model in plain polders. Procedia Environmental Sciences, 10, 2050-2057.

Maeno, K., & Tanaka, S. (2009). Is juvenile hormone involved in the maternal regulation of egg size and progeny characteristics in the desert locust?. Journal of insect physiology, 55(11), 1021-1028.

Magor, J. I., Lecoq, M., & Hunter, D. M. (2008). Preventive control and Desert Locust plagues. Crop Protection, 27(12), 1527-1533.

Manel, S., Williams, H.C. & Ormerod, S.J. (2001) Evaluating presence–absence models in ecology: the need to account for prevalence. Journal of Applied Ecology, 38, 921–931.

Marmion, M., Parviainen, M., Luoto, M., Heikkinen, R.K. &Thuiller, W. (2009) Evaluation of consensus methods inpredictive species distribution modelling. Diversity andDistributions, 15,59–69.

Mateo, R. G., Croat, T. B., Felicísimo, Á. M., & Muñoz, J. (2010). Profile or group discriminative techniques? Generating reliable species distribution models using pseudo‐absences and target‐group absences from natural history collections. Diversity and Distributions, 16(1), 84-94.

MathWorks (2018). https://es.mathworks.com/matlabcentral/fileexchange/14178-nash-sutcliffe-model-accuracy-metric (Accessed at 03/04/2018).

Maxwell-Darling, R. C. (1936). The outbreak centres of Schistocerca gregaria, Forsk., on the Red Sea coast of the Sudan. Bulletin of Entomological Research, 27(1), 37-66.

McCullagh, P. and Nelder, J. A. (1989). Generalized linear models. Chapman and Hall.

McPherson, J. M., Jetz, W., & Rogers, D. J. (2004). The effects of species' range sizes on the accuracy of distribution models: ecological phenomenon or statistical artefact?. Journal of applied ecology, 41(5), 811-823.

Melesse, A. M., Weng, Q., Thenkabail, P. S., & Senay, G. B. (2007). Remote sensing sensors and applications in environmental resources mapping and modelling. Sensors, 7(12), 3209-3241.

Melgani, F., & Bruzzone, L. (2004). Classification of hyperspectral remote sensing images with support vector machines. IEEE Transactions on geoscience and remote sensing, 42(8), 1778-1790.

Mellert K.H., Fensterer V., Küchenhoff H., Reger B., Kölling C., Klemmt H.J. and Ewald J. (2011). Hypothesis-driven species distribution models for tree species in the Bavarian Alps. Journal of Vegetation Science 22: 635-646.

Meynard, C. N., Gay, P. E., Lecoq, M., Foucart, A., Piou, C., & Chapuis, M. P. (2017). Climate-driven geographic distribution of the desert locust during recession periods: subspecies' niche differentiation and relative risks under scenarios of climate change. Global change biology.

Mi, C., Huettmann, F., Guo, Y., Han, X., & Wen, L. (2017). Why choose Random Forest to predict rare species distribution with few samples in large undersampled areas? Three Asian crane species models provide supporting evidence. PeerJ, 5, e2849.

Michalski, R. S., Carbonell, J. G., & Mitchell, T. M. (Eds.). (2013). Machine learning: An artificial intelligence approach. Springer Science & Business Media.

Minka, N. S., & Ayo, J. O. (2014). Influence of cold–dry (harmattan) season on colonic temperature and the development of pulmonary hypertension in broiler chickens, and the modulating effect of ascorbic acid. Open Access Anim Physiol, 6, 1-11.

Mohammad, F. S., & Adamowski, J. (2015). Interfacing the geographic information system, remote sensing, and the soil conservation service-curve number method to estimate curve number and runoff volume in the Asir region of Saudi Arabia. Arabian Journal of Geosciences, 8(12), 11093-11105.

Monserud, R. A., & Leemans, R. (1992). Comparing global vegetation maps with the Kappa statistic. Ecological modelling, 62(4), 275-293.

Monteiro, J. A., Strauch, M., Srinivasan, R., Abbaspour, K., & Gücker, B. (2016). Accuracy of grid precipitation data for Brazil: application in river discharge modelling of the Tocantins catchment. Hydrological processes, 30(9), 1419-1430.

Monteith, J. L. (1965). Evaporation and the environment: The state and movement of water in living organism, XIXth Symposium.

Moretti, G., & Montanari, A. (2008). Inferring the flood frequency distribution for an ungauged basin using a spatially distributed rainfall-runoff model. Hydrology and Earth System Sciences, 12(4), 1141-1152.

Nachtergaele, F., van Velthuizen, H., Verelst, L., Batjes, N. H., Dijkshoorn, K., van Engelen, V. W. P., ... & Montanarela, L. (2010). The harmonized world soil database. In *Proceedings of the 19th World Congress of Soil Science, Soil Solutions for a Changing World, Brisbane, Australia, 1-6 August 2010* (pp. 34-37).

Naujokaitis-Lewis, I.R., Curtis, J.M.R., Tischendorf, L., Bad-zinski, D., Lindsay, K. & Fortin, M-J. (2013). Uncertaintiesin coupled species distribution-metapopulation dynamicsmodels for risk assessments under climate change. Diversityand Distributions, 19, 541–554.

National Center for Atmospheric Research Staff (2017). "The Climate Data Guide: Climate Forecast System Reanalysis (CFSR)." https://climatedataguide.ucar.edu/climate-data/climate-forecast-system-reanalysis-cfsr. (Accessed at 29/11/2017).

NRCAN, Natural resources of Canada (2015). http://www.nrcan.gc.ca/node/14639 (Accessed at 07/03/2018).

Neitsch, S. L., Arnold, J. G., Kiniry, J. R., & Williams, J. R. (2011). Soil and water assessment tool theoretical documentation version 2009. Texas Water Resources Institute.

Nevo, D. (1996). The desert locust, Schistocerca gregaria, and its control in the land of Israel and the Near East in antiquity, with some reflections on its appearance in Israel in modern times. Phytoparasitica, 24(1), 7-32.

Nicholson, S. E., & Farrar, T. J. (1994). The influence of soil type on the relationships between NDVI, rainfall, and soil moisture in semiarid Botswana. I. NDVI response to rainfall. Remote Sensing of Environment, 50(2), 107-120.

Nicholson, S. (2005). On the question of the “recovery” of the rains in the West African Sahel. Journal of arid environments, 63(3), 615-641.

Nishide, Y., Tanaka, S., & Saeki, S. (2015). Adaptive difference in daily timing of hatch in two locust species, Schistocerca gregaria and Locusta migratoria: the effects of thermocycles and phase polyphenism. Journal of insect physiology, 72, 79-87.

Nishide, Y., & Tanaka, S. (2016). Desert locust, Schistocerca gregaria, eggs hatch in synchrony in a mass but not when separated. Behavioral ecology and sociobiology, 70(9), 1507-1515.

Nishide, Y., Suzuki, T., & Tanaka, S. (2017). The hatching time of Locusta migratoria under outdoor conditions: role of temperature and adaptive significance. Physiological Entomology, 42(2), 146-155.

Nix, H. A., & Busby, J. (1986). BIOCLIM, a bioclimatic analysis and prediction system. *Annual report CSIRO. CSIRO Division of Water and Land Resources, Canberra.*

Noy-Meir, I. (1973). Desert ecosystems: environment and producers. Annual review of ecology and systematics, 4(1), 25-51.

Osborne, P.E., Foody, G.M. & Suarez-Seoane, S. (2007) Non-stationarity and local approaches to modelling the distribu-tion of wildlife. Diversity and Distributions, 13, 313–323.

Pablos, M.; Gonzalez-Zamora, A.; Sanchez, N. and Martinez-Fernandez, J (2018). Assessment of Root Zone Soil Moisture Estimations from SMAP, SMOS and MODIS Observations. Remote Sens 10, 981

Parker, J. R. (1930). Some effects of temperature and moisture upon Melanoplus mexicanus mexicanus Saussure and Camnula pellucida Scudder (Orthoptera). Some Effects of Temperature and Moisture upon Melanoplus mexicanus mexicanus Saussure and Camnula pellucida Scudder (Orthoptera)., (223).

Pearce, J., & Ferrier, S. (2000). Evaluating the predictive performance of habitat models developed using logistic regression. Ecological modelling, 133(3), 225-245.

Pedgley, D. (1981). Desert locust forecasting manual (Volume 1 of 2). Centre for Overseas Pest Research.

Pekel, J. F., Ceccato, P., Vancutsem, C., Cressman, K., Vanbogaert, E., & Defourny, P. (2011). Development and application of multi-temporal colorimetric transformation to monitor vegetation in the desert locust habitat. IEEE Journal of selected topics in applied earth observations and remote sensing, 4(2), 318-326.

Pener, M. P., & Yerushalmi, Y. (1998). The physiology of locust phase polymorphism: an update. Journal of Insect Physiology, 44(5-6), 365-377.

Pener, M. P., & Simpson, S. J. (2009). Locust phase polyphenism: an update. Advances in Insect Physiology, 36, 1-272.

Perez de Ayala, J.M. (2011). Mauritanie : nature et paysage = Mauritania : naturaleza y paisaje. IUCN-2011-051 Pag: 1-40

Phillips, S. J., Dudík, M., & Schapire, R. E. (2004). A maximum entropy approach to species distribution modeling. In Proceedings of the twenty-first international conference on Machine learning (p. 83). ACM.

Phillips, S. J., Anderson, R. P., & Schapire, R. E. (2006). Maximum entropy modeling of species geographic distributions. Ecological modelling, 190(3-4), 231-259.

Phillips SJ, Dudik M, Elith J, Graham CH, Lehmann A, et al. (2009) Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. Ecological Applications 19: 181–197.

Phillips, S. (2017). A Brief Tutorial on Maxent. Available from url: http://biodiversityinformatics.amnh.org/open_source/maxent/ (Accessed at 29/3/18).

Piou, C., Lebourgeois, V., Benahi, A. S., Bonnal, V., el Hacen Jaavar, M., Lecoq, M., & Vassal, J. M. (2013). Coupling historical prospection data and a remotely-sensed vegetation index for the preventative control of Desert locusts. Basic and applied ecology, 14(7), 593-604.

Piou, C., Bacar, M. E. H. J., Babah Ebbe, M. A. O. B., Chihrane, J., Ghaout, S., Cisse, S., ... & Halima, T. B. (2017). Mapping the spatiotemporal distributions of the Desert Locust in Mauritania and Morocco to improve preventive management. Basic and Applied Ecology, 25, 37-47.

Platts, P. J., Ahrends, A., Gereau, R. E., McClean, C. J., Lovett, J. C., Marshall, A. R., ... & Marchant, R. (2010). Can distribution models help refine inventory-based estimates of conservation priority? A case study in the Eastern Arc forests of Tanzania and Kenya. Diversity and Distributions, 16(4), 628-642.

Popov, G. B. (1958). Ecological studies on oviposition by swarms of the Desert Locust (Schistocerca gregaria Forskal) in eastern Africa. Ecological Studies on Oviposition by Swarms of the Desert Locust (Schistocerca gregaria Forskal) in eastern Africa., (31).

Popov, G. B., Duranton, J. F., & Gigault, J. (1991). Etude écologique des biotopes du criquet pèlerin# Schistocerca gregaria (Forskal, 1775) en Afrique Nord-Occidentale: mise en évidence et description des unités territoriales écologiquement homogènes. CIRAD-PRIFAS.

Puschendorf, R., Carnaval, A. C., VanDerWal, J., Zumbado-Ulate, H., Chaves, G., Bolaños, F., & Alford, R. A. (2009). Distribution models for the amphibian chytrid Batrachochytrium dendrobatidis in Costa Rica: proposing climatic refuges as a conservation tool. Diversity and Distributions, 15(3), 401-408.

R Development Core Team (2012) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

R Development Core Team (2016) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Rao, K.S., Chandra, G.I.R.I.S.H., & RAO, P. N. (1987). The relationship between brightness temperature and soil moisture Selection of frequency range for microwave remote sensing. International Journal of Remote Sensing, 8(10), 1531-1545.

Rast, M., Bezy, J. L., & Bruzzi, S. (1999). The ESA Medium Resolution Imaging Spectrometer MERIS a review of the instrument and its mission. International Journal of Remote Sensing, 20(9), 1681-1702.

Recknagel, F. (2003). Ecological applications of adaptive agents. In Ecological Informatics (pp. 73-88). Springer, Berlin, Heidelberg.

Reddy S, Davalos LM (2003). Geographical sampling bias and its implications for conservation priorities in Africa. Journal of Biogeography 30: 1719–1727.

Reichle, R., G. De Lannoy, R. D. Koster, W. T. Crow, and J. S. Kimball. (2017). SMAP L4 9 km EASE-Grid Surface and Root Zone Soil Moisture Geophysical Data, Version 3. [Indicate subset used]. Boulder, Colorado USA. NASA National Snow and Ice Data Center Distributed Active Archive Center. doi: http://dx.doi.org/10.5067/B59DT1D5UMB4. (Accessed at 03/05/2018).

Renier, C., Waldner, F., Jacques, D. C., Babah Ebbe, M. A., Cressman, K., & Defourny, P. (2015). A dynamic vegetation senescence indicator for near-real-time desert locust habitat monitoring with MODIS. Remote Sensing, 7(6), 7545-7570.

Rhee, J., & Im, J. (2017). Meteorological drought forecasting for ungauged areas based on machine learning: Using long-range climate forecast and remote sensing data. Agricultural and Forest Meteorology, 237, 105-122.

Ridgeway G (1999). The state of boosting. Computing Science and Statistics, 172-181.

Ripley, B. D. (1996). Pattern recognition and neural networks. Cambridge Univ. Press.

Robinson, T. P., Wint, G. W., Conchedda, G., Van Boeckel, T. P., Ercoli, V., Palamara, E., & Gilbert, M. (2014). Mapping the global distribution of livestock. PloS one, 9(5), e96084.

Rodriguez, E., Morris, C. S., & Belz, J. E. (2006). A global assessment of the SRTM performance. Photogrammetric Engineering & Remote Sensing, 72(3), 249-260.

Ruete, A., & Leynaud, G. C. (2015). Goal-oriented evaluation of species distribution models' accuracy and precision: True Skill Statistic profile and uncertainty maps (No. e1478). PeerJ PrePrints.

Saha, S., Moorthi, S., Pan, H. L., Wu, X., Wang, J., Nadiga, S., ... & Liu, H. (2010). The NCEP climate forecast system reanalysis. Bulletin of the American Meteorological Society, 91(8), 1015-1058.

Saha, Suranjana, and Coauthors, (2014): The NCEP Climate Forecast System Version 2. J. Climate, 27, 2185-2208

Sanchez-Zapata, J. A., Donázar, J. A., Delgado, A., Forero, M. G., Ceballos, O., & Hiraldo, F. (2007). Desert locust outbreaks in the Sahel: resource competition, predation and ecological effects of pest control. Journal of Applied Ecology, 44(2), 323-329.

Schmidt, H., & Karnieli, A. (2000). Remote sensing of the seasonal variability of vegetation in a semi-arid environment. Journal of arid environments, 45(1), 43-59.

Schuol, J., Abbaspour, K. C., Srinivasan, R., & Yang, H. (2008). Estimation of freshwater availability in the West African sub-continent using the SWAT hydrologic model. Journal of hydrology, 352(1-2), 30-49.

Schwarz, Gideon E. (1978). "Estimating the dimension of a model", Annals of Statistics, 6 (2): 461–464, doi:10.1214/aos/1176344136, MR 0468014

Selley, R.C., 1997. The basins of Northwest Africa: Structural evolution. In: R. C. Selley (ed.), African Basins. Sedimentary Basins of the World 3. Elsevier, Amsterdam, 17-26.

Showler, A. T. (2008). Desert Locust, Schistocerca gregaria Forskål (Orthoptera: Acrididae) Plagues. In Encyclopedia of Entomology (pp. 1181-1186). Springer, Dordrecht.

Showler, A.T. (2018). The Desert Locust in Africa and Western Asia: Complexities of War, Politics, Perilous Terrain, and Development. https://ipmworld.umn.edu/showler-desert-locust (Accesed 01 April 2018)

Shulov, A., & PENER, M. P. (1963). Studies on the development of eggs of the desert locust (Schistocerca gregaria Forskǻl) and its interruption under particular conditions of humidity. Studies on the development of eggs of the desert locust (Schistocerca gregaria Forskǻl) and its interruption under particular conditions of humidity., (41).

Simpson, S. J., McCAFFERY, A. R., & HAeGELE, B. F. (1999). A behavioural analysis of phase change in the desert locust. Biological Reviews, 74(4), 461-480.

Simpson, S. J., Despland, E., Hägele, B. F., & Dodgson, T. (2001). Gregarious behavior in desert locusts is evoked by touching their back legs. Proceedings of the National Academy of Sciences, 98(7), 3895-3897.

Simpson, S. J., Sword, G. A., & Lo, N. (2011). Polyphenism in insects. Current Biology, 21(18), R738-R749.

Sivapalan, M. (2003). Prediction in ungauged basins: a grand challenge for theoretical hydrology. Hydrological Processes, 17(15), 3163-3170.

Skaf, R., Popov, G. B., & Roffey, J. (1990). The Desert Locust: an international challenge. Phil. Trans. R. Soc. Lond. B, 328(1251), 525-538.

Soil Conservation Service (SCS) (1972). National engineering handbook. section 4.

Song, H., Foquet, B., Mariño-Pérez, R. and Woller, D.A. (2017). Phylogeny of locusts and grasshoppers reveals complex evolution of density-dependent phenotypic plasticity. Scientific Reports 7: 6606. doi:10.1038/s41598-017-07105-y.

Springuel, I., Sheded, M. & Murphy, K.J. (1997). The plant biodiversity of the Wadi Allaqi Biosphere Reserve (Egypt): impact of Lake Nasser on a desert wadi ecosystem. Biodiversity and Conservation,6:1259. doi:10.1023/B:BIOC.0000034012.93599.c0

Srinivasan, R., Ramanarayanan, T. S., Arnold, J. G., & Bednarz, S. T. (1998). Large area hydrologic modeling and assessment part II: model application. JAWRA Journal of the American Water Resources Association, 34(1), 91-101.

Srinivasan, R., Zhang, X., & Arnold, J. (2010). SWAT ungauged: hydrological budget and crop yield predictions in the Upper Mississippi River Basin. Transactions of the ASABE, 53(5), 1533-1546.

Stower, W. J. (1958). Oviposition behaviour and egg mortality of the desert locust (Schistocerca gregaria Forskal) on the coast of Eritrea. Anti-Locust Research Centre.

Strahler, A. N. (1981). *Physical geology* (No. QE28. 2. S87 1981.).

Subyani, A. M. (2004). Use of chloride-mass balance and environmental isotopes for evaluation of groundwater recharge in the alluvial aquifer, Wadi Tharad, western Saudi Arabia. Environmental Geology, 46(6-7), 741-749.

Sun, B. Q., Zhang, Q., Dong, A. X., & CHEN, S. Y. (2005). Evolution feature on the moisture of soil for Loess Highland in Gansu. Advance in Earth Sciences, 9.

Sutherst, R. W., & Maywald, G. F. (1985). A computerised system for matching climates in ecology. *Agriculture, Ecosystems & Environment*, *13*(3-4), 281-299.

Sword, G. A., Simpson, S. J., El Hadi, O. T. M., & Wilps, H. (2000). Density-dependent aposematism in the desert locust. Proceedings of the Royal Society of London B: Biological Sciences, 267(1438), 63-68.

Syfert, M. M., Smith, M. J., & Coomes, D. A. (2013). The effects of sampling bias and model complexity on the predictive performance of MaxEnt species distribution models. PloS one, 8(2), e55158.

Symmons, P. (1992). Strategies to combat the desert locust. Crop Protection, 11(3), 206-212.

Symmons, P. M., & Cressman, K. (2001). Desert locust guidelines: biology and behaviour. FAO, Rome.

Tappan, G. G., Moore, D. G., & Knausenberger, W. I. (1991). Monitoring grasshopper and locust habitats in Sahelian Africa using GIS and remote sensing technology. International Journal of Geographical Information System, 5(1), 123-135.

Teklu, G. W. (2003). Habitats and spatial pattern of solitarious desert locusts (Schistocerca gregaria Forsk.) on the coastal plain of Sudan.

Thuiller, W., Lavorel, S., Sykes, M.T. & Araujo, M.B. (2006). Using niche-based modelling to assess the impact ofclimate change on tree functional diversity in Europe.Diversity and Distributions, 12,49–60.

Thuiller, W., Lafourcade, B., & Araujo, M. (2009). ModOperating manual for BIOMOD. Thuiller W, Lafourcade B (2010) BIOMOD: species/climate modelling functions. R package version, 1-1. (accesed 17 November 2017)

Thuiller, W., Georges, D., Engler, R., Breiner, F., Georges, M. D., & Thuiller, C. W. (2016). Package 'biomod2'.

Topaz, C. M., D'Orsogna, M. R., Edelstein-Keshet, L., & Bernoff, A. J. (2012). Locust dynamics: behavioral phase change and swarming. PLoS computational biology, 8(8), e1002642.

Torres, L. G., Read, A. J., & Halpin, P. (2008). FINE-SCALE HABITAT MODELING OF A TOP MARINE PREDATOR: DO PREY DATA IMPROVE PREDICTIVE CAPACITY. Ecological Applications, 18(7), 1702-1717.

Trape, S. (2009). Impact of climate change on the relict tropical fish fauna of Central Sahara: threat for the survival of Adrar mountains fishes, Mauritania. Plos one, 4(2), e4400.

Tratalos, J. A., & Cheke, R. A. (2006). Can NDVI GAC imagery be used to monitor desert locust breeding areas?. Journal of arid environments, 64(2), 342-356.

Tsoar, A., Allouche, O., Steinitz, O., Rotem, D. & Kadmon,R. (2007). A comparative evaluation of presence-only meth-ods for modelling species distribution. Diversity and Distri-butions, 13, 397–405.

Tucker, C. J., Hielkema, J. U., & Roffey, J. (1985). The potential of satellite remote sensing of ecological conditions for survey and forecasting desert-locust activity. International Journal of Remote Sensing, 6(1), 127-138.

USDA-NRCS (1986). Urban Hydrology for Small Watersheds TR-55.

http://www.professorpatel.com/uploads/7/6/5/6/7656897/tr55.pdf (Accessed at 30/03/2018)

Uvarov, B. P. (1957). The Aridity Factor in the Ecology of Locusts and Grasshoppers of the Old World. The Aridity Factor in the Ecology of Locusts and Grasshoppers of the Old World.

Uvarov, B. P. (1966). Grasshoppers and Locusts: A Handbook of General Acridology. Vol. 1, Anatomy, Physiology, Development, Phase Polymorphism, Introduction to Taxonomy. Published for the Anti-Locust Research Centre at the University Press.

Uvarov, B. (1977). Grasshoppers and locusts. A handbook of general acridology Vol. 2. Behaviour, ecology, biogeography, population dynamics. Centre for Overseas Pest Research.

Vaclavík, T. & Meentemeyer, R.K. (2012) Equilibrium ornot? Modelling potential distribution of invasive species indifferent stages of invasion. Diversity and Distributions, 18,73–83.

Van den Berg, R. D., & Feinstein, O.N. (2011). Evaluating climate change and development (Vol. 8). Transaction Publishers.

Van Der Werf, W., Woldewahid, G., Van Huis, A., Butrous, M., & Sykora, K. (2005). Plant communities can predict the distribution of solitarious desert locust Schistocerca gregaria. Journal of Applied Ecology, 42(5), 989-997.

Van Huis, A., Cressman, K., & Magor, J. I. (2007). Preventing desert locust plagues: optimizing management interventions. Entomologia Experimentalis et Applicata, 122(3), 191-214.

Vandewiele, G. L., & Elias, A. (1995). Monthly water balance of ungauged catchments obtained by geographical regionalization. *Journal of hydrology*, *170*(1-4), 277-291.

Venugopal, V., & Baets, W. (1994). Neural networks and statistical techniques in marketing research: A conceptual comparison. Marketing Intelligence & Planning, 12(7), 30-38.

Verlinden, H., Badisco, L., Marchal, E., Van Wielendaele, P., & Broeck, J. V. (2009). Endocrinology of reproduction and phase transition in locusts. General and comparative endocrinology, 162(1), 79-92.

Vermote, E. F., Tanré, D., Deuze, J. L., Herman, M., & Morcette, J. J. (1997). Second simulation of the satellite signal in the solar spectrum, 6S: An overview. IEEE transactions on geoscience and remote sensing, 35(3), 675-686.

Viera, A. J., & Garrett, J. M. (2005). Understanding interobserver agreement: the kappa statistic. Fam Med, 37(5), 360-363.

Villeneuve, M., 2005. Paleozoic basins in West Africa and the Mauritanide thrust belt. Journal of African Earth Sciences, 43(1), 166-195. http://dx.doi.org/10.1016/j.jafrearsci.2005.07.012 (Accessed at 02/10/2017)

Voss, F., & Dreiser, U. (1997). Mapping of desert locust habitats using remote sensing techniques. In New Strategies in locust control (pp. 37-45). Birkhäuser Basel.

Waloff, Z. (1966). upsurges and recessions of the desert locust plague; an historical survey.

Wagner, W., Dorigo, W., de Jeu, R., Fernandez, D., Benveniste, J., Haas, E., & Ertl, M. (2012). Fusion of active and passive microwave observations to create an essential climate variable data record on soil moisture. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences (ISPRS Annals), 7, 315-321.

Waldner, F., Babah Ebbe, M. A. B., Cressman, K., & Defourny, P. (2015). Operational monitoring of the Desert Locust habitat with Earth Observation: An assessment. ISPRS International Journal of Geo-Information, 4(4), 2379-2400.

Wan, Z., & Li, Z. L. (1997). A physics-based algorithm for retrieving land-surface emissivity and temperature from EOS/MODIS data. IEEE Transactions on Geoscience and Remote Sensing, 35(4), 980-996.

Wan, Z., Zhang, Y., Zhang, Q., & Li, Z. L. (2002). Validation of the land-surface temperature products retrieved from Terra Moderate Resolution Imaging Spectroradiometer data. Remote sensing of Environment, 83(1), 163-180.

Wang, S. C. (2003). Artificial neural network. In Interdisciplinary computing in java programming (pp. 81-100). Springer, Boston, MA.

Wang, L., & Qu, J. J. (2009). Satellite remote sensing applications for surface soil moisture monitoring: A review. Frontiers of Earth Science in China, 3(2), 237-247.

Wang, X., & Kang, L. (2014). Molecular mechanisms of phase change in locusts.

Wang, T., Wedin, D. A., Franz, T. E., & Hiller, J. (2015). Effect of vegetation on the temporal stability of soil moisture in grass-stabilized semi-arid sand dunes. Journal of Hydrology, 521, 447-459.

Ward, D., Feldman, K., & Avni, Y. (2001). The effects of loess erosion on soil nutrients, plant diversity and plant quality in Negev desert wadis. Journal of Arid Environments, 48(4), 461-473.

Wheatley, N. (1995). Where to Watch Birds in Africa. Christopher Helm. pp. 233–235. ISBN 0-7136-4013-8.

Williams, J. R. (1969). Flood routing with variable travel time or variable storage coefficients. Transactions of the ASAE, 12(1), 100-0103.

Williams, J. N., Seo, C., Thorne, J., Nelson, J. K., Erwin, S., O'Brien, J. M., & Schwartz, M. W. (2009). Using species distribution models to predict new occurrences for rare plants. Diversity and Distributions, 15(4), 565-576.

Wierenga, B. & J. Kluytmans (1994). Neural nets versus marketing models in time series analysis: A simulation study, Proceedings of the 23th Annual Conference of the European Marketing Academy.

Wisz, M. S., Hijmans, R. J., Li, J., Peterson, A. T., Graham, C. H., & Guisan, A. (2008). Effects of sample size on the performance of species distribution models. Diversity and distributions, 14(5), 763-773.

Wisz, M. S., & Guisan, A. (2009). Do pseudo-absence selection strategies influence species distribution models and their predictions? An information-theoretic approach based on simulated data. BMC ecology, 9(1), 8.

Woldewahid, G. (2004). Habitats and spatial pattern of solitarious desert locust (Schistocerca gregaria Forsk.) on the coastal plain of Sudan PhD thesis. Wageningen University, The Netherlands.

Woodman, J. D. (2015). Surviving a flood: effects of inundation period, temperature and embryonic development stage in locust eggs. Bulletin of entomological research, 105(4), 441-447.

Worqlul, A. W., Yen, H., Collick, A. S., Tilahun, S. A., Langan, S., & Steenhuis, T. S. (2017). Evaluation of CFSR, TMPA 3B42 and ground-based rainfall data as input for hydrological models, in data-scarce regions: The upper Blue Nile Basin, Ethiopia. Catena, 152, 242-251.

Wright, J. B. (1985). Introduction to sedimentary basins. In Geology and Mineral Resources of West Africa (pp. 75-78). Springer, Dordrecht.

Xia, J., Ning, L., Wang, Q., Chen, J., Wan, L., & Hong, S. (2017). Vulnerability of and risk to water resources in arid and semi-arid regions of West China under a scenario of climate change. Climatic Change, 144(3), 549-563.

Yackulic CB, Chandler R, Zipkin EF, Royle JA, Nichols JD, et al. .. (2012) Presence-only modelling using MAXENT: when can we trust the inferences? Methods in Ecology and Evolution: early view.

Yadav, M., Wagener, T., & Gupta, H. (2007). Regionalization of constraints on expected watershed response behavior for improved predictions in ungauged basins. *Advances in Water Resources*, *30*(8), 1756-1774.

Yu, G., Shen, H., & Liu, J. (2009). Impacts of climate change on historical locust outbreaks in China. Journal of Geophysical Research: Atmospheres, 114(D18).

Zaniewski, A. E., Lehmann, A., & Overton, J. M. (2002). Predicting species spatial distributions using presence-only data: a case study of native New Zealand ferns. Ecological modelling, 157(2), 261-280.

Zhang, X. S., Srinivasan, R., & Van Liew, M. (2008). Multi-site calibration of the SWAT model for hydrologic modeling. Transactions of the ASABE, 51(6), 2039-2049.

Zhang, Z., Cazelles, B., Tian, H., Stige, L. C., Bräuning, A., & Stenseth, N. C. (2009). Periodic temperature-associated drought/flood drives locust plagues in China. Proceedings of the Royal Society of London B: Biological Sciences, 276(1658), 823-831.

Zhang, Z. (2018). Artificial neural network. In Multivariate Time Series Analysis in Climate and Environmental Research (pp. 1-35). Springer, Cham.

Zheng, G., & Moskal, L. M. (2009). Retrieving leaf area index (LAI) using remote sensing: theories, methods and sensors. Sensors, 9(4), 2719-2745.

Chapter 7. List of publications

7.1. National Conferences

- 3ª Reunión del Grupo de Limnología (AET, Asociación Española de Teledetección)

  Oral presentation: "El agua como factor crítico en el desarrollo de la langosta del desierto y su detección mediante teledetección"

  Date: 04/03/2017

  Venue: UNED, c/ Bravo Murillo, 38 (Madrid, Spain)

  Link: http://www.aet.org.es/?q=glimnologia

7.2. International conferences

- Satellite Soil Moisture Validation and Application Workshop and the CCI Soil Moisture User Workshop

  Poster and short oral presentations of posters on display: "Soil Moisture influence in Desert Locust development"

  Date: 18/09/2018

  Venue: Vienna University of Technology, Gußhausstraße 27-29, 1040 (Wien, Austria)

  Link: https://smw.geo.tuwien.ac.at/

7.3 Publications in International Journals

- Journal: Journal of Applied Remote Sensing (JARS)

  Impact Factor (JCR, 2017) = 0.976

  Article: "Machine learning approach to locate desert locust breeding areas based on ESA CCI soil moisture"

  Date of publication: 28/08/2018

  DOI: J. of Applied Remote Sensing, 12(3), 036011 (2018).

  https://doi.org/10.1117/1.JRS.12.036011

- Journal: Journal of Remote Sensing

  Impact Factor (JCR, 2017) = 3.406

  Article: "Detecting Areas Vulnerable to Sand Encroachment Using Remote Sensing and GIS Techniques in Nouakchott, Mauritania"

  Date of publication: 25/09/2018

  DOI: Remote Sens. 2018, 10(10), 1541

  https://doi.org/10.3390/rs10101541