

# PRODUCTION DES CARTES DE PROBABILITE DE PRESENCE DES CRIQUETS PELERINS SUR LE TERRITOIRE MAROCAIN A PARTIR DES DONNEES DE TELEDETECTION

Mohammed Layelmam<sup>1</sup>, Mourad Bouziani<sup>1</sup>, Mohammed Faouzi Smiej<sup>2</sup>, El Hassane Semlali<sup>1</sup>,  
Saïd Ghaout<sup>3</sup>, Cyril Piou<sup>3,4,5</sup>

1 : Filière de Formation en Sciences Géomatiques et Ingénierie Topographique, IAV Hassan II, Rabat, Maroc.

E-mail : layelmam@gmail.com

2 : Centre Royal de Télédétection Spatiale (CRTS), Maroc

3: Centre National de Lutte Antiacridienne (CNLAA), BP125, Inezgane, Maroc

4: CIRAD, UMR CBGP, F-34398 Montpellier, France

5 : Université Ibn Zohr, Agadir, Maroc

## Résumé

Le criquet pèlerin, *Schistocerca gregaria*, est l'une des espèces acridiennes les plus redoutées en raison de l'ampleur des dégâts qu'il occasionne sur les cultures. Les pays touchés par cet insecte mobilisent des équipes de prospecteurs pour surveiller les habitats favorables à leur développement. Dans le but de faciliter les interventions des prospecteurs, nous proposons pour le cas du Maroc une méthode statistique qui permet de produire des cartes de probabilité de présence des criquets pèlerins. Cette méthode est basée sur l'historique des prospections et les variables explicatives calculées à la base des données de télédétection. Les données de télédétection utilisées (indice de végétation par différence normalisée (NDVI), température de surface (LST) et pluviométrie (P)) sont issues du capteur MODIS et la technique CMORPH (NOAA CPC MORPHING Technique). Durant cette étude, nous avons testé deux modèles statistiques (régression logistique et forêts aléatoires) afin de trouver la meilleure approche à utiliser. La comparaison des indicateurs extraits à partir de la matrice de confusion de chaque modèle nous a permis de conclure que la prévision par l'approche de forêts aléatoires semble meilleure que la régression logistique. En utilisant la valeur 0.5 comme seuil de probabilité, la validation des cartes produites par le modèle forêts aléatoires a permis d'avoir une moyenne de 69% de correspondance entre les cartes de prédiction obtenues (neuf cartes pour l'année 2013) et les signalisations réalisées durant la même période. Cette étude est une étape supplémentaire dans l'élaboration des cartes qui vont aider les prospecteurs à mieux orienter les missions de terrain et faciliter la localisation des zones favorables au développement des criquets pèlerins.

**Mots clés** : criquet pèlerin, Maroc, prédiction, télédétection, forêts aléatoires, régression logistique.

## Abstract

The desert locust, *Schistocerca gregaria*, is one of the most feared insect, because it causes significant damage on crops. The affected countries mobilize survey teams to search for suitable area for desert locust development. To facilitate the interventions of survey teams, we have proposed for Morocco area a statistical method to produce probability maps of the desert locusts presence. This method is based on historical surveys and physical values calculated from remote sensing data. The remote sensing data used (Normalized Difference Vegetation Index (NDVI), Surface Temperature (LST) and Rainfall (P)) are derived from the MODIS sensor and the CMORPH (NOAA CPC MORPHING Technique) technique. In this study, we tested two statistical models (Logistic Regression and Random Forests) to find the best approach to produce prediction maps. The comparison of indicators extracted from the confusion matrix of each model allowed us to conclude that the Random Forests approach is better than logistic regression in the predicting. By using the threshold 0.5, the validation of the maps produced by the random forest model allowed to have an average of 69% of correspondence between the prediction maps obtained (nine maps for the 2013 year) and the signaling of prospection in the same period. This study is a supplementary step in the elaboration of maps that will help survey teams to guide field missions and locate favorable areas of desert locust development.

**Keywords**: desert locust, Morocco, prediction, remote sensing, random forest, logistic regression.

## 1. Introduction

Les criquets pèlerins, *Schistocerca gregaria*, sont des insectes qui occupent les régions de l'Afrique de l'Ouest jusqu'à l'Inde (Ould Babah, 2003; Guendouz-Benrima et al., 2007; Issoufou, 2011). Ces insectes changent de comportement d'un état solitaire à un état grégarien en réponse à une augmentation de la densité résultant des changements climatiques ou environnementaux (Pener and Simpson, 2009; Cisse et al., 2013).

Les criquets pèlerins migrent sous forme de bandes larvaires ou d'essaims d'ailés. Ils causent en période d'invasion des dégâts importants sur les cultures et la végétation. A titre indicatif, 570 millions de dollars américains ont été dépensés pour lutter contre ces insectes dans la région occidentale au cours de l'invasion de 2003-2005 (Deshormes, 2011).

Pour traiter le risque d'invasion dès le début et afin d'empêcher les criquets pèlerins de former des grands essaims, le Maroc et les pays de l'Afrique de l'Ouest et du Nord ont adopté une stratégie de lutte préventive qui consiste à surveiller les zones de développement des populations à l'aide des équipes de prospecteurs (Layid, 1991; Magor et al., 2008). Cette mission de surveillance n'est jamais une mission facile car les équipes chargées des prospections doivent couvrir des superficies larges, et les endroits visités sont parfois accidentés et risqués.

Afin d'aider les prospecteurs à localiser les zones favorables au développement des criquets pèlerins, plusieurs initiatives ont été mises en œuvre par des chercheurs pour intégrer de nouvelles technologies. Parmi ces technologies on trouve l'utilisation et l'interprétation des données de télédétection. Ces données ont été utilisées pour guider les prospecteurs (Keita, 2009), pour détecter les zones de végétation verte qui peuvent attirer les populations acridiennes (Despland et al., 2004; Anyamba et al., 2005; Tratalos and Cheke, 2006; Keita, 2009), et pour identifier les conditions favorables au développement des criquets (Hielkema and Snijders, 1994; Lazar et al., 2015; Veran et al., 2015).

D'autres types de méthodes ont été utilisés pour la localisation des zones favorables au développement des criquets pèlerins à savoir les modèles statistiques. Dans ce contexte, Piou et al., (2013) ont utilisé des variables composites extraites spatialement et temporairement d'un indice de végétation issu de la télédétection (NDVI-MODIS). Ces travaux utilisent un

modèle de régression logistique pour relier la structure et la qualité de la végétation à la présence ou à l'absence des criquets pèlerins.

Dans le cadre du projet LDAS-Maroc du Centre Royal de Télédétection Spatiale (Smiej et al., 2015), l'équipe projet a utilisé l'approche statistique de forêts aléatoires en se basant sur les données de prospections acridiennes et les données satellitaires (végétation, température de surface et pluviométrie) pour générer des cartes de probabilité de présence des criquets pèlerins à une résolution de 25x25 km. Cette résolution reste difficilement interprétable sur le terrain par les enquêteurs.

Dans la continuité des travaux de Piou et al., (2013) et Smiej et al., (2015), nous proposons dans cette étude de : (1) Comparer les deux méthodes statistiques permettant de prédire la présence et l'absence des criquets pèlerins. (2) Démontrer la possibilité de générer des cartes de prédiction de la présence des criquets pèlerins à une résolution de 1x1 km en se basant sur des variables de télédétection.

## 2. Matériels et méthodes

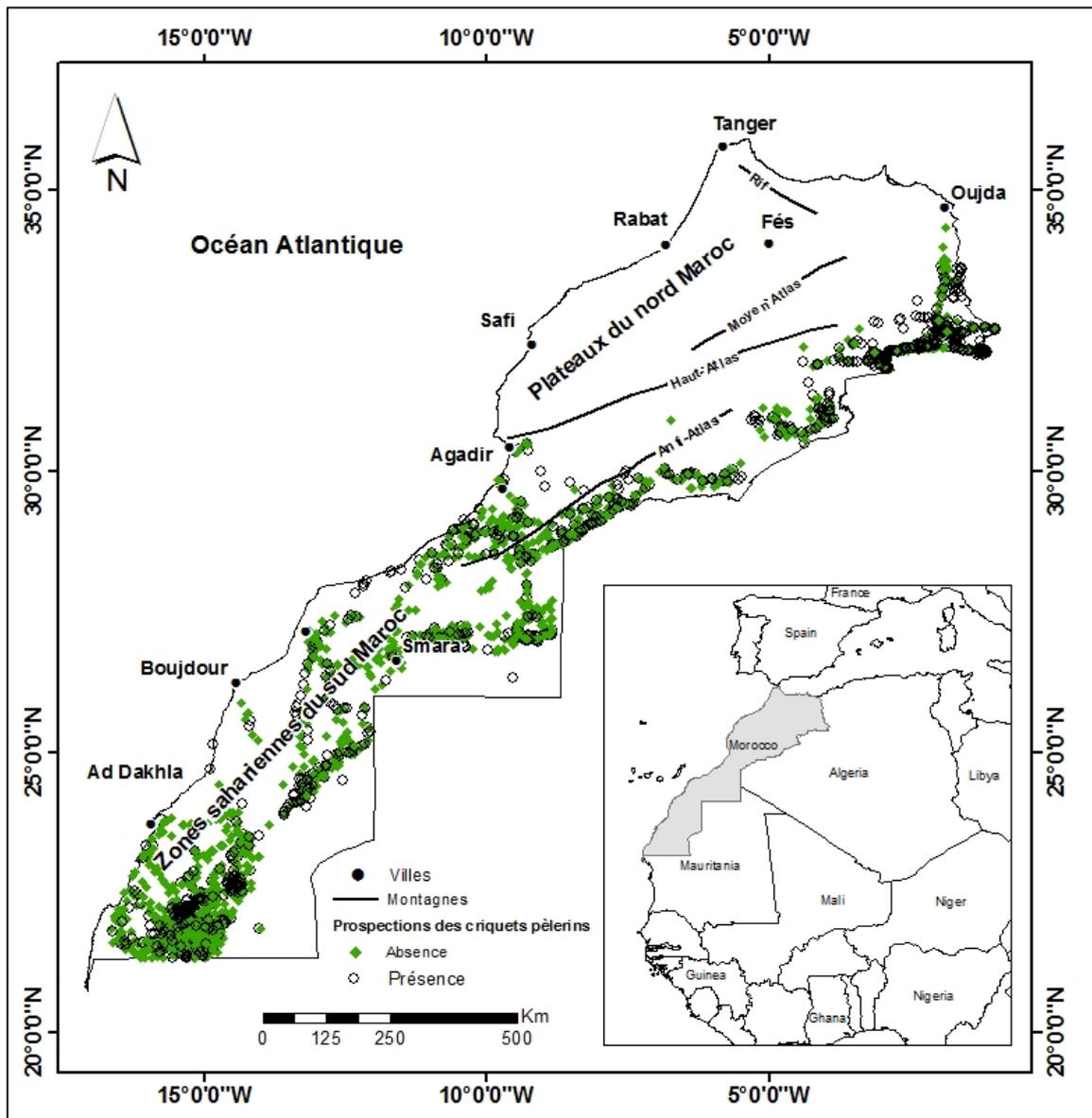
### 2.1. Zone d'étude

La présente étude a été réalisée à l'échelle du Maroc. Ce pays est caractérisé par un relief complexe qui est composé des chaînes de montagnes du Rif au nord, des Moyen et Haut Atlas au centre, de l'Anti Atlas et de l'Atlas Saharien au sud. Les chaînes montagneuses de l'Atlas délimitent deux grandes régions : les plateaux du nord (Plateau des Phosphates, Hauts Plateaux de l'Oriental) et les zones sahariennes du sud (Hamadas). (Voir Figure 1).

### 2.2. Données utilisées

#### 2.2.1. Données acridiennes

Les données acridiennes utilisées dans cette étude sont celles collectées et préparées par les équipes du Centre National de Lutte Antiacridienne au Maroc. Ces données couvrent la période 2006-2013. Elles contiennent 4825 prospections, 2692 avec absence de criquets et 2133 où les criquets pèlerins étaient présents. La majorité des prospections sont en dessous des chaînes montagneuses de l'Atlas. La Figure 1 montre la distribution géographique des prospections acridiennes.

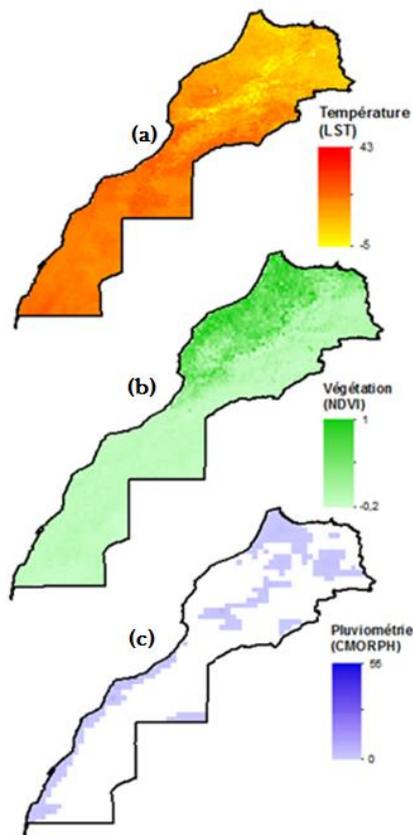


**Figure 1** : Zone d'étude avec la distribution géographique des prospections utilisées, collectées par le centre national de lutte anti acridienne du Maroc (CNLAA) entre 2006 et 2013.

### 2.2.2. Données de télédétection

Les variables de modélisation utilisées dans cette étude ont été identifiées à l'aide des données de télédétection ci-dessous :

- La température de surface (LST, MOD11A2, résolution 1x1 km), ainsi que l'indice de végétation par différence normalisée (NDVI, MOD13Q1, résolution 250x250 m) de MODIS (<https://lpdaac.usgs.gov>). Ces données ont été téléchargées gratuitement à partir de la base de données de la NASA pour les années 2006 – 2013. Les données LST représentent les valeurs moyennes sur 8 jours et les données de NDVI sont disponibles avec des valeurs maximales sur 16 jours (Figure 2).
  - Les données de pluviométrie : elles sont produites à travers une méthode nommée CMORPH (CPC MORPHING Technique). Cette méthode utilise un ensemble d'algorithmes basés sur les instruments suivants (Joyce et al., 2004) : Micro-ondes passives à bord du DMSP 13, 14 & 15 (SSM / I), la NOAA-15, 16, 17 & 18 (AMSU-B), le AMSR-E & TMI à bord d'Aqua de la NASA et le TRMM. Les données de pluviométrie utilisées dans le cas de cette étude se caractérisent par une couverture décennale à une résolution de 0,25 x 0,25 degré (Figure 2). Elles ont été téléchargées depuis le site <http://www.cpc.ncep.noaa.gov/>.
- Les données de télédétection utilisées pour la préparation des variables explicatives ont été téléchargées dans des formats traités géométriquement.



**Figure 2 :** Exemple de données de télédétection utilisées dans cette étude. (a) : Synthèse hebdomadaire de la température de surface (LST) au jour 361 de l'année 2012, (b) : Synthèse bi-hebdomadaire de l'indice normalisé de végétation (NDVI) au jour 353 de l'année 2012, (c) : Cumul décadaire de la pluviométrie (CMORPH) au jour 2012/12/21.

Pour homogénéiser les résolutions spatiales des variables explicatives issues des données de télédétection, nous avons rééchantillonné les variables NDVI et pluviométrie à une résolution de 1x1 km. La méthode de rééchantillonnage utilisée est la méthode bilinéaire. Cette dernière se base sur les pixels les plus proches en entrée (4 pixels), pour déterminer la valeur sur le raster en sortie. La valeur de pixels en sortie correspond à la moyenne pondérée des quatre valeurs.

Le choix des variables extraites des données de télédétection (Tableau 1) est justifié par un ensemble de critères liés au cycle de développement des criquets pèlerins.

- **Végétation (NDVI) :** La végétation représente pour les criquets pèlerins non seulement une source de nourriture (Popov et al., 1990), mais constitue aussi un milieu d'abri et peut favoriser la concentration de ces insectes, condition favorable à la grégarisation (Despland et al., 2004; Cisse et al., 2013).

- **Variation temporelle de la végétation (NDVI32\_48) :** La variabilité temporelle de l'état de la végétation peut conditionner la sédentarisation ou le déplacement du criquet pèlerin en solitaires ou en groupes. Cette variabilité est caractérisée par la différence des indices de végétation entre deux périodes successives. Les travaux de Piou (Piou et al., 2013) ont montré que la différence NDVI32\_48 entre les périodes 32 jours et 48 jours précédant la date de l'analyse est celle qui a le plus de poids sur la présence/absence du criquet. Nous avons donc adopté cette différence pour caractériser la variabilité temporelle de la végétation.

- **Température (LST) :** Dans notre étude nous avons utilisé la température de surface LST comme variable explicative de la présence du criquet pèlerin pour deux raisons. D'une part, la LST conditionne la température de l'air environnant le sol à travers les échanges thermiques qui s'opèrent entre eux. Il est bien reconnu que la température de l'air est l'un des facteurs clés des durées de plusieurs stades de développement des criquets pèlerins (Popov et al., 1990). D'autre part, la température de surface caractérise l'aridité du sol et peut être ainsi utilisée comme un indicateur de l'humidité de sol.

- **Pluviométrie :** Les femelles des criquets pèlerins pondent dans des sites caractérisés par un sol humide qui évite que les œufs se dessèchent au cours de leur développement (Duranton and Lecoq, 1990). L'eau est le facteur clé du développement des œufs ainsi que la croissance de la végétation qui assurera la nourriture pour le criquet pèlerin.

Après avoir identifié les variables clés pour la caractérisation des habitats favorables aux criquets pèlerins, nous avons procédé à la préparation d'un tableau qui combine les types de prospections (présence/absence) avec les variables explicatives explicitées plus haut. Le tableau est structuré selon les attributs suivants : Date, Prospection (présence/absence), Valeur NDVI, Valeur NDVI32\_48, Valeur LST, Valeur P.

A chaque date et chaque point de prospection sont associées des valeurs calculées à partir des variables explicatives NDVI, NDVI32\_48, LST et pluviométrie. Les valeurs associées aux points de prospection correspondent à la moyenne des valeurs des pixels se situant dans un rayon de 500m autour de ces points. Les valeurs récupérées pour chaque point de prospection représenteront par la suite l'échantillon de données pour la construction des modèles.

| Variable   | Description  | Résolution spatiale | Résolution temporelle | Intervalle de valeurs de pixels | Facteur d'échelle (*) |
|--|--|---------------------|-----------------------|---------------------------------|-----------------------|
| <i>NDVI</i><br>(Indice de végétation par différence normalisée)          | $NDVI = \frac{PIR - R}{PIR + R}$ <i>PIR</i><br>: Réflectance dans le proche infrarouge<br><i>R</i> : Réflectance dans le rouge   | 250 x 250 m         | 16 jours              | [0, 5472.65]                    | 0.0001                |
| <i>NDVI<sub>32-48</sub></i><br>(Variation du NDVI)<br>(Piou et al. 2013) | $NDVI_{32-48} = \frac{2(NDVI_{32} - NDVI_{48})}{NDVI_{32} + NDVI_{48}}$ <i>NDVI<sub>32</sub></i> : Végétation 32 jours avant<br><i>NDVI<sub>48</sub></i> : Végétation 48 jours avant | 250x250 m           | 16 jours              | [-0.79, 1.15]                   | 1                     |
| <i>LST</i><br>(Température de surface)                                   | LST  | 1x1 km              | 8 jours               | [0,16487]<br>(Kelvin)           | 0.02                  |
| <i>P</i><br>(Pluviométrie)   | $P = \sum_{i=1}^{10} P_i$ <i>P<sub>i</sub></i> : Précipitation journalière   | 0,25x0,25 Degré     | 10 jours              | [2, 79]<br>(mm)                 | 1                     |

**Tableau 1** : Variables utilisées pour la modélisation de la présence/absence des criquets pèlerins (\* : le facteur d'échelle est un coefficient liant la valeur de la variable avec la valeur stockée dans l'image).

### 2.3. Modélisation du risque

#### 2.3.1. Approches statistiques

Dans le but de choisir la méthode de modélisation la plus appropriée et afin d'analyser la relation qui existe entre la présence/absence des criquets pèlerins et les variables issues des données de télédétection, nous avons utilisé deux approches de modélisation statistique qui consistent à définir des relations mathématiques entre la variable Y (Y = présence/absence) et les variables explicatives observées (Variables de télédétection) :

- La première approche est basée sur un modèle de régression logistique (GLM) qui vise à prédire et à expliquer les valeurs d'une variable binaire Y (variable à expliquer) à partir d'une des variables X (variables explicatives) (Dominguez-Almendros et al., 2011). La variable à expliquer (Y) prend deux modalités : présence "1" ou absence "0". La liaison fonctionnelle qui lie les variables Y et X est de la forme :

$$\text{logit}(P(Y = 1)) = \log\left(\frac{P(Y = 1)}{1 - P(Y = 1)}\right) = a_0 + \sum_{i=1}^n a_i X_i \quad (1)$$

Avec P (Y = 1) est la probabilité que Y soit égale à 1, Xi : Variables explicatives et ai : Poids.

- La deuxième approche est nommée forêts aléatoires (RF). C'est une méthode statistique qui génère plusieurs échantillons pour la construction de plusieurs arbres de décision aléatoirement (Genuer, 2010).

La construction d'un arbre de décision se compose des étapes suivantes : (1) choix d'une variable de partitionnement parmi les attributs qui décrivent les données d'apprentissage. (2) choix d'une ou de plusieurs valeurs de coupure de cette variable pour définir la partition. (3) itération des étapes 1 et 2 avec chacun des nœuds fils qui ne remplissent pas les critères pour devenir des feuilles.

La méthode d'arbre de décision est fortement dépendante de l'échantillon initial, ce qui peut provoquer des problèmes si certaines données sont fausses. Même si les données sont exactes, il est possible que l'ajout de quelques données à l'échantillon change complètement l'arbre ainsi construit. Pour faire face à cette difficulté, Breiman (2001) a proposé un autre modèle qui intègre de l'aléatoire dans la construction des arbres de décision. De cette manière, on ne construit plus un seul arbre de décision mais (m) arbres de décision. Pour chaque arbre, les résultats de probabilité sont agrégés par une fonction majoritaire.

Les forêts aléatoires dépendent de deux paramètres : le nombre d'échantillons choisis aléatoirement pour chaque nœud des arbres (N), et le nombre d'arbres aléatoires à induire dans la forêt (m). Dans notre cas N= 3 et m = 500.

#### 2.3.2. Construction et validation des modèles

La construction des modèles a été réalisée selon les étapes suivantes :

- Tirage d'un échantillon d'apprentissage (80% de données) pour construire les modèles de prédiction (GLM et RF). Pour le modèle GLM nous avons fait

une sélection pas à pas directe et inverse (Stepwise forward and backward) afin de choisir le meilleur modèle (BestGLM).

- Sur les 20% d'observations restantes (l'échantillon test), nous avons élaboré la matrice de confusion en confrontant les valeurs observées et les valeurs prédites (méthodologie de comparaison est présentée dans la Figure 3).

La méthode de sélection pas à pas directe et inverse est une procédure statistique qui permet d'effectuer à chaque pas des tests de signification basés sur le coefficient CAIC (Consistent Akaike Information Criterion) pour choisir les variables explicatives significatives (Bozdogan, 1987). La procédure s'arrête quand on ne peut plus ajouter ni enlever des variables. Le modèle qui a une faible valeur de CAIC est choisi comme meilleur modèle.

La mesure CAIC est un critère d'information sur la qualité du modèle. Il est donné par la formule suivante

$$CAIC = -2 \log_e [L(\hat{\theta})] + K[\log_e(n) + 1] \quad (2)$$

Avec  $L(\hat{\theta})$  est la fonction de maximum de vraisemblance du vecteur  $\theta$ ,  $K$  est le nombre de variables estimées et  $n$  est la taille de l'échantillon. (Pour notre étude le vecteur  $\theta = (NDVI, NDVI_{32-48}, LST, P)$ ,  $K = 4$  et  $n =$  nombre de prospection).

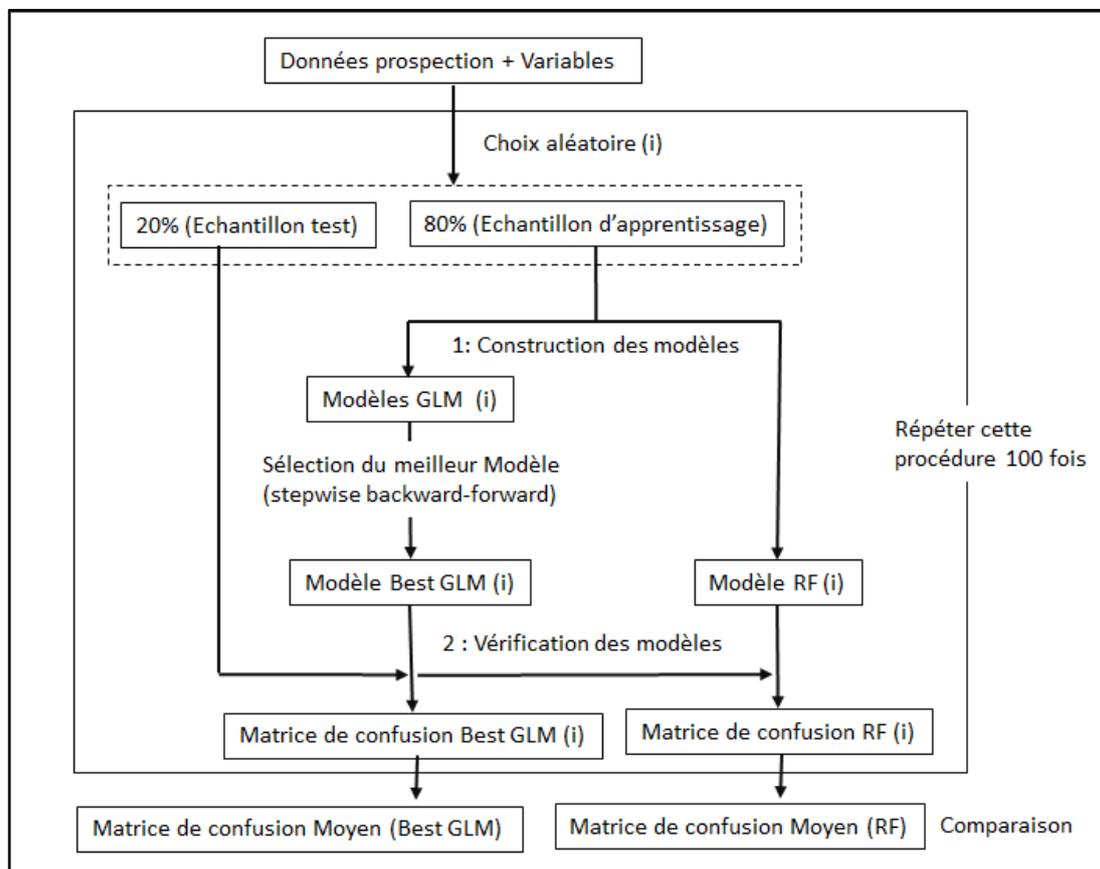


Figure 3 : Procédure de comparaison entre l'approche de régression logistique (GLM) et des forêts aléatoires (RF).

Pour comparer la capacité prédictive des deux modèles (Best GLM et RF) nous avons utilisé des matrices de confusion (Tableau 2), ces matrices comparent les valeurs prédites de chaque modèle avec les valeurs observées.

|                          | Prédiction 0 (Absence) | Prédiction 1 (Présence) |
|--------------------------|------------------------|-------------------------|
| Observation 0 (Absence)  | Vrai négatif (VN)      | Faux Positif (FP)       |
| Observation 1 (Présence) | Faux Négatif (FN)      | Vrai positif (VP)       |

**Tableau 2 :** Structure d'une matrice de confusion.

En se basant sur la matrice de confusion nous avons tiré des mesures statistiques tel que : l'erreur (E), le rappel (R), la précision (P) et le F-mesure (F) (Tableau 3).

| Indicateurs          | Formules  | Approches de prédiction   |  |
|----------------------|---|---|--|
|                      |   | « BestGLM »   | « RF »   |
| Matrice de confusion | $\begin{pmatrix} VN & FP \\ FN & VP \end{pmatrix}$ <p>Avec <math>n = FP + VP + VN + FN</math></p> | $\begin{pmatrix} 383.11 & 232.83 \\ 40.88 & 58.18 \end{pmatrix}$ <p>Avec <math>n = 715</math></p> | $\begin{pmatrix} 330.65 & 139.15 \\ 93.34 & 151.86 \end{pmatrix}$ <p>Avec <math>n = 715</math></p> |
| Erreur (E)           | $E = \frac{FN + FP}{n}$   | 0.38  | 0.32   |
| Rappel (R)           | $R = \frac{VP}{VP + FN}$  | 0.59  | 0.62   |
| Précision (P)        | $P = \frac{VP}{VP + FP}$  | 0.2   | 0.52   |
| F-mesure (F)         | $F = 2 \cdot \frac{P \cdot R}{P + R}$   | 0.3   | 0.56   |

**Tableau 3 :** Comparaison des indicateurs statistiques basés sur la matrice de confusion pour les deux modèles (Best GLM et RF).

### 3.2 Démonstration de la possibilité de générer des cartes de prédiction de la présence/absence des criquets pèlerins

Afin de produire des cartes de prédiction de la présence/absence des criquets pèlerins pour une période précise, nous avons utilisé l'approche des forêts aléatoires (RF). Cette dernière calcule pour chaque pixel (Taille 1x1 km) la probabilité de présence/absence en se basant sur les variables de télédétection de la même période.

Dans le cadre de cette étude, la construction du modèle RF est basée sur des données datant de 2006 à 2012, alors que la prédiction est lancée pour neuf périodes à partir de mois de janvier 2013. Les prédictions des cartes sont produites pour une périodicité de 16 jours en utilisant les données de la même période.

Pour notre cas la procédure de comparaison entre l'approche de régression logistique (Best GLM) et des forêts aléatoires (RF) a été répétée 100 fois pour que le résultat final soit représentatif. La matrice de confusion finale est la moyenne des matrices pour l'ensemble des tests (Voir Figure 3).

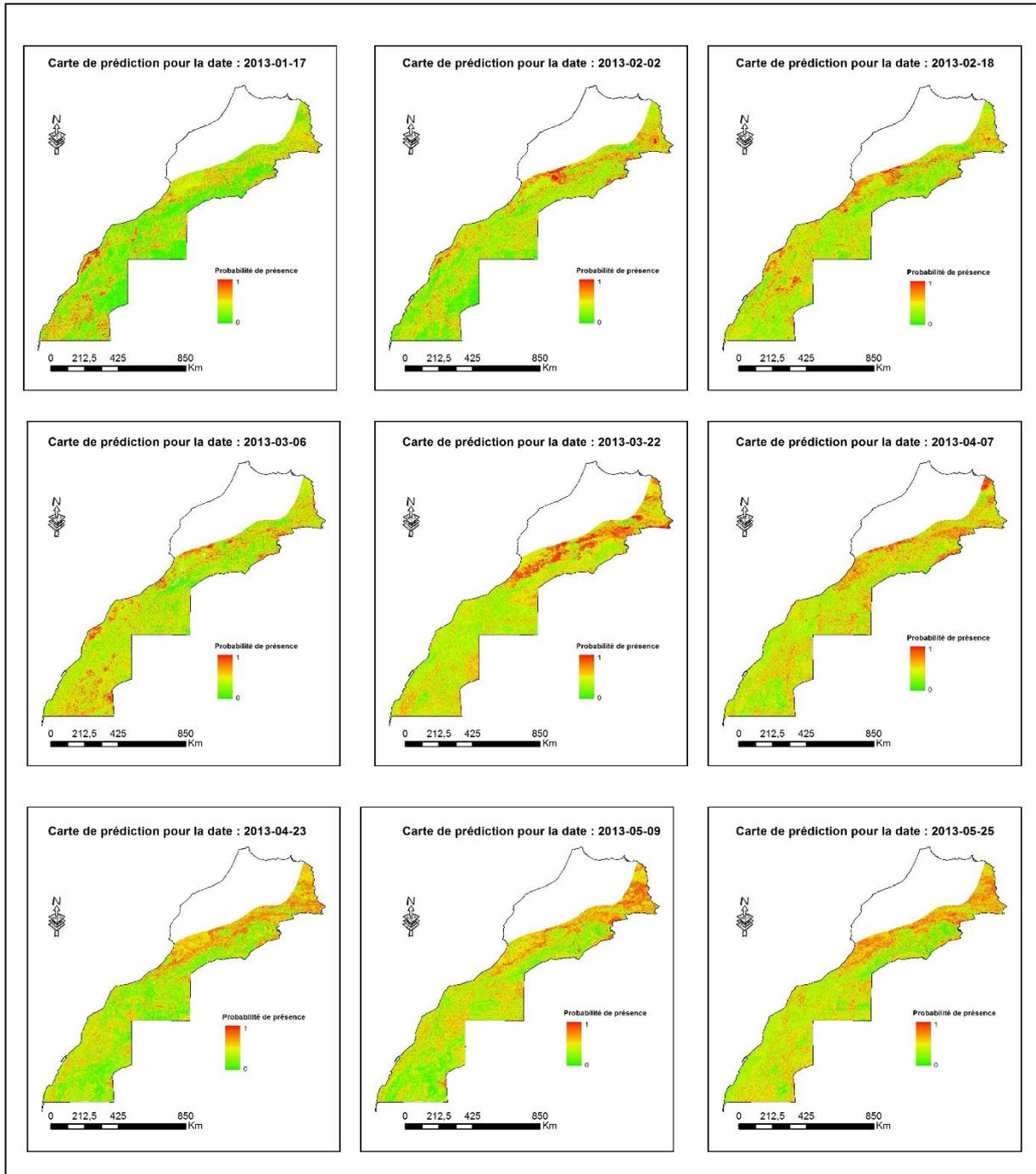
## 3. Résultats

### 3.1 Comparaison des deux approches

Les indicateurs statistiques extraits à partir de la matrice de confusion de chaque modèle (Tableau 3), nous a permis de constater que l'approche de forêts aléatoires (RF) se caractérise par une erreur faible et un F-Mesure élevé. Pour cela, nous déduisons que l'approche « RF » semble meilleure que la régression logistique.

Les cartes de probabilité de présence des criquets pèlerins dépendent directement de la résolution des variables explicatives. Puisque nous avons rééchantillonné toutes les variables à 1x1 km les cartes de probabilité ont été générées à la même résolution. Dans un premier temps, nous avons effectué des tests avec la résolution de 250 m, mais malheureusement nous avons eu des taux de correspondance très faible. C'est la raison pour laquelle nous avons décidé par la suite de continuer avec la résolution la plus faible qui est 1x1 km.

La Figure 4 présente un exemple de cartes de probabilité de présence des criquets pèlerins pour plusieurs périodes de l'année 2013. Nous signalons que nous avons appliqué un masque pour le Nord du



**Figure 4 :** Probabilité de présence des criquets pèlerins pour des périodes de 16 jours de l'année 2013 (Période de 17 Janvier au 25 Mai). Ces cartes sont le résultat de l'approche forêts aléatoires.

Pour mieux apprécier la coïncidence entre les cartes de prédiction (Figure 4), et les signalisations de présence/absence réalisées au cours de la même période nous avons essayé de chercher l'existence d'une correspondance entre les signalisations et les valeurs de pixels de la carte.

Le tableau 4 présente le pourcentage de correspondance pour les neuf périodes utilisées et ceci pour chaque seuil de probabilité. Le seuil 0.5 par exemple permet d'avoir une moyenne de 69% de prospections des criquets pèlerins qui ont été prédites par le modèle RF (79% pour la présence et 59% pour l'absence),

| Périodes   | Signalisations | Seuil de probabilité |      |      |      |     |     |     |      |      |      |
|------------|----------------|----------------------|------|------|------|-----|-----|-----|------|------|------|
|            |                | 0.1                  | 0.2  | 0.3  | 0.4  | 0.5 | 0.6 | 0.7 | 0.8  | 0.9  | 1    |
| 17/01/2013 | Présence       | 100%                 | 99%  | 89%  | 77%  | 57% | 43% | 22% | 9%   | 1%   | 0%   |
|            | Absence        | 4%                   | 12%  | 32%  | 46%  | 66% | 82% | 93% | 99%  | 99%  | 100% |
| 02/02/2013 | Présence       | 97%                  | 97%  | 97%  | 76%  | 41% | 27% | 14% | 14%  | 3%   | 0%   |
|            | Absence        | 5%                   | 16%  | 27%  | 51%  | 60% | 78% | 96% | 98%  | 100% | 100% |
| 18/02/2013 | Présence       | 100%                 | 100% | 98%  | 85%  | 73% | 56% | 33% | 21%  | 10%  | 0%   |
|            | Absence        | 1%                   | 3%   | 16%  | 38%  | 52% | 68% | 87% | 94%  | 97%  | 100% |
| 06/03/2013 | Présence       | 100%                 | 100% | 100% | 100% | 84% | 56% | 32% | 20%  | 4%   | 0%   |
|            | Absence        | 1%                   | 4%   | 18%  | 38%  | 58% | 80% | 88% | 97%  | 100% | 100% |
| 22/03/2013 | Présence       | 100%                 | 100% | 100% | 100% | 97% | 86% | 71% | 48%  | 22%  | 7%   |
|            | Absence        | 0%                   | 2%   | 9%   | 25%  | 51% | 73% | 87% | 92%  | 100% | 100% |
| 07/04/2013 | Présence       | 100%                 | 100% | 99%  | 88%  | 81% | 73% | 54% | 29%  | 8%   | 2%   |
|            | Absence        | 0%                   | 7%   | 9%   | 24%  | 43% | 65% | 85% | 100% | 100% | 100% |
| 23/04/2013 | Présence       | 100%                 | 100% | 98%  | 95%  | 93% | 80% | 71% | 54%  | 27%  | 5%   |
|            | Absence        | 0%                   | 0%   | 0%   | 33%  | 83% | 83% | 83% | 100% | 100% | 100% |
| 09/05/2013 | Présence       | 100%                 | 100% | 98%  | 95%  | 91% | 77% | 61% | 45%  | 20%  | 2%   |
|            | Absence        | 0%                   | 0%   | 25%  | 63%  | 75% | 75% | 88% | 100% | 100% | 100% |
| 25/05/2013 | Présence       | 100%                 | 100% | 98%  | 96%  | 91% | 85% | 74% | 57%  | 35%  | 20%  |
|            | Absence        | 0%                   | 11%  | 11%  | 33%  | 44% | 44% | 67% | 78%  | 100% | 100% |

**Tableau 4 :** Pourcentage de correspondance entre les cartes de prédiction et les signalisations des criquets pèlerins sur le terrain. Ces pourcentages ont été calculés pour plusieurs seuils de probabilité.

#### 4. Discussion

Les données de télédétection ont un rôle très important dans le domaine de lutte contre les invasions acridiennes, notamment pour les caractéristiques suivantes : couverture géographique, résolution spectrale, et possibilité de télécharger gratuitement ces données sur internet. (Tappan et al., 1991; Despland et al., 2004; Ji et al., 2004; Sivanpillai and Latchininsky, 2007; Zha et al., 2008; Latchininsky and Sivanpillai, 2010; Cressman, 2013).

En d'autres termes, les données de télédétection NDVI ont été beaucoup utilisées dans la caractérisation des habitats de développement des criquets pèlerins, mais la résolution spatiale de ce type de données impacte la précision des résultats d'interprétation. Despland et al. (2004) ont démontré qu'à une résolution de 7,6x7,6 km les valeurs de NDVI ne sont pas des indicateurs fiables pour la localisation des zones propices aux criquets pèlerins. Les mêmes conclusions ont été confirmées par Tratalos & Cheke (2006) pour une résolution de 8x8 km. D'un autre côté, Piou et al. (2013) ont montré qu'à une résolution de 250x250 m le NDVI peut être utile pour prédire la présence/absence de criquet pèlerin. Les résultats obtenus dans le cadre de cette étude confirment les observations de Piou et al. (2013) à une résolution de 1x1 km.

Les variables issues des données de télédétection et qui font l'objet de cette étude (NDVI, NDVI32\_48, LST et pluviométrie) ont été couplées avec les prospections

acridiennes à travers deux modèles statistiques (régression logistique et forêts aléatoires), ce qui nous a aidés à identifier la meilleure approche de prédiction. Les deux approches statistiques utilisées dans cette étude ont été choisies parce qu'elles reposent sur deux principes de modélisation différents : la régression logistique (modèle simple basé sur un modèle linéaire généralisé), et les forêts aléatoires (modèle non linéaire qui utilise des arbres de classification).

La qualité et le nombre des prospections sont des facteurs essentiels pour avoir des cartes de prédiction fiables. L'opération de collecte de ces prospections sur le terrain dépend surtout de l'expérience des équipes (Cressman, 2001). Pour cette étude, nous constatons que les prospections utilisées sont généralement réparties selon les parcours des prospecteurs (réseau routier et pistes). De ce fait, il est préférable que les prospecteurs utilisent des méthodes de rééchantillonnage adaptées afin d'aboutir à des points de prospection bien distribués géographiquement.

Un autre critère qui agit sur la fiabilité des cartes de prédiction est le nombre des variables utilisées. Plus on ajoute des variables, plus on peut avoir des probabilités de présence élevées. Dans notre cas nous avons utilisé le modèle RF avec quatre variables (Tableau 1), mais il serait possible dans de futures travaux d'utiliser plus de variables telles que la topographie et l'humidité du sol. Dans notre étude, le pourcentage de correspondance entre les cartes de prédiction et les signalisations de criquets pèlerins sur le terrain est lié aux seuils de

probabilités choisies ainsi que le nombre de prospections utilisées pour la validation. Pour cela nous recommandons dans de futures études que la validation des cartes de risque soit faite en utilisant des théories statistiques telles que la théorie des ensembles flous.

La méthodologie proposée dans notre étude permet de produire des cartes de prédiction basées sur le couplage de plusieurs variables qui caractérisent mieux les zones favorables au développement des criquets pèlerins. Il sera possible dans de futures études de tester d'autres méthodes de prédiction telles que l'approche SVM (Machine à vecteurs de support) ou les réseaux neuronaux.

Les cartes de prédiction de la présence des criquets pèlerins produites dans le cadre de cette étude représentent un outil efficace pour les prospecteurs. Ce type d'outil devra être opérationnel pour aider à la prise de décision, ce qui permettra aux agents de terrain de réduire le temps, l'effort et le coût des opérations de surveillance. Nous signalons que l'automatisation de cet outil est en cours.

## 5. Conclusion

A travers les produits de télédétection on peut avoir des informations pour caractériser les zones de développement qui se situent dans des régions naturellement difficiles d'accès ou mêmes interdites pour des raisons militaires par exemple.

Dans le cadre de cette étude, les données de télédétection (NDVI, LST et pluviométrie) ont été utilisées pour extraire des variables environnementales liées au cycle de développement des criquets pèlerins. Il existe d'autres types de données tels que la verdure de végétation, les types de sols, la texture et l'humidité de sol. Ces variables peuvent être intégrées dans des prochains travaux pour produire des cartes de prédiction de la présence des criquets pèlerins.

## Remerciements

Nous tenons à exprimer nos vives reconnaissances au Centre National de Lutte Antiacridienne Au Maroc (CNLAA) et le Centre Royal de Télédétection Spatial (CRTS) qui nous ont fourni les données nécessaires pour mener cette étude.

## Références

Anyamba, A., Small, J., Tucker, C., Cressman, K., Love, T., 2005. Remote Sensing of Eco-Climatic Conditions Associated with the 2004 Desert Locust Outbreak in Northwest and Sahelian Africa.

Presented at the Pecora 16 - Global Priorities in Land Remote Sensing, Dakota du Sud, États-Unis.

Bozdogan, H., 1987. Model selection and Akaike's Information Criterion (AIC): The general theory and its analytical extensions. *Psychometrika* 52, 345–370. doi:10.1007/BF02294361

Breiman, L., 2001. Random Forests. *Mach. Learn.* 45, 5–32. doi:10.1023/A:1010933404324

Cisse, S., Ghaout, S., Mazih, A., Babah Ebbe, M.A.O., Benahi, A.S., Piou, C., 2013. Effect of vegetation on density thresholds of adult desert locust gregarization from survey data in Mauritania. *Entomol. Exp. Appl.* 149, 159–165. doi:10.1111/eea.12121

Cressman, K., 2013. Role of remote sensing in desert locust early warning. *J. Appl. Remote Sens.* 7, 075098. doi:10.1117/1.JRS.7.075098

Cressman, K., 2001. Prospection, in: Directives sur le Criquet pèlerin, 1. FAO. Organisation des Nations Unies pour l'alimentation et l'agriculture. Rome. Italie.

Deshormes, A., 2011. Etude institutionnelle pour améliorer les rôles et responsabilités des commissions de lutte contre le criquet pèlerin créées en vertu de l'article XIV. FAO-CLCPRO.

Despland, E., Rosenberg, J., Simpson, S.J., 2004. Landscape structure and locust swarming: a satellite's eye view. *Ecography* 27, 381–391. doi:10.1111/j.0906-7590.2004.03779.x

Domínguez-Almendros, S., Benítez-Parejo, N., Gonzalez-Ramirez, A.R., 2011. Logistic regression models. *Allergol. Immunopathol. (Madr.)* 39, 295–305. doi:10.1016/j.aller.2011.05.002

Duranton, J.F., Lecoq, M., 1990. Le criquets pèlerin au sahel, in: Collection Acridologie Opérationnelle Numéro 6. Niamey, p. 183.

Genuer, R., 2010. Forêts aléatoires: aspects théoriques, sélection de variables et applications. Université Paris-Sud XI, France.

Guendouz-Benrima, A., Duranton, J.F., Benhammouda, M.E.H., Mitiche, B., 2007. Distribution des populations de Criquet pèlerin *Schistocerca gregaria* Forsk 1775 (Insecta Orthoptera) en période de rémission en Algérie de 1983 à 2000. *Sécheresse* 18, 246–253.

Hielkema, J.U., Snijders, F.L., 1994. Operational use of environmental satellite remote sensing and satellite communications technology for global food security and locust control by FAO: The ARTEMIS and DIANA systems. *Acta Astronaut.* 32, 603–616. doi:10.1016/0094-5765(94)90071-X

Issoufou, D., 2011. Etude biogéographique des zones potentielles de grégarisation du criquet pèlerin

- (*Schistocerca gregaria*, Forsk 1775) au Niger sur la base des données d'archives des prospections acridienne de 1965 à 2007. Faculté des sciences, Université Ibn Zohr, Agadir, Maroc.
- Ji, R., Xie, B.Y., Li, D.M., Li, Z., Zhang, X., 2004. Use of MODIS data to monitor the oriental migratory locust plague. *Agric. Ecosyst. Environ.* 104, 615–620. doi:10.1016/j.agee.2004.01.041
- Joyce, R.J., Janowiak, J.E., Arkin, P.A., Xie, P., 2004. CMORPH: A Method that Produces Global Precipitation Estimates from Passive Microwave and Infrared Data at High Spatial and Temporal Resolution. *J. Hydrometeorol.* 5, 487–503.
- Keita, M.N., 2009. Contribution à l'amélioration de la lutte préventive par l'étude de la biogéographie du criquet pèlerin au Mali (Mémoire). IAV Hassan 2.
- Latchininsky, A.V., Sivanpillai, R., 2010. Locust Habitat Monitoring and Risk Assessment Using Remote Sensing and GIS Technologies. *Integr. Manag. Arthropod Pests Insect Borne Dis.* 163–188.
- Layid, L.C.M., 1991. Organisation de la lutte anti-acridienne au Maroc, in: *La Lutte Anti-Acridienne. La lutte anti-acridienne*. Ed. AUPELF-UREF, John Libbey Eurotext, Paris, France, pp. 33–44.
- Lazar, M., Diongue, A., Yang, J.T., Bahia, D.M., Lecoq, M., 2015. Location and Characterization of Breeding Sites of Solitary Desert Locust Using Satellite Images Landsat 7 ETM+ and Terra MODIS. *Adv. Entomol.* 03, 6–15. doi:10.4236/ae.2015.31002
- Lecoq, M., 1988. Les criquets du sahel, in: *Collection Acridologie Opérationnelle Numéro 1*. Niamey, p. 125.
- Magor, J.I., Lecoq, M., Hunter, D.M., 2008. Preventive control and Desert Locust plagues. *Crop Prot.* 27, 1527–1533. doi:10.1016/j.cropro.2008.08.006
- Ould Babah, M.A., 2003. Biogéographie du Criquet pèlerin en Mauritanie. Fonctionnement d'une aire grégarigène et conséquences sur l'organisation de la surveillance et de la lutte antiacridienne (No. AGP/DL/TS/31), Stations de recherche acridienne sur le terrain, séries techniques. FAO, Rome.
- Pener, M.P., Simpson, S.J., 2009. Locust Phase Polyphenism: An Update, in: *Advances in Insect Physiology*. Academic Press, pp. 1–286.
- Piou, C., Lebourgeois, V., Benahi, A.S., Bonnal, V., Jaavar, M.E., Lecoq, M., Vassal, J.M., 2013. Coupling historical prospection data and a remotely-sensed vegetation index for the preventative control of Desert locusts. *Basic Appl. Ecol.* 14, 593–604. doi:10.1016/j.baae.2013.08.007
- Popov, G.B., Luong, M.L., Weel, J.V.D., 1990. Les oothèques des criquets du sahel. *Collection acridologie opérationnelle*, in: *Collection Acridologie Opérationnelle Numéro 7*. p. 93.
- Sivanpillai, R., Latchininsky, A., 2007. Mapping Locust Habitats in the Amudarya River Delta, Uzbekistan with Multi-Temporal MODIS Imagery. *Environ. Manage.* 39, 876–886. doi:10.1007/s00267-006-0193-y
- Smiej, M.F., Atillah, A., Piou, C., 2015. Modélisation du risque acridien au Maroc: couplage des données de prospections acridiennes et des paramètres environnementaux (Rapport final de la composante aide à la lutte anti acridienne), Projet LDAS- Maroc. Centre Royal de Télédétection Spatiale, Rabat, Maroc.
- Symmons, P.M., Cressman, K., 2001. Biologie et comportement, in: *Directives sur le Criquet pèlerin numéro 1*, 1. FAO. Organisation des Nations Unies pour l'alimentation et l'agriculture, Rome. Italie.
- Tappan, G.G., Moore, D.G., Knausenberger, W., 1991. Monitoring grasshopper and locust habitats in Sahelian Africa using GIS and remote sensing technology. *Int. J. Geogr. Inf. Syst.* 5, 123–135. doi:10.1080/02693799108927836
- Tratalos, J.A., Cheke, R.A., 2006. Can NDVI GAC imagery be used to monitor desert locust breeding areas? *J. Arid Environ.* 64, 342–356. doi:10.1016/j.jaridenv.2005.05.004
- Veran, S., Simpson, S.J., Sword, G.A., Deveson, E., Piry, S., Hines, J.E., Berthier, K., 2015. Modeling spatiotemporal dynamics of outbreaking species: Influence of environment and migration in a locust. *Ecology* 96, 737–748. doi:10.1890/14-0183.1.sm
- Zha, Y., Ni, S., Gao, J., Liu, Z., 2008. A new spectral index for estimating the oriental migratory locust density. *Photogramm. Eng. Remote Sens.* 74, 619–624.