

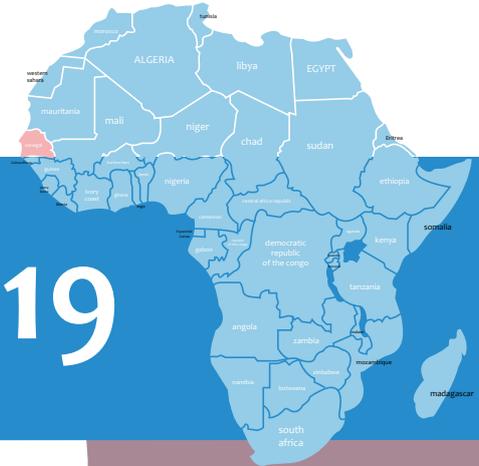


AKADEMIYA

L'expertise que nous avons. L'Afrique que nous voulons.

004

AKADEMIYA2063 - 17 août /24 août-2020



Bulletin Covid-19

Application de la télédétection et de l'intelligence artificielle pour la prévision de la production agricole en période de crise

Racine Ly Directeur, Management des Données, Produits et Technologies Numériques, AKADEMIYA2063 et Khadim Dia Chercheur Associé, AKADEMIYA2063

Dans un souci d'atténuer la propagation et l'impact du COVID-19, la plupart des pays ont, ou sont dans un processus de mise en place ou d'implémentation de mesures stratégiques. Ces dernières ont entraîné de lourdes conséquences dans différents secteurs de l'économie et exercé une pression importante sur les moyens de subsistance des communautés.

Introduction et contexte

L'une des plus grandes craintes serait que la COVID-19 et les mesures de santé publique perturbent la production agro-alimentaire, ce qui engendrerait une potentielle crise alimentaire dans de nombreuses régions du monde, notamment en Afrique.

Alors que sur le plan sanitaire les données officielles montrent que l'Afrique est moins impactée par la pandémie de la COVID-19 par rapport au reste du monde, sur le plan économique les impacts sont nombreux. Les économies

africaines sont largement dominées par un secteur informel et fortement dépendantes de la mobilité des populations. Dans le secteur agricole, des conséquences peu souhaitables sont entre autres: la rareté des intrants, la pénurie de travailleurs agricoles, un accès réduit aux marchés d'exportation et les problèmes de disponibilité de l'approvisionnement alimentaire, y compris ceux issus des importations. Une bonne planification est essentielle pour amortir les effets négatifs sur le secteur agricole et prévenir une crise alimentaire. L'évaluation précoce

Problème des données sur la production des cultures vivrières : pistes de solutions

Les effets perturbateurs de la Covid-19 sur les systèmes de production alimentaire sont réels : difficulté d'accès aux semences et aux engrais, circulation limitée des marchandises, baisse de la demande, pénurie de main-d'œuvre, en sont quelques exemples. Dans ce contexte, le défi ne consiste pas seulement à saisir l'ampleur et la complexité probables des perturbations, mais également à les identifier et à en assurer un suivi en temps réel. Contrairement à la propagation de la maladie elle-même, qui elle peut être suivie par le biais de tests et du traçage, il demeure difficile, même en temps normal, d'obtenir des informations précises sur les activités agricoles. L'introduction du confinement et des diverses mesures de restriction pour contrôler (ou limiter les effets de) la pandémie rend cette tâche encore plus ardue. Il n'existe aucun moyen de savoir si les agriculteurs ont accès aux intrants (en temps voulu et en quantité suffisante), s'ils ont été trop malades pour s'occuper de leurs fermes ou s'ils n'ont pu travailler que partiellement. Les réponses à ces questions ne pourront être obtenues qu'en fin de la saison en fonction de l'impact sur les quantités récoltées. A ce moment-là, il est déjà trop tard et il ne reste plus qu'à affronter la situation de crise.

L'absence totale d'informations sur les conditions culturales en temps de crise peut être surmontée en utilisant les technologies numériques modernes. Les données de télédétection permettent de suivre en temps réel l'évolution de la couverture végétale, les données météorologiques et bien d'autres paramètres liés aux activités agricoles. Les récents développements en matière d'apprentissage automatique et de modélisation informatique permettent de suivre et de projeter la production agricole à l'aide de ces données. Les avantages vont bien au-delà de la capacité à surmonter les obstacles à la collecte de données pendant les périodes de crise. Les nombreuses limites qui empêchent l'accès à des statistiques agricoles de bonne qualité peuvent également être surmontées grâce à ces technologies numériques, à savoir l'estimation des terres arables, des surfaces emblavées, des rendements ou la distribution géographique des quantités récoltées. Nos chercheurs utilisent ces technologies pour évaluer les changements dans les systèmes de production alimentaire durant la pandémie et ainsi fournissent des informations clés permettant de mieux lutter contre les effets de la COVID-19 au niveau des communautés locales.

Ousmane Badiane, Executive Chairperson

de la production alimentaire permettrait d'obtenir une meilleure vue d'ensemble sur la sécurité alimentaire et l'élaboration de bilans alimentaires nationaux plus précis et désagrégés.

Les pays africains publient régulièrement des prévisions de production agricole, lesquelles sont généralement évaluées au niveau national et peu désagrégées. La méthodologie communément utilisée repose sur des échantillons de données collectées dans des régions traditionnellement agricoles désignées. Ces données sont ensuite utilisées pour estimer la production au niveau régional et/ou national. Outre le fait qu'elles impliquent des collectes de données longues et coûteuses, ces techniques sont connues pour leur biais car ne tenant pas compte des nouvelles terres agricoles. Enfin, une telle méthodologie n'est pas adéquate pour la prise en compte des chocs climatiques devenus fréquents tels que les sécheresses, les inondations et les vagues de chaleur.

Les politiques restrictives visant à contrôler la pandémie ajoutent un niveau de complexité supplémentaire. Les dites restrictions compliquent l'accès aux données pour une évaluation de la croissance des cultures. En outre, le réseau complexe de perturbations résultant de la pandémie rend les techniques de prévision traditionnelles inadaptées aux chocs soudains. Notre équipe de chercheurs à AKADEMIYA2063 a développé un modèle pour contourner ces différents obstacles. Ce modèle utilise des données de télédétection et des techniques d'apprentissage automatique pour générer des cartes de prévision de la production au niveau du pixel. D'une part, les images satellites permettent d'observer les caractéristiques de la terre sur différentes longueurs d'ondes en sus de présenter d'intéressantes résolutions temporelles et spatiales. D'autre part, les techniques d'intelligence artificielle permettent un apprentissage automatique des propriétés intrinsèques des données. Le modèle a été utilisé pour prédire la production de la plupart des cultures vivrières de base en Afrique, à mesure que la pandémie de la Covid-19 évolue, et afin de permettre aux pays de mieux cibler les interventions visant à protéger les communautés vulnérables.

Résoudre les problèmes d'accès aux données grâce à la télédétection et aux cartes tierces

L'estimation de la production basée sur les données de télédétection peut être effectuée selon deux approches principales : (i) L'utilisation des données de télédétection comme entrées dans les modèles agro-météorologiques ou phyto-physiologiques ; et (ii) L'établissement d'une relation mathématique directe entre les données de télédétection et la production agricole (Huang & Han, 2014). La première approche est basée sur des descriptions « mécaniques » de la croissance, du développement et de la production des cultures simulées par des fonctions mathématiques. Ces méthodes ont donné de bons

résultats, mais ne permettent pas d'exploiter pleinement les bases de données. La deuxième approche repose sur des indicateurs dérivés de données de télédétection et sur leur corrélation avec la croissance et le rendement des cultures.

L'un des paramètres les plus utilisés pour caractériser les couvertures végétales et ainsi suivre la croissance des cultures, est l'indice NDVI¹. Il est utilisé en raison de sa corrélation avec l'ISF², la fraction de rayonnement photo-synthétiquement actif absorbée (*f*APAR) et la biomasse verte. La logique derrière le NDVI est que les feuilles des cultures absorbent le spectre rouge de la lumière visible comme source d'énergie pour les processus de photosynthèse et reflètent le spectre infrarouge. Ainsi, la différence de réflectance normalisée entre les spectre rouge et infrarouge permet d'évaluer l'état de santé d'une couverture végétale. De nombreuses études ont prédit le rendement des cultures en se basant sur les signaux NDVI (Rembold et al., 2013) ; l'évaluation du rendement du mil au Burkina Faso (Rasmussen, 1992) ; la prévision de la production de mil au Sénégal (Rasmussen, 1997).

Se fier uniquement sur le NDVI comme variable explicative pour estimer le rendement des cultures signifie que les autres signaux sont ignorés dans le calcul des prévisions. Notre

¹ Indice de végétation par différence normalisée
² Indice de Surface Foliaire

approche combine des produits de télédétection et de cartes tierces pour estimer la production. Plusieurs études menées au cours des années 1970 ont montré que le rendement final des cultures peut être lié aux indices thermiques (Leroux et al., 2018). Nous avons inclus dans le modèle une couche de température de surface des terres³, en plus des données de précipitations CHIRPS⁴. Enfin, le modèle utilise

les cartes des zones récoltées et de production de la base de données MAPSPAM⁵.

Le modèle a été appliqué pour prédire la production de mil au Sénégal en 2020. Les trois séries de cartes ci-dessous présentent les principales entrées utilisées. La première série (Figure 1) montre les valeurs moyennes NDVI, LST et de précipitations pendant la saison de croissance du mil en 2017. La deuxième série montre les superficies de mil récoltées en 2005, 2010 et 2017 (Figures 2). Chaque pixel contient une valeur qui représente la somme des superficies de mil récoltées sous ce pixel. La dernière série (Figures 3) présente la répartition de la production de mil pour les mêmes années. Chaque carte présente les valeurs par pixel de la production de mil pour l'année correspondante. Le modèle prend en entrée le NDVI, le LST, les précipitations et les zones récoltées pour prédire les valeurs de production. Les données utilisées et leurs caractéristiques techniques sont décrites au Tableau 1.

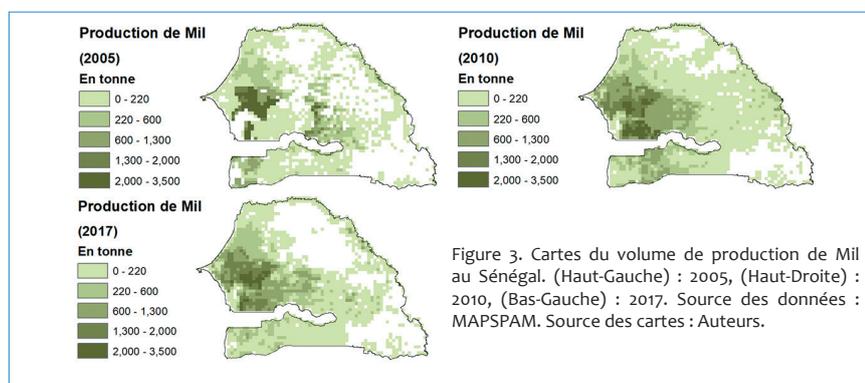
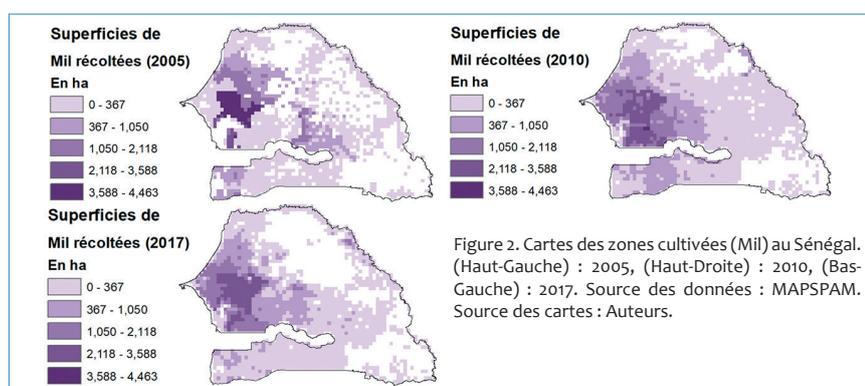
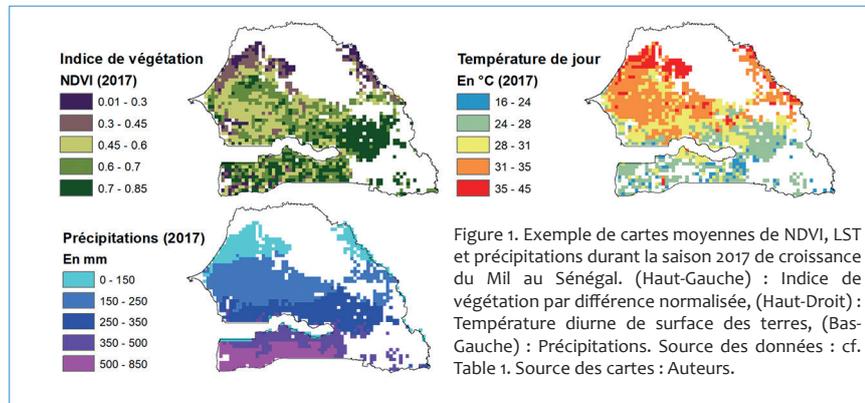


Table 1. Liste des caractéristiques de notre modèle avec leur source ou produit de télédétection et leurs résolutions temporelle et spatiale.

Caractéristiques du modèle	Source des données / produit de télédétection	Rés. temporelle	Rés. spatiale	Type
Indice de végétation par différence normalisée	MODIS ⁶ (MOD13A2.006)	16 jours	1 km	Entrée
Température diurne de surface des terres	MODIS (MOD11A2.006)	8 jours	1 km	Entrée
Précipitations	CHIRPS	Mensuelle	0.05°	Entrée
Zones récoltées	MapSpam	Annuelle	10 km	Entrée
Production	MapSpam	Annuelle	10 km	Sortie

³ LST - Land Surface Temperature
⁴ Climate Hazards group InfraRed Precipitation with Station data

⁵ <https://www.MAPSPAM.info/>
⁶ Spectroradiomètre à imagerie de résolution moyenne (MODIS)

conçues pour extraire le plus d'informations possible des données afin « d'apprendre » aux programmes informatiques à imiter leurs relations intrinsèques.

L'apprentissage automatique peut s'effectuer par trois techniques principales : l'apprentissage supervisé ; l'apprentissage non supervisé et l'apprentissage par renforcement. Le type de technique à utiliser dépend du type de données disponibles et du problème. Lorsque des variables explicatives et expliquées sont disponibles pour construire un modèle prédictif, l'apprentissage supervisé sera préconisé. Lorsque seules des variables explicatives sont disponibles, l'alternative sera d'identifier la structuration sous-jacente des données en les regroupant par caractéristiques similaires. Enfin, l'apprentissage par renforcement est une sorte de combinaison des deux techniques susmentionnées : les valeurs des variables de réponse sont générées selon un processus d'approximation successif. Dans notre cas, nous avons utilisé un algorithme d'apprentissage supervisé pour prédire la production en utilisant le NDVI, le LST, les précipitations et les superficies récoltées comme variables explicatives.

Cependant, afin de faire correspondre la résolution temporelle de la production (annuelle) et celle des variables explicatives, nous les avons annualisés en considérant leur moyenne durant la saison de croissance. Pour une culture, les résultats finaux correspondent à 15 raster moyens

(ou annuels) qui correspondent à 3 formats rasters (2005, 2010 et 2017) x 5 variables (4 explicatives et 1 expliquée).

Prédiction des variables explicatives

Les valeurs moyennes des variables explicatives ne sont pas disponibles au début de la saison de croissance pour l'année en cours. Nous les avons prédites en utilisant la méthode de régression par *Random Forest* (RF) construite sur la base des données historiques des 20 dernières années. La Figure 4.a. illustre un exemple de comparaison entre les données réelles et prédites sur la base de données de test isolées pour le NDVI ; et la Figure 4.b. montre la valeur moyenne prédite pour 2020 pour un pixel choisi aléatoirement. Les résultats des prévisions sont ensuite utilisés comme entrées pour le modèle de production.

Figure 4a. Valeurs moyennes prédites et observées de l'indice NDVI avec une série de données de tests pour le mil au Sénégal³.

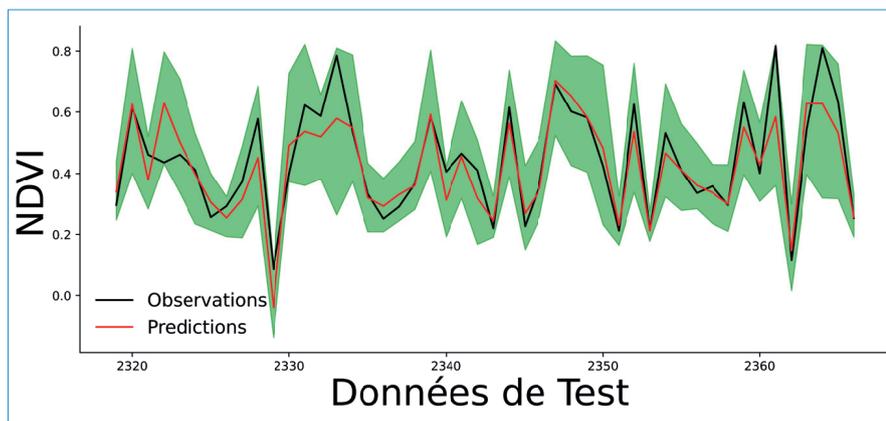
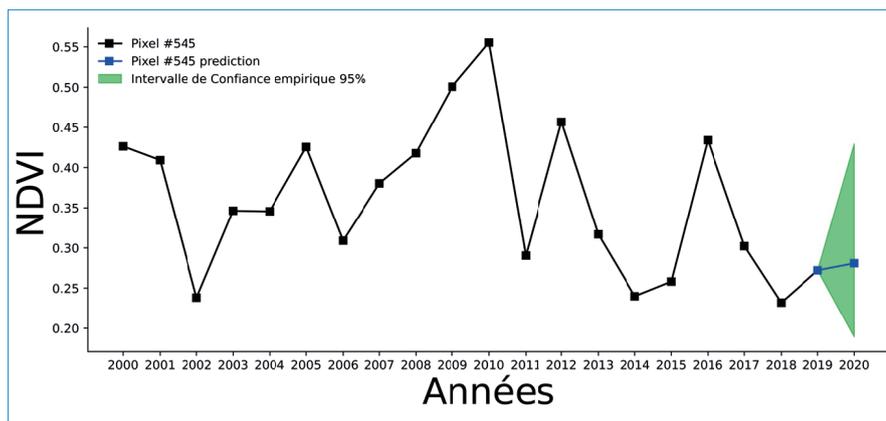


Figure 4b. Valeurs moyennes historiques et prédites (2020) de l'indice NDVI pour le mil au Sénégal³.



Architecture du modèle prédictif

Pour notre modèle, un réseau de neurones artificiels (RNA) été construit. Il s'agit d'un réseau d'unités appelé neurones ou perceptrons qui effectuent chacune une tâche spécifique. Un réseau se compose d'une couche d'entrées qui héberge toutes les données d'entrées, d'une couche de sortie qui fournit la prédiction et d'une ou de plusieurs couches cachées qui traitent les informations. Chaque couche de neurones est complétée par un neurone dit de *biais* dont le but est de simplifier

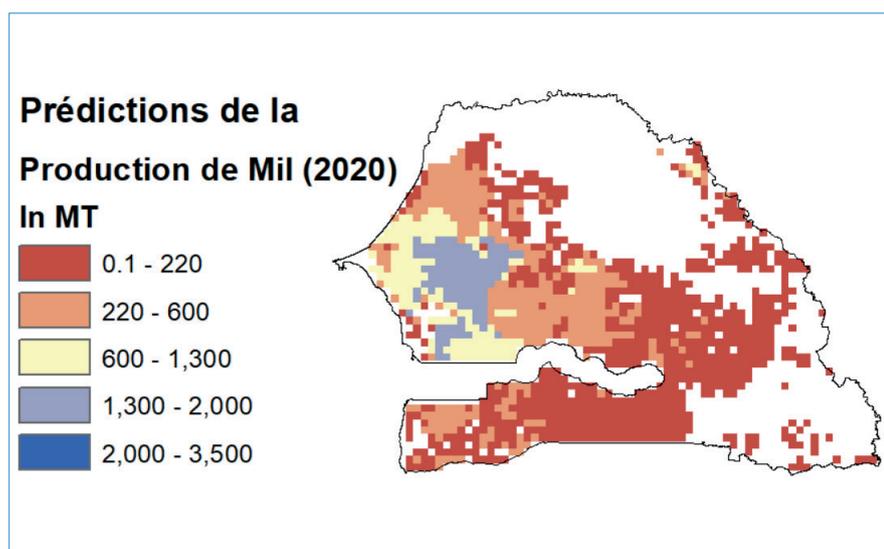
l'écriture algorithmique. Pour notre modèle de production, nous avons utilisé un réseau de neurones à quatre couches : une couche d'entrées avec 4 neurones qui correspond aux 4 variables explicatives – deux couches cachées comprenant 50 neurones chacune – une couche de sortie avec un neurone pour prédire la valeur de production d'un pixel.

Pour notre modèle de production, nous avons utilisé la fonction-objectif d'erreur quadratique moyenne durant l'apprentissage ainsi que la méthode d'optimisation ADAM (*Adaptive Moment Estimation*) (Kingma & Ba, 2014). Pour l'évaluation de la précision, nous avons évalué le modèle sur une série de données de tests qui correspond à 10 % de la base de données totale. Une erreur quadratique moyenne de 0,080 a été obtenue pour notre exemple du mil au Sénégal.

Carte de prévision de la production : Exemple du mil au Sénégal

La pandémie de la Covid-19 rend difficile la collecte de données terrain – pour évaluer la production de cultures vivrières – vu les restrictions et les risques potentiels pour la santé. En période de crise comme celle de la Covid-19, l'évaluation de la localisation des bassins de production est un atout précieux. Les pays produisent depuis longtemps de telles informations au niveau national, mais pour les crises « sensibles à la géographie » comme la Covid-19, la désagrégation est un élément clé pour atténuer les effets sur les communautés. Par conséquent, la production d'une telle carte de la production de cultures vivrières permettrait d'orienter davantage les réponses, en termes de politiques agricoles, au niveau communautaire.

Figure 5. Production nationale estimée de mil au Sénégal avec notre modèle pour la saison 2020. Les unités de production sont en tonnes métriques et les pixels sont de taille 10 km x 10 km.



Les données d'entrée pour l'année en cours (2020) ont été produites en utilisant le modèle RF. Elles correspondent à la moyenne estimée du NDVI, du LST et des données sur les précipitations pendant la saison de croissance du mil au Sénégal. Ces valeurs ont été introduites

dans le modèle pour prédire la production de mil. La carte illustrée à la Figure 5 est le résultat de notre modèle et représente la production de mil pour la saison en cours de la République du Sénégal. Chaque pixel contient une valeur de production prédite qui correspond à la quantité de mil attendue à cette position géographique.

Conclusion

Le modèle que nous avons présenté dans ce bulletin utilise des produits de télédétection, des cartes tierces et des techniques d'apprentissage automatique pour prédire la production vivrière pour une culture et un pays spécifique. Alors que la production vivrière a longtemps été estimée au niveau national ou administratif, le principal avantage de notre modèle est sa capacité à prédire la production au niveau du pixel. Cette méthode ouvre de nouvelles perspectives en matière d'évaluation de la sécurité alimentaire, du bilan alimentaire et de la conception de politiques ciblées, en particulier dans le contexte actuel marqué par la pandémie en cours.

L'exemple de la culture du mil au Sénégal a été choisi pour illustrer succinctement la méthodologie. Dans les semaines et les mois à venir, d'autres pays seront couverts à travers le continent et les prévisions de production vivrière de base seront disponibles dans nos prochains bulletins et notes.

