

HAI408I

Rapport de stage de recherche scientifique

**Analyse des méthodes de calibrations
entre deux satellites d'observation de la Terre de nuit**

COLIBEAU Jean-Matthieu

L2 informatique
Faculté des Sciences
Université de Montpellier
LIRMM



Stage encadré par M. Robin JARRY

Du 27 mai au 28 juin 2024



Table des matières

Table des matières	2
1 Introduction	3
1.1 Contexte de la recherche	3
1.2 Présentation de la structure d'accueil	3
1.3 Contexte du stage	3
1.4 Objectifs du stage	4
2 Méthodologie et matériels	5
2.1 Données	5
2.2 Méthode de comparaison	5
2.2.1 Sélection de zones à étudier et hypothèses	5
2.2.2 Récupérer les données ILN	6
2.2.3 Développer un outil si nécessaire	6
2.2.4 Mener une étude comparative	7
2.2.5 Conclure	7
2.3 Outils	8
2.4 Algorithme des K-moyennes	8
2.4.1 Accès à une machine virtuelle	8
2.4.2 Explication de la bibliothèque Matplotlib	9
2.4.3 Explication de la bibliothèque folium	9
2.5 Planning	9
3 Résultat	9
3.0.1 Ma contribution : visualisation sur carte	9
3.1 Résultats communs	12
3.2 Bilan des compétences acquises	15
3.3 Bilan des problèmes rencontrés	15
4 Discussion	16
4.1 Analyse critique des résultats	16
4.2 Apports pour l'entreprise et bilan personnel	16
5 Conclusion	17
6 Bibliographie	17

1 Introduction

1.1 Contexte de la recherche

Les **intensités lumineuse nocturne** (ILN) obtenue à l'aide des satellites d'observation de la Terre sont utilisées pour étudier le développement des activités humaines (elvidge et al. 1997). Ces observations permettent d'analyser divers aspects tels que l'évolution de la consommation d'électricité, les activités socioéconomiques, la pollution lumineuse et les écosystèmes urbains. Historiquement, la majorité des études se sont appuyées sur les données fournies par le *Defense Meteorological Satellite Program-Operational Linescan System* (DMSP-OLS). C'est un système de capteurs attachés à de satellites lancés par les États-Unis en 1992. Ces satellites ont une résolution spatiale de 30 secondes d'arc, ce qui correspond à environ 1 km^2 à l'équateur (chaque pixel de l'image correspond à un carré de 1 km^2). Ils capturent la radiance qui est ensuite ramenée ensuite à une valeur entre 0 et 63. Cependant, le dernier satellite a cessé son activité en 2013.

La nouvelle génération d'observations nocturnes repose désormais sur le *Visible Infrared Imaging Radiometer Suite* (VIIRS), un système de capteur monté sur les satellites météorologiques Suomi NPP lancé en 2012. Les capteurs VIIRS offrent une résolution spatiale améliorée de 15 secondes d'arc , ce qui correspond à environ 500m^2 à l'équateur. Ils mesurent bien mieux la radiance, ce qui la rend directement exploitable. Ils fournissent ainsi des données plus précises, notamment en ce qui concerne la localisation des lumières urbaines. Malgré ses meilleures performances, le capteur VIIRS, en raison de son temps de fonctionnement encore relativement court, n'est pas encore utilisable pour des études d'évolutions sur un long terme.

Afin de pouvoir analyser l'évolution de l'ILN sur une longue période temporelle allant jusqu'à aujourd'hui, on ne peut choisir l'un ou l'autre, il nous faut utiliser les deux. Pour cela, il faut trouver une méthode afin d'harmoniser ces deux jeux très différents. Ici, nous allons nous intéresser à deux méthodes en particulier : celle de [Li et al., 2020] et celle de [Chen et al. 2021]. La première utilise une régression entre les données VIIRS simulant ainsi DMSP-OLS avec une fonction sinusoïdale. La deuxième augmente artificiellement les données DMSP-OLS simulant ainsi VIIRS en utilisant un auto-encodeur.

1.2 Présentation de la structure d'accueil

Le stage se déroule au sein du Laboratoire d'Informatique, de Robotique et de Microélectronique de Montpellier (LIRMM), qui est une unité de recherche dépendant de l'Université de Montpellier et du CNRS.

1.3 Contexte du stage

Notre stage a été effectué dans l'équipe "Image & Interaction" (ICAR) du LIRMM qui se concentre sur quatre grands axes de recherche : Analyse & traitement, Sécurité Multimédia,

Modélisation & Visualisation, Intelligence Artificielle pour les données visuelles. Il s'inscrit dans le travail de thèse de notre tuteur visant à quantifier l'impact socioéconomique local des aires marines protégées.

En effet, il semble que les aires marines protégées, au-delà de leur impact positif sur la biodiversité marine, ont aussi un impact sur les populations côtières qui peuvent bénéficier d'une pêche artisanale et assurer une rente régulière.

Nous devons pouvoir mesurer le niveau de pauvreté avant et après cette création. Cependant, il est impossible de récupérer ces informations sur place. Nous devons donc nous tourner vers des options à distance, comme les satellites captant les lumières nocturnes.

Il existe des modèles permettant d'estimer la richesse de lieux à partir de leur ILN, mais ces derniers ne prennent en compte qu'une valeur de l'ILN à un instant donnée, et sont donc inadaptés à estimer une *variation* de valeur. La thèse de notre tuteur vise donc à répondre à ce problème en prenant en compte ce nouveau paramètre : le temps.

Ce modèle spatio-temporel permettrait d'observer les évolutions à long terme et d'aider à conclure sur l'efficacité de cette zone de pêche réservée.

1.4 Objectifs du stage

Notre objectif est de déterminer lequel des jeux de données, entre DMSP-OLS simulé et VIIRS simulé, est le meilleur pour différentes zones et sous quelles conditions spécifiques. Afin de pouvoir décider lequel choisir pour le modèle spatio-temporel. Ce travail combine la création d'outils d'analyse et de visualisation ainsi qu'une comparaison des données d'ILN issues des capteurs VIIRS et DMSP-OLS sur des zones d'études spécifiques.

Notre groupe était composé de trois personnes, Pierre PREVOST–AVENEL, Reda AM-MARI et moi. Nous avons travaillé en autonomie tout en s'entraînant et, nous nous sommes répartis les diverses tâches. J'ai été chargé de la création des outils. J'ai codé un programme générant deux types de graphiques. Le premier montre l'évolution du nombre et de la somme des pixels allumés en fonction des années. Le deuxième est un histogramme de la répartition des ILN soit pour toutes, soit pour une année. J'ai ensuite créé une visualisation des résultats sur une carte, en utilisant notamment OpenStreetMap, où les résultats des analyses et les images des ILN peuvent être superposés. Tout ceci a pour objectif de faciliter la comparaison entre DMSP-OLS simulé et VIIRS simulé.

En parallèle, les autres membres ont utilisé ces outils pour identifier le jeu de données le plus pertinent à utiliser entre DMSP-OLS simulé et VIIRS simulé.

À partir de plusieurs zones sélectionnées, nous avons mené une comparaison visuelle entre les données DMSP-OLS simulé et VIIRS simulé selon des critères visuels objectifs tels que la régularité de l'évolution ou sa cohérence avec le contexte socioéconomique locale.

2 Méthodologie et matériels

2.1 Données

Nous appellerons DMSP, le jeu de donnée produit par la méthode de [Li et al., 2020]. Ce jeu est composé des vraies données DMSP-OLS (1992–2013) et de données VIIRS (2014–aujourd’hui) simulé en DMSP-OLS.

Nous appellerons VIIRS, le jeu de donnée produit par la méthode de [Chen et al. 2021]. Ce jeu est composé des vraies données VIIRS (après mi-2012- aujourd’hui) et de données DMSP-OLS (avant mi-2012 - 2011) simulé en VIIRS.

2.2 Méthode de comparaison

2.2.1 Sélection de zones à étudier et hypothèses

Nous nous sommes intéressés à plusieurs zones aux contextes socioéconomiques différents

- **Zones urbaines développées** : Ces zones ont déjà connu un développement, ce qui signifie qu'il n'y aura probablement pas de changements majeurs à l'avenir. Leur évolution est donc relativement constante et donc prévisible. Ce qui nous permet d'avoir une idée d'à quoi ressembler les données simulées

Paris et Londres sont des exemples de villes bien développées sur les plans économique, social et infrastructurel. Leur centre urbain est très dense et prospère, ce qui se traduit par une forte luminosité. Cette luminosité dépasse ce que DMSP est capable de mesurer. Notre hypothèse est donc que VIIRS serait plus adapté que DMSP pour étudier les zones urbaines développées, grâce à sa plage de valeurs mesurables plus grande.

- **Zones urbaines en développement** : Ce sont des villes qui connaissent des croissances de leurs activités économiques, démographique ou/et industrielles, etc. Nous nous attendons à voir une croissance globale des ILN en réponse à cette évolution. Nous avons donc sélectionné Abidjan (Côte d'Ivoire), Addis-Abeba (Éthiopie) et Delhi (Inde) qui sont des villes en croissance, avec des défis et des opportunités spécifiques.

Pour les mêmes raisons que les zones développées, nous supposons que VIIRS y sera meilleur que DMSP. Mais puisque la périphérie n'est pas aussi lumineuse que le cœur, DMSP pourrait y être pertinent.

- **Zones rurales** : Morondava (Madagascar), Cuamba (Mozambique) et Ilaka (Madagascar) sont des régions moins urbanisées, caractérisées par une économie et des infrastructures limitées.

Ce sont des lieux qui produisent peu de lumière. DMSP ne va donc pas saturer, il pourrait donc tout à fait fonctionner. VIIRS est plus précis et plus sensible à la lumière, donc il y est peut-être plus intéressant que DMSP.

- **Zones côtières** L'eau joue un rôle essentiel en tant que réflecteur de lumière, ce qui peut avoir un impact néfaste sur les données. En effet, l'eau ne génère pas de lumière

par elle-même, mais elle amplifie l'émission lumineuse provenant des sources à proximité, phénomène que l'on appelle l'éblouissement. Il est donc préférable de corriger ces effets.

Entre les données du DMPS et du VIIRS, nous cherchons à déterminer lequel est moins affecté par ce phénomène. Bien que ce ne soit pas un critère essentiel puisqu'il est possible de traiter l'image pour éliminer ce défaut. Pour ce faire, nous examinons différentes zones côtières. Morondava a été choisie pour étudier les effets dans une zone à faible indice de luminosité nocturne (ILN), tandis qu'Abidjan a été sélectionnée pour observer les effets dans une zone à ILN élevé.

Notre hypothèse est que VIIRS sera plus efficace, car il est moins sensible à l'éblouissement.

— **Zones de conflit :**

Il est logique qu'un conflit violent dans une zone entraîne des dégâts matériels et un appauvrissement de la région. En conséquence, cela pourrait provoquer une baisse des ILN. L'objectif est de vérifier si les données simulées par les satellites DMPS et VIIRS reflètent ce phénomène. Si ce n'était pas le cas, cela remettrait en question la fiabilité des données. Abidjan et la bande de Gaza sont des zones avec des tensions politiques, sociales ou géopolitiques ont conduit à des conflits. Par exemple : Abidjan a été touchée par une guerre civile des années 90 à 2011. Donc, nous devrions y observer soit une baisse, soit constance des ILN globales.

Notre hypothèse est que ces conflits seraient observables dans DMSP et VIIRS.

2.2.2 Récupérer les données ILN

Nous disposons d'une base de code nous permettant de télécharger l'imagerie optique des zones d'étude, à partir d'un rectangle que nous dessinons sur une carte interactive

Nous pouvons choisir une zone rectangulaire d'une taille maximale de $8\ 000 \times 8\ 000$ pixels, à cause des limitations de l'API *Google Earth engine*.

La requête du client vers le serveur consiste à récupérer des images satellites du lieu pour chaque année entre 2000 et 2020 incluses, puis les envoie sur le bucket Google du projet. Ce bucket est un service proposé par Google pour un projet Google Cloud qui est un serveur de stockage de données, agissant comme "tampon" entre les serveurs de données et le client. Puis, nous pouvons récupérer les ILN à partir des jeux DMSP et VIIRS préalablement disponibles sur la machine

Nous obtenons alors un dossier contenant une photographie aérienne et les ILN DMSP et VIIRS de la zone pour chaque année de 2000 à 2020.

2.2.3 Développer un outil si nécessaire

Notre tuteur nous avait fourni l'algorithme des k-moyennes(expliqué en détail ensuite). Il permet de segmenter une zone en plusieurs classes évoluant de manière similaire. Cela permet entre autres de trouver automatiquement par exemple la périphérie ou le centre urbain et

d'ensuite en étudier l'évolution.

Mais nous voulions aussi avoir une évolution plus globale des ILN afin de pouvoir étudier les effets d'événements importants globaux (conflit, pandémie, changement d'éclairage, etc). Pour cela, nous avons programmé un graphique de l'évolution du nombre et de la somme des pixels allumé. Cela nous a aussi permis de voir si les données simulées suivent les tendances des vraies. Et de vérifier qu'il y ait bien une corrélation entre le nombre et la somme de pixels allumé, c'est-à-dire que si la somme de pixels allumés reste relativement constante, alors le nombre doit l'être aussi.

Ainsi qu'un histogramme de la répartition des ILN afin de voir dans DMSP le nombre de pixels saturés, c'est-à-dire dépassant une valeur de 63.

Enfin, nous avons besoin de replacer les classes fournies par l'algorithme des k-moyennes dans l'espace et de visualiser les ILN afin d'expliquer pourquoi certaines zones ou points évoluent d'une certaine manière et pourquoi certains sont plus lumineux que d'autres. De plus, il y avait un besoin de centraliser toutes les analyses afin de les exécuter rapidement et automatiquement. Pour répondre à ces besoins, nous avons créé un affichage sur une carte des données d'analyses et des ILN tout en générant les données qui pourraient manquer. Nous nous sommes donc demandés quelles cartes utiliser ? Nous avons tout de suite pensé à *OpenStreetMap*, car elle permet de voir les route et les points d'intérêt et elle est libre d'accès et régulièrement mise à jour. Nous avons aussi choisi de prendre la carte française d'*OpenStreetMap*. Afin de mieux pouvoir observer les dénivélés dans une carte en deux dimensions, nous avons rajouté une carte topologique de *OpenTopoMap*. Nous avons pris la carte optique de *ESRI World Imagery* qui permet de mettre en valeur les zones urbaines, les zones rurales et les cours d'eau.

Le développement de ses outils était la tâche qui m'a été confié par notre groupe.

2.2.4 Mener une étude comparative

Nous avons d'abord comparé les données simulées de DMSP-OLS avec les données réelles de DMSP-OLS, ainsi que les données simulées de VIIRS avec les données réelles de VIIRS. L'objectif était de vérifier si les imitateurs reproduisaient les tendances observées dans les données originales. Ensuite, nous avons opposé les vraies données DMSP-OLS avec les données simulées de VIIRS, et vice versa, afin d'observer non seulement la similarité des tendances, mais aussi la visibilité des événements impactant les ILN. Il nous est arrivé de trouver des évolutions qui allaient à l'encontre de notre intuition et nos hypothèses et notre recherche de départ et donc nous avons fait approfondi nos recherches pour expliquer ces phénomènes.

2.2.5 Conclure

Après avoir étudié une zone, nous concluons sur quelle satellite permet d'avoir le maximum d'information avec la meilleure qualité. Ensuite, nous comparons avec d'autres zones du même type, si elles arrivent toutes à la même conclusion, alors, nous pouvons affirmer que DMSP ou

VIIRS est à utiliser dans ce cas. Si nous trouvons des conclusions divergentes, nous essayons de trouver des singularités dans les zones qui explique le phénomène.

2.3 Outils

2.4 Algorithme des K-moyennes

L’algorithme des **K-moyennes** permet pour un nombre de *clusters* donné (ou classe) de regrouper les éléments qui se ressemblent. Si nous appliquions cet algorithme sur une image coloré, nous obtiendrions, par exemple, par exemple, une classe contenant tous les pixels proches du rouge, une autre contenant tous ceux proches du bleu et une autre contenant ceux proches du vert.

L’initialisation de l’algorithme se fait en choisissant un centroïde pour chaque *cluster*.

Il peut être choisi de manière aléatoire, soit en utilisant l’algorithme des *K-means++*. L’algorithme des *K-means++* choisit un premier centroïde au hasard. Ensuite, chaque centre de cluster suivant est choisi parmi les points restants, avec une probabilité proportionnelle au carré de la distance entre ce point et le centre de cluster le plus proche.

Ici, nous utilisons *K-means++* car il permet de sélectionner des points centraux éloignés et donc de répartir les points afin d’avoir une meilleure couverture de l’ensemble. Ceci nous rapproche d’une solution optimale des clusters, réduisant le nombre d’étapes à effectuer.

De manière à avoir les centroïdes les plus optimales, nous répétons l’algorithme un certain nombre de fois. Plus le nombre d’itérations maximum est grand, moins l’aléatoire a d’impact. Ici ce nombre est 10, c’était le chiffre par défaut dans le code et les résultats obtenus étaient toujours très proche les uns des autres dont nous l’avons gardé.

Chaque élément est ensuite envoyé dans le *cluster* du centroïde dont il est le plus proche, c’est-à-dire dont la distance est la plus faible. Il existe diverse façon de calculer la distance, nous utilisons la version des K-moyennes classiques utilisant la distance euclidienne. Une fois que tous les éléments ont été associés à un centroïde, les centroïdes sont recalculées en prenant la moyenne des points dans chaque *cluster*. On réitère ces deux dernières étapes jusqu’à qu’il n’y ait que peu ou aucune variation à chaque itération. Cette stabilité peut mettre longtemps avant d’être trouvé sans pour autant grandement modifier les résultats, il est donc possible de mettre un nombre limite d’itérations.

2.4.1 Accès à une machine virtuelle

Afin de pouvoir exécuter du code lorsque nos ordinateurs n’étaient pas assez puissants, nous avions accès à une machine de l’équipe ICAR par le protocole Secure Shell (SSH). Ce protocole permet de partager de manière sécurisée des données entre ordinateurs. Nous avons dû mettre en place notre propre environnement de programmation python en installant toutes les bibliothèques dont avait besoin en évitant de modifier celui de la machine. Cet accès nous a été fourni par notre encadrant.

2.4.2 Explication de la bibliothèque Matplotlib

Matplotlib est une bibliothèque python très populaire qui permet de créer des visualisations à partir de données tel que des graphiques, des images, des animations. Sa popularité couplet à sa documentation rend son utilisation simple. Elle a été la bibliothèque le plus utilisé dans ce stage. Notamment pour faire les graphiques et les images.

2.4.3 Explication de la bibliothèque folium

Folium est une bibliothèque python qui permet de faciliter la manipulation de visualisations vectorielles, raster ou HTML riches en informations. Elle s'appuie sur la bibliothèque JavaScript leaflet qui facilite la création de cartes. Cette bibliothèque est principalement utilisée pour afficher des cartes interactives sur des pages HTML. Cette bibliothèque nous a permis de construire la structure de la page HMLT de la carte. C'est elle qui gère les déplacements de l'utilisateur dans la carte.

2.5 Planning



FIGURE 1 – Diagramme de gant sur les 5 semaines de stage

3 Résultat

3.0.1 Ma contribution : visualisation sur carte

Le programme prend en entrée une liste de zone à étudier, une année de départ et une année de fin. Lors du lancement du programme, il va pré-générer dont la carte a besoin en appelant les autres programmes d'analyses sur l'intervalle temporel donné et en transformant les ILN en images. Enfin, il génère la carte grâce à folium. Il ne reste plus qu'à ouvrir le fichier HTML de la carte. L'objectif de cette carte est de centraliser toutes les analyses et d'ensuite pouvoir les placer dans l'espace.



FIGURE 2 – Aperçu global de la carte

La carte propose deux *sliders*, l'un sélectionne l'année de la donnée que l'on veut ajouter en couche et l'autre définit son opacité. Le *slider* permettent de rapidement faire varier les années afin de pouvoir observer leur évolution visuellement. Puis, nous pouvons ensuite décider de quels jeux de données utiliser. Et enfin permet de choisir entre un affichage des analyses de l'algorithme des K-moyennes et des ILN. Le choix du nombre de clusters pour les K-moyennes est par défaut cinq et n'est paramétrable qu'à la création de la carte. Chaque zone est sélectionnable dans le gestionnaire de couche de folium, elles sont délimitées par un rectangle noir et trois marqueurs, permettant d'afficher les courbes de l'algorithme des k-moyennes, le graphique des pixels allumés et l'histogramme. Le gestionnaire permet aussi de choisir le fond de la carte.

Nous allons utiliser paris comme exemple pour regarder ce que la carte permet de faire.

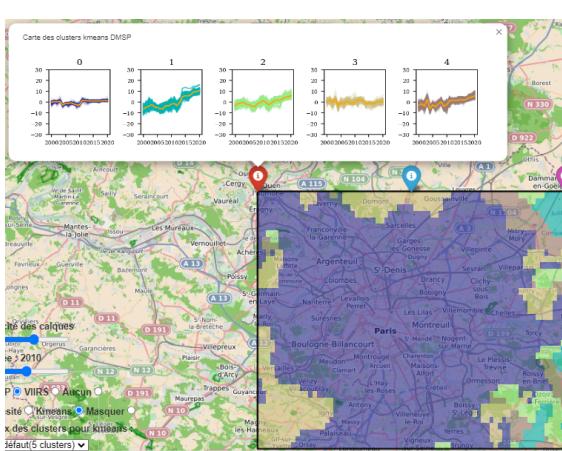


FIGURE 3 – DMSP : K-moyennes intégré à la carte

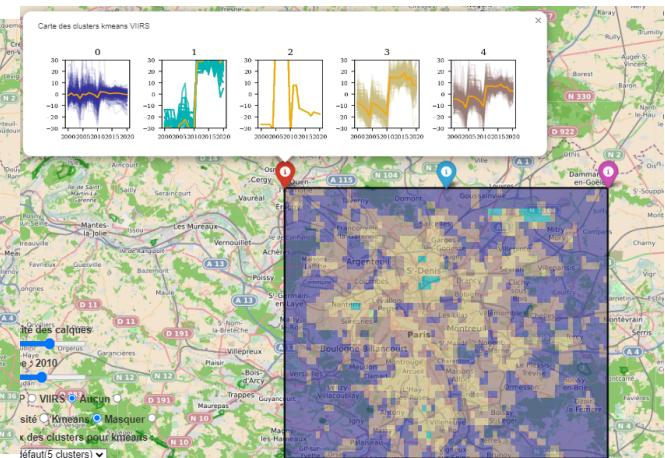


FIGURE 4 – VIIRS : K-moyennes intégré à la carte

La figure 3 est une superposition de l'analyse de l'algorithme des k-moyennes de DMSP, nous décernons en bleu foncé le centre urbain de paris, les autres couleurs représente une partie de la périphérie et c'est une partie dont les ILN ont cru.

La figure 4 est une superposition de l'analyse de l'algorithme des k-moyennes de VIIRS, nous décernons en bleu foncé la périphérie de paris et nous observons un centre urbain détaillé puisqu'il est composé de quatre classes.

Nous voyons une augmentation soudaine et inexpliquée en 2011 pour VIIRS. Ce résultat est confirmé par la figure 6, il s'agit d'une intégration dans l'un des marqueurs du résultat du programme de génération de graphique montrant l'évolution de la somme et du nombre de pixels allumé. Un pixel est considéré comme allumé s'il dépasse un certain palier d'intensité lumineuse. Ce palier est à bien choisir, afin de gommer le bruit causé par des pixels qui n'influent pas sur la tendance de la courbe, c'est-à-dire qu'ils ne sont pas suffisamment en nombre ou intense ou variant pour influencer la courbe. Ici pour l'exemple le pallier est égale à 0, mais le palier intéressant pour Pari est 30. DMSP et VIIRS n'ont pas la même échelle, nous avons donc normalisé les courbes.

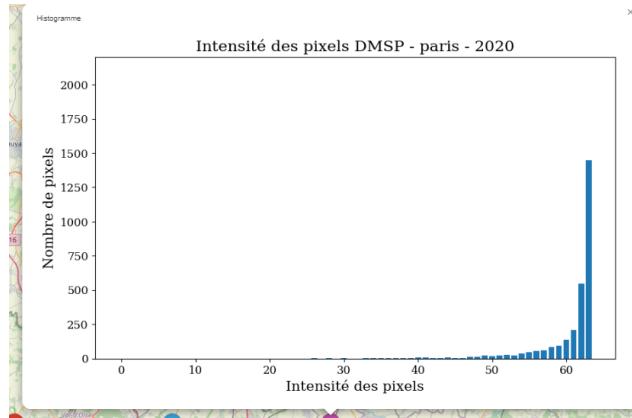


FIGURE 5 – Enter Caption

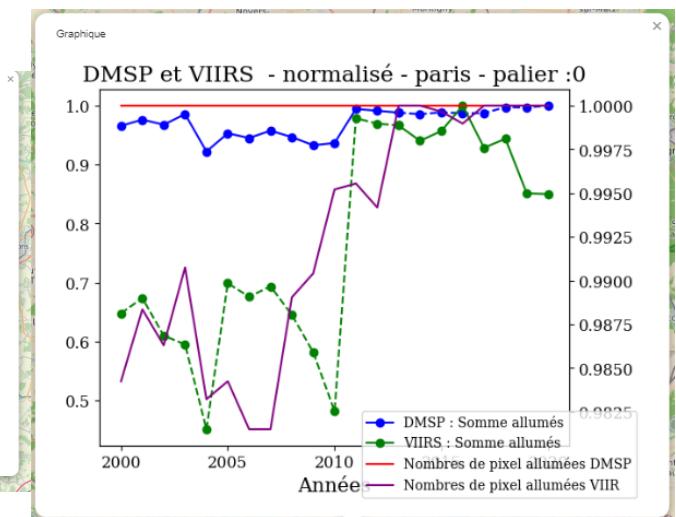


FIGURE 6 – somme et du nombre de pixels allumé

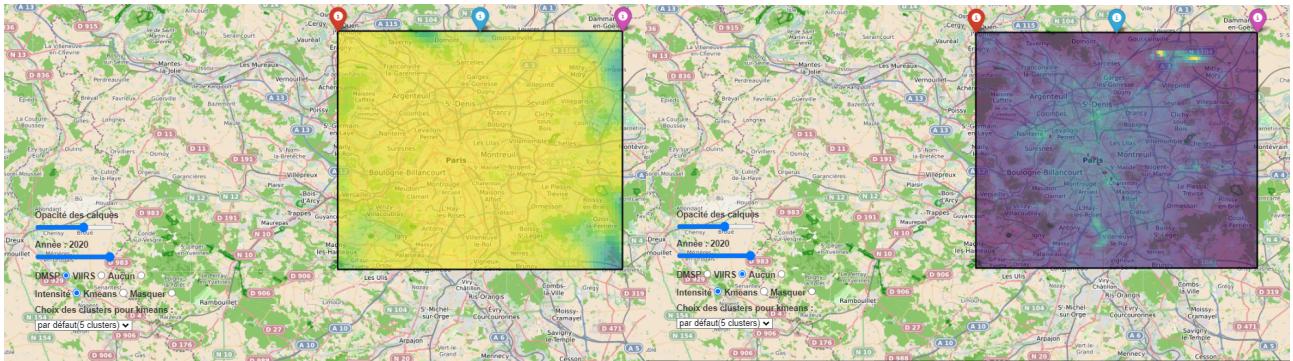


FIGURE 7 – DMSP : Intensités lumineuses nocturnes en 2020
FIGURE 8 – VIIRS : Intensités lumineuses nocturnes en 2020

Les figures 7 et 8 permettent de voir les ILN placés sur Paris en 2020, donc ce sont les vraies données des capteurs VIIRS et des données DMSP-OLS simulées. DMSP nous indique que paris est uniformément lumineux et de manière très forte. L'histogramme de la figure 5 nous indique les pixels de DMSP en 2020 à Paris sont majoritairement saturés, ce qui explique ce résultat, mais aussi explicite le gros défaut de DMSP.

VIIRS révèle une grande subtilité. Les pixels situés en haut à droite sont nettement plus lumineux que le reste, et la carte indique qu'il s'agit de l'aéroport Charles de Gaulle. Par ailleurs, nous identifions des pixels inactifs au milieu de Paris. En zoomant sur la carte, nous constaterions qu'ils correspondent à la Seine, bien que cela ne soit pas immédiatement évident. C'est pourquoi un autre fond de carte est intéressant : la carte de photographies aériennes d'*ESRI World Imagery* sur la figure 9 nous permet d'observer plus clairement le fleuve et de constater que tous les pixels lumineux se trouvent dans une zone urbaine.

3.1 Résultats communs

Nous avons réalisé un tableau comparatif entre DMSP et VIIRS à partir de critère objectif que nous avons défini avec notre tuteur. Le tableau comprenant toutes les zones que nous avons sélectionnées, il est assez grand et riche en information. J'ai décidé de garder seulement trois représentants : Paris pour les zones urbaines développées, Morondava pour les zones rurales et Abidjan pour les zones urbaines non développées (des analyses plus complètes et avec d'autre zone sont retrouvables dans notre rapport commun).

En effet, nous constatons que VIIRS est toujours précis dans le cœur urbain, mais il ne l'est jamais pour les périphéries, bien que cela soit normal pour Morondava, qui n'a pas vraiment de périphérie.

En revanche, DMSP semble plus équilibré que VIIRS : il est capable de détailler à la fois le cœur urbain et les zones rurales en développement. De plus, il capture la périphérie des villes développées et s'avère plus performant que VIIRS pour la périphérie d'Abidjan, même si les détails restent limités.

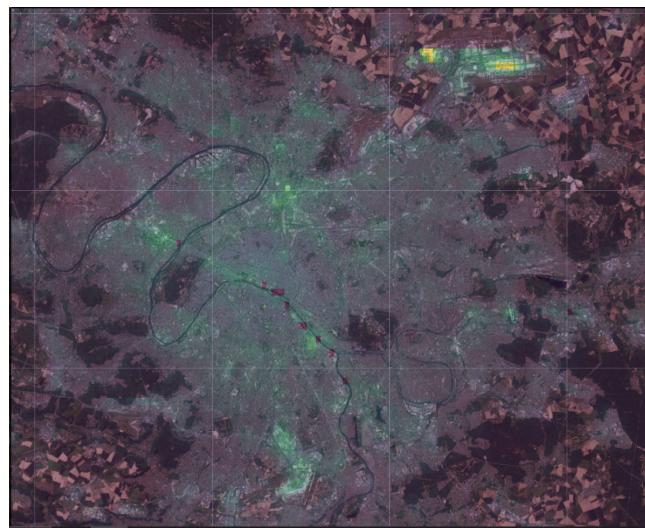


FIGURE 9 – VIIRS : ILN sur fond d'une carte de ESRI World Imagery

VIIRS et DMSP subissent l'éblouissement de l'eau pour les villes. DMSP le subit aussi à Morondava mais pas VIIRS qui est moins sensible à cet effet-là.

Nous avons observé une baisse de luminosité inexplicable à Paris entre 2004 et 2010, elle a été accompagnée d'un pic lumineux en 2011. Cette année-là, nous avons observé des données VIIRS simulées à Paris d'une qualité semblable aux vrais alors que ce n'était pas le cas avant. Elle s'observe à l'aide de l'affichage des ILN mais aussi par le graphique de la somme et nombre de pixels allumés. Nous cherchons d'en chercher la raison en essayant de trouver des indices dans d'autres zones touchés comme Morondava mais aussi d'autres que nous avons étudiées. Elle pourrait provenir d'un défaut matériel de DMSP-OLS et expliquer peut-être la mauvaise qualité des données VIIRS simulées.

L'ILN maximum et minimum ainsi que l'évolution globale des ILN ont été déterminés à l'aide du graphique des pixels allumés.

TABLE 1 – Tableau des résultats

	Paris		Morondava		Abidjan	
	DMSP	VIIIRS	DMSP	VIIIRS	DMSP	VIIIRS
Zones détaillées dans le cœur urbain	Non	Oui	Oui	Oui	Oui	Oui
Zones détaillées en périphérie	Oui	Non	Non	Non	Non	Non
Eblouissement eau		Oui, jusqu'en 2018	Oui	Non		Oui, jusqu'en 2018
Baisse de luminosité inexplicable entre 2004 et 2010	Oui, cœur urbain		Oui		Non	
Pic d'intensité lumineuse en 2011 ?	Oui		Non		Non	
Année d'intensité lumineuse minimum	2004 réel 2004	2020 réel 2004	2008 réel 2008	2015 réel 2008	2003 réel 2003	2013 réel 2010
Année d'intensité lumineuse maximum	2011 réel 2020	2016 réel 2016	2004 réel 2019	2020 réel 2020	2013 réel 2019	2020 réel 2020
Croissance / Décroissance / Stable	Stable	Décroissance	Stable	Croissance	Croissance	Croissance

3.2 Bilan des compétences acquises

Analyse et résolution de problème		
★ Savoir mettre en oeuvre une démarche expérimentale	■ ■ ■	Application
★ Développer des outils informatiques	■ ■ ■	Application
Recherche et traitement de l'information		
★ Evaluer la pertinence de l'information trouvée, l'ordonner, la hiérarchiser, la synthétiser	■ ■ ■	Notion
★ Choisir les outils et supports adaptés	■ ■ ■	Application
Niveau relationnel		
★ Coopérer pour arriver à des objectifs	■ ■ ■	Application
Communication		
★ Rédiger de manière scientifique	■ ■ ■	Application

3.3 Bilan des problèmes rencontrés

Pour appliquer l'algorithme des k-moyennes, il faut charger dans la mémoire vive les 21 images. Si les images sont trop lourdes, cela peut faire planter le programme. Pour corriger ceci, nous avons créé un code qui récupère les données d'intensité nocturnes, les réduit par un facteur paramétrable au lancement et les sauvegarde sous forme de liste dans un fichier NumPy sans modifier les données originales. NumPy est une bibliothèque qui facilite la manipulation de tableaux Python. Le chargement des données a aussi été modifié pour permettre de récupérer ces tableaux. Les données de base étant sur-échantillonnées, nous pouvons les réduire d'un facteur 32 (donc diviser la hauteur et la largeur par 32).

Nous avons relevé des incohérences entre deux visualisations du même lieu sur deux années différentes. La couleur des pixels varie entre deux images nocturnes pour deux images différentes. Nous avons modifié les visualisations des images optiques et nocturnes en fixant une échelle aux ILN commençant à zéro et finissant à la valeur maximale à travers les années afin de pouvoir comparer deux années sans ces problèmes de continuité.

Folium ne permettait pas un affichage des couches hors de leur gestionnaire de couche. Ce gestionnaire n'était pas bon pour afficher de grand nombre de couches et certaines couches

ne doivent pas être sélectionnés en même temps que d'autres (exemple : couche des ILN pour DMSP de 2010 en même temps que celle de 2011). Pour résoudre ce problème, nous avons développé notre propre système. Nous avons injecté du code JavaScript dans la carte afin de créer notre propre gestionnaire, offrant un affichage paramétrable et personnalisé.

4 Discussion

4.1 Analyse critique des résultats

Folium récupère les cartes les plus à jour des fournisseurs de carte. Il aurait été plus intéressant pour nous de regarder la carte évoluer dans le temps. Une solution était d'utiliser les cartes fournit par la NASA, mais il aurait fallu prendre le temps de modifier le système de folium pour permettre de prendre en compte le temps.

Nous avons analysé des zones à une échelle toujours assez proche. Il serait intéressant de regarder si pour des zones plus larges tout ceci reste valide.

DMSP et VIIRS sont affectés par l'éblouissement donc nous aurions pu mettre en place un système de traitement des images afin d'enlever cet effet afin d'avoir des données de meilleure qualité.

En synthèse, notre groupe a conclu que pour une étude sur une courte période, VIIRS est évidemment meilleur. Mais sur du long terme, DMSP est plus adapté dans les zones avec un ILN ne dépassant pas 63 (périphérie, zones rurales, etc.). En revanche, VIIRS se révèle efficace dans les zones très lumineuses (centres urbains, mégalopoles, etc.) grâce à ses capteurs plus performants et ses données simulées tendent à y être correcte. Nous observons une certaine complémentarité entre les deux. Ainsi, une piste de recherche future consisterait à trouver un moyen de les combiner pour bénéficier du meilleur des deux mondes.

4.2 Apports pour l'entreprise et bilan personnel

Ce stage m'a permis de découvrir le monde de la recherche à travers une étude comparative approfondie de DMSP et VIIRS. J'ai pu appliquer la méthode scientifique de manière concrète et développer des outils pour résoudre un problème réel. J'ai appris à rendre ces outils ergonomiques et à y intégrer de nouvelles fonctionnalités pour faciliter le travail de mes camarades. De plus, j'ai acquis des compétences pour présenter mes travaux de manière scientifique. Enfin, la satisfaction d'avoir répondu à une question concrète avec mon groupe, celle de déterminer lequel entre DMSP et VIIRS est le plus pertinent, a renforcé mon engagement dans ce domaine et confirme mon envie de travailler avec de l'informatique.

5 Conclusion

Notre étude visait à déterminer le jeu de données optimal entre DMSP et VIIRS pour différentes zones et conditions spécifiques. J'ai pu aider mon groupe en développé des outils d'analyse et de visualisation, et mes collègues ont évalué les données en tenant compte de critères objectifs. Cette démarche collaborative nous a permis de choisir le jeu créé par la méthode de [Li et al., 2020] (DMSP) pour le modèle spatio-temporel.

6 Bibliographie

Xuecao Li, Yuyu Zhou, Min Zhao & Xia Zhao. « A harmonized global nighttime light dataset 1992–2018 | Scientific Data », 4 juin 2020. <https://www.nature.com/articles/s41597-020-0510-y>.

Chen, Zuoqi, Bailang Yu, Chengshu Yang, Yuyu Zhou, Shenjun Yao, Xingjian Qian, Congxiao Wang, Bin Wu, et Jianping Wu. « An Extended Time Series (2000–2018) of Global NPP-VIIRS-like Nighttime Light Data from a Cross-Sensor Calibration ». Earth System Science Data 13, n 3 (5 mars 2021) : 889-906. <https://doi.org/10.5194/essd-13-889-2021>.

« Defense Meteorological Satellite Program ». In Wikipédia, 8 mars 2024.
https://fr.wikipedia.org/w/index.php?title=Defense_Meteorological_Satellite_Program&oldid=213162554.

« Équipe ICAR : Image & Interaction – LIRMM ». Consulté le 30 juin 2024.
<https://www.lirmm.fr/equipes/icar/>.

« Matplotlib — Visualization with Python ». Consulté le 30 juin 2024.
<https://matplotlib.org/>.

« Folium — Folium 0.16.1.dev76+g2921126e documentation ». Consulté le 30 juin 2024.
<https://pythonvisualization.github.io/folium/latest/>.

Ammari, Colibeau, Prevost-Avenel « rapport commun stage de recherche L2 ».
(<https://www.overleaf.com/project/665dbda7e56750c5ca315041>).