

TÉLÉDÉTECTION ET MODÉLISATION SPATIALE

Applications à la surveillance et au contrôle
des maladies liées aux moustiques

Annelise Tran, Éric Daudé, Thibault Catry, coord.



Chapitre 2

Indices spectraux et classifications d'images multispectrales pour la cartographie du risque vectoriel

ANNELISE TRAN, RENAUD MARTI, VINCENT HERBRETEAU

Les méthodes classiques d'analyse d'images de télédétection multispectrales¹ incluent la classification d'images et le calcul d'indices spectraux. Ces deux approches exploitent le lien existant entre les caractéristiques des surfaces observées et leur signature spectrale.

Les méthodes de classification permettent de regrouper des zones de l'image ayant des réponses spectrales similaires dans la même classe. Deux types de classification peuvent être utilisés pour produire des cartes d'occupation du sol à partir d'une image multispectrale. Les méthodes de classification supervisée utilisent des algorithmes (maximum de vraisemblance, *Random Forest*, etc.) pour discriminer les différentes classes d'occupation du sol présentes dans l'image (eau, zones urbaines, zones de végétation, sols nus, etc.) et préalablement définies par l'utilisateur (Rodriguez-Galiano *et al.*, 2012). Ces méthodes nécessitent donc de créer une base de données d'apprentissage pour chacune des classes à cartographier avant de lancer la classification proprement dite. Les méthodes de classification non supervisée (p. ex. algorithmes *K-means*, *Isodata*) sont des processus plus automatiques qui ne nécessitent pas de connaissances *a priori* des classes (Duda et Canty, 2002). Les deux types de méthodes peuvent être appliqués pour classer chaque pixel de l'image (classifications orientées pixel) ou des objets (classifications orientées objet) en agrégeant des pixels voisins ayant des valeurs de réflectance similaires avant la classification, par un processus de segmentation de l'image (Blaschke *et al.*, 2014). Dans les deux cas, des données de validation doivent être collectées sur le terrain pour évaluer la qualité de la classification.

Les indices spectraux exploitent la combinaison de différentes bandes d'une image multispectrale pour mettre en évidence certaines propriétés des surfaces observées (voir chapitre 1). Par exemple, la signature spectrale de la végétation révèle une forte réflectance dans le domaine du proche infrarouge et, au contraire, une faible réflectance dans le domaine du rouge (figure I.1). Pour mettre en évidence les zones en végétation d'une image multispectrale (figure 2.1a), les indices de végétation utilisent

1. Voir Introduction générale.

la différence (comme le NDVI, *Normalized Difference Vegetation Index*, cf. chapitre 1) ou le rapport entre les valeurs de réflectance dans le proche infrarouge (PIR) et le rouge (R) [figure 2.1b] :

$$NDVI = \frac{PIR - R}{PIR + R} \quad (\text{équation 1})$$

Pour les zones en eau, dont la signature spectrale révèle des valeurs de réflectance décroissantes depuis les longueurs d'onde du domaine du bleu vers l'infrarouge, les indices d'eau (figure 2.1c) exploitent les différences entre les valeurs de réflectance dans le domaine du vert (V) et le proche ou moyen infrarouge (MIR), comme pour le MNDWI (*Modified Normalized Difference Water Index*) :

$$MNDWI = \frac{V - MIR}{V + MIR} \quad (\text{équation 2})$$

Le seuillage des indices spectraux permet ensuite l'identification et la cartographie des zones d'intérêt écologique pour les moustiques (figure 2.1d).

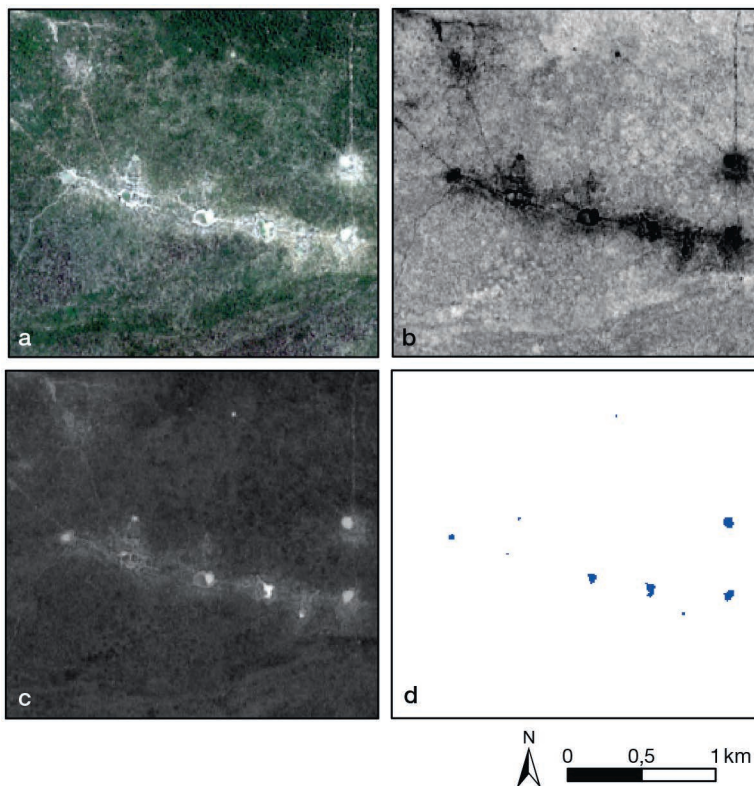


Figure 2.1. Exemples de calcul d'indices de végétation et d'eau à partir d'une image multi-spectrale.

(a) Extrait d'une image Sentinel-2, composition colorée en couleurs naturelles. (b) Indice de végétation. (c) Indice d'eau – du noir, zones à faible valeur d'indice, au blanc, zones avec les valeurs d'indice les plus élevées. (d) Résultat du seuillage de l'image de l'indice d'eau – les surfaces en eau apparaissent en bleu.

Dans ce chapitre, différents exemples d'application de ces méthodes classiques d'interprétation d'images multispectrales sont présentés. Le premier est l'utilisation d'une carte d'occupation du sol dérivée d'une image multispectrale à haute résolution spatiale pour cartographier la distribution de moustiques *Anopheles* en Camargue, à une échelle locale. Dans le deuxième exemple, un indice d'eau est calculé à partir d'images à moyenne résolution spatiale et utilisé dans un modèle prédictif du risque d'occurrence de cas humains de la fièvre du Nil occidental (*West Nile fever*) en Europe, à une échelle continentale. Enfin, nous présentons un exemple d'outil en ligne, Sen2Extract, qui permet d'extraire facilement des séries temporelles d'indices spectraux.

► Cartographie de l'occupation du sol par télédétection pour modéliser la distribution des moustiques *Anopheles* en Camargue

Contexte

La Camargue est située dans le delta du Rhône dans le Sud de la France (figure 2.2a), avec un climat méditerranéen caractérisé par des étés chauds et secs, des hivers humides et doux et des précipitations importantes en automne et au printemps. Peu peuplée, cette région est une mosaïque de paysages de marais, sous l'influence des entrées d'eau salée par le sud, et de zones agricoles plus au nord, composées principalement de rizières et prairies.

Les différentes zones humides présentes en Camargue sont particulièrement propices au développement des moustiques, qui y sont très abondants, et à certaines maladies qu'ils transmettent. Le paludisme y était endémique jusqu'au début du xx^e siècle (Jetten, Takken, 1994), puis il a été éliminé dans les années 1950 avec l'usage des insecticides (essentiellement le DDT) et la chimiothérapie/prophylaxie à base de quinine. Plus récemment, différents travaux de recherche ont étudié le risque de réémergence du paludisme dans cette région dans un contexte de réchauffement climatique (Ponçon *et al.*, 2008). Ces études ont identifié l'espèce *Anopheles (Anopheles) hyrcanus* (Pallas), dont les rizières constituent le principal gîte larvaire, comme vecteur potentiel (Ponçon *et al.*, 2007).

Dans l'exemple présenté ici, les liens entre l'occupation du sol et des relevés entomologiques sur la présence de larves d'*An. hyrcanus* ont été analysés, afin de cartographier la distribution de cette espèce sur toute la Camargue (Tran *et al.*, 2008).

Données

Données entomologiques

Un total de 80 gîtes larvaires potentiels (différents types de marais, rizières) ont été visités mensuellement pendant la période d'activité des moustiques d'avril à octobre 2006 par les équipes de l'Institut de recherche pour le développement (IRD). Les larves de moustiques ont été collectées selon les protocoles entomologiques standards et identifiées au niveau de l'espèce (Ponçon *et al.*, 2007).

Données de télédétection

Deux images Landsat 7 *Enhanced Thematic Mapper* (ETM+, résolution spatiale de 30 m) ont été acquises à deux périodes contrastées (en juillet et en octobre) pour cartographier les principaux types d'occupation du sol dans la zone (figure 2.2a).

Méthodes

Classification de l'occupation du sol

Une classification supervisée orientée objet (logiciel eCognition²) a été appliquée, à partir d'un jeu de données d'entraînement construit sur la base d'observations terrain et en concertation avec des entomologistes, afin de cartographier les habitats de plusieurs espèces retenues pour leur intérêt, pour cinq principaux écotypes (figure 2.2b). Cette typologie inclut l'ensemble des gîtes larvaires potentiels pour les moustiques *Anopheles* dans la région : eau libre, rizières, roselières, marais à scirpes, marais à joncs.

Analyses statistiques

Les corrélations entre la présence ou l'absence de larves d'*An. hyrcanus* (variable à expliquer) et le type de gîte larvaire et la distance aux rizières (variables explicatives) ont été étudiées à l'aide d'un modèle de régression logistique.

Principaux résultats

Les rizières, les marais à scirpes et les roselières ont été identifiés comme significativement corrélés à la présence de larves d'*An. hyrcanus* par le modèle de régression logistique. Ces résultats ont permis de confirmer le potentiel d'autres biotopes que les rizières comme gîtes larvaires d'*An. hyrcanus*, ce qui explique la présence de l'espèce en fin d'été et en automne lorsque les rizières sont à sec. D'autre part, une carte de probabilité de présence de l'espèce a pu être générée sur toute la zone d'étude à partir de la carte d'occupation du sol (figure 2.2c) et cette carte a pu être validée par des données de terrain sur la distribution des moustiques *An. hyrcanus* adultes. Il faut noter que la même approche appliquée à des données d'autres espèces de moustiques *Anopheles* et *Culex* a montré que la distribution des gîtes larvaires en zone humide impactait la répartition spatiale des espèces à l'état adulte, lorsque celles-ci étaient fortement inféodées à un type de gîte larvaire (Cailly *et al.*, 2011).

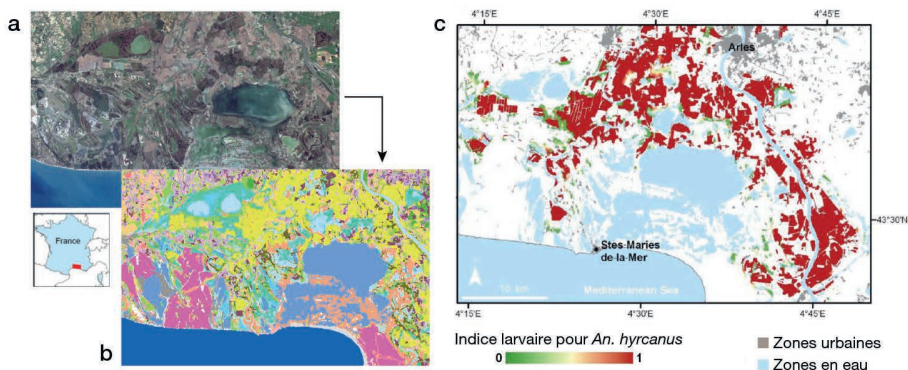


Figure 2.2. Cartographie d'un indice des zones favorables au moustique *Anopheles hyrcanus* en Camargue, à partir d'une image satellite optique. Adapté de Tran *et al.*, 2008.

(a) Extrait d'une image Landsat 7 ETM+. (b) Résultat de la classification en différents types d'occupation du sol. (c) Indice larvaire pour l'espèce *An. hyrcanus*, variant de 0 (zone très peu favorable) à 1 (très favorable).

2. <https://fr.geospatial.trimble.com/products-and-solutions/trimble-ecognition>

► Indices spectraux dérivés d'images de télédétection, comme facteurs environnementaux des cas humains de fièvre du Nil occidental en Europe

Contexte

Le virus du Nil occidental (*West Nile virus*, WNV), responsable de la fièvre du Nil occidental (*West Nile fever*, WNF), est transmis par piqûre de moustiques du genre *Culex*, avec comme hôtes principaux des espèces d'oiseaux domestiques et sauvages, et parfois une transmission à l'humain ou au cheval, considérés comme des hôtes accidentels et des impasses épidémiologiques (*dead-end host*). En Europe, la circulation du WNV a été confirmée épisodiquement ces dernières décennies dans plusieurs pays du bassin méditerranéen, mais durant l'été 2010, particulièrement chaud, un très grand nombre de cas humains de WNF a été rapporté dans des zones jusqu'alors indemnes (Paz *et al.*, 2013). L'étude résumée ici visait à identifier les principaux déterminants environnementaux du risque d'occurrence de cas humains de WNF en Europe, en testant en particulier des indices spectraux issus de la télédétection (Tran *et al.*, 2014).

Données

Données épidémiologiques

Les données sur les cas humains confirmés de WNF ont été répertoriées à l'échelle du district en Europe sur la période 2002-2013. Pour chaque année, chaque district ($n = 1\,113$) était considéré comme « infecté » si au moins un cas humain de WNF avait été rapporté dans ce district et « non infecté » dans le cas contraire.

Données de télédétection

Une série temporelle d'images MODIS a été téléchargée afin de couvrir toute l'Europe, et ce, avec une fréquence de 8 jours pour toute la période considérée.

Autres données spatialisées

D'autres sources de données à référence spatiale ont été utilisées pour caractériser, à l'échelle du district, des facteurs de risque potentiels d'occurrence de cas humains de WNF : températures, population, routes de migration aviaire et zones humides.

Méthodes

Calculs d'indices de végétation et d'eau

Les indices de végétation (NDVI) et d'eau (MNDWI) présentés précédemment (équations 1 et 2) ont été calculés à partir de la série temporelle d'images MODIS.

Analyses statistiques

Les corrélations entre le statut d'infection par le WNV (variable à expliquer) et la population, la présence de zones humides, la présence de routes de migration aviaire, les valeurs d'anomalies de température, de NDVI et de MNDWI (variables explicatives) ont été étudiées par une analyse univariée (en testant les variables une à une), puis par

une analyse multivariée, combinant les variables les plus significatives à l'aide d'un modèle de régression logistique. Les données épidémiologiques de la période 2001-2011 ont été utilisées pour construire le modèle et les données épidémiologiques de la période 2012-2013, pour tester sa capacité prédictive.

Résultats

Le meilleur modèle permettant d'expliquer le statut infecté/indemne de cas humains de WNF par district inclut comme variables explicatives les anomalies de température au mois de juillet et d'indice d'eau MNDWI début juin (9-16 juin), la circulation du WNV l'année précédente, la présence de zones humides, le type de route migratoire aviaire et la population humaine (figure 2.3a); toutes ces variables étant positivement et significativement corrélées avec la probabilité d'infection (Tran *et al.*, 2014). À partir de ce modèle de régression logistique, des cartes de la probabilité d'infection par le WNV par district peuvent être produites annuellement (figure 2.3b). Ces résultats mettent en évidence le rôle des zones en eau dans le risque de transmission du WNV en Europe et la pertinence d'un indice d'eau simple (le MNDWI dérivé des images MODIS) pour détecter des surfaces en eau au-dessus des moyennes saisonnières au mois de juin, susceptibles de favoriser la prolifération des moustiques durant les mois d'été et la transmission du virus à l'humain en fin d'été-début d'automne.

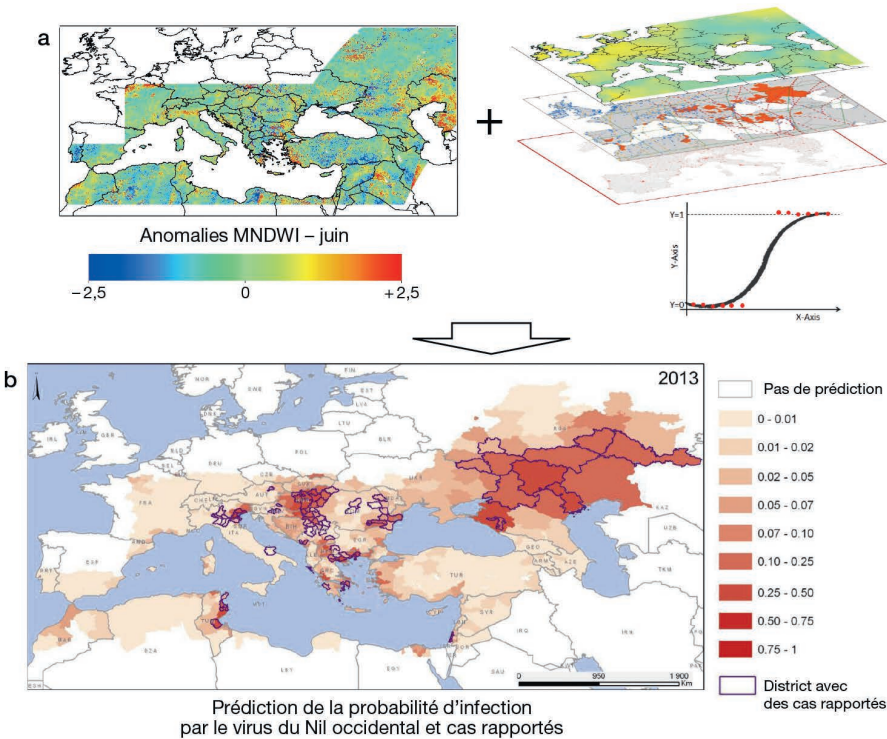


Figure 2.3. Cartographie des zones à risque de transmission du virus du Nil occidental en Europe basée sur des indicateurs environnementaux et météorologiques. Adapté de Tran *et al.* (2014).

(a) Analyse d'un indice d'eau et de différents prédicteurs environnementaux du virus du Nil occidental par un modèle de régression logistique. (b) Cartographie du risque d'infection.

► Production automatisée d'indices spectraux : exemple de l'outil Sen2Extract

La télédétection reste un domaine technique demandant une bonne connaissance des méthodes, des logiciels et algorithmes de traitement d'images et du large éventail des données spatiales disponibles. Il n'est donc pas étonnant que l'essentiel de l'usage des images satellite dans le domaine de la santé se soit fait indirectement *via* la mise à disposition de produits pré-calculés, tels que les indices spectraux, gratuitement et simplement accessibles par téléchargement. Ces produits ont permis d'atteindre une communauté d'utilisateurs, essentiellement des épidémiologistes et des statisticiens, sans compétences en télédétection, qui ont pu les intégrer dans leurs études. Bien que ces indices ne soient pas forcément les mieux adaptés aux caractéristiques des zones d'étude ou à la problématique, dans bien des cas leur disponibilité constante dans l'espace et le temps en a fait de précieux indicateurs des fluctuations saisonnières.

C'est à partir de ce constat que des outils ont été développés pour automatiser la production d'indices spectraux à partir des images des satellites Sentinel-2A et 2B du programme Copernicus de l'Agence spatiale européenne (ESA). Comme leur nom le suggère, ces deux satellites ont été conçus à des fins de surveillance environnementale et ils peuvent donc être indirectement utiles pour la surveillance de maladies liées à l'environnement. En effet, ils fournissent des images optiques ayant une résolution spatiale de 10 m, disponibles gratuitement tous les 5 jours, depuis mi-2015, sur la quasi-totalité du globe. Parmi les outils développés, Sen2Extract est une application en ligne³ qui permet d'extraire simplement des séries temporelles d'indices spectraux sur une période et des sites choisis selon les besoins de l'utilisateur. Une interface simple et intuitive (figure 2.4) invite à indiquer son/ses sites d'intérêt en chargeant un fichier au format vectoriel (points ou polygones, format shapefile zippé), puis à choisir la période et les indices à calculer (indice de végétation NDVI et différents indices d'eau). Dès que le calcul est réalisé, un e-mail est envoyé pour indiquer le lien de téléchargement du résultat de la requête, qui prend la forme d'un tableau de données au format csv. Pour chaque entité spatiale, polygone ou point, et chaque date, des statistiques descriptives des valeurs de l'indice choisi sont calculées (moyenne, minimum, maximum, écart-type, médiane, premier et troisième quartiles, nombre de pixels ennuagés ou sans données). Le fichier final est donc très léger (quelques kilo-octets), comparé au poids d'une image Sentinel-2 et de ses traitements qui peuvent représenter près d'un gigaoctet. Ceci facilite l'utilisation des images dans des pays où l'accès à internet est limité.

Sen2Extract est l'interface utilisateur d'un ensemble d'algorithmes développés dans le cadre du projet S2-Malaria (appel à projets de recherche Tosca, financé par le Cnes, de 2017 à 2020) pour produire des données d'observation satellitaire utiles à la surveillance de maladies. Ces algorithmes ont ensuite été optimisés pour le suivi des dégâts liés aux passages des cyclones (Projet Interreg *Renovrisk-impact* de 2018 à 2020). En amont de Sen2Extract, la chaîne Sen2Chain, codée en Python, est au cœur de la production des indices. Elle permet de télécharger les images Sentinel-2 sur deux des principaux catalogues, le *Copernicus Open Access Hub*⁴ de l'ESA et la plateforme d'exploitation

3. <https://web.seas-oi.org/sen2extract/>

4. <https://scihub.copernicus.eu/>

des produits Sentinel⁵ (Peps) du Cnes. Elle intègre l'algorithme Sen2Cor de l'ESA pour effectuer les corrections atmosphériques, de terrain et des cirrus sur les images brutes (niveau 1A) afin d'obtenir le niveau 2A, nécessaire à la comparaison des dates et des scènes. Elle permet ensuite de produire différents indices spectraux — il est aussi possible d'en intégrer de nouveaux selon les besoins —, puis d'extraire les valeurs des pixels pour calculer leurs statistiques descriptives sur des entités spatiales définies, à chaque date. Les codes de Sen2Chain sont ouverts⁶ et peuvent donc être réutilisés et améliorés librement. Sen2Extract se greffe en aval pour offrir l'interface de requête et lancer le calcul des séries temporelles d'indices selon les besoins. Cet outil est codé en R⁷ et l'interface en ligne repose sur le package R Shiny⁸. Les codes sont également ouverts pour pouvoir être réimplémentés⁹. Ils permettent d'envoyer les requêtes vers la chaîne Sen2Chain, de paramétrer les extractions et les calculs statistiques puis d'envoyer le lien par e-mail. L'interface accessible en ligne est installée sur les serveurs de la station de réception satellitaire SEAS-OI à La Réunion. Les requêtes de séries temporelles peuvent se faire sur l'emprise des images qui ont été acquises sur la région océan Indien essentiellement. Pour d'autres zones, il convient donc d'implémenter ces outils sur un nouvel ordinateur ou serveur.

L'ensemble des tâches proposées par ces chaînes est entièrement automatisable. Ceci permet de produire des séries temporelles en routine pour le besoin d'outils de surveillance. C'est le cas de l'application Lepto Yangon¹⁰ qui peut réaliser des cartographies des environnements favorables à la transmission de la leptospirose sur l'agglomération de Yangon au Myanmar, dès qu'une nouvelle image est disponible (tous les 5 jours). Une telle automatisation de la production d'indices spectraux dans des temps courts serait envisageable pour le suivi d'environnements favorables aux espèces de moustiques pour lesquelles les liens entre ces indices et leur présence ou abondance ont pu être établis.

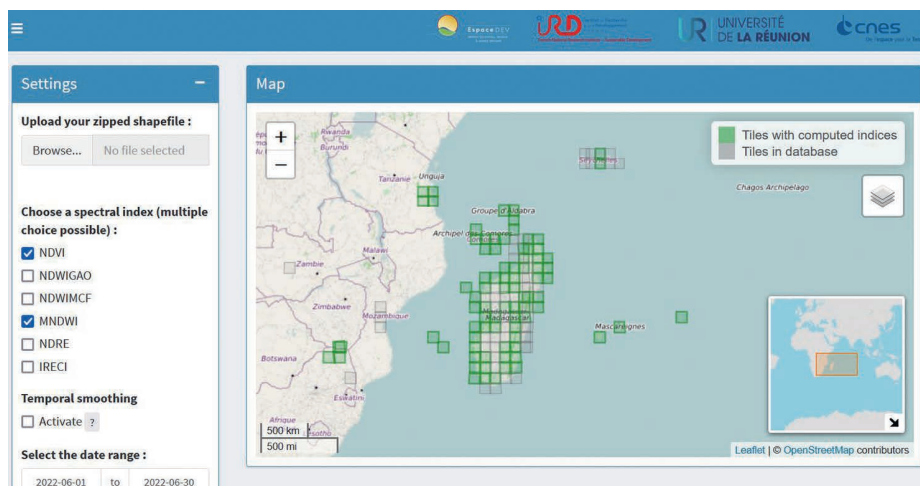


Figure 2.4. Interface de l'outil en ligne Sen2Extract (<https://web.seas-oi.org/sen2extract/>).

5. <https://peps.cnes.fr>

6. <https://framagit.org/espace-dev/sen2chain>

7. <https://www.R-project.org/>

8. <https://shiny.rstudio.com/>

9. <https://framagit.org/espace-dev/sen2extract>

10. <https://leptoyangon.geohealthresearch.org/>

» Références

- Blaschke T., Hay G.J., Kelly M., Lang S., Hofmann P., Addink E., Queiroz Feitosa R., Van der Meer F., Van der Werff H., Van Coillie F., Tiede D., 2014. Geographic Object-Based Image Analysis – Towards a new paradigm. *ISPRS Journal of Photogrammetry and Remote Sensing*, 87, 180-191.
- Cailly P., Balenghien T., Ezanno P., Fontenille D., Toty C., Tran A., 2011. Role of the repartition of wetland breeding sites on the spatial distribution of *Anopheles* and *Culex*, human disease vectors in southern France. *Parasit Vectors*, 4, 65.
- Duda T., Canty M., 2002. Unsupervised classification of satellite imagery: Choosing a good algorithm. *International Journal of Remote Sensing*, 23 (11), 2193-2212.
- Jetten T.H., Takken W., 1994. Anophelism without malaria in Europe, A review of the ecology and distribution of the genus *Anopheles* in Europe, *Wageningen Agricultural University, papers*, 94-5, 69 p.
- Paz S., Malkinson D., Green M.S., Tsioni G., Papa A., Danis K., Sirbu A., Ceianu C., Katalin K., Ferenczi E., Zeller H., Semenza J.C., 2013. Permissive summer temperatures of the 2010 European West Nile fever upsurge. *PLoS One*, 8 (2), e56398.
- Ponçon N., Toty C., L'Ambert G., Le Goff G., Brengues C., Schaffner F., Fontenille D., 2007. Biology and dynamics of potential malaria vectors in Southern France. *Malaria Journal*, 6, 18.
- Ponçon N., Tran A., Toty C., Luty A.J., Fontenille D., 2008. A quantitative risk assessment approach for mosquito-borne diseases: malaria re-emergence in southern France. *Malaria Journal*, 7, 147.
- Rodriguez-Galiano V.F., Ghimire B., Rogan J., Chica-Olmo M., Rigol-Sanchez J.P., 2012. An assessment of the effectiveness of a random forest classifier for land-cover classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, 67, 93-104.
- Tran A., Sudre B., Paz S., Rossi M., Desbrosse A., Chevalier V., Semenza J.C., 2014. Environmental predictors of West Nile fever risk in Europe. *International Journal of Health Geographics*, 13, 26.
- Tran A., Ponçon N., Toty C., Linard C., Guis H., Ferre J.B., Lo Seen D., Roger F., La Rocque S. de, Fontenille D., Baldet T., 2008. Using remote sensing to map larval and adult populations of *Anopheles hyrcanus* (Diptera: Culicidae) a potential malaria vector in Southern France. *International Journal of Health Geographics*, 7, 9.