JOURNÉE SCIENTIFIQUE

& INTELLIGENCE ARTIFICIELLE BIODIVERSITÉ

MARDI 30 MARS 2021 / 9H30 - 17H

PROGRAMME

9h30-9h40 : Introduction et mot de bienvenue par les LabEx Cemeb et NUMEV

9h40-10h30 : Modèles linéaires structurés : apport de l'optimisation et des statistiques
Joseph Salmon, UM *(Introduction et modération : Nadine Hilgert)*

10h30-11h20 : Pl@ntNet : quels enjeux scientifiques au-delà de la collecte de données
Alexis Joly, Inria *(Introduction et modération : Julien Renoult)*

11h20-11h30 : PAUSE

11h30-12h05 : SESSION 1 - IA ET FAUNE AQUATIQUE (présentations de 6 min)

1. **Laura Mannocci**, MARBEC, LIRMM : Deep learning et réseaux sociaux pour la détection de la mégafaune marine rare.

Le deep learning devient un outil incontournable pour l'automatisation du suivi des populations animales par capture et analyse d'images. Cependant, l'obtention d'un grand nombre d'images pour l'entrainement des algorithmes est un défi majeur pour les espèces rares. C'est le cas notamment de la mégafaune marine qui, de plus, est peu visible à la surface et présente dans des zones isolées. Pour ces espèces rares mais charismatiques, les images accumulées sur les réseaux sociaux pourraient permettre de constituer des bases d'entrainement pour les algorithmes lorsque les images capturées dans une zone sont limitées. Dans cette étude nous évaluons cette possibilité pour les algorithmes de détection de dugong en Nouvelle Calédonie. Nous évaluons d'abord la performance d'un Faster R-CNN entrainé avec 1303 images de dugongs issues des réseaux sociaux (images WEB) à détecter les dugongs sur des images acquises par ULM en Nouvelle Calédonie. Nous incorporons ensuite un nombre croissant d'images ULM dans l'entrainement du Faster R-CNN (représentant 2 à 12% des images WEB) et évaluons l'amélioration de la capacité de détection du dugong sur les images ULM. L'algorithme entrainé uniquement avec les images WEB montre une capacité de détection du dugong conséquente sur les images ULM, avec seulement 38% de faux négatifs. L'incorporation dans la base d'entrainement de seulement 136 images ULM permet de diviser par 2 le nombre de faux négatifs. Nos résultats soulignent l'intérêt des réseaux sociaux pour la détection d'espèces charismatiques rares dans une zone donnée par deep learning.

2. **Guilhem Marre**, Andromède Océanologie : Reconnaissance d'espèces du coralligène par des réseaux de neurones convolutifs.

La surveillance de l'état écologique des habitats naturels est essentielle au processus de conservation, car elle permet la mise en œuvre de politiques de conservation efficaces. Les récifs coralligènes sont un habitat sous-marin Méditerranéen d'une importance particulière, et d'une complexité similaire à celle des récifs coralliens. Ils sont surveillés dans les eaux françaises depuis 2010 à l'aide de quadrats photographiques annotés manuellement (réseau de surveillance RECOR). Sur la base de cette vaste base de données, nous avons entraîné des réseaux neuronaux convolutifs à reconnaître automatiquement les principales espèces du coralligène. Nous avons pour cela conçu un ensemble de réseaux basé sur une architecture classique, capable de distinguer 61 classes avec une précision de 72,59 %. Nos résultats ont montré que les erreurs de

confusion étaient pour la plupart taxonomiquement cohérentes, avec 84,47 % de bonnes classifications lorsque la tâche était simplifiée à seulement 15 classes, surpassant ainsi la précision humaine quantifiée sur un problème similaire en milieu corallien. Nous avons par ailleurs construit un outil semi-automatisé permettant de rejeter les résultats incertains et ainsi de réduire le risque d'erreur. Enfin, nous avons utilisé notre modèle pour évaluer la biodiversité et le statut écologique des récifs coralligènes avec l'indice d'assemblage coralligène (CAI) et l'indice de Shannon. En conclusion, on peut affirmer que l'approche décrite par cette étude offre un outil rentable et efficace en termes de temps et de coût pour l'analyse des assemblages coralligènes, qui peut être intégré dans un réseau de surveillance à grande échelle de cet habitat.

3. **Sébastien Villéger**, MARBEC : Algorithmes pour la détection et l'identification des poissons sur des vidéos sous-marines.

Les écosystèmes coralliens sont de plus en plus affectés par les changements globaux alors qu'ils abritent une biodiversité unique (e.g. >6000 espèces de poissons). Or les protocoles d'observations des poissons par des plongeurs présentent des limites de rendement et de réplicabilité. Pour pallier à ces contraintes, les protocoles reposant sur des caméras sous-marines fixes sont de plus en plus utilisés mais l'analyse de ces vidéos requiert un important temps de travail par des experts. Les laboratoires MARBEC et LIRMM ont collaboré depuis 2015 pour développer des algorithmes de type « Deep Learning » capables de détecter et identifier les poissons. Nous présenterons nos travaux sur 1) l'acquisition et l'annotation des vidéos pour la construction des bases de données, 2) l'optimisation des algorithmes de classification et de détection, et 3) les tests de performance des algorithmes pour des problématiques écologiques. Enfin nous présenterons les limites des algorithmes actuels et nos perspectives de recherche pour les rendre encore plus efficaces en écologie.

4. **Bruno Guinand**, ISEM : Reconnaissance d'images de truite (*Salmo trutta*) et deep learning.

La truite (*Salmo trutta*) est une espèce patrimoniale à intérêt conservatoire possédant une forte variabilité phénotypique de ses patrons de coloration et, dans une moindre mesure, de sa morphologie. Ces caractéristiques peuvent servir à identifier l'origine géographique des individus et sont appréhendables sur des photographies. Dans une optique exploratoire, nous avons développé une approche basée sur le deep learning afin d'estimer l'origine géographique de truites ('atlantiques' vs 'méditerranéennes') à partir de photos disponibles dans une base de données (OSU OREME) ou mise à disposition par une fédération de pêche. Dans ma présentation, je soulignerai les succès de cette approche, ainsi que certaines des limites actuelles notamment dans l'étape de validation. Je présenterai en quoi cette approche pourrait compléter les analyses moléculaires traditionnellement entreprises pour rendre compte de l'«intégrité» des populations de truite face par exemple au réempoissonement (truites atlantiques déversées dans des cours d'eau méditerranéens).

5. **Pauline Stephan**, MARBEC : Utilisation des "self-organizing maps" (SOM) pour la détection de clusters spatio-temporels de captures dans la pêche à la senne thonière de l'Atlantique tropical.

La pêche à la senne sous dispositifs de concentration de poissons (DCPs) dérivants génère d'importantes captures de juvéniles de thon albacore et de thon obèse, ainsi que des captures accessoires d'espèce sensibles comme le requin soyeux. Afin d'identifier les zones de hotspots de captures dans l'Atlantique tropical, nous utiliserons les « self-organizing maps » (SOM) sur les données géoréférencées de pêche commerciale de la Commission Internationale pour la Conservation des Thonidés de l'Atlantique (CICTA) et sur les variables environnementales associées. Ce type de réseau neuronal non-supervisé permet de détecter des clusters spatio-temporels sans a priori. Le résultat attendu est une aide à la définition de moratoires, i.e. de zones d'accès limité à la pêche sous DCPs dérivants dans l'Atlantique tropical.

12h05-12h30 : DISCUSSION *(Modération : François Rousset)*

12h30-13h30 : PAUSE DEJEUNER

13h30-14h20 : SESSION 2 - IA ET FAUNE TERRESTRE (présentations de 6 min)

1. **Simon Boitard**, CBGP : Prédiction génomique de l'adaptation de *Drosophila suzukii* à différentes plantes hôtes.

La sélection génomique pratiquée depuis une vingtaine d'années sur les animaux de rente a démontré qu'il était possible de prédire efficacement le potentiel génétique d'un individu au regard d'une contrainte sélective donnée. Avec le développement des technologies de séquençage à haut-débit, cette approche prédictive commence à être appliquée à l'échelle de populations y compris pour des espèces non modèles, afin d'évaluer notamment la capacité d'adaptation de ces populations aux changements climatiques à venir. Nous proposons ici de développer une approche de prédiction génomique populationelle de ce type dans le cadre de la lutte contre la mouche *Drosophila suzukii*, une espèce invasive originaire d'Asie qui cause des dommages majeurs sur les cultures fruitières (fraise, cerise, framboise, mûre et autres fruits rouges) notamment en Europe et en Amérique. Bien que *Drosophila suzukii* soit une espèce généraliste, des travaux en cours dans notre unité montrent qu'elle est capable de s'adapter à un fruit hôte particulier en quelques générations et que cette spécialisation peut la conduire à devenir moins performante sur ses autres hôtes potentiels. Cette adaptation ayant une base génétique, on peut ainsi espérer construire des modèles prédictifs permettant de quantifier le risque d'établissement d'une population donnée sur différents fruits à partir du génome de cette population. Ce type de prédiction pourrait servir d'outil diagnostic afin de quantifier les risques représentés par une population observée donnée pour les cultures avoisinantes, permettant ainsi d'adapter les pratiques agricoles de protection des cultures en fonction des prédictions. Nous présenterons un jeu de données qui pourrait servir pour l'apprentissage d'un tel modèle prédictif, à savoir des données de polymorphisme pan-génomique pour plusieurs populations naturelles et expérimentales adaptées à différents fruits. Nous présenterons également deux approches envisagées pour construire ce modèle et aimerions discuter d'approches alternatives issues de l'intelligence artificielle.

2. **Laëtitia Lemière**, AMAP : Détection, caractérisation et quantification des attaques de frelon asiatique sur l'abeille domestique par des méthodes de deep learning et de vidéo-surveillance.

Le frelon asiatique, introduit en France en 2004, est rapidement devenu une espèce envahissante en Europe. La cible préférée de ce prédateur, qui chasse en vol, est l'abeille domestique, un maillon essentiel de la biodiversité et principal pollinisateur de nombreuses cultures vivrières. Alors que ce frelon semble participer au syndrome d'effondrement des colonies d'abeilles, son comportement de vol qui fait son succès de prédation est peu étudié et son impact sur les ruchers reste difficilement mesurable. Développer une méthodologie fiable et automatisée permettant de détecter, suivre et compter les frelons, les abeilles butineuses et les événements de prédation à l'entrée des ruches permettrait de quantifier précisément l'impact du frelon sur les ruchers et d'analyser son évolution dans le temps, ainsi que le moment et les conditions favorables à l'intervention ou au piégeage. Un travail exploratoire, soutenu en 2020 par l'i-Site MUSE (Key Initiative Data & Life Sciences), a consisté à chercher à dénombrer automatiquement à partir de données de vidéosurveillance le ratio captures/attaques dont sont victimes les abeilles par tranche horaire. Nous avons évalué l'intérêt des méthodes de deep learning pour traiter des données collectées en environnement non contrôlé, et notamment le couplage du réseau de neurones profond YoloV4 au système de suivi V-OUI. L'étude a exploité 360 heures de films basse résolution pris entre 2016 et 2018. Le réseau de neurones profond a permis de détecter différentes catégories d'insectes avec des taux de prédictions appréciables (variant de 85 à 90% selon les classes d'objet) mais leur suivi s'est avéré plus problématique principalement à cause de la vitesse d'acquisition trop faible des caméras utilisées. Nous discutons des solutions permettant d'améliorer les stratégies d'entrainement et de validation, ainsi que des algorithmes de suivi d'insectes en mouvement au comportement en apparence erratique.

3. **Noa Rigoudy**, CEFE, LBBE, ENS Lyon : Automatisation de la reconnaissance des espèces de la faune française dans les données de piège- photographiques par apprentissage profond (deep learning).

Les pièges photographiques sont devenus essentiels au recensement et au suivi des espèces dans la gestion et la préservation des patrimoines naturels (Burton et al., 2015). Face à l'explosion récente de la quantité d'images à traiter, les modèles d'apprentissage profond (deep learning) ont révélé toute leur puissance en permettant l'automatisation du tri et la classification de ces données à grande échelle (Norouzzadeh et al. 2018, Baraniuk et al. 2020, Høye et al. 2021). Néanmoins, les modèles existants, entraînés sur des images de faunes Sud-Africaine ou Nord-Américaine, restent difficilement généralisables et ne peuvent être directement appliqués au système français (Beery et al. 2019, Schneider et al. 2020). Cette étude a pour but de développer un modèle d'apprentissage profond de reconnaissance des espèces mammifères clés dans les données de pièges photographiques récoltées à travers la France hexagonale. Pour ce, nous utilisons la méthode de « transfer learning » qui consiste à développer des modèles de reconnaissance en adaptant l'architecture interne de modèles performants préexistants (Pan & Yang 2010). Nous essaierons d'identifier les forces et faiblesses de notre approche.

4. **Paul Taconet**, MIVEGEC : Améliorer les connaissances sur les liens entre environnement et vecteurs du paludisme à l'aide de méthodes d'apprentissage automatique.

Afin de lutter plus efficacement contre les vecteurs du paludisme, il est important de mieux comprendre leur biologie à fine échelle spatio-temporelle : en particulier, d'identifier les principaux déterminants environnementaux de leurs densités agressives, les éventuelles interactions entre ces facteurs et les valeurs seuils conduisant à un risque accru de piqûre. Toutefois, le système environnement - vecteur à ces fines échelles est complexe et les connaissances théoriques seules peuvent être insuffisantes pour identifier ces éléments. Les algorithmes d'apprentissage automatique sont particulièrement adaptés à l'extraction de connaissances dans ces situations, de par leur capacité intrinsèque à détecter des associations complexes entre variables (interactions, relations non-linéaires ou non-monotones). Dans cette présentation, nous montrons comment nous avons utilisé des modèles d'apprentissage automatique (forêts aléatoires) et des outils d'interprétation de ces modèles ("interpretable machine learning") pour améliorer notre compréhension du système vecteur - environnement à fine échelle spatio-temporelle dans une zone rurale du Burkina Faso. Notre variable à expliquer est le nombre de piqûres de moustiques par nuit et par personne, et nos variables explicatives sont des paramètres paysagers et climatiques (issus d'images satellitaires) que l'on sait influencer les dynamiques des populations de vecteurs. Nous discutons de l'intérêt de cette approche, et nous ouvrons sur les points importants à considérer lors de l'utilisation de méthodes d'apprentissage automatique à des fins explicatives (et non prédictives).

5. **Simon Chamaillé-Jammes**, CEFE : Détection automatique du comportement d'alimentation d'animaux à partir de données audio analysées par deep learning.

Nous présenterons ici nos travaux autour de la détection de comportements d'animaux à partir de données de biologging analysées par deep learning. Après une présentation rapide des motivations initiales au projet et du développement d'un capteur audio embarqué sur l'animal ayant bénéficié d'un soutien financier Cemeb/Numev, nous présenterons les modèles d'apprentissage profond mis en œuvre sur les 3500h de données audio collectées lors du déploiement pilote du logger sur des zèbres dans une réserve sud-africaine. Nous présenterons une comparaison des performances respectives des différents modèles. De manière générale, nous démontrerons dans cette présentation la capacité des approches de deep learning à produire une inférence précise sur un comportement clé des animaux – l'alimentation, sur des jeux de données massifs à partir d'un faible nombre d'annotations.

6. **Etienne Loire**, ASTRE : Retrouve mon chien (errant) à Bali et au Cambodge.

Les chiens sont des réservoirs pour de nombreux pathogènes et parasites zoonotiques, comme la rage en Asie. La gestion des chiens nécessite une meilleure estimation de la taille et de la distribution des populations,

notamment des chiens errants. Dans le cadre d'une étude sur la socio-écologie des chiens en Indonésie et au Cambodge, nous avons collecté des données de terrain dans deux villages afin d'analyser par un modèle de capture-marquage-recapture la dynamique des populations de chiens. Chaque session de « capture visuelle » consiste en une série de photos obtenues pendant 3-5 jours consécutifs par 12-15 pièges photographiques à déclenchement automatique, positionnés dans un village à l'issue d'un atelier participatif pour choisir les emplacements des caméras-traps. Une banque de photos de référence de chiens « marqués » est également collectée à chaque session (approx. 50-100 chiens), à l'aide d'une enquête porte-à-porte. Les sessions de collecte de données de terrain (3 sessions en 2019-2020) ont été interrompues par la crise COVID-19, mais viennent de reprendre en Indonésie, et probablement au Cambodge dans les mois qui viennent. Nous recherchons un partenaire pour établir une méthode semi-automatique d'analyse des images, afin de filtrer les gros volumes d'images brutes collectées (e.g. >32000 photos pour une session 3 jours), et de tester la faisabilité et l'efficacité d'une reconnaissance individuelle des chiens par intelligence artificielle.

7. **Nicolas Silva**, CEFE : Utilisation de MASK RCNN pour automatiser la prise de mesure de traits phénotypiques et de comportement.

En écologie et évolution, obtenir des mesures individuelles et nombreuses sur les traits phénotypiques des individus constitue la première étape de beaucoup études. La plupart du temps, ces mesures sont acquises manuellement en attrapant l'organisme d'étude. Cette prise de mesure est un travail long et fastidieux. Depuis quelques années, les méthodes d'automatisation de tâches via des réseaux de neurones se multiplient et permettent un gain de temps considérable. Nous présentons ici nos projets sur l'utilisation de MASK RCNN pour automatiser la prise de mesure de la taille d'un trait coloré et d'un comportement chez le républicain social (*Philetairus socius*), une espèce d'oiseau social d'Afrique du Sud. Le trait phénotypique étudié est la bavette noire située sur la gorge des individus. Il a été montré que cette bavette joue un rôle important dans les interactions de dominance entre individus. Pour étudier maintenant son association avec des mesures de sélection sexuelle et naturelle, des photos de bavette sont prises sur 600 individus chaque année. Nous présentons ici le réseau de neurones de type Mask RCNN qui nous permet d'extraire la taille de la bavette de façon automatique à partir de photos prises lors de capture sur le terrain. Le trait comportemental étudié est le comportement de construction du nid. Ces oiseaux vivent en colonie dans un nid gigantesque de plus de 5m de long qu'ils construisent et entretiennent régulièrement. Afin d'étudier le comportement de construction du nid, nous présenterons notre projet d'utiliser un autre réseau de neurones pour détecter les individus construisant sur le dessus de la colonie et de le coupler à un réseau de neurones existant pour connaître l'identité des individus.

8. **Olivier Gimenez**, CEFE : Erreurs d'annotation en deep learning, et quantification des interactions prédateurs-proies. Pérégrinations sous R.

Les interactions prédateurs-proies façonnent les communautés animales. Le piégeage photographique permet d'étudier ces relations à partir de l'identification des espèces sur les photographies prises en conditions naturelles. Pour annoter automatiquement de grandes quantités de photographies, le deep learning est de plus en plus utilisé. Deux difficultés se posent toutefois. Premièrement, la reconnaissance n'est pas parfaite, et les taux d'erreur ne sont pas nuls. Deuxièmement, le principal langage du deep learning, Python, est peu connu des écologues. Dans cette communication, on se pose la question de l'effet du taux d'erreur dans l'identification des espèces sur l'inférence des interactions entre elles. L'analyse est faite entièrement sous R, un langage bien connu des écologues. Ce travail bénéficie des ressources mises à disposition par le Labex Cemeb, et plus précisément des machines GPU de la plate-forme MBB.

14h20-14h50 : DISCUSSION *(Modération : Pascal Poncelet)*

14h50-15h00 : PAUSE

15h00-15h45 : SESSION 3 - IA ET FLORE (présentations de 6 min)

1. **Clara Gritti**, Conservatoire botanique national méditerranéen de Porquerolles : Cartographie d'habitats par télédétection en Méditerranée : exemple du site Natura 2000 « Complexe lagunaire de Salses-Leucate ».

Le renouvellement de la cartographie des sites Natura 2000 est d'ordinaire réalisé par le biais de méthodes mêlant observations de terrain et photo-interprétation. Sur le site du « Complexe lagunaire de Salses-Leucate », trois organismes partenaires ont mis leurs compétences en commun afin de réaliser une cartographie par télédétection : la société I-SEA, spécialisée en télédétection, le Conservatoire botanique national méditerranéen de Porquerolles, expert en sciences de la végétation ainsi que le Syndicat Mixte RIVAGE, animateur du site Natura 2000. Cette cartographie est issue du service de cartographie BIOCOAST développé par la société I-SEA. Ce service repose sur l'usage de méthodes d'apprentissage automatique, appliquées à la reconnaissance des habitats à partir de séries temporelles d'images satellites et de données de référence relevées sur le terrain. De nombreux échanges et itérations entre les différents experts sont au cœur du processus de cartographie. Cette discussion constante est nécessaire pour aboutir à une cartographie à la fois représentative des réalités de terrain et pertinente pour la gestion du site. Ce type d'approche est innovant pour les sites Natura 2000 méditerranéens. Les différentes étapes méthodologiques de cartographie seront présentées ainsi que les problèmes rencontrés et les solutions apportées. Les trois partenaires de l'étude vous proposent leurs retours d'expérience sur ce projet.

2. **Joaquim Estopinan**, LIRMM : Une approche prédictive de la détermination du statut de conservation conjoint des espèces.

La liste rouge IUCN est aujourd'hui la référence mondiale répertoriant les risques d'extinction des espèces. L'évaluation du statut de conservation d'une espèce a été unifiée par l'application de 5 critères synthétiques concernant les tailles de populations, les distributions géographiques et leurs dynamiques. Environ 86% des mammifères (5792 sur 6500) ont désormais un statut de conservation attribué, contre moins de 8% des 352 000 espèces de plantes estimées dans le monde. Il y a donc urgence à évaluer les espèces végétales pour aider à correctement orienter les politiques de conservation de la flore face au changement climatique. Le processus d'évaluation IUCN manuel étant chronophage, le recours à des évaluations automatiques est une solution prometteuse. La communauté scientifique se saisit actuellement des outils de l'intelligence artificielle pour espérer atteindre des méthodes robustes de classification du risque d'extinction. Seulement, de nombreux problèmes clefs méritent une attention particulière avant de prétendre d'approcher fidèlement la méthode manuelle. Ils concernent principalement la qualité des données en entrée des algorithmes : représentation des menaces encourues par les espèces, données de présence uniquement, biais d'observation ou encore distributions d'espèces très inégales, et les stratégies compensatoires possibles de machine learning : apprentissage profond, par transfert, multitâche, méta-apprentissage entre autres.

3. **Pierre-Olivier Cheptou**, CEFE : Mise au point d'un observatoire de la flore par l'analyse automatisée du pollen récoltée dans la ruche.

Le projet Superbeelive (MUSE) est un projet multidisciplinaire autour de l'abeille associant des recherches sur la santé de l'abeille, des recherches sur la socialité (superorganisme) et des recherches en écologie (observatoire de la floraison). Il se déploie autour du développement d'une ruche « prototype » instrumentée (caméra, mesure…) située sur les terrains d'expériences du CEFE-CNRS pour la récolte en temps réel de différents types de données, en collaboration avec des informaticiens/roboticiens. Partant de l'idée simple que l'abeille butine les plantes au pic de floraison (optimal foraging theory), l'abeille nous renseigne sur la composition des communautés végétales ainsi que sur la date précise de la floraison de chaque espèce butinée. La ruche constitue donc un observatoire très précis des communautés végétales, potentiellement généralisable à grande échelle (via les sciences participatives par exemple). L'ambition du projet est d'analyser de manière systématique le pollen collecté par les butineuses (de manière non-invasive) et ainsi d'identifier les espèces en fleur. Dans le cadre des changements planétaires et notamment climatiques, ces

données sont en mesure de renseigner les évolutions de phénologie et les dynamiques d'abondance des espèces dans le temps. L'idée est d'utiliser les propriétés macroscopiques identifiables de la pelote de pollen par caméra par une approche non invasive. Durant le printemps 2020, un travail expérimental a été mené. A l'aide de caméras et d'un traitement d'images, des essais ont été réalisés tout d'abord sur la pelote de pollen, puis sur des butineuses en utilisant l'autofluorescence du pollen, les intensités lumineuses dans différentes longueurs d'ondes pour attribuer à chaque pelote de pollen l'espèce florale dont elle provient. Ces tests ont confirmé l'intérêt de l'utilisation de l'autofluorescence comme paramètre déterminant dans la discrimination des pelotes de pollen. Dans cette présentation, je présenterai les résultats préliminaires obtenues, les difficultés rencontrées et les perspectives que nous envisageons.

4. **Olivier Argagnon**, Conservatoire botanique national méditerranéen de Porquerolles : Les données du système d'information en usage au Conservatoire botanique national méditerranéen de Porquerolles.

Le Conservatoire botanique national méditerranéen de Porquerolles a pour mission de faire l'inventaire de la flore sauvage et des végétations de son territoire. Ce dernier s'étend sur l'ancienne région Languedoc-Roussillon et les départements de Vaucluse, des Bouches-du-Rhône, du Var et des Alpes-Maritimes. Dans ce cadre il produit de nombreuses données qu'il administre conjointement aux données de ses partenaires et aux données flore des SINP régionaux de son territoire. Le jeu de données produit par le conservatoire botanique contient plus d'un million d'observations géoréférencées et validées sur le plan taxonomique. Une partie de ce jeu de données contient des relevés de végétation c'est-à-dire des données de co-occurence d'espèce dans un milieu homogène. Ce volet végétation concerne environ 340 000 observations pour 17 000 relevés. Lorsque c'est possible ces relevés de végétation sont rattachés à un type de milieu naturel selon les typologies usuelles (EUNIS, Natura 2000, nomenclature phytosociologique). Ce lien entre milieux naturels et composition floristique pourrait faire l'objet d'approches de type apprentissage machine.

5. **Camille Garcin**, IMAG : Top-k loss for plant classification.

L'application Plantnet téléchargée plus de 10 millions de fois, permet aux utilisateurs d'identifier les plantes qu'ils rencontrent. Grâce aux millions d'images collectées par les utilisateurs, nous avons créé un dataset public comprenant plusieurs milliers d'espèces et plus de 100.000 images. La base de données ainsi construite présente diverses ambiguïtés : certaines classes (espèces de plantes) sont visuellement très similaires, une photo peut être partielle (par exemple seulement la fleur est prise en photo), etc. Ainsi, certaines photos pouvant vraisemblablement correspondre à plusieurs classes, il apparaît pertinent de proposer à l'utilisateur un nombre de classes possibles supérieur à 1, par exemple k=3. On a alors une erreur si la vraie classe de la plante n'est pas dans les k premières suggestions de l'algorithme. On parle de top-k error. Les réseaux de neurones sont très souvent entraînés avec une loss cross entropy, qui à première vue ne semble pas indiquée pour minimiser cette top-k error. Dans cette présentation nous présenterons des loss spécifiquement conçues pour minimiser cette top-k error et présenterons les résultats obtenus sur notre dataset de plantes.

6. **Loïc Lehnoff**, BAGAP : Identification d'espèces végétales d'intérêt écohydrologique dans les paysages agricoles : apports du deep learning et de l'IA.

Identifier des espèces végétales sur le terrain est important pour détecter précocement les mauvaises herbes, analyser la biodiversité d'un paysage ou encore caractériser les propriétés physiques des couverts vis-à-vis des écoulements en hydrologie. Cette activité a été longtemps réalisée par des botanistes via des relevés visuels, limitant de ce fait l'exhaustivité et/ou la taille des zones couvertes. Pouvoir automatiser cette activité d'identification des espèces pour l'appliquer à l'échelle d'un paysage est un véritable défi méthodologique. Un jeu de données préliminaires a été collecté sur des espèces végétales spontanées trouvées dans des fossés agricoles. Le jeu de données est constitué d'un relevé botanique cartographié et d'une campagne de mesure par proxidétection (4000 photos prises selon des angles différents). Un traitement photogrammétrique des images a permis la création d'une orthoimage à résolution sub-centimétrique. A l'aide de méthodes de classification classiques (Machine Learning) mais aussi de méthodes

de deep learning, la classification de l'ortho-image par identification des espèces des fossés est réalisée. L'intégration de données multivues est envisagée pour apporter une plus-value à la classification.

7. **Nicolas Verzelen**, MISTEA : Méthodes de Biclustering pour des inventaires de diversité avec prise en comptes de covariables.

On considère des inventaires de diversité réalisé sur différents sites. Les latent block models (LBM) permettent de réaliser simultanément une typologie (i.e. un clustering) de sites ayant des compositions spécifiques similaires et une typologie d'espèces ayant tendance à être co-occurente. Comme cette approche s'inscrit dans le cadre des modèles linéaires généralisés, on peut naturellement prendre en compte des covariables de contrôle dans l'estimation des clusters. Les fonctions du packages seront illustrés au travers d'un exemple d'inventaire de diversité cultivée sur les pentes du Mont Kenya.

15h45-16h15 : DISCUSSION *(Modération : Pierre Bonnet)*

16h15-16h30 : La Key Initiative MUSE (KIM) "Data & Life Sciences" - Jean-Michel Marin

16h30-16h45 : L'Institut de Science des Données de Montpellier (ISDM) - Anne Laurent

16h45-17h00 : L'appel à projets ANR-AFD "Challenge IA-Biodiv" - Anne Lieutaud, ANR

(Modération : Emmanuel Le Clézio)

17h00 : FIN DE LA JOURNEE

SALON DE DISCUSSION ouvert de 12h30 à 17h30
Vous pouvez accéder librement à une deuxième salle Zoom si vous souhaitez poursuivre les échanges après la fin des sessions thématiques.

Comité d'organisation : *Francois Rousset (ISEM), Julien Renoult (CEFE), Sébastien Villéger (MARBEC), Pierre Bonnet (AMAP), Corine Delesalle (Labex Cemeb), Andrea Parmeggiani (L2C), Emmanuel Le Clézio (IES), Nadine Hilgert (MISTEA), Pascal Poncelet (LIRMM), Camille Couralet (Labex NUMEV).*