Studying variations in pollen signals from the Mediterranean vegetation and its key taxa: contribution of automated pollen analysis

Betty Gimenez

THÈSE POUR OBTENIR LE GRADE DE DOCTEUR DE L'UNIVERSITÉ DE MONTPELLIER

En Écologie, Evolution, Ressources Génétique, Paléobiologie

École doctorale GAIA

Unité de recherche ISEM

Étudier les variations des signaux polliniques de la végétation méditerranéenne et ses taxons clés : apport de l'IA pour des analyses palynologiques automatisées

Présentée par Betty GIMENEZ
Le 09 Décembre 2024

Sous la direction de Odile PEYRON et Céline DEVAUX

Devant le jury composé de

Emilie GAUTHIER, Professeure, Université de Franche-Comté	Rapporteuse
Carlos JARAMILLO, Chercheur, Smithsonian Tropical Research Institute	Rapporteur
Alexis JOLY, Directeur de recherche, INRIA	Examinateur / président du jury
Ola OLSSON, Maître de conférence, Lund University	Examinateur
Céline DEVAUX, Maître de conférence, ISEM	Co-directrice de thèse
Odile PEYRON, Directrice de recherche, ISEM	Directrice de thèse
Sébastien JOANNIN, Chargé de recherche, ISEM	Encadrant / invité
Jérôme PASQUET, Maître de conférence, Université Paul-Valéry	Invité

UNIVERSITÉ DE MONTPELLIER

Résumé en français

Les changements globaux dans le bassin Méditerranéen occidental menacent ses écosystèmes et fragmentent ses habitats, menaçant ainsi les espèces qui en dépendent, telles que la vigne sauvage. Cette liane héliophile a une grande valeur écologique, archéologique et agronomique, en tant qu'ancêtre sauvage de la vigne domestiquée. Cette thèse visait à (1) décrire des communautés végétales Méditerranéennes, à partir d'un suivi de pluies polliniques mené au Pic Saint Loup (Hérault), dans une zone incluant divers habitats Méditerranéens typiques, une population sauvage de vigne, et un vignoble, et (2) comparer les signatures polliniques associées à la vigne sauvage et cultivée à partir du suivi et de références d'individus de collection (Vassal-Montpellier). Ces objectifs reposent sur l'analyse d'une grande quantité d'échantillons polliniques, et sur un grand nombre de pollen par échantillon pour mieux représenter le rare pollen de vigne dans les assemblages polliniques. Pour traiter ces données polliniques et tenter de différencier le pollen de vigne sauvage et domestiquée, l'objectif primaire de la thèse était de développer des méthodes automatiques pour l'analyse pollinique. L'analyse pollinique automatisée est un objectif de longue date en palynologie, mais au début de la thèse les méthodes n'étaient pas adaptées aux échantillons environnementaux tels que les pollen issus de pièges passifs contenant de multiples taxons, une diversité de qualités des images, et de multiples débris. Un pipeline entièrement automatisé a été développé à partir d'intelligence artificielle, et comprenant : l'acquisition d'images (Chapitre I), la détection, pour localiser les grains de pollen dans des images comprenant de multiples particules (Chapitre II), et l'identification de taxons Méditerranéens clés, y compris *Vitis*, à partir des particules automatiquement détectées (Chapitre III). Ce pipeline a ensuite été appliqué pour estimer la production et la taille des pollen d'individus de vigne sauvage et domestiquée de collection et pour tenter la classification de leurs pollen fertiles pourtant très similaires (Chapitre IV). La détection à partir de deep learning est efficace même sur des images complexes et n'a pas introduit de biais majeurs entre les taxons. Les jeux de données de détection n'étant pas optimaux pour générer des annotations adaptées à la classification, cette étape a été conduite séparément, mais dans l'objectif de traiter l'ensemble des particules détectées. L'identification des taxons clés était efficace à l'exception des taxons incluant des parents proches avec des pollen similaires (e.g. *Phillyrea* vs *Fraxinus*). Pour le monitoring, les comptages automatisés étaient en accord avec les observations de terrain, et ont mis en évidence des différences entre communautés végétales et au cours du temps. Des différences entre les signatures polliniques de la vigne sauvage et domestiquée ont été identifiées in situ, à partir de leurs cortèges floristiques et de l'abondance du pollen fertile et stérile produit par la vigne, et dans les références, avec plus de grains de plus petites tailles produits par les individus domestiques. Les modèles CNN n'ont pas différencié le pollen fertile des deux sous-espèces. Ces résultats soulignent le potentiel du suivi de la végétation pour étudier des changements spatiaux et temporels de taxons clés en Méditerranée, et invitent à une étude plus approfondie du signal pollinique de la vigne sauvage et cultivée en tant qu'outil pour retracer son histoire. Cette thèse a contribué à faire progresser la palynologie automatisée vers des méthodes applicables au réel, à rendre ces approches plus accessibles en utilisant des outils abordables et faciles d'utilisation, et en mettant en ligne les jeux de données. Ces méthodes automatiques, pour compter du pollen ou étudier des taxons spécifiques, peuvent déjà soutenir la recherche en palynologie ou plus généralement basée sur le pollen.

Court résumé en français

L'intense aridification et anthropisation en Méditerranée menacent ses écosystèmes et sa biodiversité, e.g. la vigne sauvage. Cette thèse visait à automatiser l'analyse pollinique pour suivre la végétation d'habitats Méditerranéens typiques où une population de vigne sauvage est établie, et étudier le pollen de vigne, considérant sa valeur écologique, historique et agronomique. Un pipeline entièrement automatisé a été développé pour acquérir des images, détecter et identifier des grains de pollen par intelligence artificielle. En appliquant ce pipeline, on trouve que les pluies polliniques diffèrent entre sites et années. Des différences entre la vigne sauvage et domestiquée in situ et à partir de références ont été mises en évidence, mais l'intelligence artificielle (IA) n'a pas permis de discriminer les pollen fertiles des deux sous-espèces. L'analyse pollinique automatisée peut déjà être exploitée pour soutenir des activités de recherche basées sur du pollen.

English abstract

Intense global changes are threatening the ecosystems in the Western Mediterranean basin, leading to habitat fragmentation and loss that put species at risk of extinction, e.g. the wild grapevine. This heliophilous liana, holds high ecological, archaeologic and agronomic values, being a relative of the domesticated grapevine. In that context, the thesis aimed to (1) assess spatial and temporal changes in key Mediterranean plant taxa from pollen rains, and (2) investigate differences in the pollen signature of wild and domesticated grapevines, notably given its potential for paleo-ecology. To do so, pollen samples from a vegetation monitoring conducted in an area harboring typical Mediterranean plant communities, a relic population of wild grapevines, and a commercial vineyard (Pic St Loup, France), and also from grapevine reference individuals from the Vassal-Montpellier collection were analyzed. To process the large amount of samples, and generate extensive counts to detect the scarce grapevine in pollen assemblages, the primary objective of this thesis was to develop automated pollen analysis methods. Automating pollen analysis has been a long-standing objective in palynology, but when this PhD started, developments were not fully adapted to environmental samples originating from passive traps containing an uncontrolled diversity of pollen taxa, pollen of varying preservation states, and occurring with many non-pollen particles, such as spores and debris. To process the samples from the vegetation monitoring, a fully automated pollen analysis pipeline was developed, including : image acquisition (Chapter I); detection, i.e. locating pollen grains in images containing multiple pollen and non-pollen particles using the YOLOv5 algorithm (Chapter II); classification, i.e. identifying key Mediterranean taxa, including *Vitis*, among all automatically detected particles using the ResNet algorithm (Chapter III). From the predictions of the pipeline, pollen production, pollen sizes were estimated. The pipeline was adjusted to discriminate similar pollen morphotypes of wild and domesticated grapevines from known individuals (Chapter IV). Detection proved very efficient even in noisy images (<5% error including largely covered or damaged pollen), and errors were similar across taxa, thus not introducing major biases. Classification could not be conducted jointly with detection due to inadequacy of the detection dataset for that task and was developed separately but considering the full range of detected particles. Automated classification was efficient for 15 target Mediterranean taxa, except for taxa occurring alongside relatives with similar morphotypes (e.g. *Phillyrea* vs *Fraxinus*). Automated pollen counts were consistent with field observations, and displayed distinct patterns informing on main vegetation types and on few temporal changes. Distinct pollen signatures of the wild and domesticated grapevines were observed in situ from their accompanying vegetation and the abundance

of its fertile and sterile morphotypes, and in collection, with larger production of smaller grains in domesticated individuals. CNN models did not differentiate the fertile pollen from both subspecies. These results highlight the potential of the monitoring to assess spatial and temporal changes in key Mediterranean habitats, and call for further investigation concerning the wild grapevine pollen signature as a tool to trace its history in the Mediterranean region. This thesis contributed to advance automated palynology for the analyses of environmental pollen samples, making also the automated methods accessible, affordable and user-friendly. This thesis takes part in the latest stage in developments targeting methods that can be applied to the real-world. Automated pollen data, to count pollen or monitor specific taxa, can already help pollen-based research.

Short English abstract

The Mediterranean basin is facing intense global change, threatening its ecosystems and biodiversity. In this thesis, we used artificial intelligence to analyze pollen rains in typical Mediterranean habitats, harboring the endangered wild grapevine, and to discriminate pollen of wild grapevines to that of domesticated grapevines, as this former species holds high ecological, historical and agronomical value. A fully automated pipeline was developed including automated image acquisition, automated pollen detection and automated identification. Pollen rains displayed distinct compositions informing on main vegetation types, which also varied through time. Distinct pollen signals were identified between the wild and domesticated grapevines in situ, and from reference samples, but their fertile pollen could not be differentiated. This thesis contributed to advance automated palynology for real-world applications.

Contents

Remerciements (Acknowledgements)

Et me voilà déjà arrivée au bout de ces trois années sans même l'avoir réalisé. Cette thèse est l'aboutissement d'un travail regroupant un grand nombre de personnes, de plusieurs équipes et laboratoires, et je vais donc tenter de remercier toutes les personnes qui m'ont aidées et accompagnées dans cette aventure.

Tout d'abord je tiens à remercier les membres de mon jury qui ont acceptés de relire ma thèse, de se déplacer, pour certains de loin, et participer à la soutenance, Emilie Gauthier, Carlos Jaramillo, Alexis Joly et Ola Olsson. *First, I would like to thank the jury members who accepted to read the thesis and travel to Montpellier, for some of them from far away, to attend the defense, Emilie Gauthier, Carlos Jaramillo, Alexis Joly and Ola Olsson.*

Je souhaite remercier très grandement et chaleureusement Odile Peyron et Céline Devaux, même si de simples remerciements semblent si limités pour refléter tout votre investissement, votre soutien, vos conseils, et votre accompagnement sur le plan scientifique, administratif, mais aussi au delà, que vous m'avez apporté au cours de ces trois dernières années. Je vous remercie du fond du coeur pour vos mots d'encouragements, votre écoute toujours bienveillante, vos explications et conseils, ainsi que pour la confiance que vous m'avez accordée ces trois dernières années. Votre soutien m'a permis de me dépasser et a été déterminant pour l'accomplissement de mon travail et je vous en suis extrêmement reconnaissante. Je tiens également à remercier très vivement Sébastien Joannin, ton rôle dans mon encadrement a aussi été crucial. Merci pour ton aide précieuse, tes conseils et ton engagement durant ces trois années (et pour ces heures passées à regarder du pollen!). Je vous remercie pour votre soutien et votre patience, on a traversé de multiples explorations, des moments de doute, des essais divers, des réunions intenses, car il faut le dire, cette thèse c'était bien une aventure dans l'inconnu, et ça n'a pas été une tâche facile pour aucun d'entre nous! Chacun de vos encadrements ont été très importants et complémentaires, et m'ont permis d'arriver jusqu'ici, et je vous en remercie infiniment. Je tiens aussi à vous remercier tous.tes les trois pour votre écoute, votre compréhension et votre soutien dans des moments difficiles, et qui m'ont réellement permis de mener ce projet à son terme. J'ai été vraiment chanceuse d'être encadrée par vous, autant sur le plan scientifique que personnel.

Je tiens aussi à remercier Doris Barboni, qui m'a ouvert des portes inattendues sur un monde nouveau, échanger dans un cadre académique (et venir!) en Inde a été une expérience extrêmement enrichissante. Merci pour ton accueil chez toi en France et au CEREGE, chez toi à Pondichéry et à l'IFP, et surtout pour tout tes conseils et ton aide. Je tiens aussi à remercier Luc Beaufort, Yves Gally, Thibault de Garidel-Thoron ainsi que Françoise Challie, pour avoir partagé votre expertise et l'accès au microscope automatique, et sans qui cette thèse n'aurait pas été possible. Je vous remercie aussi pour votre accueil au CEREGE. Un grand merci aussi à Benjamin Bourel pour ses conseils.
Je tiens aussi à remercier Jérôme Pasquet pour ses conseils et son regard très affûté sur la question de l'analyse d'images avec les algorithmes de deep learning, qui ont largement orientés les choix dans cette thèse.

I also thank very warmly Sourish Das, Pranabendu Misra and Pascal Weil, for your welcome at the CMI and in Chennai, for including me in the dynamics of the team during my stay and for your advice and expertise. I also thank the students, Shubhashish, Anuraj, as well as the PhD students from the CMI, thank you for your hospitality and sympathy

during my journey in Chennai. Finally, I want to thank Alka, I am very happy that I met you, I remain inspired by all your work to supporting others in scientific academia, which makes me look beyond, and I hope our paths will cross again.

Je tiens à remercier toutes les personnes investies dans le projet Pollimed sur lequel se base une grande partie de ce travail de thèse. Merci Bertrand Limier pour toutes ces conversations, ces histoires de volcans entre bien d'autres, pendant les terrains au Pic Saint Loup. Merci Sarah Ivorra pour tes photos, mais surtout tes petites, régulières, et très appréciées (!) interventions pendant l'écriture du manuscript. Merci Sandrine Canal pour ces échanges musicaux, l'atelier couture qui bien que je ne couture pas (je sais c'est pas comme ça qu'on dit) anime positivement le bureau. Merci à vous pour la collecte et le traitement des échantillons que j'ai utilisé pendant la thèse. Merci aussi à Laurent Bouby et Jean-Frédéric Terral, pour vos conseils sur la vigne et sur le suivi du Pic Saint Loup. Merci Nathalie Combourieu-Nebout pour tes retours sur le Pic Saint Loup, ton aide sur les données palynologiques, et tous tes comptages.

Je tiens à remercier très chaleureusement Marie Challe, pour toutes ces journées partagées sur le terrain, enfin les terrains, parce que des endroits pour trouver les bonnes plantes, il y en a eu! Merci aussi Agnès, pour tes conseils et tes messages d'encouragements.

Je souhaite aussi remercier Maud Tenaillon, Thierry Lacombe et Cécile Marchal, pour le matériel biologique de la vigne de Vassal, sans qui toute une partie de la thèse n'aurait pas été possible. Je remercie aussi Anna Maria Mercuri pour son accueil dans son laboratoire à Modène.

Je tiens à remercier les membres de mon comité de thèse, Nicolas Galtier, Franck Richard, Colin Fontaine, Anna Maria Mercuri et Marc Chaumont, pour vos conseils. Je remercie aussi l'école doctorale GAIA pour le financement.

Je tiens finalement à remercier tout les membres du département PAST. Merci tout particulièrement à tout ceux et celles des bureaux des doc du 21 et du 22, déjà partis ou arrivés après, Marine, Angèle, Mary, Cyprien, Léa, Adeline, Ana B., Lucas, Fiona, Juliette D., ainsi que Katy, Olga, Amandine, Reyes, Miguel, Juliette T., Marie H.-L., Fatima, Marie L., Virgina, Eva, Ana P.. Merci pour la super ambiance d'aide, et de motivation, et de détente, et de travail, et d'écoute, et la liste est encore longue donc je m'arrêterai ici, établie par tout le monde. Je suis très heureuse de vous avoir rencontrés et j'espère partager encore pleins de chouettes moments avec vous.

Je remercie tout les membres de l'administratif pour prendre soin de nos demandes, car c'est vrai qu'on peut parfois savoir écrire des pages entières de trucs sur du pollen, et galérer à remplir (ou même à trouver!) un formulaire administratif. Merci aussi Yannick et Dimitry pour m'aider à remettre en route le PC quand plus rien ne marche ! et merci Iago pour tes conseils sur python.

Un grand merci à l'ensemble du personnel de maintenance qui veille à garder nos espaces de travail et nos espaces communs propres.

Finalement, je tiens à remercier mes ami.es. Merci Lucile pour tes mots qui ne manquent jamais de me faire chaud au cœur, merci Pyrène on se suit depuis si longtemps et j'espère que ça va continuer, merci Bérangère pour tout les supers moments que l'on passe ensemble, pour toujours répondre présente pour tout évènement, qu'il soit sérieux ou détente, Sakina, j'attends avec impatience le moment où on sera de nouveau dans le même pays !!! Je suis très heureuse que notre belle bulle d'amitié continue d'exister.
Merci les copain.es parisiens qui ne sont pas (tous.tes) parisiens mais se reconnaîtront, Raphaëlle merci pour tes petits messages, et ton super colis glacés (oui il en faut peu pour me rendre heureuse), Gauthier, Stéphanie, Claire, Alexia, MS, Alexandra, merci pour vos lettres, merci pour tout les moments passés ensemble, chaque fois vous voir me redonne un max d'énergie. Bon courage Raphaëlle, et Marie (et Swann!!) pour votre parcours de thèse, vous allez y arriver!
Merci Nicolas pour nos balades, nos discussions, tout les supers moments qu'on passe ensemble. Merci Héloïse, je suis très heureuse que notre amitié se poursuive, je ne me lasse pas de nos conversations toujours passionnantes et des découvertes cinéma que j'ai faites à tes côtés!
Merci Fanny, infiniment, pour toi je n'ai pas de mots, les années durant lesquelles on s'est suivies ne se comptent plus depuis longtemps, notre rêve de petite de devenir des mamies amies et devenu un rêve d'adulte, et le jour où il se concrétisera (je suis confiante), il faudra trinquer mamie! Blague à part, tu es pour moi comme une autre sœur, j'adore tout les moments passés avec toi, des plus sérieux aux plus drôles, et je t'en remercie. Merci aussi Olivier pour tes conseils et nos échanges, bon courage pour la suite!

Je tiens finalement à remercier mes parents je vous aime très très fort, et je vous remercie d'être toujours là pour moi et de me soutenir dans mes choix. A ma sœur Ninko et mon frère Raphaël, je vous aime si fort, il n'y a rien de mieux pour moi que de passer un bon moment ensemble tous les trois, les personnes qui vous côtoient (dont moi hehe!) sont très chanceuses de vous connaître, merci d'exister!!

Parce que le plus important, ce n'est pas ce que l'on fait dans la vie, mais avec qui on le fait.

Long résumé en français

0.1 Introduction et Objectifs principaux

Le bassin Méditerranéen est écologiquement unique, avec 4.3% des espèces de plantes mondiales et un fort taux d'endémisme, il est considéré comme le troisième hotspot de biodiversité au monde (Myers et al., 2000 ; Cuttelod et al., 2008 ; International Union for Conservation of Nature, 2016). Sa biodiversité est le résultat des conditions climatiques Meditérranéennes uniques (Lionello et al., 2006), et de la longue présence humaine qui a façonné les paysages et les espèces, notamment par le biais de processus tels que la domestication (Blondel et Aronson, 1999). Les ressources naturelles du bassin Méditerranéen ont ainsi une grande importance au sein des sociétés Méditerranéennes présentes et passées, autant sur le plan économique que culturel. Aujourd'hui, cette biodiversité est sous forte pression due à un changement climatique et une aridification intense (Giorgi, 2006 ; Lionello et Scarascia, 2018), alors que ces écosystèmes sont particulièrements sensibles à ces perturbations (Giorgi et Lionello, 2008 ; Newbold et al., 2020). Ces facteurs, combinés à une urbanisation intense, entrainent une intense érosion et une fragmention des écosystèmes Mediterranéens, mettant en danger de nombreuses espèces, et faisant de la conservation des écosystèmes du bassin Méditerranéen une priorité (Myers et al., 2000 ; Cramer, 2018 ; Newbold et al., 2020). La vigne sauvage (*Vitis vinifera subsp. sylvestris*) est une liane héliophile qui dépend largement de ces écosystèmes, et dont la distribution a été drastiquement réduite au cours des deux siècles derniers du fait de maladies mais aussi de la perte de ses habitats naturels, et elle est aujourd'hui rare en Méditerranée occidentale. La vigne sauvage représente pourtant une grande importance, autant d'un point de vue écologique en tant qu'espèce présente dans les écosystèmes Méditerranéens depuis le Pléistocène, mais aussi archéologique et historique en tant qu'ancêtre sauvage de la vigne domestiquée (*Vitis vinifera subsp. vinifera*), qui a façonné les paysages, les échanges sociaux-culturels et économiques des sociétés Méditerranéennes sur des millénaires.
Dans ce contexte, les objectifs ultimes de cette thèse étaient d'étudier les changements spatiaux et temporels de taxons Méditerranéens clés à partir de l'analyse de pluies polliniques, et d'étudier les signatures polliniques associées à la vigne sauvage et cultivée. Pour cela, deux types d'échantillons ont été considérés : (1) ceux issus d'un monitoring de la végétation conduit depuis 2019 au Pic Saint Loup, dans un site caractérisé par un habitat Méditerranéen typique, hébergeant une population de vigne sauvage, incluant un vignoble commercial, et mené à partir de pièges passifs annuels, mensuels et hebdomadaires pendant la période de floraison de la vigne (projet Pollimed), (2) et ceux issus d'individus de vigne sauvage et cultivée de la collection ampélographique de Vassal-Montpellier. Dans le but de traiter et de préparer l'analyse de la grande quantité d'échantillons collectés dans le monitoring de la végétation, collectés pour avoir une meilleure représentation de la vigne, dont le pollen est rare, et pour analyser en détails les pollens de la vigne sauvage et domestiquée, l'objectif premier de la thèse a été de développer des méthodes automatiques pour l'analyse pollinique, basées sur de l'intelligence artificielle (IA). Cette thèse a donc tout d'abord consisté à effectuer des développements pour construire un pipeline entièrement automatisé pour l'analyse pollinique à partir d'un microscope automatisé et d'algorithmes de deep learning, et dans

le but d'extraire : des comptages par taxon dans des pièges passifs à pollen, ainsi qu'une quantification de la quantité de grains de pollen et de leurs tailles pour discriminer des grains morphologiquement très similaires.

Les développements ont reposé sur des outils différents, en considérant les objectifs biologiques visés pour chaque étape, et ont été conduits à partir de différents états de l'art au moment des études, sachant qu'ils changent extrêmement rapidement dans ce domaine.

0.2 Résultats principaux

0.2.1 Développements méthodologiques

Le pipeline a été développé en quatre grandes étapes qui ont été automatisées : (1) l'acquisition d'images de lames polliniques par un microscope optique automatisé ; (2) la détection des grains de pollen dans ces images à partir de l'algorithme YOLOv5 ; (3) la classification de particules détectées pour obtenir l'identification de 15 taxons clés à partir de l'algorithme ResNet152 ; (4) l'intégration de l'ensemble de ces étapes pour obtenir un méthode applicable à de nouveaux échantillons.

0.2.1.1 Acquisition d'images

La méthode d'acquisition d'images se base sur un microscope optique équipé d'une platine motorisée et d'une caméra, tous deux contrôlés par un logiciel développé au CEREGE par des collaborateurs. Ce système d'acquisition consiste à photographier plusieurs champs de vue, et pour chaque champ de vue de prendre une pile d'images à différents focus, images qui sont ensuite combinées pour former une image composite. Différents tests ont été effectués pour développer des méthodes afin d'obtenir des lames polliniques adaptées à ce système d'acquisition, et d'ajuster les paramètres d'acquisition pour obtenir des images des grains de pollen présents dans les lames. Pour le montage des lames, le protocole final se base sur des lames fixes montées à partir de glycérine gélifiée. Pour l'acquisition des images, et pour chaque champ de vue, 13 images sont prises, espacées de 8 µm chacune, correspondant à un compromis pour obtenir un nombre suffisant de niveaux de focus par grain de pollen, pour capturer des images nettes de grains de pollen se trouvant à plusieurs niveaux de profondeurs dans les lames, tout en limitant le nombre d'images à acquérir pour chaque champs de vue, et qui contribue largement au temps d'acquisition et à l'espace de stockage nécessaire pour chaque lame.

0.2.1.2 Détection des grains de pollen dans des images issues d'échantillons environnementaux

Les développements sur l'étape de détection ont été guidés par deux constats faits à partir de la littérature au moment de l'étude : (1) les développements sur la détection restaient rares et les méthodes n'étaient pas encore adaptées à des images bruitées et typiques dans les échantillons environnmentaux contenant de multiples grains de pollen, d'autres particules telles que les débris ou les spores, mais aussi des particules superposées. Pourtant, les erreurs générées par la phase de détection se propagent dans un pipeline entièrement automatisé, et cette étape a donc le potentiel de biaiser les données polliniques générées automatiquement. (2) Les méthodes automatiques proposées dans la littérature restaient limitées à certains chercheurs, et avaient un accès et une prise en

main difficile, notamment pour les palynologues, souvent non-experts en machine learning, et ainsi limitant le potentiel de ces développements pour la recherche en palynologie. Dans ce cadre, la première étude visait à développer une méthode de détection applicable aux images issues des pièges passifs à pollen, évaluée pour extraire des données polliniques utilisables pour adresser des questions en écologie, basée sur des outils accessibles, et présentée de façon à être prise en main facilement par d'autres chercheurs intéressés par les études basées sur du pollen.

Dans ce but, l'algorithme YOLOv5 a été sélectionné, du fait de ses performances sur d'autres objets d'études (e.g. cellules anormales dans des images médicales, bétail dans un champs) et de sa prise en main facilitée par une interface dédiée et accompagnée de guidelines (github.com/ultralytics/yolov5, Jocher, 2020). Il s'agit d'un convolutional neural network (CNN) supervisé, son utilisation repose donc sur une phase d'entrainement et un jeu de données d'images annotées et représentatives de l'application visée. Dans ce cas, le jeu de données d'entrainement correspondait à des images issues des acquisitions par le microscope automatique (champs de vue de 214 µm × 214 µm) de lames montées à partir des pièges à pollen annuels. Les annotations des images ont consisté à placer des boîtes englobantes autour des grains de pollen, à partir du logiciel LabelImg (T. Lin, 2015). Dans l'objectif d'explorer les façons les plus optimales d'utiliser cet algorithme, e.g. pour répondre à divers questionnements à partir des mêmes données polliniques, plusieurs façons d'annoter les images ont été testées incluant : (1) l'annotation de tous les grains de pollen sans indiquer d'information taxonomique les concernant, ce qui permet à l'algorithme de détecter les grains de pollen et de les différencier des autres particules, (2) l'annotation de tous les grains de pollen en indiquant l'information taxonomique pour 5 taxons clés dans les échantillons (Pinaceae, Oleaceae, *Buxus*, Poaceae, *Quercus*) les autres étant annotés sans information taxonomique, ce qui permet à l'algorithme de détecter tous les pollen et d'effectuer une étape de classification, et (3) l'annotation des grains de pollen appartenant à un ou plusieurs taxons clés uniquement, ce qui permet à l'algorithme de générer uniquement des comptages polliniques associés à ce ou ces taxons. Pour chaque stratégie d'annotation, les erreurs de détection ont été comparées, incluant une analyse des grains de pollen non détectés et des particules autres que du pollen ou non visées et faussement détectées. Les erreurs de détection ont aussi été analysées en fonction des caractéritstiques des grains de pollen (taxon et donc morphologie, ainsi que la qualité des images), afin d'évaluer les causes des erreurs et d'estimer les potentiels biais par taxon entre les assemblages polliniques détectés et annotés dans les images.

Un total de 4 096 images issues de 16 pièges passifs annuels collectés durant trois années et sur six sites distincts a été annoté, générant un total de 12 531 annotations de grains de pollen et de *Lycopodium*, utilisé pour la calibration des comptages ; ces annotations ont été utilisées pour entrainer et évaluer les modèles. L'annotation la plus simple a obtenu les meilleures performances de détection, avec 5% de grains non détectés, et 5% de particules non polliniques faussement détectées. Les erreurs de détection ont principalement concerné les grains de pollen avec une mauvaise visibilité (ex. couvert par des débris), sachant qu'uniquement 1% des grains de pollen bien visibles n'étaient pas détectés. Les erreurs de détection étaient aussi similaires entre taxons, à l'exception du *Lycopodium* et du pollen de Pinaceae, qui a une forme se distinguant des autres pollens annotés, et qui étaient détectés un peu plus fréquemment. Les erreurs pour ces taxons restant relativement faibles (<1.3% pour le *Lycopodium*, <3% pour le pollen de Pinaceae, en comparaison de <0.2% pour les autres taxons). Le fait d'annoter les images avec davantage de détails taxonomiques n'a pas amélioré les performances de détection mais a

pris davantage de temps pour construire le jeu de données. Cela a cependant permis la classification des taxons annotés ayant une morphologie typique, par exemple, les grains de *Lycopodium*, *Buxus* et Pinaceae, qui étaient identifiés avec plus de 90% de réussite. Les autres taxons avec des morphologies plus semblables aux autres grains de pollen étaient moins bien identifiés avec cette approche, confirmant le fait de conduire une étape de classification indépendante pour identifier l'ensemble des taxons détectés à cette étape. De même, annoter uniquement un ou certains taxons dans les images s'est montré uniquement efficace pour les taxons avec une morphologie spécifique par rapport aux autres pollen présents dans le jeu de données. Finalement, la méthode de détection des grains de pollen était efficace quand appliquée sur des images issues des pièges à pollen collectées durant de nouvelles années. L'ensemble de ces tests et de leurs résultats ont finalement permis de proposer quelques conseils d'utilisation, le principal étant qu'il est plus efficace en terme de temps et de performance d'utiliser cette méthode uniquement pour la détection et d'opérer l'étape de classification en une deuxième étape distincte (Gimenez et al., 2024).

0.2.1.3 Classification des grains de pollen issus d'échantillons environnementaux

L'objectif pour le suivi de végétation du Pic Saint Loup a été d'analyser les pluies polliniques totales, et les pluis polliniques pour 15 taxons Méditerranéens clés, incluant (1) *Quercus ilex*, *Quercus pubescens*, *Buxus*, *Juniperus*, *Cupressus*, *Pistacia*, *Phillyrea*, *Olea*, et *Pinus*, en tant qu'espèces clés de la guarrigue Méditerranéenne ouverte et fermée, (2) *Poaceae* et *Plantago* en tant qu'indicateurs de milieux plus ouverts et herbacés, et antropisés pour *Plantago*, (3) *Fraxinus* (*ornus* and *excelsior*) en tant qu'indicateurs de conditions plus humides, et finalement (4) le pollen de *Vitis*, considérant les deux morphotypes stériles et fertiles produits par cette espèce. L'objectif était donc de développer une méthode capable de traiter l'ensemble des particules générées par la méthode de détection précédente et appliquée aux images issues des pièges passifs à pollen pour en extraire les identifications associées aux taxons clés. L'ensemble des particules détectées par la méthode de détection appliquée aux échantillons environnementaux considérés (pièges annuels passifs) comprend une grande diversité de taxons polliniques, issus d'une diversité non contrôlée, des grains de pollen pouvant apparaître dans de multiples états (abîmés, couverts par des débris, flous), et parfois avec des critères d'identification non visibles, et des particules non polliniques faussement détectées telles que des débris ou des spores.

L'algorithme de classification supervisé ResNet152 a été sélectionné pour traiter la grande diversité d'images produites (He et al., 2015). Afin d'obtenir des modèles capables de traiter l'ensemble des particules considérées, la stratégie a été de construire un jeu de données d'entrainement incluant au maximum la diversité des particules (multiples taxons, mais aussi debris, spores, bulles d'airs) attendues lors de l'application sur l'ensemble des pièges à pollen du monitoring, et incluant un grande nombre d'images pour chaque type de particules attendues. Dans ce cadre, les particules détectées à partir des images issues d'une partie des échantillons du monitoring ont été vérifiées manuellement, identifiées et classées par taxon. Une classe pour les particules non polliniques, deux classes pour les grains de pollen indéterminables car recouverts par d'autres particules ou flous dans les images ont aussi été crées. Ensuite, des images issues de lames de référence, principalement de la collection de palynologie de l'ISEM ont été ajoutées en vue d'augmenter le nombre d'images par classe, et d'ajouter des classes pour des taxons attendus mais non trouvés dans les images, ou difficiles à identifier du fait de la présence d'autres taxons similaires dans les images (ex. *Betula* ou *Corylus* en vue polaire). Le jeu

de données de classification était donc finalement composée d'images dites 'environnementales' (issues des acquisitions des pièges à pollen du monitoring), des images dites de référence (issues de la collection), et de 83 classes pour un total de 56 101 images. Dix modèles indépendants ont été entrainés (cross-validation) sur ce jeu de données incluant images environnementales et de référence. Les performances de la classification ont ensuite été évaluées uniquement sur des images environnementales afin d'obtenir une évaluation plus représentative des conditions attendues pour le monitoring. L'évaluation a été conduite en deux étapes, tout d'abord en attribuant comme classe prédite pour un grain de pollen uniquement celle avec le meilleur score (top-1), et ensuite en attribuant comme classes prédites pour un grain de pollen celles avec les quatre meilleurs scores.

Les taxons clés ont été bien reconnus par les modèles de classification avec des performances au delà de 90% pour plusieurs des taxons clés (*Lycopodium*, *Buxus*, Poaceae, *Plantago*, *Pistacia*, *Vitis*). Les erreurs de classification étaient principalement causées par la confusion entre taxons issus de la même famille et présentant des morphotypes polliniques similaires. Par exemple, *Phillyrea* et *Fraxinus* (Oleaceae) présentaient chacun autour de 10 à 40% de confusions entre ces taxons, et de même pour *Cupressus* et *Juniperus* (Cupressaceae). Considérer, pour chaque image, les quatre meilleures classes prédites, et pas seulement la meilleutre classe prédite, a permis d'obtenir la classe correcte pour 98.8% des images.

0.2.1.4 Application du pipeline entièrement automatisée et application pour obtenir des comptages polliniques

Ce procédé a permis de générer 10 modèles cross-validés, entrainés pour l'identification de 83 classes et générant la bonne classe par les 4 premières predictions pour la très majorité des images. Dans le développement du pipeline automatisé complet, pour chaque image, les quatre premières prédictions générées par les 10 modèles ont donc été combinées afin d'obtenir l'identification de l'image. Le choix, à partir des $4 \times 10 = 40$ classes générées par ce processus, a consisté à selectionner la classe qui a été générée le plus fréquemment, et à ajouter ensuite un filtre pour exclure les classes prédites avec moins de 9 occurences et alors considérées comme incertaines, limitant ainsi les fausses attributions.

Ce pipeline entièrement automatisée, incluant l'acquisition d'images, la détection, la classificaion et l'intégration des prédictions de 10 modèles de classification cross-validés, a ensuite été appliqué à l'ensemble des images générées à partir des échantillons du Pic Saint Loup collectés sur 5 années dans 6 sites. Pour chaque échantillon, les comptages ont été calibrés à partir du nombre de grains de *Lycopodium* identifiés. Les données générées automatiquement ont tout d'abord été comparées avec des comptages effectués par une palynologue au microscope, sur 13 échantillons en communs, mais sur des lames différentes du fait du type de montage différent entre les deux méthodes. Les comparaisons entre les deux méthodes de comptage ont mis en évidence des résultats cohérents lorsque les taxons étaient comparés au niveau de la famille, mais également des résultats parfois différents, comme par exemple pour *Quercus* qui était sous-estimé dans les comptages automatiques par rapport aux comptages fait par l'experte. Lorsque les taxons étaient comparés à l'espèce, notamment pour les taxons d'Oleaceae, de grandes différences ont été mises en évidence entre les comptages de *Fraxinus* et *Phillyrea*, comme attendues des erreurs de classification observées dans les matrices de confusion évaluées sur le jeu de données annoté.

Les données ont ensuite été analysées au regard des observations faites sur le terrain, tout d'abord en comparant les assemblages polliniques moyens sur les années collectées et obtenus sur chacun des sites du monitoring. Différentes compositions des assemblages polliniques, et des abondances relatives de chaque taxon ont été identifiées et correspondaient avec le type de végétation sur chaque site, avec davantage de pollen d'arbres dans les sites situés en bordure ou au sein de la forêt de chênes, et davantage de pollen d'herbacées au sein du vignoble. De même, les taxons les plus abondants dans chacun des assemblages polliniques correspondaient aux mêmes taxons abondants retrouvés sur le terrain à côté de chaque piège à pollen, et incluaient des taxons indicateurs tel que *Plantago* retrouvé en large abondance sur un piège en bordure de vignoble, *Fraxinus* retrouvé en majorité dans les échantillons collectés près d'une zone ripicole intermitente, et *Phillyrea* retrouvé en majorité dans le site de bordure forestière et de guarigue Méditerranéenne sur le pente du massif du Pic Saint Loup, et *Vitis* retrouvé en majorité au centre du vignoble. Des changements dans les assemblages polliniques au cours du temps ont aussi été identifiés et validés par des observatiosn de terrain, tels qu'une grande réduction du pollen de *Buxus* sur un site, en lien avec la Pyrale du Buis qui décime cette espèce. La comparaison entre les assemblages polliniques générés par la méthode automatique et les observations de terrain indique que la méthode développpée permet déjà de refléter les tendances générales dans les assemblages polliniques présents dans les échantillons environnementaux étudiés.
Toutefois, des erreurs ont pu être mises en évidence et montrent les limites de la méthode automatique à partir des développements effectués pour le moment. Par exemple, dans les sites où le pollen de *Phillyrea* était identifié en abondance dans les assemblages polliniques, le pollen de *Fraxinus* était aussi identifié en proportion non négligeable, et vice versa, ce qui est le signe de possibles erreurs de classification de la méthode automatique et confirmées par la matrice de confusion). De même, certains taxons étaient observés en faible quantité sur des sites où ils n'étaient pas attendus : ce pourrait être du à leur présence rare mais aussi à un faux signal généré par la méthode automatique, empêchant, pour le moment, l'analyse de ces signaux.

0.2.1.5 Adaptation de la méthode automatique pour comparer des échantillons de pollen de collection de la vigne sauvage et cultivée

La méthode précédemment développée avec YOLOv5, permettant la détection des grains de pollen et des spores de *Lycopodium* et de leurs identifications respectives, a ensuite été exploitée pour l'analyse des échantillons de référence de la vigne et ainsi estimer la quantitié de pollen produit par des fleurs d'individus des deux sous-espèces *Vitis vinifera subsp vinifera* et *Vitis vinifera subsp sylvestris*. Ensuite la taille des objets prédits par le modèle de détection a été utilisée en tant que proxy de la taille des grains de pollen, afin d'évaluer des éventuelles différences entre les deux sous-espèces. Finalement, les images détectées ont ensuite permis de constituer un jeu de données de classification, afin de tester la possibilité d'utiliser l'algorithme de classification ResNet152 pour différencier le pollen fertile produit par la vigne sauvage et par la vigne cultivée.

0.2.2 Exploitations des développements pour étudier les variations dans les pluies polliniques et les signaux polliniques associés à la vigne sauvage et cultivée

Le monitoring est encore récent et la quantité des données polliniques accumulées reste encore trop faible pour évaluer des tendances dans les changements des assemblages polliniques, et de les mettre en liens avec des perturbations climatiques. Ceci étant, le monitoring a déjà permis de mettre évidence des variations entre les assemblages polliniques issus de sites d'échantillons différents, et malgré leurs proximités géographiques (tous les sites sont dans un rayon de 1 km), ainsi que des variations temporelles telles que la perte de *Buxus*. Des pics dans les quantité de Quercus sur certaines années et communs à l'ensemble des sites ont aussi été mise en évidence. La quantité de pollen totale produite a montré une faible augmentation entre les années 2020 et 2023, mais ces changements restent faibles et un plus long suivi est nécessaire pour mieux estimer les tendances, notamment sachant qu'une bien plus grosse quantité de pollen avait été collecté dans les pièges en 2019.

Les assemblages polliniques permettent d'identifier certaines différences entre la vigne sauvage et domestiquée. Tout d'abord, les différences observées dans les assemblages polliniques entre les sites informent sur la végétation associée à la vigne. Ensuite, des différences ont été observées sur le signal pollinique du pollen de vigne uniquement, notamment une bien plus grande quantitié de pollen de vigne trouvée en centre du vignoble en comparaison des sites en bords de vignoble et près des individus sauvages, et, une plus grande abondance du pollen stérile de vigne trouvé sur le site près de l'individu sauvage femelle, et obtenu dans des quantités similaires au pollen fertile de vigne. L'analyse des échantillons de référence a montré une plus grande production de grains de pollen de plus petites tailles chez les individus de vigne domestique en comparaison des individus de vigne sauvage. Le modèle de classification n'a en revanche pas différencié les images des pollen fertiles des sous-espèces.

0.3 Discussion générale

0.3.1 Discussion sur les développements méthodologiques

Le pipeline de reconnaissance automatisé développé pendant cette thèse a permis une utilisation double : (1) pour traiter des échantillons issus de pièges passifs à pollen collectés sur le terrain et effectuer un suivi de végétation en Méditerranée, et notamment de la vigne sauvage et domestiquée, et (2) pour traiter des échantillons de référence d'individus de collection de vigne sauvage et cultivée, et comparer la production pollinique et les morphotypes polliniques associés à la vigne sauvage et domestiquée. Ces deux axes - obtenir des comptages par taxon dans des échantillons environnementaux, et analyser en détail du pollen de référence et son morphotype - correspondent à des axes de recherche fondamentaux et complémentaires en palynologie.

En 2021, lorsque cette thèse a commencé, les méthodes pour la reconnaissance automatique permettant de générer des comptages polliniques sur des échantillons tels que ceux du monitoring, i.e. des échantillons dits 'environnementaux' avec beaucoup de taxons polliniques et de multiples débris et particules non polliniques, étaient encore rares et ne proposaient pas de méthodes entièrement automatisées. Des premiers développements avaient été effectués sur des images de pollen issues d'échantillons

collectés sur le terrain, telles que du pollen fossile (Bourel et al., 2020), ou collecté sur des insectes (Olson et al., 2021). Ces études correspondent à une nouvelle étape d'importance, après les développements sur des images issues de lames de référence. Cela-dit, de nombreux verroux étaient encore à présents, car ces études se concentraient sur des jeux de données largement manipulés pour éliminer la plupart des verrous liés à l'application de terrain, en reconstruisant des jeux de données présentant une diversité taxonomique contrôlée, sans débris ni pollens indéterminables et sans évaluation détaillées des méthodes lorsque ces cas étaient rencontrés. Durant ces trois dernières années, la discipline a cependant subit une grande mutation avec un déplacement des axes de recherche vers la reconnaissance automatisée pour des échantillons de terrain, tels que rencontrés lors des applications réelles, ou de routine en palynologie, pour l'analyse d'échantillons de carottes fossiles (Barnes et al., 2023 ; Teuerkhauf et al., 2023, Durand et al., 2024), et pour le monitoring de la végétation, tout d'abord avec une étude présentant encore quelques manipulations sur les jeux de données d'évaluation (Punyasena et al., 2022), et dans le cadre de cette thèse (Gimenez et al., 2024-chapitre II, et chapitre III). Ce changement de cadre d'étude pour la reconnaissance automatisée a aussi largement été marqué l'aéropalynologie pour l'allergologie, pour laquelle les études se sont multipliées depuis 2022 (Levetin et al., 2023 ; Tummon, Bruffaerts et al., 2024). Les travaux de cette thèse sont donc en adéquation avec les développements et les recherches en cours dans la discipline à l'échelle mondiale, et sont en phase avec la transition qui s'opère aujourd'hui, vers des méthodes applicables dans le réel pour générer des données polliniques utilisables, et utilisées, au delà du développement de la méthode.

0.3.1.1 détection

Dans le cadre de la méthode de détection développée (Gimenez et al., 2024 ; Chapitre II), deux constats principaux faits à partir de la littérature ont orienté les développements. Premièrement, l'étape de détection était très souvent occultée dans les études conduites jusqu'alors, ou alors son évaluation n'était pas conduite, ou conduite de façon très succincte. Pourtant, les erreurs générées par l'étape de détection se propagent aux étapes suivantes dans le cadre d'un pipeline automatisé complet. Les travaux pour la détection des grains de pollen étaient ainsi en retard par rapport à ceux effectués pour leur classification. Les développements restaient aussi majoritairement effectués avec des algorithmes de machine learning standard (sans deep learning), et ne permettant pas d'atteindre des performances optimales pour une utilisation sur des échantillons de terrain, tels que des pièges passifs à pollen ou des sédiments lacustres, typiques en écologie ou paléoécologie. Deuxièmement, les études présentées dans la littérature étaient souvent axées sur la présentation des métriques obtenues, et proposaient souvent des méthodes décrites de façon assez générale, parfois avec des algorithmes modifiés, et ainsi sans réelle possibilité pour des non-experts de machine learning de prendre en main ces méthodes. La première étude a donc été menée dans l'optique de pallier ces verrous identifiés sur l'étape de détection mais aussi de rendre plus accessibles les méthodes d'analyse pollinique automatisées pour l'ensemble de la communauté des palynologues. Pour cette première étude, il s'agissait de développer la méthode de détection et de conduire une analyse détaillée des erreurs générées, afin de donner une meilleure compréhension de la qualité des données polliniques générées par la méthode dans son ensemble. Il s'agissait dans un second temps de proposer une méthode simple d'utilisation, accessible et accompagnée de quelques guidelines dans l'optique d'une prise en main de ce type de méthodes par la communautés des palynologues, et ne reposant pas sur le besoin de grandes connaissances en informatique. Dans cette optique, l'algorithme de détection YOLOv5 a été sélectionné

(Jocher, 2020). Cet algorithme est utilisable facilement grâce à une interface développée à cet effet (github.com/ultralytics/yolov5) et il est accompagné de guidelines accessibles pour son utilisation générale. L'étude menée dans le cadre de la thèse a ensuite visé à explorer les performances et les potentialités d'utilisation de cet algorithme pour une utilisation spécifique en palynologie appliquée aux sciences environnementales, et plus particulièrement pour l'analyse d'images obtenues à partir des échantillons du monitoring de végétation. Les diverses expérimentations menées ont été partagées par le biais d'une publication, illustrant différentes façons d'utiliser cet algorithme pour une utilisation en palynologie, et associées à des performances obtenues sur notre jeu de données et de quelques guidelines formulées à partir des tests menés. Notamment, l'objectif initial avec cet algorithme était d'automatiser directement l'étape de détection et de classification, d'identification de l'ensemble des taxons polliniques présents dans les pièges, car ces deux étapes sont effectuées conjointement par cet algorithme. Finalement, les expérimentations menées ont mis en évidence que l'étape de classification ne pouvait pas être automatisée de cette façon pour ce type de données du fait des difficultés à construire et annoter un jeu de données adapté. Finalement, le jeu de données d'images polliniques annotées en considérant uniquement la détection, et la classification uniquement pour quelques taxons très abondants et caractéristiques, ainsi que les modèles entraînés ont aussi été rendus accessibles (at doi : 10.5281/zenodo.11126431). La méthode développée dans cette étude est utilisable de façon primaire pour la détection. Cela dit, elle est aussi déjà adaptée pour l'identification de certains taxons polliniques. Par exemple, dans le cas de l'analyse d'échantillons de pollen de référence, elle est utile pour quantifier la production de grains de pollen par fleur en différenciant du pollen du *Lycopodium*, comme exploitée dans le chapitre III. La méthode de détection développée s'est aussi montrée efficace pour la détection de l'ensemble des pollens dans les échantillons environnementaux complexes, tels que les pièges passifs annuels, et elle permet aussi de donner des comptages du total pollinique à partir d'échantillons environnementaux complexes, permettant déjà d'informer sur la production pollinique d'un écosystème donné. Sa capacité de généralisation a peut-être été sur-évaluée, sachant que toutes modifications dans les conditions d'acquisition des images par exemple, i.e. de différences entre les données d'entraînement et de tests par rapport aux données d'application, peuvent affecter les performances de la méthode.
L'objectif de proposer une méthode facilement utilisable par d'autres, menée dans cette thèse, surtout dans le cadre de la première étude (Gimenez et al., 2024, chapitre III), est en accord avec un effort commun visant à proposer des méthodes en ligne, accessibles par tous et toutes pour obtenir des analyses polliniques automatisées. Dans ce cadre, divers outils ont été proposés en ligne, comme l'outil Tofsi-POST (Trainable object finder, selector and identifier for pollen, spores and other things) proposé récemment et qui permet la préparation des jeux de données d'entraînement, d'entraîner les modèles et de les appliquer sur de nouvelles données (Theuerkauf et al., 2023). De façon plus générale, la détection a fait l'objet de multiples développements durant ces deux dernières années, permettant de combler les limites initialement observées au début de cette thèse et mettant l'utilisation des algorithmes de deep-learning, et notamment les 'object-detection algorithms', au rang de méthode standard et qui remplacent maintenant largement les méthodes de machine-learning utilisées auparavant.

0.3.1.2 classification multi-classes

La deuxième étape méthodologique de la thèse a ensuite consisté à mettre en place l'étape d'identification des particules détectées pour obtenir les comptages polliniques par taxon. La méthode de classification a tout d'abord été développée dans le cadre du suivi des

pluies polliniques Méditerranéennes et dans l'optique d'obtenir les comptages pour une sélection de 15 taxons. Cette méthode d'identification repose sur un algorithme de classification standard (ResNet152) et l'intégration de plusieurs modèles au sein du pipeline complet de l'analyse pollinique, i.e. incluant l'acquisition des images, la détection et l'obtention des comptages polliniques par échantillon collecté dans le cadre du monitoring. La méthode visait donc à traiter l'ensemble des cas pouvant être rencontrés dans ce type d'application, prenant en compte les cas complexes, par exemple, les images générées avec une mauvaise qualité du fait du système d'acquisition d'images automatisé (eg. images floues), les images de particules n'étant pas du pollen du fait d'erreurs de la détection automatisée, ou encore les images de pollen de taxons non attendus du fait de l'application sur de nouveaux échantillons de pluies polliniques collectées pendant une année entière au sein d'un écosystème avec une grande diversité taxonomique en Méditerranée. En raison de la récente évolution de l'intérêt de la communauté pour les méthodes applicables au réel, les méthodes de classification qui visent à traiter des taxons clés parmi des pollen de taxons inconnues ou indéterminables et des débris sont le sujet de recherches en cours très intenses ces dernières années, en allergologie (Zhuang et al., 2022), mais aussi pour des disciplines de sciences environnementales et notamment pour la paléoecologie (Theuerkauf et al., 2023 ; Durand et al., 2024). En dehors de l'allergolgoie, les méthodes n'étaient jusqu'à présents pas encore prêtes pour les utilisations de routine dans le réel, la deuxième étude proposée dans cette thèse sert ainsi de preuve de concept sur le fait que les méthodes sont maintenant prêtes à être utilisées. L'étude menée dans le cadre de cette thèse est ainsi à l'intermédiaire entre l'utilisation et la preuve de concept pour ce type d'utilisation réelle. En effet, dans cette étude, la méthode développée est directement employée pour récolter l'ensemble des données polliniques collectées par le monitoring. Les données générées automatiquement sont aussi analysées pour effectuer une analyse spatio-temporelle des pluies polliniques de l'écosystème étudié.

Ainsi les méthodes permettent maintenant de gérer les difficultés liées à la diversité des particules et à la qualité des images rencontrées dans les applications entièrement automatisées pour des échantillons issus du terrain. Dans le cadre d'une utilisation réelle entièrement automatisée, un dernier verrou a maintenant été identifié pour ce type d'application et pour laquelle la communauté a encore à adresser et à résoudre. Ce verrou est lié aux limites des capacités de généralisation des algorithmes de deep learning supervisés, pour une application au delà des données d'entraînement et de tests. Dans la première étude de cette thèse qui explorait l'étape de détection (chapitre II), nous avons analysé les capacités de généralisation de l'algorithmes à des images issues du même monitoring, générées avec les mêmes protocoles et matériels, collectés sur les même sites mais collectés d'autres années et non incluses dans les données d'entrainement. Les résultats avaient mis en évidence une bonne capacité de généralisation dans ces conditions, i.e. avec la seule variable année de collecte et où tous les autres facteurs étaient similaires (traitement chimique, milieu de montage, microscope pour l'acquisition). Cependant, l'extrapolation de la méthode de détection sur des images issues de nouvelles acquisitions par un autre matériel a généré davantage d'erreurs (transfert de la méthode du chapitre II au chapitre III), reflétant tout de même les limites de la généralisation de la méthode en dehors des jeux des données d'entraînement et de tests. Par ailleurs, les capacités de généralisation des modèles de deep learning ont aussi affecté les résultats, et affectent donc la méthode, dans le cadre de l'étape de classification. Cela a limité l'optimisation de l'utilisation des images issues des lames de référence de la collection de palynologie de l'ISEM (chapitre III), qui n'ont pas toujours permis d'améliorer l'identification des pollen des même taxons mais retrouvés dans les pièges à pollen, i.e. obtenues avec un autre traitement chimique, montées sur des lames avec un protocole différent, et conservées depuis plus ou moins longtemps. Ainsi des petites différences sur la

qualité finale des images, que ce soit du pollen mais aussi du contraste de l'image par exemple, peuvent faire par exemple qu'utiliser des références de collections de palynologie n'est pas toujours efficace. De même, la distribution des classes dans le jeux de données d'entraînement affecte également les méthodes développées par deep learning, et ainsi va tendre à influencer les résultats dans les échantillons de terrain pour lesquels la méthode de classification est appliquée, et pour lesquels les distributions des espèces peuvent varier au cours de temps ou entre différents sites d'échantillonnage. Ces problèmes sont bien sûr connus pour les développements du deep learning lorsque l'on passe à des utilisations réelles, et ont notamment été largement traités pour l'utilisation pour l'imagerie médicale. Des méthodes sont ainsi en cours de développement pour gérer ce problème, nommées sous le terme de 'domain adaptation' dans la littérature anglophone (Guan et Liu, 2022 ; Farahani et al., 2021 ; Kumari et Singh, 2024), et proposent des solutions qui passent notamment par l'utilisation d'autres algorithmes, plutôt que des traitements de calibration effectuées post-hoc dans d'autres études (Guiot et al., 2023), notamment en palynologie (Olsson et al., 2021 ; Punyasena et al., 2022), ou qui visent à gérer les déséquilibres de classes, mais restent dans une configuration pour laquelle les prédictions du modèles sont tirées dans une distribution qui n'est pas forcément celle rencontrée dans le réel.

0.3.1.3 classification binaire de morphotypes similaires

Le même pipeline et la même méthode générale de classification (acquisition, détection et classification par CNN) a aussi été utilisé pour comparer des grains de pollen avec des morphotypes très similaires. L'approche testée a été simplement d'entraîner un algorithme de classification CNN à partir de deux classes, une classe pour le pollen de vigne domestiquée et une classe pour le pollen de vigne sauvage (morphotype fertile uniquement). Cette méthode n'a pas permis de différencier les grains de pollen produits par les deux sous-espèces de vigne. Ce résultat peut mener à deux hypothèses. Soit les pollens ne sont pas différenciées morphologiquement - cela est discuté dans la section suivante sur les résultats biologiques -, soit la méthode n'était pas adaptée pour capturer des différences présentes de façon très fines. En général les études qui s'intéressent à la description fine de morphotypes polliniques se basent sur de la microscopie électronique qui permet d'observer de multiples détails ; c'est souvent à partir de ce type d'images que les différents types de textures (folvéolé, rugulés) ont été observés pour des pollens de *Vitis* ou des Vitaceae (Cartaxo-Pinto et al., 2017 ; Gallardo et al., 2009 ; İnceoğlu et al., 2000 ; Marasali et al., 2005 ; Mercuri et al., 2021, bien qu'aucune étude n'a pour le moment comparé les deux sous-espèces de *Vitis vinifera* (incluant *Vitis vinifera subsp sylvestris* et *Vitis vinifera subsp vinifera*). Par ailleurs, la discrimination ou l'analyse détaillée des morphotypes polliniques par intelligence artificielle est aussi souvent utilisée par d'autres techniques de microscopie que la micrscopie optique, comme détaillé plus haut, comme la microscopie confocale particulièrement favorable pour ce genre d'approches (Adaïmé et al., 2024 ; Collevatti et al., 2024 ; Kong et al., 2016 ; Romero et al., 2020). Finalement, d'un point de vue méthodologique, il est possible que la classification de simples images 2D directement à partir de neurones conventionnels et d'une approche supervisée peut aussi être une méthode qui peut-etre remplacée par d'autres approches plus optimales, telles que l'utilisation de piles d'images avec différents types d'assemblages implémentées dans un CNN (testé avec de la microscopie confocale, Romero et al., 2020).

0.3.1.4 limites

La méthode d'analyse pollinique automatisée développée se base sur l'acquisition d'images en microscopie optique. La méthode d'acquisition repose sur un microscope standard, équipé d'une platine automatisée à laquelle est associée une caméra, et pour chaque champs de vue, l'acquisition de plusieurs photographies à différentes profondeurs, ou focus, sont ensuite compilées. Ce processus génère des images en niveaux de gris, largement grossies, et pouvant contenir plusieurs objets microscopiques chacune (pollen, spores). Ce système à l'avantage de se baser sur un microscope optique standard, et donc davantage accessible par rapport à d'autres outils exploités dans d'autres études, telles que, en microscopie optique, les scanners de lames (ex. Nanozoomer, Punyasena et al., 2022), ou encore d'autres techniques de microscopie telles que la microscopie électronique (C. Li et al., 2023, la microscopie en flux (Barnes et al., 2023), la microscopie par fluorescence (Brdar et al., 2023), la microscopie confocale (Adaïmé et al., 2024). La résolution obtenue avec ce système d'acquisition (optique $\times 63$ à immersion), utilisée tout au long de la thèse, permet de générer des images avec une bonne visibilité des caractéristiques morphologiques des grains pour de la microscopie optique, ce qui a notamment facilité la construction des jeux de données d'entraînement, qui reposent sur des identifications manuelles faites à partir de l'observation des pollen dans les images. Cependant, les images sont compressées lors de leurs implémentations par les algorithmes, et donc la résolution n'est pas exploitée pour l'analyse automatique des images. Par ailleurs, les grains de pollen ont une taille autour de 10-60µm en moyenne, auquel s'ajoute certains larges taxons tels que les pollen de Pinaceae qui peuvent dépasser les 100µm, alors que les images ne couvrent un champs de vue que de 214µm $\times$ 214µm. Ainsi, une grande quantité des grains de pollen se retrouvent coupés sur les bords des images acquises avec cette méthode d'acquisition. Ce système d'acquisition est très coûteux en temps et en espace de stockage, il nécessite environ 24h pour scanner une lame complète et plusieurs centaines de Giga pour la sauvegarde des images, ce qui pose une contrainte significative pour un déploiement efficace de la méthode. Tout en restant en microscopie optique et avec ce même équipement, ce problème pourrait être résolu en utilisant un optique au $\times 40$. L'utilisation d'une autre technique d'acquisition, telle que le Nanozoomer qui peut scanner une lame complète en moins d'une heure, pourrait aussi optimiser cette étape d'acquisition d'images.
En termes de technique d'imagerie, la cytometrie en flux a aussi été mise en évidence comme un technique à haut potentiel pour ce type d'approches (Dunker et al., 2021), et notamment pour l'analyse d'échantillons collectés sur le terrain, par exemple pour des sédiments lacustres fossiles (Barnes et al., 2023). Cette technique a l'avantage de ne pas reposer sur une étape de détection, évitant ainsi les erreurs générées à cette étape. Elles évitent aussi certains cas problématiques rencontrés avec l'utilisation de lames fixes comme utilisées ici, telles que les images de pollens recouverts par des débris ou flous, et les rendant ainsi indéterminables. Ces approches pourraient ainsi s'avérer particulièrement adaptées pour un monitoring des pluies polliniques, mais le protocole de collecte et de traitement des échantillons serait alors à revoir, notamment le traitement et la préparation des échantillons, pour éviter les gros débris. En parallèle, la microscopie confoncale est quant à elle en train de s'établir en tant que technique de référence pour l'analyse détaillée des pollens, de leurs morphotypes et de leurs comparaisons, en vue d'améliorer la taxonomie et de retracer la phylogénie des espèces (Adaïmé et al., 2024 ; Collevatti et al., 2024 ; Kong et al., 2016 ; Romero et al., 2020). Ces approches ont un très grand potentiel dans le cadre de l'étude présentée dans le chapitre IV et visant à évaluer des potentielles différences entre les grains de pollen produits par la vigne sauvage et par la vigne domestiquée.

0.3.2 Discussion sur l'analyse des pluies polliniques en Méditerranée (Pic St Loup) et le signal pollinique associé à la vigne

Le suivi des pluies polliniques au Pic St Loup (Pollimed) est récent, et donc la série temporelle est encore trop courte pour analyser des tendances globales et les relier aux conditions météorologiques de chaque année et les perturbations climatiques globales ; ce sera l'objectif des prochaines études menées sur ce monitoring. Cela dit l'étude des pluies polliniques issues des cinq premières années de ce suivi ont déjà mis en évidence des différences dans les signaux polliniques associées aux divers types de végétation rencontrés sur les sites d'études et associés des taxons Méditerranéens clés, et notamment à la présence de la vigne sauvage et domestiquée. Les taxons majoritaires dans chacune des sites d'échantillonnages ont mis en évidence des différences entre les milieux ouverts et fermés, ainsi que les milieux plus de forêts 'naturelles' et les milieux plus transformés. Des changements au cours du suivi de cinq années ont aussi été mis en évidence pour certaines espèces, telles que la disparition de *Buxus*, des pics de production de pollen de *Quercus* certaines années sur l'ensemble des sites étudiés. De même, il semble que les changements temporels dans les pluies polliniques sont beaucoup plus marqués pour les sites échantillonnés en forêt plutôt que sur les sites plus de milieux ouverts et davantages anthropisés.

Au niveau de la vigne, les pluies polliniques ont mis en évidence des différences dans les taxons majoritaires, caractérisant des milieux plus ou moins ouverts et plus ou moins anthropisés. Ces différences sont caractéristiques de la présence de la vigne sauvage, dans son habitat naturel qui nécessite donc la présence d'arbres et l'établissement d'une végétation typique méditerranéenne (davantage d'arbres et d'arbustes), par rapport à la vigne domestiquée, cultivée dans des champs ouverts avec des sols plus tassés et travaillés (avec davantage d'herbacées annuelles). Des indicateurs de présence locale d'un cours d'eau intermittent a aussi pu être mis en évidence à partir de l'analyse des pluies polliniques (e.g. du fait de la présence de *Fraxinus*). Plus précisément la présence du pollen de vigne dans les pluies polliniques a démontré des patrons distincts dans le cas de la présence de la vigne domestiquée, de la vigne sauvage mâle et de la vigne sauvage femelle. Par ailleurs l'analyse des références du pollen de vigne est en accord avec ces résultats. Premièrement, une plus grande quantité de grains de pollen a été retrouvé dans les fleurs d'individus de *Vitis vinifera* en comparaison de *Vitis sylvestris*. Bien que des facteurs s'ajoutent à la production, tels que la dispersion et surtout le nombre d'individus en fleurs (très grand pour la vigne domestiquée comparé à la sauvage), ces éléments vont dans le sens d'un plus grande quantité de pollen de vigne retrouvé dans le monitoring au sein du vignoble. Ainsi l'analyse des pluies polliniques a mis en évidence des différences dans le cortège associé à la vigne ainsi que spécifiquement l'abondance du pollen de vigne, et le ratio entre leurs deux morphotypes polliniques (stérile et fertile). Dans le cadre du monitoring, la présence du pollen de vigne stérile, du moins en quantité supérieur à 1%, n'a été observée qu'en présence de l'individu sauvage femelle, ce qui va dans le sens d'une étude précédente proposant la présence de ce morphotype comme un indicateur de présence d'une population de vigne sauvage (Mercuri et al., 2021). Ces résultats sont donc une preuve de concept pour montrer que ce type d'analyse a le potentiel de discriminer la présence de la vigne sauvage et de la vigne domestiquée, dans un contexte où des observations de terrain pour effectuer les vérifications sur les individus sont impossibles, par exemple dans des zones géographiques difficiles d'accès, ou dans le passé. A noter cependant que certaines variétés culitvées par le passé correspondent à des individus fonctionnellement femelles produisant le morphotype pollinique stérile (ex. variété Picolit, Cargnello et al., 1980), et comme démontré par la présence de deux échantillons de pollens issus d'individus domestiqués de la collection ampélographique utilisée pour la

référence (chapitre III) avec ce morphotype. Ces variétés ne sont aujourd'hui que très rarement cultivées mais elles ont pu être cultivées par le passé, et devront être prises en compte dans ce type d'analyse.

Finalement, la méthode de classification des références pour le pollen fertile produit par des individus sauvages et domestiqués n'a pas été concluante. Il est possible que cela soit lié à une limite méthodologique, qui n'a pas su capter des différences très fines entre ces morphotypes, et comme discuté dans la section dédiée à la méthode. Cependant il est aussi possible qu'elle résulte d'une absence réelle de différences entre les pollen de vigne sauvage et domestiquée, i.e. la variabilité entre sous-espèces est du même ordre que la variabilité entre individus de même sous-espèce ou au sein des individus. Cela est possible étant donnée la courte durée de leur évolution, i.e. environ 11000 ans (Dong et al., 2023 ; Grassi et al., 2008), et les flux génétiques qui se sont opérés au cours du processus de domestication de la vigne, incluant potentiellement multiples évènements d'introgression (Arroyo-García et al., 2006 ; Dong et al., 2023).

0.4 Conclusion

Cette thèse a permis de développer une méthode de reconnaissance automatique et a pu être appliquée pour traiter des questions d'ordre biologique. L'objectif biologique premier de cette thèse était d'analyser le suivi des pluies polliniques sur un site avec une végétation Méditerranéenne typique et hébergeant une population de vigne sauvage ainsi qu'un vignoble commercial, dans le cadre des enjeux actuels de conservation de ce type d'habitats et de la vigne sauvage, espèce en danger sur la zone d'étude face aux pressions exacerbées que fait peser le changement climatique sur le bassin Méditerranéen. Il s'agissait aussi d'améliorer la connaissance du signal pollinique associé à la vigne sauvage et domestiquée, notamment au regard d'applications en paléoécologie. Dans ce cadre, l'analyse des pluies polliniques a été complétée par l'analyse de références issues d'individus d'une collection ampélographique. Ces études reposent sur l'analyse de grandes quantités d'échantillons, d'une part pour le suivi de la végétation, qui n'est établit que depuis 2019 mais vise a être poursuivi sur le long terme, et pour affiner la description du pollen de vigne car il est retrouvé en très faible quantité dans les échantillons polliniques, et d'autre part pour l'analyse de référence qui repose sur l'analyse d'une grande quantité de pollen issus de multiples individus. Finalement l'objectif premier est donc devenu de développer une méthode de reconnaissance automatique pour les échantillons polliniques grâce à l'intelligence artificelle (IA). L'investigation de l'IA visait aussi à explorer des nouvelle méthodes possibles pour différencier le pollen de vigne produit par la vigne sauvage et cultivée et qui ne peut pas être différencier à l'oeil.

La reconnaissance automatisée du pollen est un objectif de longue date en palynologie. Au début de la thèse, les recherches en étaient encore au stade de développement de méthodes et de nombreux verrous persistaient pour obtenir une méthode de reconnaissance automatique applicable à des échantillons collectés sur le terrain pour traiter des questions en écologie ou en paléoécologie, donc des conditions telles que pour faire un suivi des pluies polliniques dans le cadre de la thèse.

Les développements faits pendant cette thèse ont abordé les différents verrous méthodologiques identifiés dans le cadre de ce type d'application à des conditions d'études réelles pour la palynologie dans les sciences environnementales. De façon globale, la thèse a permis de développer un pipeline complet, de la collecte des échantillons jusqu'à un comptage pollinique pour plusieurs taxons Méditerranéens clés obtenus par des pièges

passifs à pollen, et pour comparer des morphotypes polliniques proches. Dans le détail, un protocole a tout d'abord été adapté pour le montage des lames et l'acquisition des images à partir du matériel proposé par un laboratoire collaborateur, le CEREGE, matériel dont l'ISEM est maintenant aussi équipé. Différents développements ont ensuite été effectués pour l'analyse automatique de ces images, chacun pouvant contenir plusieurs grains de pollen, spores et débris, des pollen issus d'une grande diversité taxonomique, telle qu'on la trouve en Méditerranée, ainsi que des grains de pollen pouvant être abîmés, mais surtout pouvant apparaître dans les images recouverts par exemple par des débris ou parfois flous dans les images. L'approche développée devait aussi permettre l'analyse de grande quantité de pollen issus d'individus de collection.

Dans ce cadre, une nouvelle méthode reposant sur un algorithme de deep learning a été mis en place pour la première étape de détection, qui est une étape complexe sur le type d'image rencontré, et il a été évalué dans le cadre de conditions d'applications réelles tout en considérant divers questionnements ou enjeux biologiques qui peuvent être associés à de tels échantillons. Ensuite, une méthode d'identification de particules détectées a été mise en place dans le but d'obtenir le comptage du total pollinique et de 15 taxons Méditerranéens clés. Le pipeline final a permis l'obtention de comptages polliniques pour l'ensemble des échantillons polliniques collectés jusqu'alors dans le cadre du monitoring de végétation. Les données polliniques générées automatiquement pour le suivi de la végétation ont mis en évidence différentes communautés végétales, telles qu'observées sur le terrain, ainsi que des patrons différents dans la présence du pollen de vigne dans les pluies polliniques associées à la présence de vigne domestiquée, sauvage mâle (la vigne sauvage étant dioique), et sauvage femelle. De même l'analyse de références, avec des comptages sur de grands nombres de grains, et facilités par la méthode automatique, a mis en évidence des différences dans la quantité de grains de pollen produits par fleur, et une tendance dans les tailles des grains de pollen fertile de vigne sauvage et domestiquée. Par contre les algorithmes de classification standards (CNN) n'ont pas permis de discriminer les grains de pollen fertiles produits par les deux sous-espèces, due à une trop grande similarité morphologique, peut-être même à une absence réelle de différence entre les grains, du moins à partir d'observations en microscopie optique. Ces résultats posent de nouvelles clés de compréhension du signal pollinique associé à la vigne dans le cadre de ces conditions d'occurrence sur le terrain, et de ces conditions de production. Ces données sont aussi une preuve de concept pour le développement d'une possible méthode de discrimination de la présence de vigne sauvage et domestiquée à partir d'échantillons polliniques issus d'écosystèmes modernes et pouvant être valorisés pour les études en paléoécologie dans le but de reconstruire l'histoire de la viticulture.

Finalement, la méthode de reconnaissance automatique développée pendant cette thèse, est déjà en tant que telle un outil qui peut aider les palynologues dans leurs analyses, pour quantifier des grains de pollen dans des échantillons purs (de référence) ou pour effectuer des mesures morphométriques standardisées sur de grandes quantités de pollen. La méthode et son application pour suivre la végétation au Pic St Loup dans le temps fait elle preuve de concepts pour une possible utilisation pour traiter de grandes quantités d'échantillons collectés sur le terrain et obtenir des comptages polliniques pour les taxons d'intérêts. De façon générale ces méthodes automatisées de l'analyse pollinique vont apporter des façons complémentaires de traiter la donnée pollinique, et de les questionner, i.e. de nouveaux questionnements vont pouvoir être explorés à partir des nouvelles méthodes. La reconnaissance automatisée de l'analyse pollinique pour les sciences environnementales permettront certainement, et prochainement, de diversifier et d'augmenter le potentiel de la palynologie.

Chapter 0

Introduction

The stakes for ecosystems in the Mediterranean basin

The Mediterranean basin is ecologically unique, with 4.3% of the world plant species and more than ten thousands endemic plants, it has thus been identified as the third global biodiversity hotspot (Fig. 1; Myers et al., 2000; Cuttelod et al., 2008; International Union for Conservation of Nature, 2016). This biodiversity results from the unique Mediterranean climatic conditions (Lionello et al., 2006), combined with the presence of human societies that have shaped the region landscapes and species for millennia through land use (Klein Goldewijk et al., 2011) and complex processes such as domestication (Blondel and Aronson, 1999). As a result, the Mediterranean basin is also a center of diversity for many crop plants, e.g. of wild olives and grapevines, while its natural resources have had in turn a long-standing influence on sociocultural interactions, commercial exchanges and agricultural practices of Mediterranean societies. The Mediterranean basin is thus the center of a diverse and unique biological resource ranking among the world's most ecologically significant regions, while also holding substantial importance for the historical and economic value of some of its emblematic domesticated species, markers of long-term human-nature interactions.

The Mediterranean climate is however changing at a fast pace due to its transitional location between both temperate and subtropical climates, making Mediterranean areas hotspots of climate change (Giorgi, 2006; Lionello and Scarascia, 2018). The ongoing and intensifying droughts and increase in temperatures are threatening Mediterranean ecosystems with aridification (Giorgi and Lionello, 2008; Stocker et al., 2014; Polade et al., 2015; Tramblay et al., 2020) with many species already reaching their temperature and moisture limits. The area is thus threatened by high risks of contraction due to aridification (Barredo et al., 2018). Additionally, intensification of anthropogenic pressures through urbanization and agriculture is leading to extensive habitat loss and fragmentation. The consequences of such disturbances are further exacerbated by the high sensitivity of Mediterranean ecosystems, which, alongside tropical regions, rank among the most vulnerable to climate and land-use changes (Cramer, 2018; Newbold et al., 2020).

Given the current global change and its immense ecological consequences, Mediterranean ecosystems, especially in the Mediterranean basin, have thus become critical priority areas (Myers et al., 2000; Cramer, 2018; Ali, 2022).

Key Mediterranean plants: the case of the grapevine and its natural habitat in the Mediterranean area

In the context of ongoing disturbances in the Mediterranean region, it is crucial to better understand the past, present and future evolution of certain key taxa that are emblematic of the Mediterranean region (e.g. *Olea*, *Vitis*) and that have to be safeguarded. Within this broader conservation challenge, safeguarding the wild grapevine and its natural habitat stands at the intersection of multiple conservation priorities.

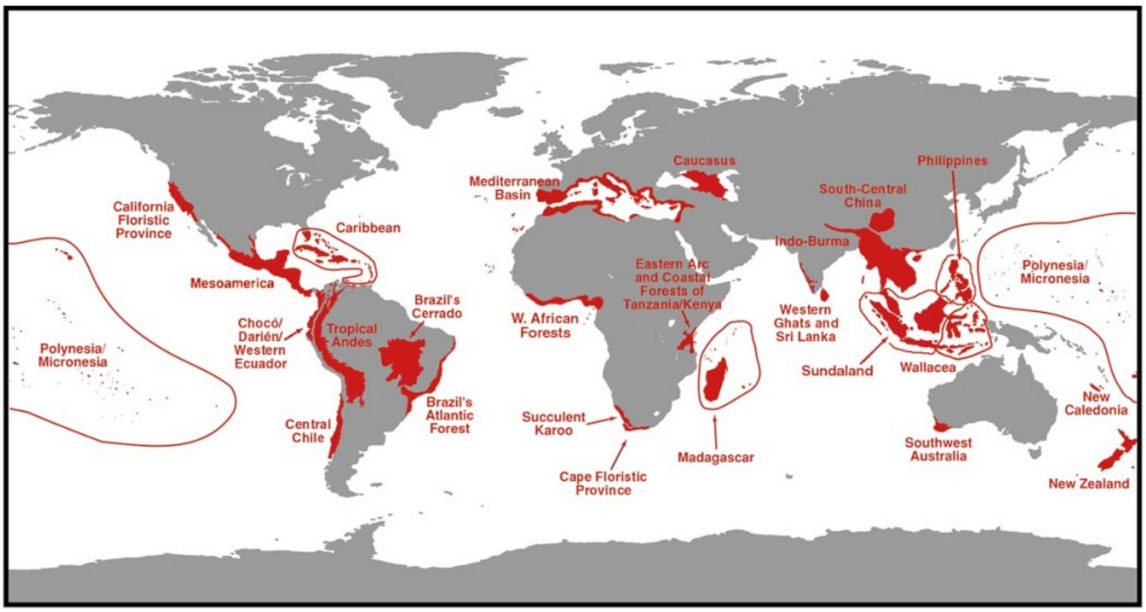


Figure 1: The 25 areas identified as hotspots of biodiversity and of conservation priorities (figure extracted from Myers et al., 2000)

The wild grapevine, *Vitis vinifera subsp. sylvestris* (later on *Vitis sylvestris*), is a dioecious heliophilous liana that grows in colluvial and alluvial forests (Naqinezhad et al., 2018) with a natural range from the Caspian sea to the Atlantic coast (Arnold, 2002; André et al. 2017). Although thriving during the Holocene (last 11700 cal years BP), *Vitis sylvestris* has faced significant decline over the last two centuries due to diseases, habitat loss, and fragmentation (This et al., 2001). Its distribution is now sporadic, especially in the western Mediterranean basin, with few relic populations (Cantos et al., 2017; Anzani et al., 1990; Clò et al., 2023; This et al., 2001; Arnold, 2002/1998), providing a first priority level of conservation. *Vitis sylvestris* has been identified as the wild relative of the commercially very economic cultivated grapevine, *Vitis vinifera subsp. vinifera* (later on *Vitis vinifera*, Zohary, 1996; McGovern et al., 1996), recognized as a distinct subspecies (Grassi et al., 2008; Dong et al., 2023). Being the wild relative of such an important crop thus gives a second conservation priority level for *Vitis sylvestris*, as a biological resource for agronomy (Castañeda-Álvarez et al., 2016).

Finally, the grapevine is a historically emblematic species for the Mediterranean basin and holds a high and needed value, as a biological resource for paleoecology and archaeology (Mercuri et al., 2021). The primary domestication of the grape is recognized to have occurred in the Caucasian area around 11,000 years ago (McGovern et al., 2017; Dong et al., 2023), with a consecutive spread of the domesticated grape to the western Mediterranean. The arrival of *Vitis vinifera* in Iberia and Southern France has been dated to about 3,000 years ago (Bouby et al., 2013; McGovern et al., 2017). Yet, much about the history of its cultivation remains unknown (Grassi et al., 2008; Arroyo-García et al., 2006). Palynology holds potential for reconstructing its history, as done for other cultivated plants (Langgut et al., 2019, Pollegioni et al., 2017, Krebs et al., 2019), but research has been hindered by limited knowledge of the grapevine pollen signature. A detailed understanding is especially critical for the grapevine since its pollen signal is always low compared to other taxa (Turner and Brown, 2004). Moreover, new approaches relying on such palynological knowledge of the grapevine have been suggested to provide key information for such studies (Mercuri et al., 2021). *Vitis sylvestris* thus holds a third conservation priority as a modern analogue for such archaeological and paleoecological studies.

Altogether, the grapevine is at the heart of many concerns regarding the Mediterranean basin, ranging from prospects in ecology, agronomy and archaeology. For these reasons, the conservation and study of *Vitis sylvestris*, and of its natural habitat, on which it is highly dependent, is of great importance.

Why monitoring pollen rains ?

There is an urgent need to better understand environmental responses to ongoing and predicted disturbances in the Mediterranean basin, including for characteristic Mediterranean habitats harboring key Mediterranean species, such as the endangered *Vitis sylvestris*. Ecological monitorings have been shown to be an efficient approach to understand responses of plant and animal communities to increasing climatic and anthropogenic disturbances. Biodiversity monitoring consists in obtaining a continuous collection through time of entities from plant or animal taxa and analyzing their changes with field-based measurements (Walther et al., 2002, Lindenmayer and Likens, 2010). Data from ecological monitoring informs on the effects of environmental changes on structural and functional ecological responses, and can also provide baseline for research, for instance to test ecological theory or develop models (Lindenmayer and Likens, 2010), as well as for conservationists, ecosystems managers and policy makers (Lindenmayer and Likens, 2010). Time-series data from ecological monitoring have for instance shown the advance of the spring and summer phases, from measurements of flowering times or first singing of birds, which have been linked with local and regional climate changes, and namely, the increase in temperatures (Parmesan, 2006, Menzel et al., 2006, Walther et al., 2002; Ellwood et al., 2013; Parmesan and Yohe, 2003).

Among the diversity of entities that can be monitored, pollen rains have been the focus of many studies. Pollen rains are mostly supplied by anemophilous plants, and their monitoring provide information on local and regional plant composition and dynamics (Sugita, 2007a; Sugita, 2007b), and reflect flowering phenologies (times and intensities; Hicks et al., 2001), which are fundamental life-history traits of plants, and are highly sensitive to both environmental disturbances. Additionally, the study of pollen rains allows a better understanding of the pollen-vegetation relationships (Prentice and Webb III, 1986; Grindean et al., 2019; Githumbi et al., 2021), which are influenced by factors such as pollen productivity (Sugita et al., 2010) and pollen dispersal (Tauber, 1965; Jackson and Kearsley, 1998; Jackson and Lyford, 1999), and are consequently of high interest for the community of paleoecologists. Finally, due to the global public health they are causing with allergenic properties of many taxa, pollen rains are also under high surveillance from allergists (Anderegg et al., 2021).

Pollen rain monitoring have been conducted through measurements of airborne pollen indexes, or measurements of pollen accumulation rates (PAR), i.e. the number of pollen grains deposited per unit area over a given period of time, using natural traps such as moss pollsters (Gaillard et al., 2008), or artificial gravimetric traps (e.g. Fig. 3), such as Tauber-type traps (Hicks et al., 2001). Studying PAR using gravimetric traps gained popularity in Europe in the late 1990s, notably with the international project "Pollen Monitoring Program" (Hicks et al., 1996, Hicks et al., 2001). Initial objectives were to collect standardized pollen data and study modern pollen-vegetation relationships to better understand these relationships in the past (e.g. Pidek et al., 2010; Filipova-Marinova et al., 2010; Sjögren et al., 2010); later on, this research focused more on assessing responses of PAR to environmental conditions such as temperatures and

precipitations (e.g. van der Knaap et al., 2010; Hattestrand, 2013). In parallel, airborne pollen monitoring, primarily driven by allergists, have produced extensive and high-resolution time series (e.g. weekly) of pollen data, informing on pollen seasons and airborne pollen indexes, and thus informing on flowering phenologies.

In the Western Mediterranean area, notably in Italy and Spain, pollen rains monitoring have been intensively conducted with many studies focusing on key Mediterranean taxa such as *Quercus* (Fernandez-Gonzalez et al., 2020; Recio et al., 2018) and *Olea* (Boullayali et al., 2024), or the more ubiquitous *Pinus* (De Linares et al., 2017). These monitorings have demonstrated shifts in pollen production and flowering phenologies in response to climate disturbances and aridification ongoing in the western Mediterranean region (e.g. Ghitarrini et al., 2017; Fernandez-Gonzalez et al., 2020; Cristofolini et al., 2024, Recio et al., 2018, De Linares et al., 2017). For instance, such studies have identified *Olea* pollen emissions to be good indicators of flowering intensities and of fruit production (e.g. Oteros et al., 2013, 2014a, 2014b), and have shown an extended pollination period of *Pinus* in Spain (De Linares et al., 2017), and rapid responses of *Quercus* pollination season (duration, intensity, and peak, timing) to climate change (Fernandez-Gonzalez et al., 2020; Recio et al., 2018).

Pollen monitoring has thus a dual potential: it can allow to track the responses of characteristic Mediterranean vegetation, including of endangered species such as the grapevines, to ongoing disturbances, and, it can improve our understanding of the pollen signal associated to characteristics Mediterranean habitats harboring the grapevine.

Figure 2: Illustration of a gravimetric pollen trap placed to monitor pollen rains

Why describing the pollen signal of the grapevine matters?

The grapevine pollen is particular due to its dimorphism and to distinct predominant reproductive systems found in *Vitis vinifera* and *Vitis sylvestris*. First, while *Vitis sylvestris* is dieocious, *Vitis vinifera* is in contrast predominantly hermaphroditic (Pratt, 1971; This et al., 2006). Very few cultivated varieties with only female flowers were identified (e.g. Cargnello et al., 1980), but they are now sporadic. Two pollen morphotypes can be produced by these grapevine individuals, and including (Fig. 1):

- a fertile tricolporate pollen morphotype, produced by the *Vitis vinifera* and by functionally male individuals of *Vitis sylvestris*;
- a sterile inaperturate pollen morphotype, produced by functionally female individuals of *Vitis sylvestris*, and the few functionally female cultivated varieties of *Vitis vinifera*.

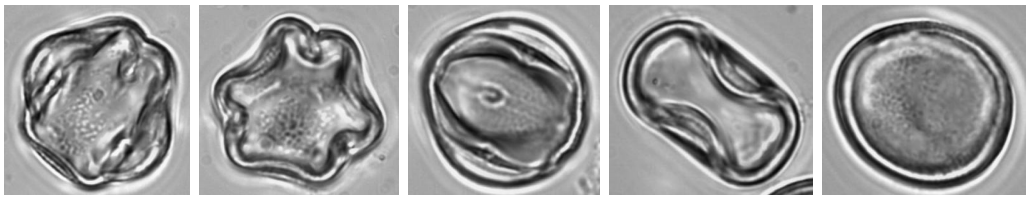


Figure 3: Examples of grapevine pollen images with the tricolporate morphotype (images 1, 2, 3) and with the inaperturate morphotype (images 4 and 5).

These differences likely result in distinct pollen signals associated to a population of wild grapevine or to a cultivated vineyard. The description of the two pollen morphotypes have been described in several studies, using pollen collected in flower anthers of reference individuals of collections or individuals from identified wild populations. However, these studies often focused on one subspecies (wild or domesticated) and rarely compared pollen from *Vitis sylvestris* and *Vitis vinifera* (as done in Gallardo, Mercuri et al., 2021), and comparison of the measured morphometric parameters between studies are otherwise limited due to distinct microscopy techniques, or protocols. Additionally, the abundances in which each morphotype are produced and released and the consequent grapevine signal in pollen rain are unknown.

A better description of the grapevine pollen signal associated to both the wild and domesticated subspecies, i.e. pollen productions and its presence in pollen rains, can thus also help research in paleoecology and archaeology research. For instance, such knowledge could be used to reconstruct the history of cultivation of the grape in the Mediterranean basin from a palynological perspective. Indeed, pollen is a major proxy to reconstitute past vegetation and has been used to study the history of cultivated plants such as *Olea* (Langgut et al., 2019), *Juglans* (Pollegioni et al., 2017), or *Castanea* (Krebs et al., 2019). This approach has never been tested on grapevine pollen due to the low amount of *Vitis* pollen collected in samples (Turner and Brown, 2004). Other approaches, relying on a more detailed understanding of the grapevine pollen signal, could thus be employed. For instance, the presence of the inaperturate morphotype in certain archaeological contexts has recently been suggested as a potential indicator of wild grapevines (Mercuri et al., 2021), while this grapevine inaperturate morphotype has never been considered in paleoecological studies, often due to a lack of awareness of its existence within the palynologist community. Alternatively, the possibility of discriminating the pollen grains from each morphotype produced by the wild and the domesticated grapevines could also be a complementary approach in this perspective.

Biological objectives and projects : Pollimed and Vassal

In that context, this PhD was conducted with two distinct objectives.

First (1), we aimed to monitor pollen rains collected from a typical, characteristic Mediterranean vegetation and harboring a relic population of wild grapevine and a nearby cultivated vineyard, to assess the composition and spatial distribution of such plant communities from a palynological perspective and assess temporal changes. The monitoring aimed to assess the response of the grapevine and of typical Mediterranean

taxa from its associated vegetation to disturbances, such as climate change, and to assess the composition of these plant communities associated to the wild and the domesticated grapevines. With such monitoring conducted in an area harboring the wild and the domesticated grapevine, we also aimed to improve our understanding of the pollen signal associated to both subspecies, regarding their associated vegetation and the presence of each grapevine pollen morphotypes in the pollen counts, and regarding prospects such as improving our understanding of the ecology and conservation of the species (flowering intensities), and as a background idea, to provide modern analogues for paleo-ecological studies. Considering the first objective, the focus was placed on a pollen rain monitoring conducted as part of the project Pollimed (funded by OSU OREME, included in the POlarise project, IRN, CNRS, INSU), conducted in an area harboring a relic population of wild grapevine and a nearby commercial vineyard of domesticated grapevine located in Southern France, associated to Mediterranean habitats including open scrubs, oak forests and meadow. This monitoring is still recent, being conducted since 2019 with the prospect of generating long-term pollen data.

Second (2), we aimed to study in details the pollen produced by *Vitis vinifera* and *Vitis sylvestris*, to describe their morphologies, their pollen production, and investigate possible methods to discriminate the pollen grains of the fertile tricolporate pollen morphotype produced by both subspecies. Considering this second objective, reference pollen samples collected from known individuals from the ampelographic collection of Vassal-Montpellier were used.

From ultimate biological objectives to proximate methodological objectives : development of AI-based methods to process pollen samples

Both biological objectives of the thesis rely on the analysis of large samples of pollen, i.e. on palynology, to identify the plant taxon -usually the botanical genus or family- from pollen grains that are traditionally observed by experts under light microscopy. This process is time-consuming, requires expertise while facing an erosion in the number of experts (Braconnot et al., 2024), and is also prone to some subjectivity and suffer from intra- and inter- operator effects. In the frame of the PhD objectives, pollen analysis thus presented several challenges. First, the monitoring aimed to be established on the long-term, and thus require analyzing large data sets, all comprising a potentially large diversity of pollen taxa. Second, the grapevine pollen is often found in low abundance in pollen samples so that focusing on this taxon requires again to analyze large amounts of data to be able to find a sufficient number of pollen grains for powerful statistical analyses. Third, no method exists yet to discriminate a pollen grain produced by the wild or the domesticated grapevine when considering its morphology (either the sterile or the fertile morphotype), so that novel approaches should be investigated. Standard pollen analysis in that frame is thus a major limiting factor. To be able to analyze large data sets and investigate alternative approaches to compare similar pollen morphotypes, the proximate objective of the PhD was thus to develop a new and automated method of pollen analysis to improve the palynological study of Mediterranean taxa.

The automation of pollen analysis can alleviate constraints of the pollen analysis (K. A. Holt and Bennett, 2014; Stillman and Flenley, 1996), and granted to developments conducted for more than two decades (France et al., 2000) and extensive advances in microscopy, digital technology (computing power), and machine learning, methods are now ready to start experimenting these methods to address biological objectives.

Consequently, the primary objective for the PhD was to develop an automated pipeline for the analysis of pollen samples collected in the Mediterranean area.

The developed method had also to fit the specific requirements of each of the two biological objectives of the PhD.

Principles and applications of palynology

Palynology relies on several characteristics related to pollen, including:

1. the highly structured and morphologically diverse external wall of pollen grains, allowing to create a taxonomy (Reille and Pons, 1990; Punt et al., 2007);
2. the ubiquity of pollen allowing to retrieve plant-related information for a wide range of environmental contexts and processes, and resulting from the diversity of dispersal mechanisms (wind, insects, but also water and various animals), along with the significant quantities of pollen released by anemophilous plants;
3. the resistance of the pollen external wall, made of sporopollenin, an extremely durable biological material (Brooks and Shaw, 1978; F.-S. Li et al., 2019), which allows identification of pollen grains collected from numerous contexts (airborne samples, peat or lake sediment, insects) and across extensive time periods (fossil cores from millions of years).

As a result, palynology contributes to many disciplines (1) ecology, with the monitoring of modern vegetation as investigated here (or e.g. van der Knaap et al., 2010), (2) biology with the study of the diet of insect pollinators (e.g. Lazar et al., 2024; Cullen et al., 2021), (3) paleoecology to reconstitute past vegetation and climate changes (e.g. Donders et al., 2021) (4) evolution to infer the history of cultivated plants (e.g. Langgut et al., 2019; Pollegioni et al., 2017; Krebs et al., 2019) or to reconstruct the phylogeny of species (Gonçalves-Esteves et al., 2022; Adaïmé et al., 2024), (5) agronomy to determine the geographic origin of honey (Von Der Ohe et al., 2004) or predict fruit harvests (e.g. of olives : Oteros et al., 2014; Aguilera and Ruiz-Valenzuela, 2019; and grapevines: Cunha et al., 2016), (7) in forensic science for criminal investigations (Coyle, 2004), and (8) allergology (Anderegg et al., 2021 ; Linneberg et al., 2016), with aeropalynology the most active field of automated palynology.

The wide range of disciplines related to palynology results in a wide diversity of pollen samples, which have distinct contents, taxa diversity, states of preservation of the pollen grains, amounts of target and non-target taxa, and amounts of non-pollen objects (debris and spores). Each type of pollen sample requires distinct chemical treatment, and analyses, and distinct levels of difficulty for the automation of the pollen analysis.

Automated pollen analysis : brief state of the art overview

First developments for automated palynology go back to the early 2000's (France et al., 2000), the field has nonetheless tremendously changed since then. At the beginning of the PhD, most widely adopted approaches relied on light microscopy images processed with machine learning algorithms (e.g., Sevillano and Aznarte, 2018; Bourel et al., 2020; Olsson et al., 2021; Punyasena et al., 2022). The image analysis process was split into two tasks : first, locating pollen grains in images containing multiple particles (referred to as detection), and second, identifying the pollen taxon from images of individual pollen grain (referred to as classification). Deep-learning algorithms had been widely adopted for the classification task while the detection was still mostly conducted using standard machine learning algorithm. Despite efforts, several challenges remained unresolved but had to be addressed to ensure that the developed method would be both efficient and adapted to our goals of analyzing pollen rains to obtain the Mediterranean taxa diversity from pollen

samples collected in gravimetrics traps and to compare morphologically similar grapevine pollen grains.

Three years ago, most developments had been conducted using reference samples and considering theoretical frameworks, i.e. not representative of conditions encountered in practice in palynology. In these studies, the developments, but most importantly the evaluation of the method performances, were conducted using datasets including only images from well preserved pollen grains and from a controlled number of taxa represented by a high number of images (e.g. K. A. Holt and Bennett, 2014, Sevillano and Aznarte, 2018 for classification; Kubera et al., 2021 for detection). While these developments represent essential first steps, they remain within a theoretical framework, and the resulting automated method cannot be used in practice to generate pollen data that can support research outside the method development, i.e. for 'real-world' studies, identified as the golden objective of automated palynology (K. A. Holt and Bennett, 2014). Indeed, the ultimate objective of automated palynology is to generate reliable pollen identification from pollen samples with typical properties encountered in each palynological discipline, collected in the field, such as airborne samples, honey samples, gravimetric traps or fossil cores.

Considering environmental-related disciplines, few studies had proposed automated methods applied to field samples to study: fossil pollen (Bourel et al., 2020), honey samples (Tsiknakis et al., 2022, Sevillano et al., 2020), pollen from insects (Olsson et al., 2021), and ecological airborne pollen monitoring (Punyasena et al., 2022). In these studies, the datasets got closer to conditions expected in routine for environmental-related disciplines, while major and crucial advances towards the automation of pollen analysis for each of these disciplines were accomplished. Nonetheless, they still presented some limitations preventing full real-world applications. First, these studies mostly focused on only the classification task, bypassing the first step of detection necessary to locate individual pollen grains in microscopic images that may contain many pollen and non-pollen particles. Second, the most complex cases were avoided by discarding rare or underrepresented taxa, by manually removing images of debris, spores, or of pollen images with poor visibility or of damaged grains. Because these cases may occur when applying the method in routine on real-world samples, an efficient automated method should deal with these challenging cases generated from such samples properties, which include :

- an uncontrolled and potentially high pollen taxa diversity;
- possible many non-pollen particles (e.g. NPPs and debris);
- pollen grains with bad image quality, damaged, crumpled, or clumped with other particles;
- morphologically similar pollen taxa;
- rare or unexpected pollen taxa.

Driven by concerns over the prevalent increase in pollinosis worldwide and the increasing need for real-time airborne pollen data (Buters et al., 2018), studies on automated pollen analysis for allergology was more advanced considering these conditions. Few solutions had thus been proposed to address such challenges, but were limited to airborne samples (Cao et al., 2020b; Oteros et al., 2019, Crouzy et al., 2022, Buters et al., 2018). These samples still contain few debris or spores, low taxa diversity (constrained by plant phenologies due to daily sampling), and well-preserved pollen compared to typical samples encountered in environmental studies that can cover large period of time or space.

Automated pollen analysis : remaining challenges for our biological objectives

In this PhD, the objective was to develop and directly apply an automated method to address our biological objectives, i.e. to assess spatial and temporal changes in pollen rains from a Mediterranean vegetation harboring the wild and domesticated grapevines, and to compare the morphological similar pollen grains produced by the two grapevine subspecies. The method had thus to be developed to handle the complexities of environmental samples from vegetation monitoring, and to provide an evaluation that would remain reliable when used to generate pollen data from new sites or years. Considering that framework, several technical and scientific challenges were identified, namely:

(1) automating the first step of detection for samples with noisy, debris-filled, overlapping, or blurred images, which are too complex for standard machine learning algorithms that had been used so far, and addressing the question of the potential impact of detection errors on the subsequent classification;

(2) automating the classification, i.e. identification, of all detected particles and thus including pollen but also debris, as well as bad quality images;

(3) developing an integrated pipeline for all steps, that can remove falsely detected debris or blurry images, and evaluating the conditions for routine application, i.e. reliable considering 'real-world' application conditions.

To address our biological questions while overcoming the methodological and technical challenges related to the automation of pollen analysis, this thesis was structured following these steps :

Chapter 1 : Main pipeline for automated pollen analysis

The first chapter details the general pipeline developed for the automated pollen analysis of environmental pollen samples. The pollen samples used at each steps of the method development were collected and processed by several members from ISEM, including Bertrand Limier, Sarah Ivorra and Sandrine Canal (Pollimed project), Marie Challe and Agnès Mignot (project Polumine), with whom I had the chance to work on the field or in the lab. Reference samples of grapevine pollen from the Vassal-Collection were also collected by a fellow researcher from the INRAE laboratory, Maud Tenaillon.

Chapter 2 : A user-friendly method to get automated pollen analysis from environmental samples

The second chapter, presented as an accepted manuscript, presents the method developed to detect pollen grains in complex images, and some guidelines for its application to real-world samples.

The use of deep-learning algorithm proved to be effective for the detection of pollen grains in noisy and challenging images with pollen grains detected. Most detection errors were related to cases of grains with bad visual qualities (blurred, obscured by debris). Slight detection differences were observed for pollen taxa of rare and singular morphotypes.

The detection step relying on the CNNs used in the chapter II was conducted under the guidance of Jérôme Pasquet.

Chapter 3 : Pollen rains in the Mediterranean area with automated pollen analysis

The third chapter both presents the developed classification of pollen grains obtained from the previous detection method, and applies it to pollen rains from the vegetation monitoring. Annual pollen accumulation rates from key Mediterranean taxa including *Vitis* are studied, also considering differences among years and sites.

This study is also a proof of concept for the automation of pollen analysis for this type of monitoring.

The automated method reflected well the composition of the main Mediterranean taxa in each sampling sites, and allowed to identify distinct compositions between the monitored sites and years for the 15 target taxa considered.

Pollen rains tended to display large discrepancies between years but no significant trends in the amount of released pollen. Pollen rains in anthropized and open sites were more stable across years compared to those collected in the oak forest, less anthropized areas. Some major changes were nonetheless noticed for some of the 15 target taxa identified, such as the total decline of *Buxus* in one sampling site, or the increase in *Quercus* pollen for specific years. The composition of the pollen rains displayed distinct patterns in the sampling sites located near *Vitis vinifera* in a cultivated vineyard, and *Vitis sylvestris* either with male or female individuals. These differences were observed in the accompanying vegetation, i.e. the target taxa, and were related to the distinct surrounding vegetation types. Differences in the ratio and abundances between the two grapevine morphotypes were also found with more of the fertile pollen grains in the middle of a vineyard compared to below a wild individual, and a balanced ratio between the two morphotypes only found near the wild female individual, as the sterile pollen grains was scarce otherwise.

Chapter 4 : Differentiating the pollen from the wild and the domesticated grapevine using automated image analysis

In the fourth and final chapter, pollen grains from reference individuals of the wild and domesticated grapevine were compared using AI models. Differences in the amount of pollen produced were found between both subspecies, which tended to correlate to size differences. The use of a standard classification Convolution Neural Network for differentiating the fertile pollen grains from *Vitis vinifera* and *Vitis sylvestris* was not effective.

The classification step relying on the CNNs used in the chapter III and IV was conducted in collaboration with a team of researcher from the Chennai Mathematical Institute (India), Dr. Pranabendu Misra and Dr. Sourish Das, as well as two master students, Shubhashish Chauhan and Anuraj Kashyap, initiated by Doris Barboni, who assisted with the thesis. As part of this collaboration, I was hosted for one month and a half at the Chennai Mathematical Institute, in 2023.

The numerical GPU from the Indian collaborators were used for the classification (Chapter III and IV), while the Jean-Zay resources were used for the detection step (Chapter II).

Chapter 1

Automated pollen analysis: state of the art and methodological developments

Betty Gimenez, Céline Devaux, Sébastien Joannin, Jérôme Pasquet, Doris Barboni, Luc Beaufort, Yves Gally, Sourish Das, Pranabendu Misra, Odile Peyron

1.1 Introduction

This PhD aimed to develop an automated pipeline for the analysis of pollen samples collected in the Mediterranean area. The ultimate goal was to generate pollen counts from samples collected in the field and address environmental-related questions. The developed method had thus to be efficient under routine application conditions, i.e. when applied on new samples, and without any manual intervention (e.g. to filter out bad quality images). A vegetation monitoring was primarily considered for the development and the application of this automated pollen analysis pipeline. This study aimed to monitor pollen rains from the anemophilous vegetation with passive pollen traps collected annually in Southern France (Hérault, Montpellier). The content of the traps was typical of those in environmental-related studies, such as moss pollsters or lake sediments (as in Barnes et al., 2023), with an uncontrolled and potentially high pollen taxa diversity, pollen grains in various states of preservation, and significant amounts of non-pollen particles of many types, such as spores and debris from plants or insects. Samples displaying this type of content are referred here as environmental samples.

At the start of the PhD, studies on automated pollen analysis had not yet addressed specific requirements for routine application on such environmental samples. Previous studies had mostly focused on a sub-task of the full process, pollen classification, and they evaluated results on datasets not fully representative of those obtained with environmental samples (e.g. Sevillano et al., 2020; Schaefer et al., 2021), i.e. on datasets from reference samples containing a determined and controlled taxa diversity and only pollen. Thus, two main challenges had to be addressed : (1) developing an integrated method including all steps required to automate pollen analysis, starting from the collection of samples in the field, and (2) evaluating the method under conditions similar to those encountered in routine with environmental samples.

This chapter aims to offer a comprehensive overview of the starting point, the available tools, and the inputs made, as well as the complementary activities undertaken such as fieldwork, to provide a clearer insight on how this PhD was conducted. Overall, this chapter follows the chronological and pivotal steps designed to achieve the automated pollen analysis pipeline for environmental pollen samples (Fig. 1):

Stage A - Getting microscopy images from environmental pollen samples

1. get pollen samples: field work, chemical treatment, and development of a new protocol to mount slides adapted to the automated microscope;

2. obtain microscopic images from the pollen samples using an automated acquisition pipeline;

Stage B - Getting individual pollen images from microscopy images each containing several pollen and non-pollen particles

3. detect (i.e. localize) and segment individual pollen grains in a bounding box from the microscopic images, including the implementation of an object-detection algorithm (YOLO);

Stage C - Getting pollen counts from individual pollen images

4. identify the class, pollen taxa or non-pollen particles, of all bounding boxes detected in the microscopic images, including the implementation of a standard classification algorithm (ResNet);

5. generate calibrated pollen counts per pollen taxa;

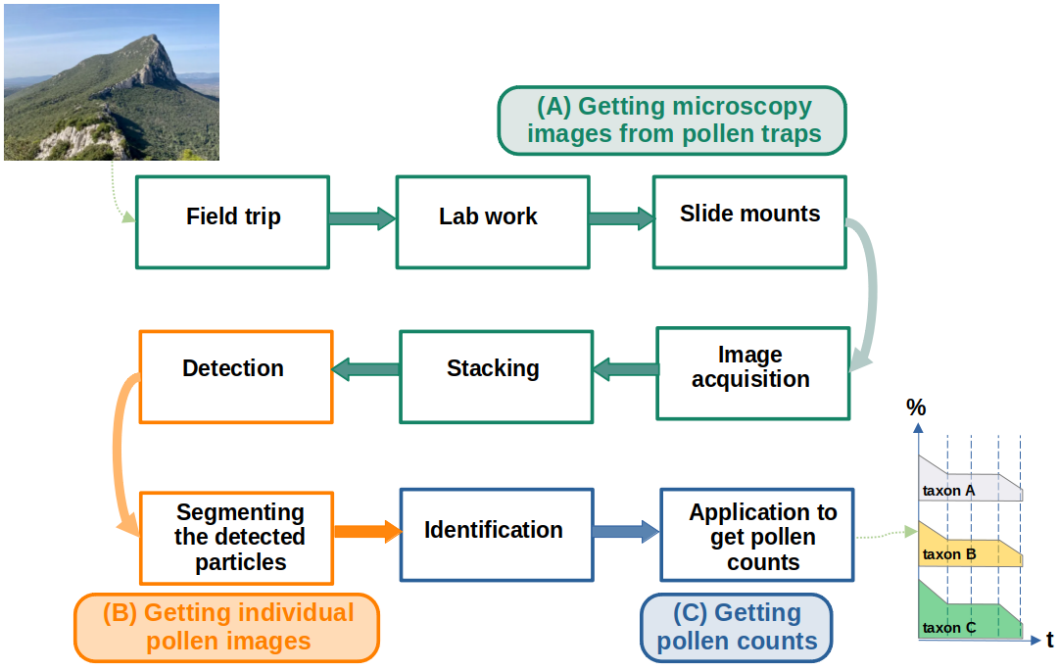


Figure 1: Graphical diagram of the steps identified for the automated pollen analysis pipeline

For each step, the process consisted in identifying its respective challenges, proposing and testing solutions adapted to environmental samples, and evaluating its performance considering constrains and objectives of an environmental study, which included the detailed analysis of errors, and how they could affect the accuracy and precision of pollen counts.

At the beginning of the PhD, no methods were available in my home laboratory. An automated analysis pipeline had been developed in a partner laboratory (CEREGE, Aix-Marseilles University) for marine microorganisms (Tetard et al., 2020a) and tests had been conducted on pollen under controlled conditions (Bourel et al., 2020). A collaboration was established with the CEREGE, and their image acquisition process as well as the general structure of their pipeline were used as a starting point for the method developed in this study on pollen. The followed approach was exploratory and interdisciplinary, and the efforts made spanned a broad range of areas, including palynology (pollen samples, wet lab, identification of Mediterranean pollen taxa), microscopy (helping install and testing the new automatic equipment in the microscopy room), AI algorithms (implementing and

testing different deep learning models), ecology (vegetation monitoring), and field work for pollen collection.

1.2 Getting microscopy images from environmental pollen samples (Stage A)

1.2.1 Pollen traps and chemical treatment

The passive pollen traps were deployed as part of the Pollimed project (OSU OREME) to monitor vegetation, and consisted in containers with a 5 cm width opening covered by a grid of 5 mm × 5 mm, inspired from Tauber traps (Tauber, 1965). These traps collected a diversity of pollen from anemophilous taxa, along with various other particles, such as spores, dust, plant debris, or small insects.

Figure 2: Annual pollen trap (PSL 11)

The Pollimed project was initiated in 2019, before the start of my PhD, and therefore the lab protocol for sample acetolysis was already established and conducted by a team member (Mme S. Canal). The chemical treatment of the samples followed these steps:

- Add *Lycopodium* tablets to the trap contents for calibration (Stockmarr, 1971) and add 5ml of 10% hydrochloric acid (HCl) for 1 hour, to dissolve the calcium carbonates of the *Lycopodium* tablets;
- Rinse with water and empty supernatant (vortex and centrifuge for 5 minutes at 3500 rpm);
- Add hydrofluoric acid (HF) at 70% for 24 hours at ambient temperature;
- Rinse with water and empty supernatant (vortex and centrifuge for 5 minutes at 3500 rpm);
- Acetolyze samples to empty and color pollen grains :
  - add acetic acid to dehydrate the medium, and empty supernatant (vortex and centrifuge for 5 minutes at 3500 rpm),
  - add the acetolitic solution for 10 minutes in a bath at 90°C,

  - add a few ml of acid acetic to stop the reaction, and empty supernatant (vortex and centrifuge for 5 minutes at 3500 rpm),
  - add acetic acid and empty supernatant (vortex and centrifuge for 5 minutes at 3500 rpm);
- If the samples are rich in organic matter, add 5ml of 10 % potassium hydroxide (KoH) for 3 minutes at 80°C to dissolve organic matter and deflocculate the particles, then empty supernatant (vortex and centrifuge for 5 minutes at 3500 rpm);
- Sieve the content with a 200µm and a 10µm filters;
- Add pure alcohol, empty supernatant (vortex and centrifuge for 5 minutes at 3500 rpm), and let the remaining layer of alcohol evaporate;
- Transfer the obtained material in ependorf tubes and add a few drops of glycerol to protect the pollen content.

This protocol is adapted for a standard pollen analysis conducted manually under a light microscope with mobile slides. It is thus designed to minimize the amount of debris while minimizing the risk of losing pollen grains; note that manual counts tolerate a certain quantity of debris. Consequently, the samples obtained from this protocol contained a significant amount of non-pollen particles along with the pollen grains and *Lycopodium* grains. Because the automated method is more sensitive to debris, adding steps to remove them further, by extending the KoH treatment for instance, could enhance automated performances. A balance must nonetheless be found to avoid excessive pollen grain loss.

1.2.2 Mounting pollen slides adapted to the automated microscope

The objective was to mount slides allowing the automated acquisition pipeline to capture a maximum number of identifiable pollen grains with a minimum number of acquisitions to optimize image acquisition time and disk storage. The method also had to apply to samples stored in glycerol, because samples were already stored in this medium. First, standard slides mounted with glycerol and sealed with Histolaque-type products were tested. This technique allows pollen grains to move under the coverslip and is best for manual counts as palynologists can observe pollen grains at different angles. Here, it made pollen grains shift slightly during the automated acquisition process, and generated blurred images, useless for identification. Therefore, we decided to mount pollen on fixed slides. Several strategies were tested and finally the below one was selected, based on glycerol jelly, which is miscible with glycerol, melts at 50°C but is solid at ambient temperature:

- add a small piece of glycerol jelly at the center of a slide;
- put the slide on a hot plate to melt the glycerol jelly;
- add $\sim$ 15 µL of the pollen sample to the drop of melted glycerol jelly;
- homogenize the content while keeping the slide hot;
- put the cover slip on the top of the drop and leave to cool down to ambient temperature;
- seal the cover slip with Histolaque.

The pollen samples contained distinct amounts of particles, depending on the amount of pollen and debris that had fallen in the trap, and the amount of glycerol added. A difficulty was thus to adjust the volume of pollen sample to add to get the expected compromise between the number of particles and pollen grain per image and the number of overlapping

objects. This method was kept throughout the PhD for consistency but had several issues: (1) in some slides, particles tended to clump together when mixing the glycerol and the glycerol jelly, (2) pollen grains could lie at different depth levels under the coverslip, making them appear sharp at different focal depths when observed under a microscope.

1.2.3 Obtaining microscopy images from the pollen samples

1.2.3.1 Starting point for the image acquisition process

The tools and methods to acquire images at the beginning of the PhD relied on our partner laboratory (CEREGE, Aix-Marseille Univ.), pioneer in the development of the automatic recognition of marine microorganisms, including foraminiferas, radiolarians, or coccoliths. Pollen had not been explored consistently with their method as palynology is not a primary focus in this research team. Their pipeline is based on a bright light microscope Leica DM6 B TL BF equipped with an oil immersion ×63 objective lens, an automated stage and an Hamamatsu ORCA FLASH camera. An interface developed at CEREGE (Tetard et al., 2020a) is used to control the stage and the camera, and acquire images automatically in a grid pattern within a slide. Before initiating the acquisition, the operator sets several parameters, including (Fig. 3):

- the number of fields of view (FOVs) to be photographed;
- the central position of the grid (x and y-axis);
- the spacing between each FOV (x and y-axis);
- the number of images to be photographed at each FOV position to create a pile of images (z-axis);
- the distance separating the images in the pile (z-axis).

After the acquisition, the pile of images photographed for each FOV is merged into one final 2D composite image, using the Hyperfocus software, which selects the sharpest areas and discards the blurred parts.

1.2.3.2 Adapting the image acquisition parameters

The z-axis step The objective was to optimize the spacing between images in the stacks taken at each FOV to capture both small and large pollen grains with sharp details, while limiting the number of images to capture considering the constrained acquisition time, and disk space. For instance, if spacing is wide, few images are photographed but small pollen grains are not photographed in focus in any image of the pile. To define this parameter, several tests were conducted on slides containing reference pollen taxa of variable morphologies and sizes ranging from 15 to 100 µm (*Noccaea caerulescens*, *Viburnum*, *Coronilla*, *Cercis*, *Cichorium*, *Centaurea*, *Mentha*, *Malva* and *Scabiosa*). The slides were photographed on the same FOVs positions, successively with z-axis steps of 2, 4, 6, 8, 12 and 16 µm, for a total range of 60 µm, considering that most Mediterranean taxa are smaller than 40 µm and that large grains are morphologically typical, and can therefore be identified even if they are not photographed in focus from top to bottom. The step chosen was 8 µm, combined with 13 images in a pile to cover an overall focal range of 104 µm. This value were selected as a compromise between the visibility of the pollen identification criteria in the final stacked images (external texture and sectional view), and the required acquisition time and disk space, each image in the FOV being photographed in about 1 second and saved as a tiff file of 8.4 MB. Observations also indicated that acquiring too many sharp focal planes for a pollen grain

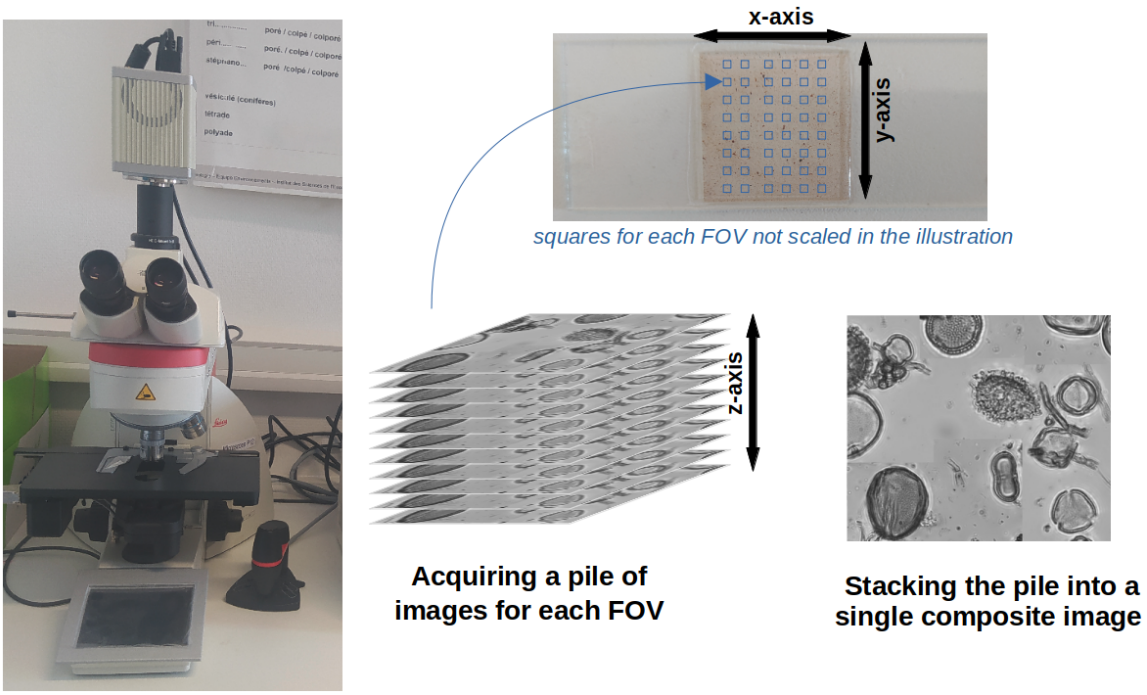


Figure 3: Illustration of the image acquisition pipeline

tended to reduce visibility of fine details in its final stacked image. The step chosen was slightly too large for the smallest pollen taxa (around 12 µm) that appeared blurry in the final stacked images, but reduced acquisition time compared to 6 µm.

Light, contrast and resolution Several settings for light intensity, contrast and exposure were tested with both a dry ×40 objective lens and an oil immersion ×63 objective lens. The ×40 objective lens did not provide sufficient precision and sharpness in the images, so we proceeded with the ×63 objective lens for all subsequent acquisitions. The settings were then chosen to maximize contrast and dark tones, prevent saturation in the highest pixel values, and thus enhance the visibility of the morphological features of pollen grains. Note that the camera used in this pipeline generates grey-level images.

1.3 Getting automated pollen counts from microscopy images : initial object of study, objectives and challenges

1.3.1 Material and main objectives

The acquisition step generates for each pollen sample grey level images covering a FOV of 214 µm × 214 µm. These images can contain many particles including pollen grains, NPPs and debris, sometimes air-bubbles, or no particle at all. Pollen grains can be well preserved or damaged (e.g. broken, crumpled), sometimes they can clump together with other particles or be obscured by other particles. Pollen grains can be cut off, at the FOV margins, or appear blurred if standing at a z-axis level that is outside the focal range captured by the pile of images for a FOV (Fig. 4).

The objective was to analyze these images to get the number of pollen grains per taxa, and the number of *Lycopodium* spores for calibration. This objective was broken down into smaller steps (Fig. 5):

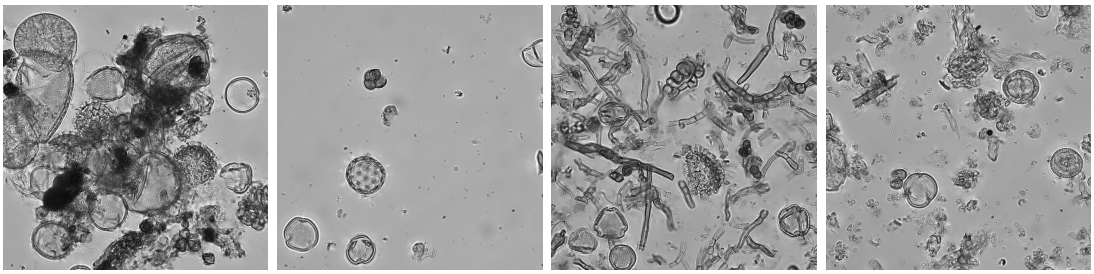


Figure 4: Examples of FOV images acquired from annual trap samples with the acquisition pipeline

- locate each particle in the images, i.e. determine their position and dimension, referred here as **detection** (Diwan et al., 2022);
- segment the initial FOV into smaller images containing each a single particle, referred here as **segmentation** (Diwan et al., 2022);
- identify the category of each segmented particle (e.g. pollen taxon), referred as **classification** (Diwan et al., 2022);
- compile all detection, segmentation and classification outputs obtained from each FOV into a **full integrated pipeline**.

Automating detection and classification, but not segmentation, are complex image analysis steps that required implementing machine learning algorithms (see details on deep learning in Supporting Information Methods S1).

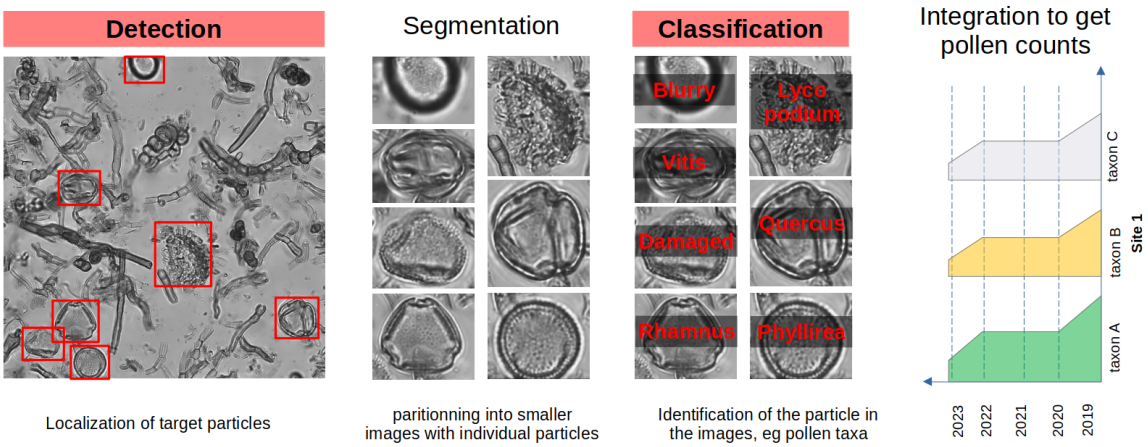


Figure 5: Illustration of the steps and expected outputs of the pipeline for the automated image analysis

1.3.2 Automated pollen detection

1.3.2.1 Relevant available methods for detection

Since the early 2000's, developments in automated pollen detection (France et al., 2000) have primarily focused on standard machine learning techniques. These techniques rely on images pixel intensity, gradients, changes, and threshold values, which are all processed by standard machine learning algorithms (e.g. watershed algorithm, Hough Circle Transform algorithm) to identify edges or patterns, and generate binary masks that distinguish objects from background (e.g. France et al., 2000; Boucher et al., 2002; G. Allen, 2006;

K. Holt et al., 2011; Johnsrud et al., 2013). These techniques have proved efficient on images with isolated pollen grains, not obscured by other particles, clear contrast between background and pollen, and devoid of non-pollen particles such as NPPs or debris. They have consequently been widely used on reference pollen samples (e.g. from collections) to build large libraries of pollen images, which are sometimes shared online (Johnsrud et al., 2013; Gonçalves et al., 2016; Tsiknakis et al., 2021), and used for training classification algorithms (Sevillano and Aznarte, 2018; Olsson et al., 2021; Tsiknakis et al., 2021). However, these techniques are not appropriate to differentiate target objects (e.g. pollen) from non-target objects (e.g. debris), and they loose significant efficiency when applied on images with aggregated objects, such as pollen grains clumped together or covered by debris (Diaz-Lopez et al., 2015). These conditions are though typical for environmental pollen samples, collected from lakes, peat, moss pollster or traps.

Deep learning algorithms, especially object-detection algorithms, such as Faster R-CNN (Ren et al., 2015), YOLO (Redmon and Farhadi, 2016) or RetinaNet (T.-Y. Lin et al., 2017), have recently revolutionized the field of computer vision due to their high efficiency even with challenging images, such as traffic signs on busy streets, animals in dense forests or sick cells in blood microscopic images, which all feature multiple types of target and non-target objects, cases of overlap and no clear distinction between target objects and background. They were first, and successfully, applied to detect pollen in 2019 (Gallardo-Caballero et al., 2019). Since then, they have been increasingly tested on pollen images, including reference samples (Kubera et al., 2022; Khanzhina et al., 2022), honey samples (Viertel et al., 2022), environmental samples (Punyasena et al., 2022; Theuerkauf et al., 2023; Durand et al., 2024; Gimenez et al., 2024 chapter II), and airborne samples, which represent the majority of studies on automated pollen analysis (Zhao et al., 2022; Jin et al., 2023; Boldeanu et al., 2022; Li et al. 2023).

1.3.2.2 Remaining challenges for our objectives

When the PhD started in 2021, the use of deep learning algorithms for the detection of pollen grains were limited to reference images (Gallardo-Caballero et al., 2019; Kubera et al., 2021) and airborne samples (Cao et al., 2020a). They had never been applied on samples from environmental-related studies, such that requirements related to routine application conditions were not available.

Environmental samples can contain a much higher pollen taxa diversity than do airborne pollen samples, notably because they cover extended periods, from years to decades, compared to daily volumetric traps used in aerobiology (Fig. 6). They also contain more debris due to the gravimetric harvesting process. The first identified challenge in detecting pollen grains from environmental samples was thus to manage the diversity of pollen morphotypes and of non-pollen particles, as it increased the risks of morphological similarities between pollen and non-pollen particles (e.g. NPPs), the occurrence of rare pollen morphologies, and of overlap between particles.

Additionally, we aimed to develop a method that could be used in routine to study environmental-related questions. The method had thus to remain efficient when applied on new environmental samples, and without manual intervention, especially between the detection and classification, for instance to filter out images of debris falsely detected or of damaged pollen grains that might be unidentifiable. It also required to assess whether the detection was biased, i.e. whether the pollen content among the detected particles were representative of the pollen content in the images, and by extension, in the pollen samples (although biases, not evaluated in this study, can add when mounting the slides and acquiring the images). Although this issue is irrelevant when using reference samples to

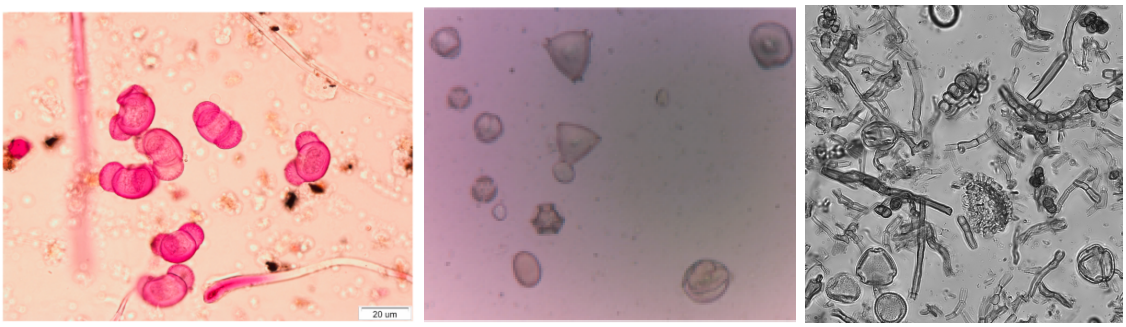


Figure 6: Examples of images from (a) reference samples (Khanzhina et al., 2022), (b) airborne volumetric samples as used in allergology (Cao et al., 2020a), and (c) passive trap samples as used to monitor vegetation in this study (Chapter II).

build a reference image database, it becomes crucial when used to generate pollen counts to study modern and past vegetation dynamics. The second identified challenge was thus to develop a method that would be efficient and correctly evaluated under routine application conditions to environmental samples.

1.3.3 Automatically classifying pollen grains

1.3.3.1 Relevant available methods

Contrary to pollen detection, the classification task has been more extensively studied in palynology, as a result of faster developments of algorithms and increased interest from palynologists. Deep learning algorithms, and deep Convolutional Neural Networks (deep CNNs), have been rapidly adopted for classifying pollen and have now become standard. These techniques have thus a strong background in palynology, and include implementations in allergology (Sevillano et al., 2020; Oteros et al., 2020; Tummon, Adams-Groom, et al., 2024), melissopalynology (Tsiknakis et al., 2021), pollination ecology (Olsson et al., 2021), ecology (Punyasena et al., 2022; this PhD chapter III) and paleoecology (Bourel et al., 2020; Theuerkauf et al., 2023; Durand et al., 2024). Although first studies only focused on the classification of reference pollen images (Sevillano et al., 2020; Khanzhina et al., 2018), the discipline has now shifted towards a use under routine application conditions to solve 'real-world' problems, the ultimate goal of automated palynology (K. A. Holt and Bennett, 2014). Similar to pollen detection, pollen classification performed on environmental samples is made difficult by to the presence of an uncontrolled taxa diversity, of pollen grains that can be distorted or damaged, and that occur along with many other non-pollen particles (Diaz-Lopez et al., 2015).

In the literature, several solutions have been tested to address these challenges, i.e. to manage debris, indeterminable pollen grains, doubtful identifications, and new pollen taxa that can occur in routine but were not included in model training :

- including a class for indeterminable pollen, a class for non-target pollen taxa and a class of non-pollen particles, alongside with targeted pollen taxa (as in Battiato et al., 2020a; Schaefer et al., 2021; Boldeanu et al., 2022; Theuerkauf et al., 2023; Brdar et al., 2023);
- using thresholds on confidence probability of the predictions (Crouzy et al., 2022);
- using morphological criteria to classify new pollen taxa to high taxonomic ranks, based on guided deep learning (Barnes et al., 2023);

- combining several trained models to classify pollen at different taxonomic levels and remove uncertain predictions using decision-trees (Bourel et al., 2020; Li et al., 2023), or to improve the robustness of predictions using ensemble learning strategies such as max and average voting (Tsiknakis et al., 2021);
- pre-processing images to improve their quality, for instance Gaussian filters to deal with blurred pollen images (Chen and Ju, 2022);
- extracting several layers of information (e.g. standard deviation, minimum intensity, and extended focus) from image stacks, then used as separate channels for classification (Li et al., 2023).

In most studies, standard classification CNNs are implemented, including ResNet (He et al., 2015), DenseNet (Huang et al., 2017), VGG (Simonyan and Zisserman, 2015) or AlexNet (Krizhevsky et al., 2012). Some alternative approaches were also tested, for instance by developing custom light-weights CNNs (Schiele et al., 2019; Khanzhina et al., 2018; Khanzhina et al., 2022), training multiple binary classifiers for the differentiation of a target class from the other classes (Khanzhina et al., 2018; Zhao et al., 2022), or using the recently developed Vision Transformers (Rostami et al., 2023; Zu et al., 2024). So far, these alternative approaches do not show high improvements in classification performance compared to using standard CNNs.

Usually, studies on pollen classification rely on custom pollen datasets tailored to the specific working conditions of each research team and discipline (Tsiknakis et al., 2022; Punyasena et al., 2022; Theuerkauf et al., 2023). These datasets are usually built by small team of people, and are consequently relatively limited compared to the large datasets required to efficiently train deep learning algorithms. To solve this issue, most studies on pollen classification rely on models already pre-trained on large open-source datasets, such as ImageNet (Deng et al., 2009b), which is composed of more than a million of images of many types, including bikes and cats to streets and landscapes. Models pre-trained on this dataset are available online and are used as a starting point for the training stage, during which they are fine-tuned to the custom pollen dataset. This approach, referred as transfer learning, has proved to perform better compared to training models from scratch (Sevillano and Aznarte, 2018; Geus et al., 2019; Punyasena et al., 2022; Rostami et al., 2023), and is now becoming state-of-the-art. Transfer learning with models pre-trained on the ImageNet dataset also outperformed those pre-trained on the open-source POLLEN13K dataset (Battiato et al., 2020a) for classifying reference pollen images (Kubera et al., 2021), most probably because of the smaller size of the pollen dataset POLLEN13K.

1.3.3.2 Remaining challenges considering our objectives

At the beginning of the PhD, the switch towards routine application conditions was very recent and limited to allergology. A first identified challenge was thus to build an identification method that would process the full range of particles obtained from the detection method when used in routine on environmental samples, including abundant and rare pollen taxa, pollen grains of bad visual quality that are not always identifiable (e.g. covered or blurry), and non-pollen particles (e.g. NPPs, debris, air-bubbles) that might have been incorrectly detected. Similarly to the detection, the second challenge was to assess the performance of the classification under conditions as representative as possible to routine application, to give a reliable assessment of the quality of the automatically generated pollen data.

1.4 Implementing deep learning on environmental samples

1.4.1 General pipeline for both detection and classification

In this study, separate and independent algorithms were used to perform detection and classification (see Chapter II for results showing this strategy provides better results than joint detection and classification). To reach our objectives, i.e. developing a method efficient under routine application on environmental samples and evaluated under such conditions, below requirements were identified for both the detection and classification:

1. Select an algorithm that is efficient on complex images;
2. Build a dataset for training and validating the models that is representative of the environmental samples;
3. Optimize the training stage while avoiding over-fitting, and considering the application of the method to new samples;
4. Evaluate performances on images that are representative of routine application conditions, and proceed to a comprehensive and detailed analysis of errors, namely, assess the causes for errors, and the potential biases;
5. Set rules for the routine application based on all previous steps.

1.4.2 Selecting deep learning algorithms

1.4.2.1 Algorithm selected for pollen detection

In the pipeline developed at the CEREGE lab, a particle segmentation method based on ImageJ and a thresholding approach was used to locate and isolate the studied particles from the stacked images. However, this method was not adequate for our images due to the characteristics of the environmental samples. A different approach to locate and detect the pollen grain in the images, and relying on deep learning, was thus tested and implemented for our automated pollen analysis pipeline.

The object-detection algorithm YOLO (Redmon and Farhadi, 2016) was selected for the detection task, using its latest available version at the time, the version 5 proposed by Ultralytics (Jocher, 2020). Its speed and accuracy as well as its user-friendly interface had made it very popular in the early 2020's. It had thus already proved efficient for a wide range of applications involving complex images with noisy backgrounds, in the agro-industry, e.g. to detect cattle (Brown et al., 2022), kiwi flowers (K. Li et al., 2022), grape (Sozzi et al., 2022), or mold on food surfaces (Jubayer et al., 2021); in medical microscopy images, e.g. to detect leukemia in leukocytes (Abas et al., 2022), malaria (Krishnadas et al., 2022; Zedda et al., 2022; Rocha et al., 2023), or *Escherichia coli* (Sun et al., 2022); or in environmental sciences to detect soil invertebrates (Pruvost et al., 2022), animals (Tan et al., 2022), forest health indicators (Hussain et al., 2021), fishes (Knausgard et al., 2022; Connolly et al., 2022), river floating debris (Zhuang et al., 2022) or algae (Abdullah et al., 2022). Its version 5 had also outperformed other related object-detection algorithms for a wide range of applications (Diwan et al., 2022; e.g. Connolly et al., 2022; Tan et al., 2022; Datar et al., 2018), and including in microscopy images, mainly in medicine (Waithe et al., 2020; Rocha et al., 2023), but also in palynology (Kubera et al., 2022).

The YOLO algorithm is set to perform the detection task jointly with the classification task, i.e. each detected object is also associated to a label. From an input image, the algorithm predicts bounding boxes that frame the detected and identified target objects, e.g. pollen grains. These bounding boxes are rectangles that tightly enclose objects and

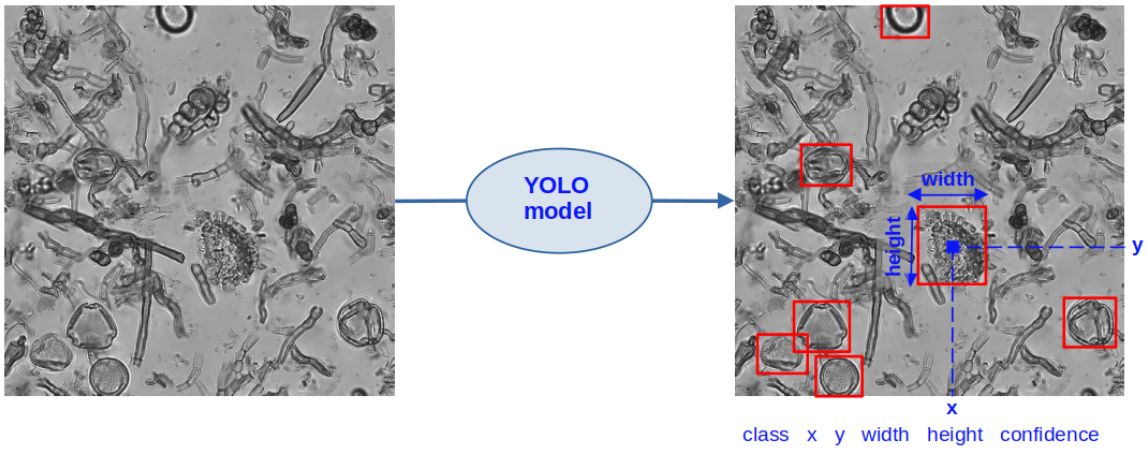


Figure 7: Illustration of predictions from the detection algorithm YOLO

are defined by the coordinates of their center (x, y), their dimensions (width and height), and a label corresponding to the predicted object class (Fig. 7). For each bounding box, a confidence score ranging from 0 to 1 is associated to a prediction. When using the trained model for routine application, bounding boxes with confidence scores above a predefined threshold are typically kept, while those below this threshold are discarded, the threshold being determined a posteriori during performance evaluation.

1.4.2.2 Algorithm selected for pollen classification

The algorithm ResNet152 was selected (He et al., 2015). The Residual Networks (ResNets) have been one of the most tested and approved architecture for pollen classification (Tsiknakis et al., 2022; Rostami et al., 2023; C. Li et al., 2023; Mahmood et al., 2023). In several pollen classification studies, ResNet algorithms reached similar to better performances compared to other related algorithms such as VGG (16,19), AlexNet, MobileNet, Xception, InceptionV3, and including for the classification of the POLLEN73S dataset (Mahmood et al., 2023), of Urticaceae pollen grains (C. Li et al., 2023), of taxa from the Nevada desert (Rostami et al., 2023), or of tropical taxa from a vegetation monitoring (Punyasena et al., 2022).

The input for ResNet152 consists in a bounding box previously detected by YOLOv5, and segmented by a custom python script (Fig. 5). The ResNet model output for each bounding box is a vector of scores, with one score for each pre-defined class, and scores summing up to 1. Typically, the class with the highest score is chosen as the model prediction.

1.4.3 Building datasets to train and evaluate the detection and classification algorithms

1.4.3.1 Building the detection dataset

The objective was to build a dataset representative of the environmental samples and covering the range of image types in such samples. We thus aimed to include as many environmental pollen samples, and thus pollen slides, as possible. This would allow to cover distinct sample contents, in which the types and quantities of spores, debris, and pollen taxa can differ. It would also allow to account for variations in image quality, influenced by both slide mounting (if particles clump together or disperse evenly) and by the image acquisition process (if images are out of focus). This strategy was applied to the annual

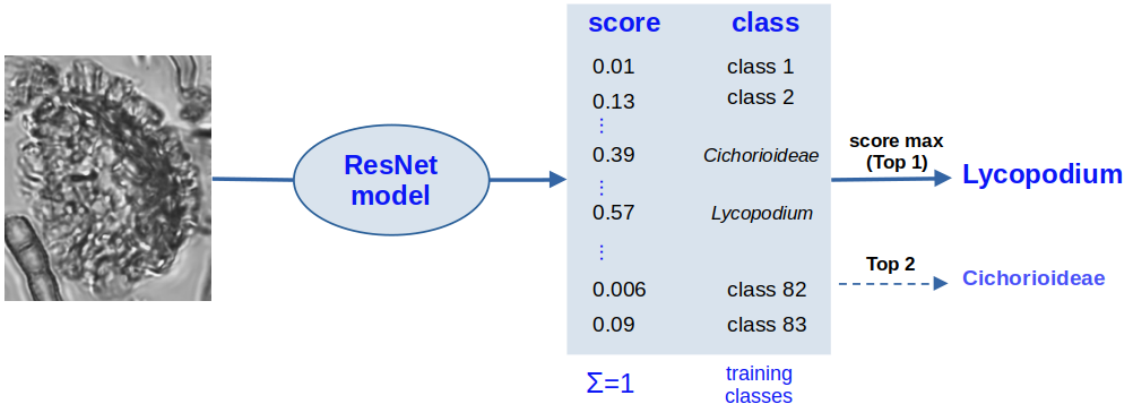


Figure 8: Illustration of predictions from the classification algorithm ResNet

pollen trap samples collected from the vegetation monitoring in Pic Saint Loup. All slides mounted and photographed at the time of the PhD were selected, and one forth of the images from each slide were annotated for the detection dataset, yielding a total of 4096 images from 16 distinct slides and samples (Supporting Information Table S1, see details in Chapter II).

Initially images were annotated to the highest taxonomic level possible, because we aimed to automate both detection and classification with the YOLOv5 algorithm. However, taxa abundances were constrained by natural occurrences, and some taxa were too scarce in the selected images to train the model effectively for their identification (e.g. 7 pollen grains of *Juglans* were found in the 4096 selected FOV images). Additionally, many pollen grains in the images could not be identified with confidence to the desired taxonomic level (e.g. when partially covered), resulting in poorly defined classes that would not be adequate for classification. Due to both of these issues, we decided to use the YOLO algorithm solely for the detection.

Considering the detection task only, all pollen grains and *Lycopodium* spores found in the images selected for training were annotated to the YOLO format by positioning bounding boxes around them, with LabelImg (T. Lin, 2015). Several annotation strategies were tested (see details in Chapter II). The most efficient strategy to detect pollen was finally implemented in the general pipeline. This strategy consisted in annotating all pollen and *Lycopodium* grains, regardless of their state of preservation, their visibility in the image (i.e. also including cut, blurred and covered grains), and using three labels : *Lycopodium* to be able to calibrate counts, pollen of any taxon, and pollen cut on the edge of the FOV.

1.4.3.2 Building the dataset for the classification

Similar to the detection, and to reflect routine conditions on environmental samples, the classification dataset was first built using images generated from the annual pollen samples from the vegetation monitoring. Pollen grains that could be manually identified in the same 16 images used for training the detection stage were selected and sorted by taxon (Reille and Pons, 1990; Punt et al., 2007), partly using Particle Trieur (Marchant et al., 2020b). A classification model only generates predictions among the pre-determined training classes, even if the processed image does not belong to any of them. Thus, to ensure the classification could handle the diversity of particles detected in the previous stage, and limit false positives for target taxa, classes were also built for blurred pollen

grains, pollen grains obscured by debris, and non-pollen objects (e.g. debris, NPPs, air bubble, Fig. 9).

In order to be properly trained, classification CNNs rely on datasets with a large number of examples for each class. The number of images available from the environmental samples were insufficient for many taxa, either due to their low abundance or due to the difficulty in distinguishing them from morphologically similar taxa in 2D images (e.g. *Corylus* and *Betula* in polar view). The images were also missing some pollen taxa expected for routine application. Pollen images from reference samples, i.e. collected on flower anthers from known plant individuals, were thus added to complement the classification dataset. Both environmental and reference images were combined in the final classification dataset, but some taxa were still represented by few images. To cover a maximal range of expected particles while avoiding overly small classes, these images were kept but pooled into a single 'other' class. Meanwhile, the largest classes were cut down to a maximal number of images to avoid excessive differences with the smallest classes (Supporting Information Table S2, see details in chapter III).

The ResNet algorithm works on square images of standard dimensions. To prevent distortion of the particles, a black padding was added to both smallest sides of all images to convert them to squares (Fig. 9).

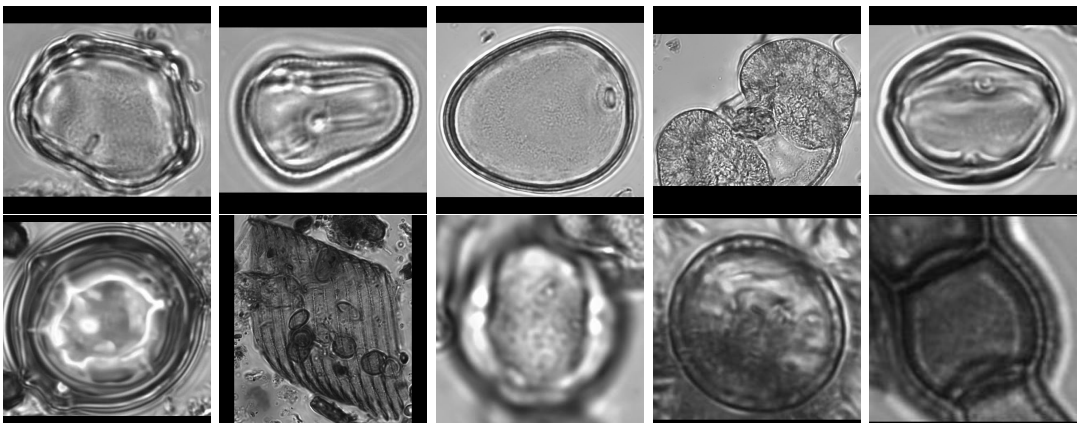


Figure 9: Examples of images from distinct pollen and non-pollen classes included in the classification dataset, resized and padded to squares

1.4.4 Training the models

1.4.4.1 Splitting the dataset and cross-validations

The training stage relies on a training set and on an independent validation set (Supporting Information Methods S1 and Fig. M2). In this study, we also evaluated performances on a third test set of images, set aside and reserved prior to the training stage. When training a model, the datasets were thus consistently divided in three, i.e. into a training set (60%), a validation set (20%), and a test set (20%). Then, models were cross-validated several times by interchanging the images used for training, validation or testing (Fig. 10). This approach allows an efficient use of the dataset, a robust evaluation of the performance, and it improves the method performances when combining the models during application.

Two cross-validation strategies were tested. First, images were divided in three, while ensuring an equal representation of all image types and classes in all three subsets ('standard split', Fig. 10). The models used in the final pipeline for routine application were trained and evaluated based on this approach. The second approach consisted in putting in the test set all images from a specific context, for example a year, or a site, or a type of slides (see Chapter II and IV), while using the other images to train and validate the models. This approach allowed to further assess the generalization abilities of the models ('Leave-One-Out', Fig. 10).

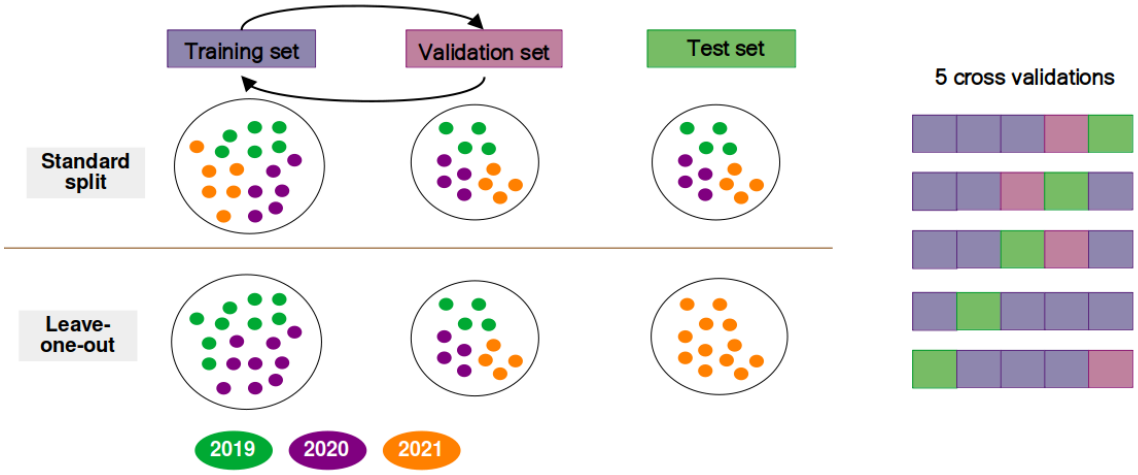


Figure 10: Illustration of the approaches used to split the datasets and cross-validate the models

1.4.4.2 Model training

The YOLOv5 algorithm has a user-friendly interface and most training parameters needed to start model training are already set up. Few parameters have to be chosen, including the number of training epochs, the dimension to which images are resized and the batch size (Supporting Information Methods S1 and Fig. M2). Here, the image and batch sizes were selected following the guidelines provided by Ultralytics on github (https://github.com/ultralytics/yolov5). The number of epochs were adjusted after several tests to keep a number of epoch for which model performance and loss curves plateau. Models were pre-trained on the large-scale detection dataset COCO (T.-Y. Lin et al., 2014). The training were conducted using the Jean-Zay Nvidia V100 GPU (IDRIS, CNRS). The YOLOv5 model was implemented using the PyTorch library.

For the classification, implementing the ResNet algorithm required more coding. The standard suite of data augmentation techniques were applied to the training set, including rotations, flips, shearing, focus changes and saturation level changes. Model training was conducted over 150 epochs with an early-stopping callback of 30 epochs to stop training when performance plateaued. Transfer learning was used, based on the ImageNet classification dataset (Deng et al., 2009b). Implementation was conducted using the Keras library and the models were run on a graphic card with access provided by researchers from the Chennai Mathematical Institute, with whom we collaborated for the development of the pollen classification task.

1.4.5 Evaluation of the performance

1.4.5.1 Performance metrics

To evaluate the performance of the trained models, predictions were compared to manually labeled ground truths to determine the following statistics: (1) True Positives (TP) for correct predictions that match the ground truths, (2) False Positives (FP) for incorrect predictions that do not match the ground truths, and (3) False Negatives (FN) for ground truth instances that were not correctly predicted. For the detection, FP refers to the false detection of a non-target object, e.g. a debris, and FN refers to the absence of detection for a target object, e.g. a pollen grain (Fig. 11). For the classification, a FP for class i, e.g. for *Buxus*, refers to an object belonging to another class j, e.g. *Olea*, but assigned to class i, and a FN refers to an object of class i but assigned to class j (Fig. 11).
From these statistics, the following performance metrics can be measured :

- the recall R, which evaluates the rate of false negatives, with

  $R = \frac{TP}{TP+FN}$

- the precision P, which evaluates the rate of false positives, with

  $P = \frac{TP}{TP+FP}$

- the F1-score, which provides a balanced measure of the two previous metrics, with

  $F1_{score} = 2 \times \frac{precision \times recall}{precision+recall}$

- the accuracy A, which quantifies the proportion of correct predictions out of the total number of predictions, using TP and FN for all images regardless of classes, with

  $A = \frac{TP}{TP+FN}$

For the classification task, accuracy can be evaluated in two ways. The microaccuracy corresponds to the overall accuracy across all images, regardless of class. The macroaccuracy corresponds to the average of the accuracies measured for each class separately (i.e. average of the recall rates per class), giving an equal weight to all classes despite potential imbalance in class distribution.

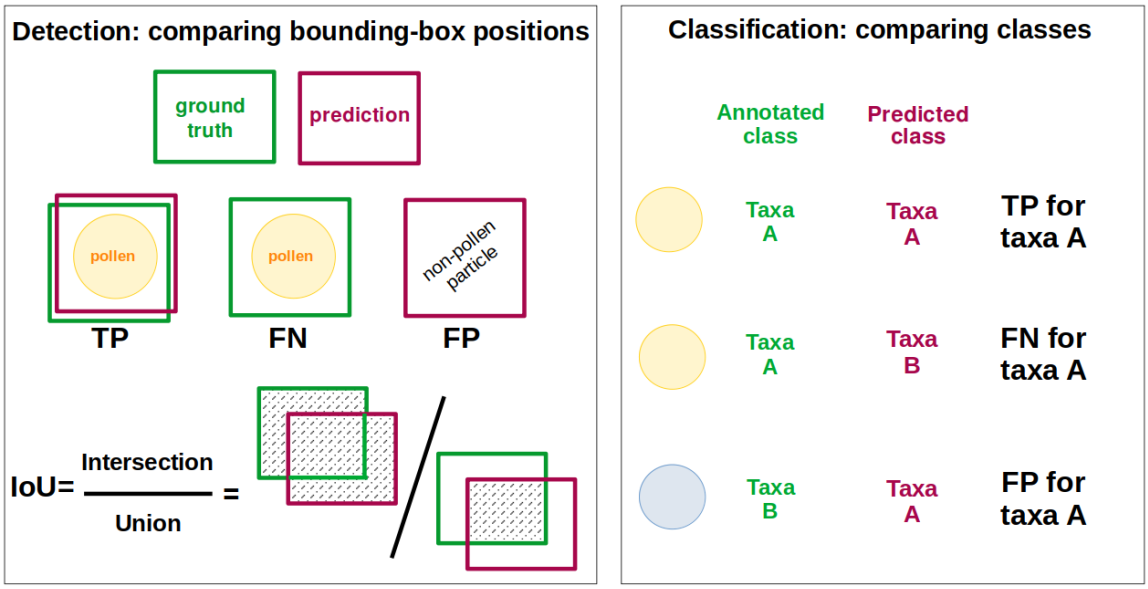


Figure 11: Illustration of True Positives (TP), False Negatives (FN) and False Positives (FP), both for the detection and the classification tasks

1.4.5.2 Evaluation for the detection task

For detection, comparing ground truths and model predictions consists in evaluating the overlap between predicted and manually located bounding boxes, which is done using a threshold for the IoU (Intersection Over Union) metric (Fig. 11). For each image, manually labeled and predicted bounding boxes are compared in pairs, allowing to determine matches (TP) if the IoU is above a pre-determined threshold, and mismatches if an annotated bounding box is not overlapping with any prediction (FN), or if a predicted bounding box is not overlapping with any annotation (FP).

Each predicted bounding box is associated to a confidence score ranging from 0 to 1. The TP, FP and FN statistics can be determined for distinct threshold values used to filter predictions for evaluation. Receiver Operating Characteristic (ROC) curves can then be generated by plotting precision and recall metrics measured across a range of confidence score thresholds, from 0 (for which all predictions are kept, resulting in maximal recall and minimal precision) to 1 (for which all predictions are discarded, leading to minimal recall and maximal precision). These curves provide a comprehensive view of the model performance and illustrate the trade-off between false positives and false negatives. These curves are useful to select a confidence score threshold tailored to the specific objectives of each study.

1.4.5.3 Evaluation for the classification task

Predictions from a classification model for an image consists in a vector of scores. Usually, the class predicted with the highest score is used as prediction for the evaluation. The predicted class can then directly be characterized as TP, FP or FN by comparing it to the ground truth class (Fig. 11). From these statistics, the standard and previously described performance metrics (section 4.5) can be measured, and a confusion matrix can be generated. A confusion matrix visually gives the number of correctly and incorrectly classified images to identify and analyze misclassification errors, and for instance to determine which classes are frequently misclassified with each other.

Another approach to evaluate performances consists in considering a few classes predicted with the highest scores, i.e. the i^{th} top-scoring predictions. In that approach, a prediction is considered correct (true positive) if the ground-truth class is among the i^{th} top-scoring predictions. Performance metrics can then be determined when sequentially increasing the number of considered top-scoring classes, to find out the best i. On a practical point of view, this approach can help to determine how many top-scoring classes to keep to maximize the chance of including the correct class among the predictions. For the developed pipeline used in routine, the top-4 scoring classes were used to generate a prediction for each new pollen image (see details Chapter III).

1.4.6 Setting rules for routine application and implementation of full pipeline

In the final detection method for routine application, we use the 5 cross-validated models trained on the simplest annotation strategy (see Chapter II for more details). When processing a new FOV image, all 5 models were successively applied and their predictions were combined to keep the most confident predictions. The coordinates and dimensions of the predicted bounding boxes were used to segment one new image per bounding box that contained a single particle (Fig. 12). To ensure that the entire detected object is included within the segmented image, a padding of 10 pixels was added to all its sides. This detection method was applied to all the FOVs images obtained when scanning new slides.

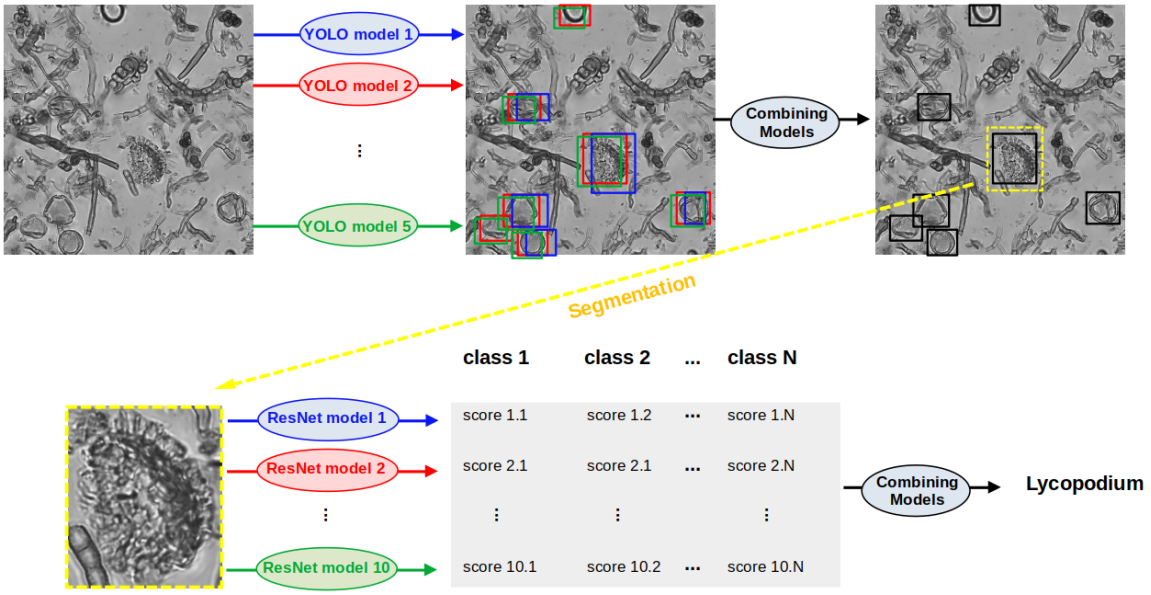


Figure 12: Final pipeline to identify the pollen content in the FOV images

All detected particles were then processed with the classification method. Similarly to detection, all 10 cross-validated models were sequentially applied to each image, and their predictions were combined to generate a single final class, using the top-4 evaluation approach, and an ensemble modeling strategy referred to as 'max voting' (Mohammed and Kora, 2023) and based on the occurrences of classes within the predictions (see details in Chapter III).

1.4.7 A pipeline tailored to address environmental-related questions

For both the detection and classification steps, the focus was to evaluate performances with emphasis on routine application on environmental samples. Performance evaluations were conducted beyond standard performance metrics, to provide a deeper understanding of the causes of errors and how errors were distributed. The objective was not only to generate an automated pollen analysis pipeline but to generate one tailored to the specific challenges and requirements of environmental-related disciplines relying on pollen data. Analyses thus mainly consisted in investigating whether errors could bias the results and affect the accuracy of the final pollen counts produced by the automated pipeline. This notably included a thorough analysis of detection errors, which had never been previously conducted for automated pollen analysis (see Chapter II). For classification, evaluation were especially conducted for critical taxa and classes relevant to environmental-related studies in the Mediterranean area (see Chapter III).

Overall, developments and their evaluation aimed to assess whether the automated method could reliably support scientific research related to environmental monitoring and environmental research in the Mediterranean area.

1.5 Conclusion

The developments of each method sections were very exploratory, many tests were conducted in different scientific disciplines, from the laboratory, the microscopy image acquisition, the pollen identification, to the implementation of deep learning algorithms.

Finally, the method for the automated analysis of the slide images is finally based on the combination of the YOLOv5 detection algorithm (Jocher, 2020) and the ResNet153 classification algorithm (Chollet, 2015) to detect and classify pollen grains in images containing 78 Mediterranean pollen taxa and many different types of debris (grouped in one class), sampled from annual pollen traps.

When compiling all the developed steps, the fully integrated automated pollen analysis pipeline follows these steps :

1. Mounting fixed slides;
2. Scanning the slides with the automated microscope;
3. Applying the trained and evaluated detection models (YOLOv5) to detect all pollen grains;
4. Segmenting each detected bounding box from their original FOV image to generate images containing a single particle each;
5. Applying the trained and evaluated classification models (ResNet153) and processing the top-4 predictions from all cross-validated models to generate one class per image and remove uncertain predictions;
6. Compiling all predictions obtained for each slide to generate calibrated pollen counts for each new environmental sample.

Use of artificial intelligence for other tasks

Artificial intelligence was also used to help in the writing process for the manuscript for translation between the English and French languages by implementing sentences or portion of sentences and then manually editing the translated portions; providing examples of reformulations by implementing sentences or portions of sentences and then manually editing the reformulated portions; and for finding synonyms (deepl.com and chatgpt.com).

Supporting Information (SI)

Methods S1 : General details on the architecture of deep learning algorithms

Deep learning algorithms are advanced techniques of machine learning, designed to model complex patterns in data (Goodfellow et al., 2016), they are very powerful to automate tasks for image analysis. This field is largely dominated by neural networks, i.e. algorithms characterized by an architecture composed of multiple layers of interconnected nodes, commonly referred to as neurons. Their architecture typically includes an input layer, several hidden layers, and an output layer. The input layer receives initial values, for instance the pixel intensities for an image. These values are then processed through each hidden layer, which performs a combination of linear operations (involving weights and biases) and non-linear transformations (using activation functions). The final output layer generates the prediction, for instance a class for an image when using a classification algorithm (Fig. M1.b).

Deep Convolutional Neural Networks (deep CNNs) are a specialized type of neural networks designed for computer vision tasks, and relying on convolutional products (between 2D matrices referred as kernels and the input image to produce 3D feature maps, i.e. several smaller 2D matrices, Fig. M1.a). This process allows to effectively extract features from input images, such as edges, shapes, as well as complex and more abstract patterns (LeCun et al., 1989).

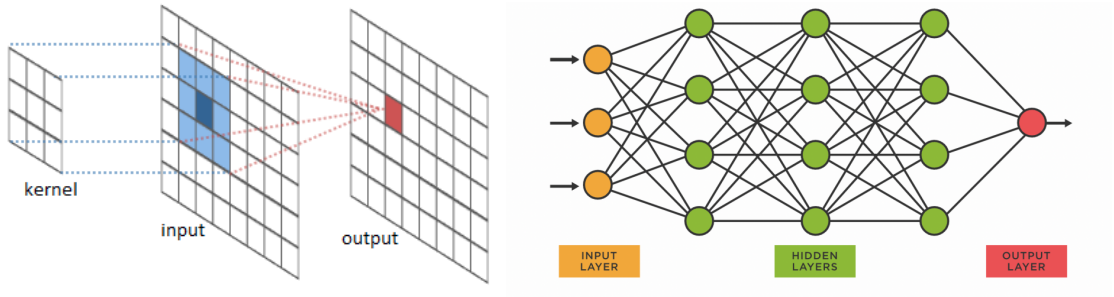


Figure M1: Simplified illustrations of (a, left) convolutional filters, and (b, right) the architecture of a deep neural network (online images)

Deep learning algorithms rely on a training stage, and a training dataset composed of labeled images, i.e. associated to their ground truth. The training stage consists in iterative steps during which the model parameters, namely the model weights and biases are fitted to the training dataset (Fig. M2). Initially, the model weights and biases are set randomly, or are based on those obtained from a previous training when using transfer learning.

At each iteration, referred to as an epoch, input images are fed to the model, which generates predictions based on the current values of its weights and biases. Differences between the predicted outputs and the ground truths are quantified with a loss function, and the measured error values, or loss, is used to adjust the model weights and biases. To do so, the errors are propagated backward through the network using a gradient descent algorithm that calculates values (gradients) associated to each weights and biases (back-propagation). These gradients are then combined with a second value, referred to as learning rate, to modify each weights and biases, and generate new predictions with expected reduced errors (forward propagation). Additionally to the value of loss measured at each epoch, the model is also evaluated on a separate 'validation' set, to get performances on unseen data and

assess over-fitting, i.e. whether the model overspecializes on training data and loses ability to generalize.

Overall, this process consists in adjusting via iterations the model parameters until the loss converges to a minimum, i.e. the model predictions match the ground truths. It involves alternating learning and evaluation stages, and relies on two subsets of annotated images, referred to as training and validation sets. At each iteration, all training images are fed to the model but separately per batch, i.e. subsets of the data.

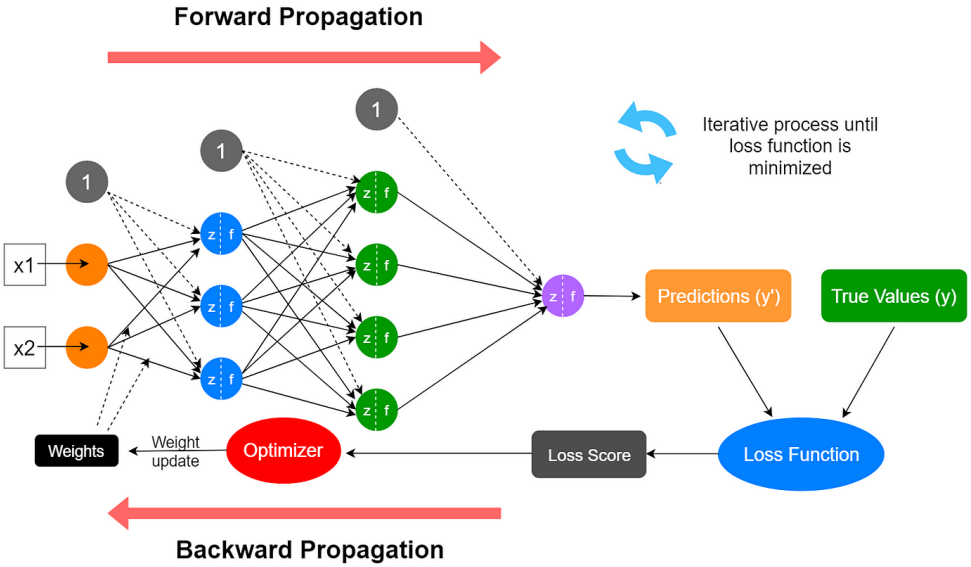


Figure M2: Simplified illustration of the training iterative stage of a deep learning algorithm (online image)

Table S1 : Summary of the annotations for the detection dataset

Number of images of field of views (FOVs) annotated, and total number of bounding boxes annotated, from the slides mounted from annual passive traps collected for the vegetation monitoring in the Pic Saint Loup area, and used to set the detection method (chapter II).

Table S1

site	year	nb fov	nb of bounding box
MAU2	2021	256	1775
MAU2	2020	256	349
MAU1	2021	256	1041
PSLAE	2019	256	1316
PSL2	2021	256	858
PSL11	2020	256	482
PSLH	2020	258	669
PSL2	2020	256	864
PSLH	2019	256	844
PSLAE	2021	256	740
PSL11	2021	256	167
PSLAE	2020	256	675
MAU1	2020	256	400
PSL11	2019	256	724
PSL2	2019	256	531
PSLH	2021	256	1096
total of 6 sites	total of 3 years	4096	12531

Table S2: Number of images per class in the classification dataset

Number of individual pollen images, i.e. detected and segmented from the original FOV image, compiled in the classification dataset (chapter III), and including images manually identified from the FOVs images from the passive trap samples collected near the Pic Saint Loup, i.e. environmental images ('env'), and from slides from collection, i.e. reference images ('ref'). 'Indetermined tpi' refers to pollen grains of good visual quality belonging to a single pollen type, but from which the taxon could not be identified.

Table S2

Class	nb env	nb ref	total
Acacia	1	58	59
Acer	89	842	931
Alisma	1	349	350
Alnus	105	454	559
Amarantaceae	81	1218	1299
Apiaceae	7	0	7
Artemisia	33	1235	1268
Asphodelus	7	0	7
Betula	11	1289	1300
Brassicaceae	181	1119	1300
Buxus	926	374	1300
Cannabaceae	0	115	115

Table S2: (Continued)

Class	**nb env**	**nb ref**	**total**
Dipsacaceae tp.	4	39	43
Viburnum-Sambucus tp.	21	366	387
Carduus	74	910	984
Carpinus	0	56	56
Caryophyllaceae	0	685	685
Castanaea	0	1279	1279
Centaurea	4	250	254
Cichorioideae	102	1198	1300
Cistus	0	124	124
Convolvulus	29	320	349
Cornus	0	1228	1228
Corylus	36	1098	1134
Cupressus	432	556	988
Cyperaceae	52	86	138
Echium	1	1161	1162
Ericaceae	63	653	716
Euphorbiaceae	0	297	297
Fabaceae	0	1029	1029
Fagus	4	40	44
Fraxinus excelsior	225	106	331
Fraxinus ornus	54	1246	1300
Gallium	58	530	588
Geraniaceae	10	23	33
Hedera	7	0	7
Helianthemum	11	25	36
Hypericum	0	279	279
Ilex aquifolium	0	36	36
Indeterminable pollen blurry	440	0	440
Indeterminable pollen covered	487	0	487
Juglans	7	185	192
Juniperus	421	720	1141
Lamiaceae	16	446	462
Ligustrum	0	410	410
Liliaceae	6	448	454
Linum	0	99	99
Lonicera	0	38	38
Lotus	0	51	51
Lycopodium	1300	0	1300
Lysimachia	0	115	115
Malva	0	17	17
Moraceae	0	642	642
Myrtaceae	0	547	547
Non pollen	1162	138	1300
Olea	326	974	1300
Ostrya	0	321	321
Papaveraceae	0	80	80
Phillyrea	263	1038	1301
Pinaceae	954	331	1285
Pistacia	230	817	1047

Table S2: (Continued)

Class	**nb env**	**nb ref**	**total**
Plantago	345	956	1301
Platanus	70	539	609
Poaceae	956	342	1298
Polygonaceae	3	0	3
Populus	0	1300	1300
Quercus cerris	0	868	868
Quercus coccifera	0	1300	1300
Quercus deciduous	121	1177	1298
Quercus ilex	677	624	1301
Ranunculaceae	10	839	849
Rhamnus	96	260	356
Rosaceae	385	915	1300
Rumex	52	1056	1108
Salix	7	258	265
Sanguisorba	48	10	58
Scabiosa	10	55	65
Sedum	0	234	234
Tamarix	0	528	528
Tilia	1	50	51
Tsuga	0	8	8
Ulmus	21	428	449
Urtica	0	1248	1248
Valeriana	7	0	7
Vitis fertile morph	572	729	1301
Vitis sterile morph	183	1078	1261
Xanthium-Ambrosia tp.	34	1264	1298
Indetermined tp1	3	0	3
Indetermined tp2	13	0	13
Indetermined tp3	15	0	15
Indetermined tp4	75	0	75

Chapter 2

A user-friendly method to get automated pollen analysis from environmental samples

This chapter corresponds to a published article in the review New Phytologist (2024). doi: 10.1111/nph.19857.

Betty Gimenez, Sébastien Joannin, Jérôme Pasquet, Luc Beaufort, Yves Gally, Thibault de Garidel-Thoron, Nathalie Combourieu-Nebout, Laurent Bouby, Sandrine Canal, Sarah Ivorra, Bertrand Limier, Jean-Frédéric Terral, Céline Devaux∗, Odile Peyron∗
∗These authors contributed equally to this work.

Key words : artificial intelligence, automated pollen analysis, deep learning, detection errors, environmental real-world samples, guidelines, Mediterranean vegetation monitoring, YOLOv5

2.1 Introduction

Pollen is a major tool for ecological, paleo-environmental and evolutionary studies, used to monitor plant responses to environmental changes (van der Knaap et al., 2010), inform on plant-pollinator interactions (Morente-López et al., 2018), reconstruct past vegetation and climate (Peyron et al., 2017), anticipate allergology (Anderegg et al., 2021), infer honey origin (Corvucci et al., 2015) or predict harvests (Oteros et al., 2014). Counting and identifying pollen is traditionally performed manually by experts, using a slide under light microscopy. These tasks are time consuming, and limit the size and replicability of pollen studies. Automation of pollen analysis, a long-standing objective in palynology, can help open new research avenues by extending spatial and temporal resolution of pollen studies, and help obtain standardized data comparable among years, sites and research teams (Stillman and Flenley, 1996; K. A. Holt and Bennett, 2014).
The rapid development of Convolutional Neural Networks (CNNs) for image analysis now makes routine automated pollen analysis achievable, as demonstrated in recent studies (e.g. Barnes et al., 2023; Khanzhina et al., 2022; Punyasena et al., 2022). When applied on slides scanned under light microscopy, automation of pollen analyses relies on (1) the detection of pollen grains, i.e. finding their position in an image, and (2) the classification of the detected objects into predefined classes, e.g. pollen taxon (Diwan et al., 2022). Classification is usually performed separately from detection, on images containing a single pollen grain (Olsson et al., 2021; Punyasena et al., 2022; Viertel et al., 2022). In contrast to classification, interest and progress for automated pollen detection remains limited, yet, detection is a crucial first step and a challenging task, especially when applied to images containing many pollen and non pollen objects of many kinds, as in environmental samples. The CNNs-based object-detection algorithms, notably

Fast-RCNN (Ren et al., 2015), RetinaNet (T.-Y. Lin et al., 2017), or YOLO (Redmon et al., 2016), and one of its latest versions YOLOv5 (Jocher, 2020), can now perform detection jointly with classification at high speed and accuracy, on almost any object, e.g. molds in fruits (Jubayer et al., 2021), arctic benthic fauna (Marini et al., 2022), Plasmodium falciparum (Zedda et al., 2022), and even on complex images with dense and heterogeneous backgrounds (Diwan et al., 2022; Jiang et al., 2022). In palynology, the implementation of these algorithms is recent (Gallardo-Caballero et al., 2019), and remains mainly limited to allergology, on images containing few pollen taxa and a uniform background, or on airborne samples that contain fresh and well preserved pollen with few debris (Gallardo-Caballero et al., 2019; Khanzhina et al., 2022; Kubera et al., 2022). Under such conditions, studies report excellent results, e.g. (1) for 11 pollen types, the combination of Faster R-CNN and RetinaNet correctly detected 98.54% of the annotated pollen grains, and wrongly detected as little as 0.25% of non pollen objects (Gallardo-Caballero et al., 2019); (2) for 13 allergenic pollen species, a modified RetinaNet ("BayesianRetinaNet network") correctly detected 96.32% of pollen grains, and also correctly classified 97.66% of them (classification F1-score) (Khanzhina et al., 2022), (3) for three Betulaceae pollen taxa, YOLOv5 correctly detected and classified 89.7 to 98.9% of the pollen grains, and 91.7 to 97.8% of the predictions were correct (Kubera et al., 2022).

These results however do not apply for pollen-based research typically relying on environmental samples from the "real-world", such as gravimetric pollen traps, moss pollsters, and sediment records, which can contain many debris and damaged pollen grains, with an uncontrolled and potentially high diversity of pollen taxa. For these environmental samples, the methods for the automation of pollen analyses are still in development. The first two attempts to automate pollen analysis on such samples recently achieved promising results: (1) for gravimetric traps placed in a tropical forest, 83.7% of the pollen grains were correctly detected, and 89.5% of detected grains were correctly classified into 25 selected pollen taxa (Punyasena et al., 2022); (2) for pollen samples from lake sediments, from 87.2 to 99.1% of pollen grains were correctly detected, 84% were correctly classified into 11 selected pollen taxa, and 7% were incorrect classifications (Theuerkauf et al., 2023). These results pave the way for new investigations, especially to understand the impact of detection errors on the accuracy of automated pollen analysis, because any error at this step will inevitably propagate, e.g. to the classification step.

In this study, we aim to (1) improve the process of pollen detection in environmental samples containing large amounts of debris and pollen taxa, by analyzing the variation in detection errors, (2) find general implementation guidelines applicable to any pollen study, and (3) evaluate the joint detection and classification errors in a full automated analysis. We do so using gravimetric pollen trap samples collected annually in the Mediterranean area. As we aim to make automated pollen detection accessible to non-experts of pollen or deep learning, we rely on simple and common tools: mounted slides scanned under light microscopy, and the open-source and user-friendly algorithm YOLOv5. We search for the best annotating strategies that balance workload and performance for studies of a single taxon or an assemblage of taxa, and for studies further extended in time or space, e.g. as in long-term plant monitoring. We also study in detail the causes for the detection errors, using information on pollen morphology and the image quality. We finally assess automated and joint detection and classification on five pollen taxa common in our dataset, and also compare automated results to the ones made by an expert palynologist.

2.2 Materials and Methods

2.2.1 Pollen samples and image acquisition

We used pollen samples from gravimetric traps collected in 2019, 2020 and 2021 for a project monitoring vegetation in 6 locations in a Mediterranean massif. The pollen traps consisted of containers with a 5 cm width opening, and a 5 mm mesh containing glycerin and thyme essential oil to retain and preserve pollen grains, and avoid fungus growth. Traps were placed on the ground or attached to a tree in early January, and collected one year later. To calibrate pollen counts, tablets of *Lycopodium* marker spores (*Lycopodium* clavatum L.) were added to the pollen samples (Stockmarr, 1971). The samples were then chemically treated to remove calcium carbonates and silicates, and were acetolysed for 6 to 8 minutes. For image acquisition, one fixed slide per sample was mounted with glycerin jelly, under 16 mm × 16 mm cover slides. Two samples were discarded due to insufficient amount of pollen. We therefore worked with 16 slides mounted from 16 samples.
Microscopic images of each slide were acquired with an automated bright light microscope Leica DM6 B TL BF (63× magnification under oil immersion). The imaging was done by an Hamamatsu ORCA FLASH camera with a 2048 × 2048 pixel camera sensor. We used the image acquisition pipeline developed by Tetard et al. (2020), including a LabVIEW interface. We scanned only 17% of each slide to decrease acquisition time of images, but covered the entire and potentially heterogeneous distributions of pollen within each slide by acquiring images in 16 squared areas, arranged in a 4 × 4 grid; each scanned area consisted of 64 (8 × 8) fields of views (FOV) of 214 × 214 µm with an overlap of 10 µm (smaller than the smallest pollen taxon here). For each FOV, 11 images were taken along the z-axis, spaced 8 µm apart, to produce a stack capturing the vertical details of the pollen grains (Fig. S1 step 1). The depth resolution was a compromise between acquisition time and the need for details of pollen grains, the sizes of which range from ∼12 µm to ∼150 µm. For each of the 16 slides, piles of images were automatically taken for 1,024 fields of views in ∼2h30. Each pile of images was then stacked using Helicon Focus 7 (Tetard et al., 2020b), which selects the sharpest areas and discards the unfocused areas to create a single final composite 2D image, hereafter called FOV (Fig. S1 step 1). The FOVs from 4 out of the 16 scanned areas, from the top-left / bottom-right diagonal of each slide, were selected to train and evaluate the YOLOv5 algorithm, totaling 4,098 FOVs for the full dataset (16 slides × 4 scanned areas × 64 FOVs, with 2 exceptions). The FOVs from the other 12 scanned areas were never used during training, but used to generate the automated pollen counts compared to those obtained by an expert palynologist, on distinct slides mounted from the same samples.

2.2.2 Optimization of the object-detection algorithm for detecting pollen

We selected the light version of the algorithm YOLOv5 among its several releases (Jocher, 20200) as it performs as well as the heavy version on pollen (Kubera et al., 2022). More details on YOLOv5 are available in the Supporting Information (Methods S1). The algorithm YOLOv5 performs detection and classification in a single step, by (1) predicting bounding boxes (defined by their width, length, and location within an image) around the detected objects, and jointly (2) predicting a label, among predefined classes, for the objects within the bounding boxes. Confidence scores are also provided for all predictions. The image dataset used to train the algorithm had thus to be manually annotated by (1) tagging the targeted pollen and *Lycopodium* grains, i.e. placing bounding boxes that frame them, and (2) labeling these grains, e.g. as a pollen taxon. We used the labelImg software (T. Lin, 2015), and annotated a total of 12,531 pollen and *Lycopodium* grains in

the 4,098 FOVs of the dataset (Fig. S1 step 2). We split this annotated dataset into 60% for training, 20% for validation and 20% for testing, unless otherwise mentioned, with an equal contribution of all 16 slides in each subset. We performed a five-fold cross validation, by interchanging the FOVs from the training, validation and test datasets for each analysis (Fig. S1 step 3). Models were trained for 150 epochs, i.e. training iterations, and on images resized to 640 pixels, which took less than 3 hours per training using Jean-Zay Nvidia V100 GPU (IDRIS, CNRS). The number of epochs was chosen empirically to enhance model performance without inducing over fitting. To evaluate the performance, we applied the model saved after the last training epoch on the test datasets made of annotated FOVs never seen before. We eliminated the predicted bounding boxes that overlapped using an IoU (Intersection over Union) threshold of 0.7 (Supporting Information Methods S2), and that had a confidence score below 0.45 (Supporting Information Methods S3). To evaluate the performance of the sole detection, we then compared the manually annotated and the predicted bounding boxes, without taking into account the labels predicted by the algorithm through its joint classification (Supporting Information Methods S4). We used an IoU of 0.5 between annotated and predicted bounding boxes to determine (1) a true positive (TP) for a bounding box both predicted and manually annotated, e.g. a pollen grain correctly detected, regardless of its classification (2) a false negative (FN) for a bounding box manually annotated but not predicted, e.g. a pollen grain not detected, and (3) a false positive (FP) for a bounding box predicted but not manually annotated, e.g. a debris falsely detected as pollen. Combinations of these statistics produced (i) the recall, percentage of correctly detected grains among all true grains, (ii) the precision, percentage of correctly detected grains among all detected objects, (iii) F1-score, the harmonic mean of recall and precision, and (iv) the ROC (Receiver Operating Characteristics) curves, which all inform on the power of the models to discriminate pollen grains from the background; see details in the Supporting Information (Methods S5).
To evaluate the adequacy of the size of our dataset (4,098 FOVs including 12,531 annotations), we trained the models on an increasing number of FOVs. We split each annotated dataset into 80% for training, and 20% for validation, with five cross-validations, but, we systematically tested the trained models on the same annotated dataset (Fig. S1 step 3b). For these tests, we used the models from both the last training epoch, and from the epoch providing the best performance on validation. F1-scores showed that the best and last epoch models provided similar performances, we therefore chose to use only last epoch models for all below analyses (Supporting Information Fig. S1c, Notes S1).

2.2.3 Optimization of the annotation strategies tailored for distinct pollen studies

We evaluated the performance of the three following annotation strategies, a priori common in any pollen-based studies, to provide useful and efficient guidelines on how to balance performance and workload (Fig. S1 step 2, Supporting Information Methods S6):
- all-0taxon considered the simplest annotation strategy consisting of only 3 labels associated to the tagged bounding boxes, with no information on pollen taxon: *Lycopodium*, pollen of any taxon, and grains of either pollen or *Lycopodium* cut on the FOVs edge with less than a quarter of their surface visible (Fig. S1 step 2a);
- all-5taxa used the same above dataset but divided the aforementioned pollen category into 6 labels: Pinaceae (Pinaceae sp.), *Buxus* (*Buxus sempervirens* L.), Poaceae (Poaceae sp.), *Quercus* (*Quercus* sp.), Oleaceae (Oleaceae sp.), and the label pollen for other taxa,

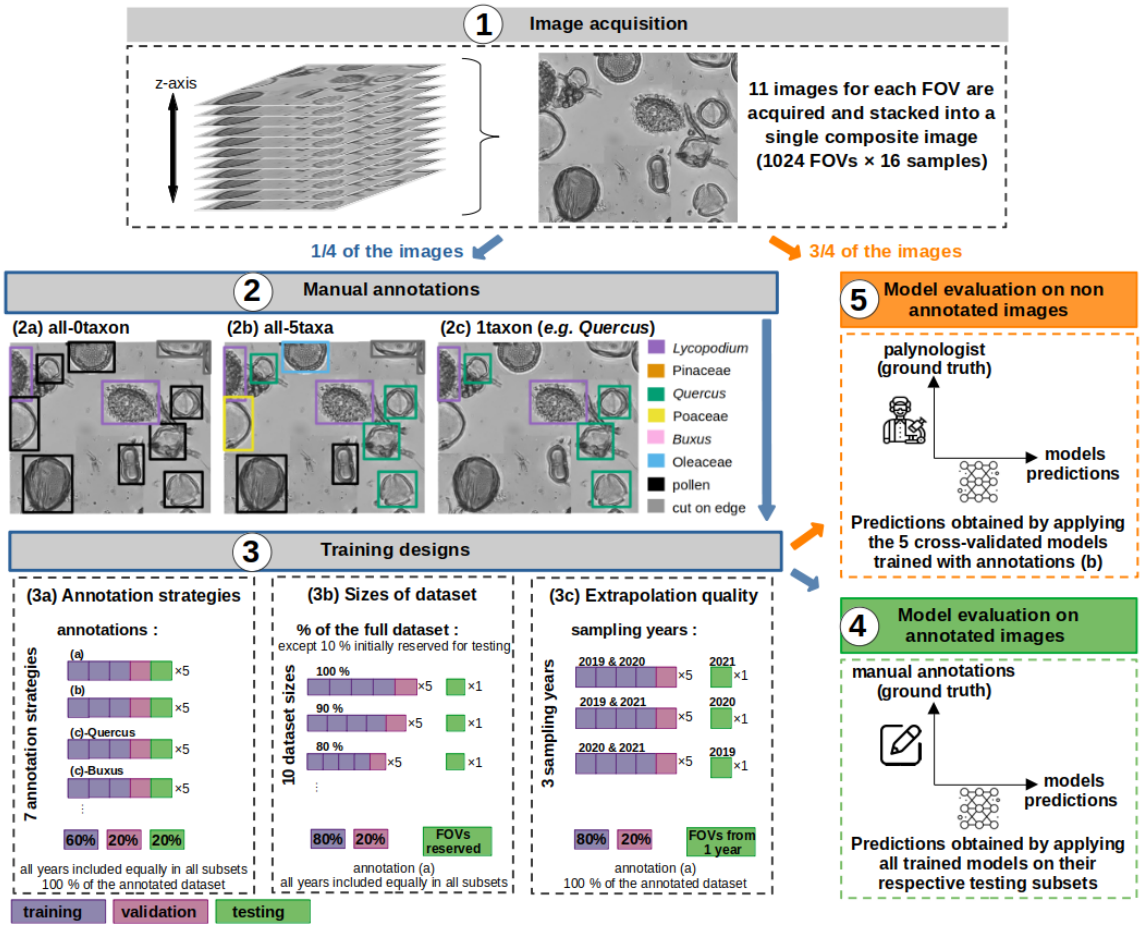


Figure S1: Experimental design. (1) A pile of 11 images at distinct focal depths is acquired for each field of view (FOV), and then stacked into a single 2D composite image. (2) One fourth of all images (1024 FOVs × 16 samples × $\frac{1}{4}$ = 4'096) are annotated following 3 distinct annotation strategies (2a-2c). The annotation strategies are illustrated on a composite image: (a, all-0taxon) tagging 12 bounding boxes with 3 labels (*Lycopodium*, pollen and grain cut on edge); (b, all-5taxa) tagging 12 bounding boxes with the same 3 previous labels and 5 extra labels (Pinaceae, *Buxus*, Poaceae, *Quercus*, Oleaceae); (c, 1taxon-1taxon, e.g. *Quercus*) tagging 6 bounding boxes with 2 labels (*Lycopodium* and a model taxon); see details in Supporting Information (Methods S6). (3) The annotated datasets are split into subsets to train, validate and test the models with a 5-fold cross-validation, and following distinct designs (3a-3c) tailored to distinct questions. (4) The performances are evaluated on the testing subsets by comparing manual annotations with predictions obtained with the respective trained models. (5) The models trained with the most detailed annotation strategy 2b (5 models from 5 cross-validations) are applied on non annotated images (1024 FOVs × 16 samples × 3/4 = 12'288), and predictions are compared with manual counts made by a palynologist on distinct slides mounted from the same samples.

could they be determined or not. We kept the labels *Lycopodium* and grains cut on the FOVs edge, thus increasing the number of labels from 3 to 8 to test for the effect of including information on taxonomy. These five taxa selected were the most frequent ones in the samples, together representing 43% of the tagged grains (Fig. S1 step 2b);
- 1taxon-1taxon restricted pollen detection and thus annotation to a single model taxon, thus labeling bounding boxes only for *Lycopodium* and alternatively one of the five taxa

mentioned above (Fig. S1 step 2c).
Lycopodium, used for calibration of pollen counts, was systematically tagged and labeled in all annotation strategies, while debris were never annotated. We compared the detection performance and analyzed the detection errors for each of the tagged pollen taxon under the three annotation strategies. At most 12,531 grains were tagged in the 4,098 FOVs, among which 984 were labeled as *Lycopodium*, 5,402 as one of the five taxa, 4,615 as pollen with no associated taxon, and 1,530 as pollen cut on the FOVs edge. Pollen grains were labeled with a taxon only when they could be confidently determined; otherwise they were categorized as pollen, meaning that pollen from the 5 previously mentioned taxa could have been left in this pollen category.

Finally, we assessed the performance of the trained models when applied to new samples never seen by the algorithm. We trained the algorithm on FOVs from all but one sample year (split randomly with a ratio 80:20 for training and validation), and evaluated its performance on 820 random FOVs from the sampling year left apart. We used the simplest all-0taxa annotation strategy, and cross-validated results five times for each year (5×3 in total), by interchanging the training and validation datasets, and while keeping the test sets unchanged for each of the three sample years (Fig. S1 step 3c). The sizes of the training datasets (2186 ± 99 FOVs) were larger than the dataset size at which detection performance plateaus ($\sim$1,770 FOVs; Supporting Information Fig. S2).

2.2.4 Analyses of the detection errors

We analyzed the variation in the percentage of pollen grains left undetected (false negative), using the 5 following categorical variables, which were included during the annotation process but never used for training: (1) identification, with the two levels determined or not for $\sim$35 taxa, (2) taxon, with the five levels *Buxus*, Pinaceae, Poaceae, Oleaceae or *Quercus* (the five most common taxa), (3) visible section, with the two levels fully visible within the FOVs or not, (4) image quality, with the two levels good, or poor for grains that are unfocused, covered by a debris or damaged in the image, and (5) deterioration type for poor quality images only, with the three levels covered, unfocused, or mixed deterioration (Supporting Information Methods S7, Methods S8). False positives predicted as pollen grains were not studied as they all corresponded to debris.

2.2.5 Analyses of the joint detection and classification errors in a full automated analysis

We evaluated the performance of the joint detection and classification produced with the all-5taxa annotation strategy (Fig. S1 step 2b, Supporting Information Methods S6), and compared it to the sole detection error to evaluate whether they can or not compensate. First, we evaluated the performance of the full automation on the same annotated test sets as before (Fig. S1 step 4). We used the confusion matrix (Supporting Information Methods S4a) to compare for each of the 5 taxa, the counts from automated predictions and from manual annotations, obtained before and after calibration with *Lycopodium* counts. Second, we applied the 5 cross-validated trained models to the FOVs from the 12 scanned areas of the slides that were not annotated or used before, totaling 768 FOVs per slide. We used an IoU threshold of 0.7 to remove overlapping bounding boxes generated by applying successively the 5 cross-validated models on the same FOVs. The automated counts were calibrated with their respective *Lycopodium* counts, and finally compared to

those obtained manually from 9 common gravimetric traps but from different slides (Fig. S1 step 5). Calibrated counts corresponded to the number of pollen grains for 100 *Lycopodium* spores.

2.3 Results

Detection performances tended to plateau at ∼60% of the full dataset, i.e. ∼1,770 training FOVs (Supporting Information Fig. S1, Notes S2), suggesting that the detection performances presented below and obtained with 2452 ± 6 training FOVs are not constrained by the size of the dataset.

2.3.1 Detection of pollen grains regardless of their taxon

We tested whether the models could accurately and precisely detect pollen in the simplest configuration, i.e. with no distinction of their taxon (all-0taxon, Supporting Information Methods S4). Based on 2,506 ± 26 (mean ± 2SE) manually annotated bounding boxes, the average performance achieved was good (Supporting Information Table S1). The percentage of grains detected (recall) was 94.8 ± 0.33%, the percentage of correct predictions (precision) was 94.7 ± 0.38%, leaving 5.2 ± 0.33% of tagged *Lycopodium* or pollen grains left undetected (FN), and falsely detecting 5.3 ± 0.38% of debris. *Lycopodium* and pollen grains were left undetected with the same frequency (5.0 ± 0.25% and 4.4 ± 0.42%, respectively) while grains cut on the FOVs edge were missed twice as frequently (10.2 ± 0.98%, Supporting Information Table S1). The bounding boxes correctly predicted were well positioned, as they overlapped by 94% ± 3 (IoU) with manually tagged ones.

We analyzed the variation in false negatives, using descriptive variables for the annotated grains. Image quality contributed greatly to the detection performance. Pollen and *Lycopodium* grains in images of good quality were left undetected 10 times less frequently than grains with poor visual quality (1.1 ± 0.37% vs 10.9 ± 1.09%), especially in unfocused images (15.6 ± 2.85%) or when covered by a debris (12.1 ± 2.21%; Fig. S2a). Similarly, pollen grains for which their taxon could not be determined because of the poor image quality, which represent 72% of the impossible identifications, were also left undetected about four times more often (9.4 ± 0.50%) than grains that could be identified (2.4 ± 0.20%; Fig. S2b). Missed detections were also higher for pollen and *Lycopodium* grains not fully visible within the FOVs (8.3 ± 0.91%) compared to those fully visible within the FOVs (4.1 ± 0.26%; Fig. S2c). Disentangling the effects of taxonomy from the image quality or the visible section of grains was constrained by the uneven distributions of these variables in the dataset. For example, Pinaceae could be identified during annotation, granted to its typical morphology, in any image of bad or good quality, and even for grains not fully visible within the FOVs. In contrast, a good quality image with a full view of the grain was required to identify *Quercus*, because of its morphological similarity with other taxa (only 5.4 ± 0.48% of all grains identified as *Quercus* presented non-optimal conditions compared to 67.9 ± 3.7% for Pinaceae, Supporting Information Fig. S2). To try to get the independent effect of taxonomy on detection errors, we analyzed only optimal images, i.e. focused, with grains not covered by debris or deteriorated, and with grains fully visible within the FOVs, totaling 1,183 ± 13 annotated grains. Under these conditions, we found grains of *Buxus*, *Quercus*, Poaceae and Oleaceae were rarely missed (FN below 0.2 ± 0.37%) compared to *Lycopodium* (1.3 ± 0.54%) and

Pinaceae (3.0 ± 3.56%; Fig. S2d). These differences in detection errors among taxa were not related to the abundance of a given taxon. Similar detection was achieved for *Buxus*, *Quercus*, Poaceae and Oleaceae despite representing on average 5.3 ± 0.6 % to 27.7 ± 1.4% of the grains in those optimal images (Fig. S2d). These pollen grains share a common morphology, a circular-shaped monad, which thus is frequent in the dataset. In contrast, Pinaceae, a saccate monad, and *Lycopodium*, a triangular-shaped spore, have specific morphologies, which are thus less abundant in the dataset, accounting respectively for 3.5 ± 0.5% and 10.6± 0.6% grains of optimal images (Fig. S2d). Their shapes are also more irregular, making them more likely to be mistaken for debris. From these results, we conclude that the abundance of a given morphology, not of a given taxon, and its resemblance with debris explained the performance of detection. Using grains in optimal conditions only also reduced the percentage of grains left undetected by a factor of 10.

2.3.2 Effects of the annotation strategy

2.3.2.1 Application to new pollen samples

Detection was good when models were applied to the sampling years 2019 and 2021, not included in model training. F1-scores remained above 93.5 ± 0.45%, corresponding to 1% decrease compared to that of the reference models (Fig. S3, Supporting Information Table S2, ROC curves in Fig. S3), and the average percentage of false negatives (6.5 ± 0.45%) and false positives (6.5 ± 0.13%) increased at most by 1.2% compared to the reference models (Fig. S3b). Surprisingly, detection performance evaluated on the 2020 sampling year achieved a higher F1-score of 98.1 ± 0.12%. Detection errors decreased by at least 3.4% for both false negatives (1.81 ± 0.10%) and false positives (1.92 ± 0.16%). This pattern could result from the greater quality of the pollen images for that year. There were 1,865 annotated bounding boxes in the test set in 2020 compared to 2,787 in 2019 and 3,050 in 2021, and 58.6% of the grains were in images of good quality in 2020, compared to 46.3% in 2019 and 55.1% in 2021 (Supporting Information Fig. S4).

2.3.2.2 Increasing the labels details

Including taxonomic details when manually annotating the dataset, with labels for the five most frequent Mediterranean pollen taxa in the dataset, had no effect on the percentage of false positives, while the percentage of false negatives increased on average by 1.5% (Table 1, Fig. S4, ROC curves in Supporting Information Fig. S5a). This slight increase was mostly due to grains cut on the FOVs edge being more frequently missed (67 ± 10 compared to 34 ± 4 not detected, out of 330 ± 19 grains; Fig. S4b).

2.3.2.3 Selective annotation of a single model taxon

As expected, tagging and labeling a single model taxon decreased annotation time compared to tagging and labeling all pollen (1,385 to 2,778 bounding boxes depending on the taxon instead of 12,531). This strategy increased the percentage of false positives, and to different extents depending on taxon (Table 1, Supporting Information Fig. S5 and Fig. S6). This increase was mainly caused by the wrong detection of grains cut on the FOVs edge and not determined, and grains from other taxa than the targeted one, but not of debris falsely detected as pollen (Fig. S4a). For example, when detecting *Buxus*,

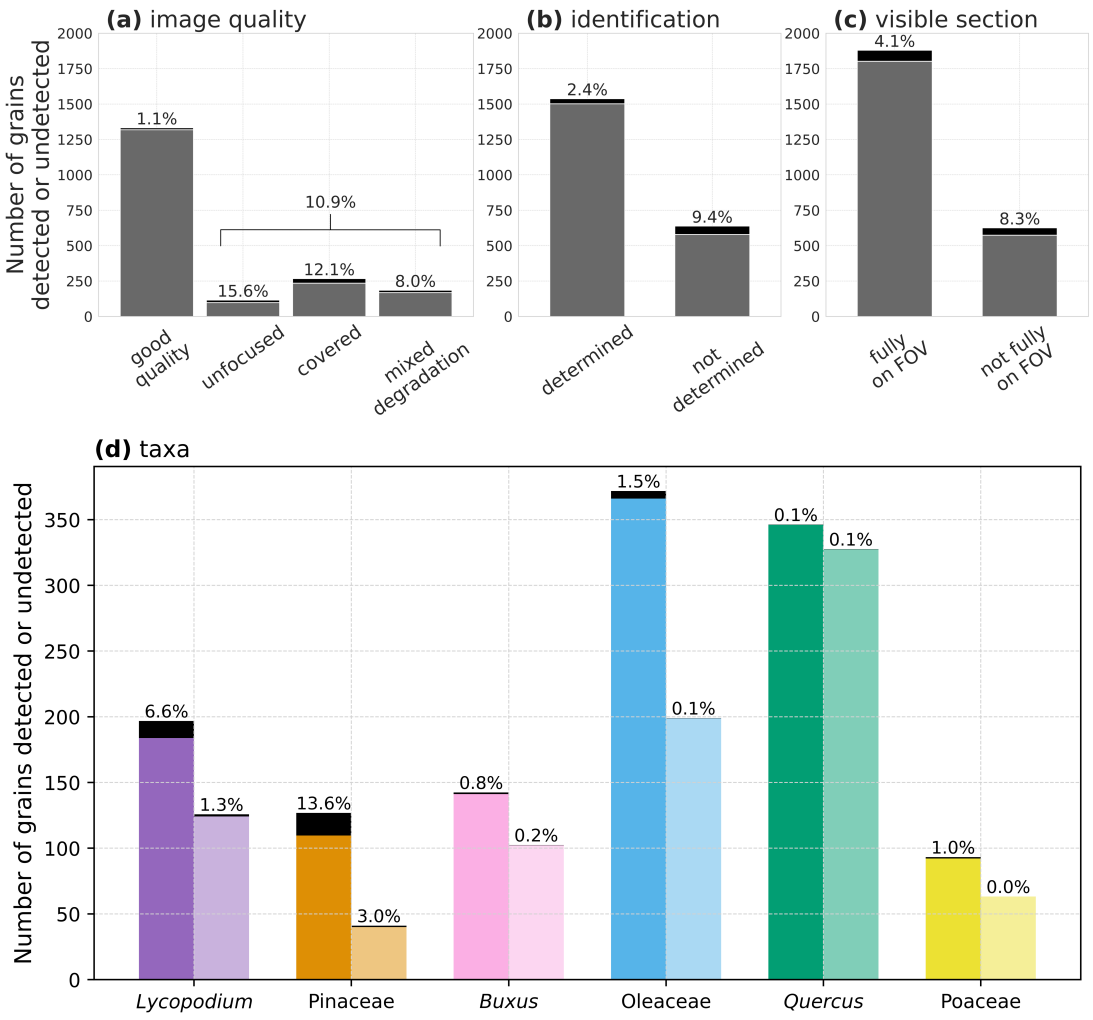


Figure S2: Analysis of detection errors. Number of pollen grains correctly detected (true positives TP, gray and colored bars) or left undetected (false negatives FN, black bars and numbers above bars) according to: (a) image quality, (b) identification, discarding the grains not fully visible within the fields of views (FOVs, a and b), (c) visible section, (d) taxon for 5 taxa and accounting for all grains (darker bars, left) and for grains in optimal conditions (lighter bars, right), i.e. with good visual quality and fully visible within the FOVs. Values represent means over the 5 cross-validation tests, and bars height correspond to the number of annotated bounding boxes.

among the 88 ± 9 falsely detected bounding boxes, only 12 ± 4 were debris while 26 ± 4 were Oleaceae pollen, which has a similar reticulated exine, and 23 ± 4 were pollen cut on the FOVs edge. The amount of undetected grains (FN) for Pinaceae and *Lycopodium* grains was not affected by the annotation strategy (Fig. S4b, Supporting Information Table S3). In contrast, pollen with common morphologies, i.e. *Quercus*, Poaceae, *Buxus* and Oleaceae, were left undetected more frequently when tagged solely with *Lycopodium* compared to when tagged along with the other pollen grains (Fig. S4b, Supporting Information Table S3).

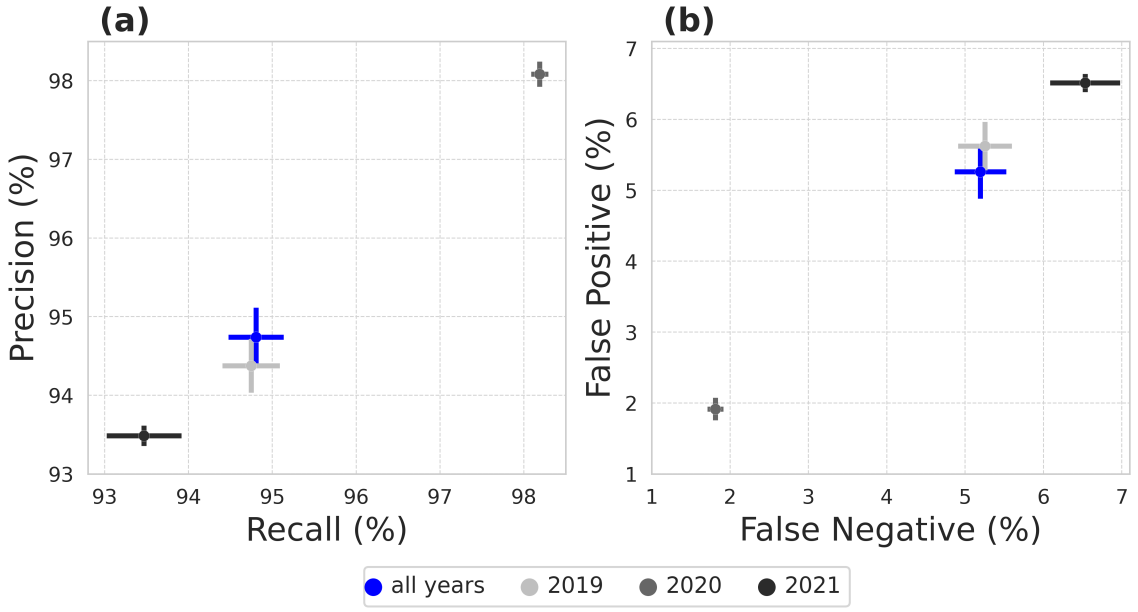


Figure S3: Performance of the detection method measured with (a) recall and precision, (b) the percentages of false positives and false negatives, when trained on all (blue), or all but one sampling years : 2019 (lighter gray), 2020 (intermediate gray), or 2021 (darker gray). Points and bars represent means and 95% confidence intervals over the 5 cross-validation tests.

Table 1 Performance metrics of the detection for all annotation strategies (mean ± 2SE for the five cross-validated test datasets).

Annotation strategy	No. of annotated bounding boxes	Percentage of false negative	Percentage of false positive	Precision	Recall	F1-score
all-0taxon	2506 ± 26	5.20 ± 0.33	5.26 ± 0.38	94.74 ± 0.38	94.80 ± 0.33	94.77 ± 0.08
all-5taxa	2506 ± 26	6.70 ± 0.40	5.16 ± 0.25	94.84 ± 0.25	93.31 ± 0.39	94.07 ± 0.11
1taxon-Pinaceae	300 ± 9	10.32 ± 1.09	15.64 ± 0.85	84.36 ± 0.85	89.68 ± 1.09	86.93 ± 0.43
1taxon-*Buxus*	326 ± 12	9.41 ± 0.68	22.96 ± 2.12	77.04 ± 2.12	90.59 ± 0.68	83.24 ± 1.15
1taxon-Poaceae	277 ± 10	10.38 ± 0.58	24.32 ± 2.05	75.68 ± 2.05	89.62 ± 0.58	82.04 ± 1.10
1taxon-Oleaceae	556 ± 19	9.38 ± 0.59	15.86 ± 0.78	84.14 ± 0.78	90.62 ± 0.59	87.26 ± 0.45
1taxon-*Quercus*	530 ± 13	12.63 ± 1.74	30.26 ± 2.15	69.74 ± 2.15	87.37 ± 1.73	77.54 ± 1.72

2.3.3 Combination of detection and classification errors in a full automated pollen analysis

Here we analyzed errors combining both detection and classification to evaluate how the method predicts pollen counts, assessing the drivers of potential biases.

2.3.3.1 Compensation between undetected grains and falsely detected debris

In this study, we chose a confidence score threshold to balance on average the percentage of undetected grains and of falsely detected debris (Supporting Information Methods S3). We found that false negatives were unevenly distributed among taxa; *Lycopodium* and Pinaceae grains were left undetected more frequently than others because of their specific morphologies. Using the classification information for falsely detected debris, we found that their distribution was also uneven among taxa, and followed the same above pattern for detection: Pinaceae and *Lycopodium* grains were also associated with a higher percentage of falsely detected debris (4.2% and 7.6%) compared to the other four taxa defined by common and regular morphologies (below 2.3%, black in Fig. S5b, Supporting Information Table S4a). Therefore, errors during detection from both undetected grains

and detected debris partially compensated each other.

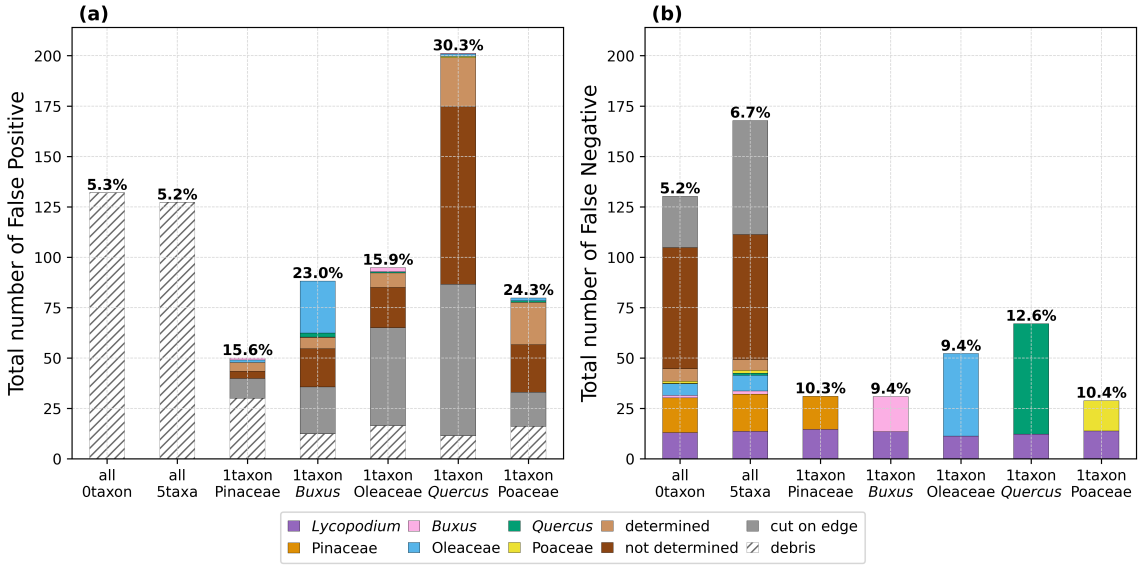


Figure S4: Effect of annotation strategy on detection errors. Number (bars height) and percentage (values above bars) of (a) false positives, i.e. detected but not annotated debris and pollen grains, and (b) false negatives, i.e. annotated but not detected pollen grains, obtained with each annotation strategy, and averaged over the 5 cross-validation tests.

2.3.3.2 Accuracy of combined detection and classification

Once detected, the 5 target taxa were almost never misclassified within the 5 categories, with the exception of ∼2% of grains tagged as Oleaceae but misclassified to *Buxus*, and 0.7% of grains tagged as *Buxus* but misclassified to Oleaceae, which both have a reticulated exine (light blue, Fig. S5, Supporting Information Table S4b). Most classification errors for the target taxa, with the exception of *Lycopodium* and Pinaceae with specific morphologies, were made to the class of pollen from other taxa simply labeled as pollen (dark blue, Fig. S5), and pollen cut on the FOVs edge (gray, Fig. S5). The large amount of grains from the pollen category and misclassified to target taxa was partly compensated by the amount of grains from the target taxa misclassified to the pollen category, thus limiting the over-estimation of the predicted counts (dark blue, Fig. S5). A post-hoc visual examination of the images showed that some misclassifications to other pollen taxa corresponded to pollen with similar morphologies, e.g. few Viburnum and Brassicaceae pollen grains were misclassified as Oleaceae, all of which have a reticulated exine, while most misclassifications were to pollen that could not be determined during annotation. We cannot exclude that some of these latter misclassifications could actually correspond to the correct pollen taxon, i.e. a true annotation error or a lack of confidence to label a grain. Similarly, grains labeled as cut on the FOVs edge and misclassified to a pollen taxon could correspond to a correct identification. Post-hoc visual inspections confirmed this hypothesis for Pinaceae. Once detected, the classification of *Lycopodium* was very efficient (Fig. S5). Less than 0.9% were misclassified, and all grains cut on the FOVs edge and classified to *Lycopodium* were actually correct. This last result confirms that automated detection of pollen using *Lycopodium* for calibration should provide accurate counts.

Finally, predicted counts for all labeled categories were systematically overestimated compared to manual annotations, except for the category of pollen cut on the FOVs edge,

which were directly classified to pollen taxa (Fig. S6a). As *Lycopodium* counts were also over-estimated, the inferred pollen counts after calibration, matched very well the calibrated pollen counts from the manual annotations (Fig. S6b).

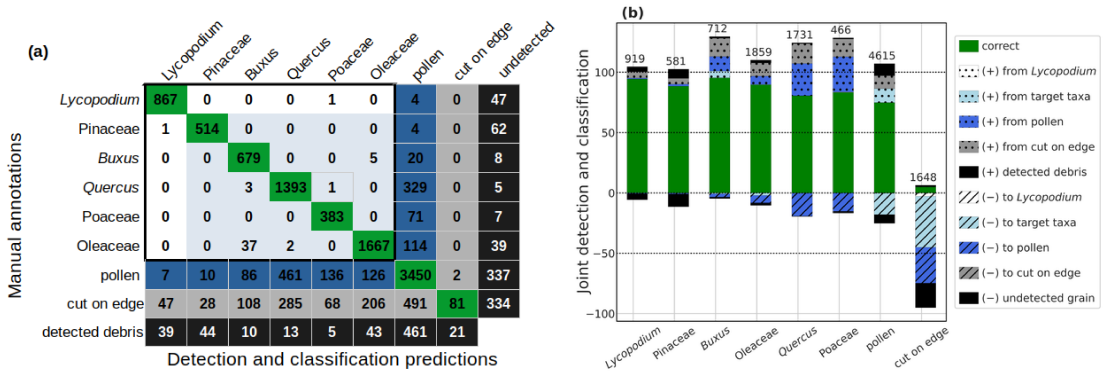


Figure S5: Performance of the joint detection and classification. Grains correctly or incorrectly detected and/or classified (a) as the confusion matrix, (b) plotted for each label separately. Results are summed over the 5 cross-validation tests from the all-5taxa strategy, and show correct detection and classification (green), detection errors (black), and classification errors from confusion with : *Lycopodium* (white), the 5 target taxa (light blue), the category pollen (dark blue), or grains cut on the edge of the fields of views (FOVs, gray). In (b), gridded bars represent errors under-estimating the class (grains undetected or classified into another class), and dotted bars represent errors over-estimating the class (detected debris or grains not belonging but classified into the class); values above bars are the number of bounding boxes annotated for each label.

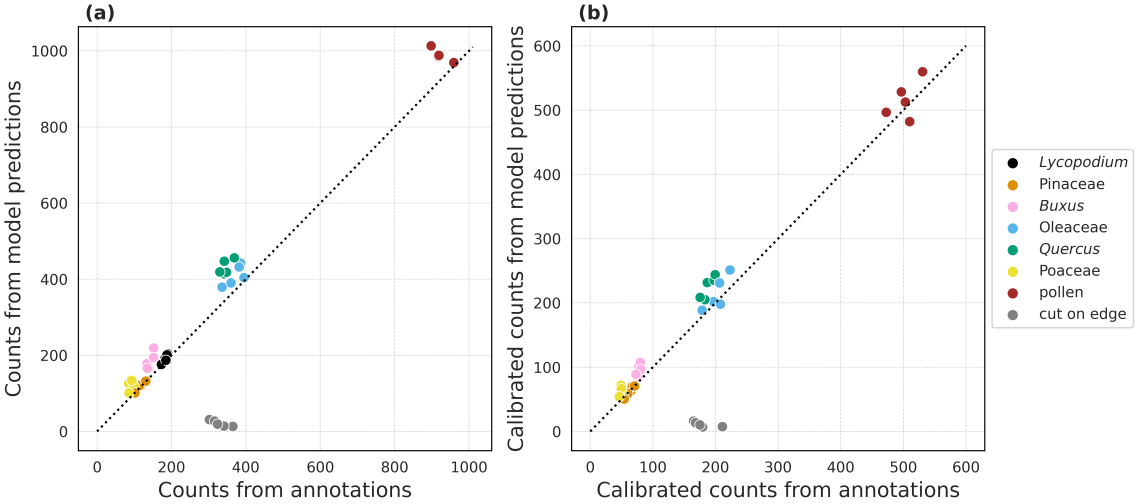


Figure S6: Comparison of automated predictions and manual annotations. (a) Counts for the 5 taxa and *Lycopodium*, and (b) Counts calibrated with *Lycopodium*; both for the manual annotations and from the model predictions in each of the 5 cross-validation test (some points overlap), and obtained with the annotation strategy all-5taxa; the dotted line is the 1:1 line.

2.3.3.3 Application of the automated method to routine conditions and comparison with a palynologist

Once the method was established and the images were acquired, predictions for each slide produced up to 5,700 detected and classified pollen grains in one step and only a few minutes, compared to ∼2-3 hours for the standard microscopy analysis by the palynologist expert. Calibrated counts for the 5 most common taxa predicted by the models and obtained by the palynologist, for the same gravimetric traps but not the same slides, were close to each other, except for *Buxus*, a rare taxa in these samples for which sampling variance is expected to be large (Fig. S7). Predicted counts for *Quercus* were systematically underestimated compared to manual counts, potentially because we labeled *Quercus* simply as pollen when we could not confidently identify it, and which was replicated by the model, while the expert could identify *Quercus* in many more conditions. Predicted counts for Pinaceae tended to be overestimated compared to manual counts, potentially because Pinaceae grains were manually annotated as $\frac{1}{2}$ when torn but counted as a full pollen grain by the model.

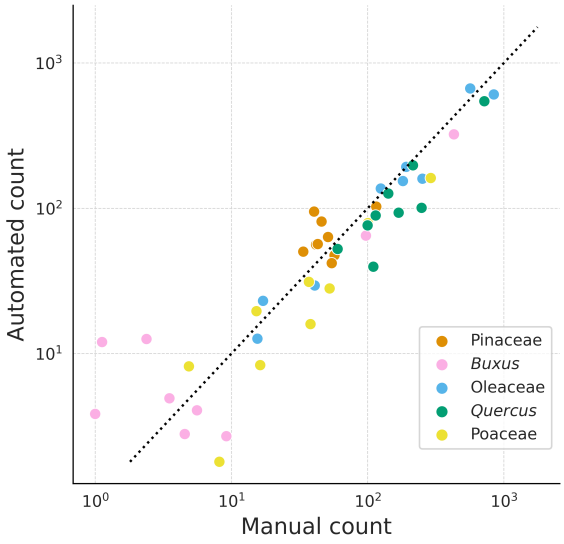


Figure S7: Comparison of automated predictions and count from palynologist. Manual (expert palynologist) vs automated pollen counts from 5 key taxa; the models used for automated counts were trained with the strategy all-5taxa and applied to images not used before (704 fields of views for each of 9 slides). Results are log transformed, the dotted line is the 1:1 line.

2.4 Discussion

2.4.1 Environmental pollen samples are no longer a barrier to automated detection

Despite challenges inherent to environmental samples, with images not manipulated before the analyses, and thus containing many debris and pollen taxa, the performance of pollen detection here is consistent with two similar and recent studies. From European lake sediment samples, 87.2% to 99.1% of pollen was retrieved for a Faster R-CNN detector on ∼1,450 annotations (Theuerkauf et al., 2023), and from pollen trap samples collected in a tropical forest, 83.7% of pollen was retrieved for a modified ResNet34

architecture on 14,764 annotations (Punyasena et al., 2022), while our method retrieved 95.6% of pollen and 95.0% of *Lycopodium* using as many as 12,531 annotations. Only 5.3% of the total detected objects were debris misidentified as pollen (Supporting Information Table S1), and the predicted bounding boxes were very well positioned (IoU of 0.94), possibly making subsequent classification efficient. Missed detections were mainly caused by grains covered by debris, unfocused grains, or grains not fully visible within the FOVs (Fig. S2). These cases are likely to occur regardless of pollen taxon, and thus not likely to bias the pollen counts in the detected assemblage. These grains with poor visual quality also lack discernible identification criteria, making them irrelevant for classification, be it by an algorithm or an expert palynologist. Excluding these grains with visual poor quality increased pollen detection to 99.4% recall, a performance consistent with that obtained on reference pollen images taken under optimal conditions, i.e. with one or very few taxa from flower anthers with no debris (Gallardo-Caballero et al., 2019; Khanzhina et al., 2022; Kubera et al., 2022).

2.4.2 How to efficiently increase detection performance

Debris generated false negatives by masking pollen grains, and also false positives, since debris could be misidentified to pollen grains. Thus any strategy that can eliminate debris will undoubtedly increase detection performance, e.g. through chemical treatment or sieving, or by diluting the material in the slides. The imaging process, fully automated here, also contributed to detection errors by generating images of unfocused pollen grains, which were left undetected more frequently than pollen in focused images (Fig. S2a). Therefore, improving the focus of pollen grains shall decrease the percentage of undetected pollen, e.g. by increasing the number of focal planes during image acquisition, or reducing the thickness of the slide preparation.

Our results also show that increasing the taxonomic details, and thus number of labels, required tedious identification effort but did not improve the performance of pollen detection. Slightly more grains were left undetected for similar numbers of debris falsely detected (Fig. S4). If the goal is to detect pollen regardless of taxon, we thus recommend to not spend time identifying and labeling taxa, but instead use a few general categories as done here (*Lycopodium*, pollen and cut on edge). Apart from the high performance this strategy can achieve, it has the advantage that it can be done by non-palynologists. We also found that the annotation, and thus detection, of a single model taxon was effective only for morphologically distinct taxa such as Pinaceae or *Lycopodium* spores (Fig. S4). This strategy can be used, as it saves time, only if the model taxon has a distinct shape and can be identified with a good confidence in all images. Lastly, although often overlooked, the trade-off between false positives and false negatives can be adjusted to the study objectives, by modifying the confidence score threshold to filter the predicted bounding boxes, for instance by increasing it to reduce the detection of debris (Supporting Information Methods S3).

Using a model pre-trained on pollen images, even if outside the data to be analyzed, can help limit the number of pollen images to annotate and increase pollen detection. Therefore, we make our best trained model here on the anemophilous flora of the Mediterranean region available to all, calling also for a large share among research teams of models and data to increase performance of all future automated pollen analyses.

2.4.3 Robustness of the automated detection method to extrapolation

Our models achieved the same detection performance whether they were applied to samples used or not to train the model (Fig. S3). This result has important consequences for many pollen-based studies, based on long-term monitoring (van der Knaap et al., 2010), or long fossil sequences (Donders et al., 2021) or at large spatial scales. For a given study, good detection performance can be achieved by training the algorithm only once on annotated images from a few samples, and without the need to reiterate the training process on new sampling years and/or new locations. Of course, the inherent variation in data between studies, such as the chemical treatment of samples, microscope settings or the sample content itself, will affect the detection performances. It should however be noted that using models pre-trained on images from any pollen study can help increase detection performances, and limit the annotation workload required to build a new training dataset.

2.4.4 Comprehensive evaluation of the joint detection and classification to achieve full automated pollen counts

The proposed method detected and classified *Lycopodium* spores and 5 pollen taxa, which is a small fraction of pollen diversity present in the samples. Our goal was to provide guidelines for other pollen studies while focusing on detection errors and on how they propagated to classification, and not to make a full analysis of the samples. The chosen taxa accounted for 43% of all pollen grains in the samples, and also corresponded to a diversity of morphologies, which thus allowed to address our goals. Detection errors were affected by the abundance of the grain morphologies and resemblance with debris, rather than by the taxon abundance. We found fewer false negatives for taxa that share a standard and thus abundant pollen morph (Oleaceae, Poaceae, *Buxus* and *Quercus*) compared to taxa that have rare and irregular morphologies (Pinaceae and *Lycopodium* spores, Fig. S2d), which were thus more underestimated in the detected pollen assemblage. Nonetheless, when detection was performed jointly with classification, grains with rarer and irregular morphologies were also assigned a higher number of falsely detected debris, thus partially compensating errors. When classification is conducted separately from detection, debris falsely detected are often processed with dedicated classes or processes (Crouzy et al., 2022; Zhao et al., 2022), and pollen grains left undetected are usually not considered in the final assessment, which may generate biases. The proposed method, based on an algorithm jointly conducting detection and classification, thus integrates detection errors through both stages, which effectively mitigates both types of detection errors.

The detected grains of *Lycopodium* and of the 5 target taxa were classified with very little confusion between themselves (Fig. S5a, Supporting Information Fig. S6). Classification errors of the target taxa mainly occurred for grains of other morphologically similar taxa, and for grains that could not be identified with confidence, which we both labeled here as simply pollen (Fig. S5a). The proposed method requires to annotate, and thus identify, pollen grains directly in 2D images containing many different taxa and debris, sometimes unfocused, and in which pollen may be not well oriented or sufficiently visible to use the appropriate discriminant criteria for identification. Such conditions make identification, when annotating the images, challenging and sometimes not possible, which will generate ambiguity when training the models, and contribute to increase detection errors. The presence of grains not identified and annotated in the test dataset also prevents the accurate evaluation of errors, as some predictions are likely to be correct, e.g. a pollen grain is indeed a *Quercus* one but we did not label it as such. Nonetheless,

misclassifications from the target taxa to the category of other grains simply labeled as pollen, and from the category pollen to target taxa partly compensated each other. The distinction of grains cut on the FOVs edge also generated many misclassifications, although some were correct; we thus recommend avoiding this label, and instead, to directly label cut pollen grains with their taxa or as simply pollen. Despite these challenges and potential biases, automated counts matched those manually done, but also counts obtained by a palynologist expert with standard microscopy analysis.
If the goal is to get pollen counts per taxon for many taxa, improving the confidence of manual identifications will for sure improve classification performances. Identifications used for training could be improved by enhancing the visibility of pollen structures in the images, e.g. by decreasing pollen density, using colored images (Dunker et al., 2021), or using images of reference pollen collected in flowers anthers. This latter suggestion though may be less effective for detecting pollen in environmental samples afterwards, as the model also trains on the background of the images, clear of debris in reference samples. Conducting a separate classification, after the detection of pollen of any taxon from their original sample images, would bypass the challenging and error-prone step of manual pollen identification and annotation on images taken from environmental samples. In that case we recommend that classification will be performed on images of pollen from known species and plant individuals. This approach should especially improve classification performance for taxa of common morphologies that are difficult to identify. It would also allow the identification of rare taxa that de facto have too few images in the samples to train the models (e.g. only 1 pollen grain of Julgans was found in our dataset). Conducting a separate classification may finally benefit from the extensive work carried out by many these last few years on pollen. Large image classification datasets such as POLLEN23E (Sevillano and Aznarte, 2018), POLLEN73S (Astolfi et al., 2020) or POLLEN13K (Battiato et al., 2020b) were made open source. Diverse new CNN-based image classification methods have recently been developed to improve the performances of pollen classification. One approach relies on both taxonomical and morphological labels to train the classification models (Barnes et al., 2023), others use multi-CNN architectures with a decision tree (Bourel et al., 2020), or add a pre-liminary image deblurring process prior to classification, combined to a multi-scale architecture to also include image sizes in the training (Chen and Ju, 2022).
Our study is based on purpose on simple tools and thus on images from slides scanned under light microscopy. Other type of pollen data have been tested for the automation of pollen analysis, and gave good classification performances, by-passing some limitations encountered in our study: pollen images acquired with flux cytometry (Barnes et al., 2023; Dunker et al., 2021), with scanning electric microscopy (C. Li et al., 2023), or with confocal microscopy then classified with a 3D-classification algorithm (Wang et al., 2021).

2.4.5 Concluding remarks

Our work represents a comprehensive attempt to assess joint detection and classification of pollen using artificial intelligence, and shows that pollen detection is a critical step for getting accurate pollen counts in pollen assemblage. Overall, our method, which relies on standard equipments, simple tools and rules, provide excellent performance on environmental samples from the "real-world" containing many debris and pollen taxa. The method can generate for a slide, once the images are acquired and labeled, and the models are trained, up to 5,700 detected and labeled pollen grains in only a few minutes, compared to ∼500 grains in 2-3 hours for a palynologist, with no subjectivity or fatigue, and can be confidently extrapolated to new samples not seen by the models.

Acknowledgements

We thank the doctoral school GAIA (Biodiversité, Agriculture, Alimentation, Environnement, Terre, Eau), the project PEPS-CNRS DETECT to CD, the project MUSE-UM to OP and CD, the OSU OREME SO POLLIMED and SO POLLUMINE (https://oreme.org/observation/), which all funded this study. We also thank three anonymous reviewers for their helpful comments. This work was performed using HPC resources from GENCI-IDRIS (Grant 2023-[AD011013554R1]). This work is a ISEM contribution (ISEM 2024-096). For the purpose of Open Access, a CC-BY public copyright license has been applied by the authors to the present document and will be applied to all subsequent versions up to the Author Accepted Manuscript arising from this submission.

Author contributions

B.G., C.D., O.P. and S.J. conceived the ideas, designed the methodology, analysed the data and led the writing of the manuscript. J.P. provided ideas on the methodology of YOLO. L.Beaufort, T.dG-T, Y.G. helped with the image acquisition process. L.Bouby, J-F.T., S.C., S.I., B.L. and B.G, collected pollen in the field, and conducted wet lab work, N.C-N. counted pollen manually on the microscope and helped with image analysis. All authors helped to the writing of the manuscript.

Competing interests

None declared. All authors gave final approval for publication.

Data availability

The data that support this study (images, annotation metadata and the weights of the YOLOv5 trained models) are openly available in Zenodo (https://zenodo.org/) at http://doi.org/[10.5281/zenodo.11126431].

New Phytologist

***New Phytologist* Supporting Information**

Article title: A user-friendly method to get automated pollen analysis from environmental samples

Authors: B. Gimenez, S. Joannin, J. Pasquet, L. Beaufort, Y. Gally, T. de Garidel-Thoron, N. Combourieu-Nebout, L. Bouby, S. Canal, S. Ivorra, B. Limier, J.F. Terral, C. Devaux*, O. Peyron*
*These authors contributed equally to this work.

Article acceptance date: 05 May 2024

The following Supporting Information is available for this article:

Fig. S1 Performance of the detection method for each sizes of the dataset from the last epoch model, and from the last vs the best epoch model

Fig. S2 Distribution of grains in optimal conditions per taxon and category

Fig. S3 Performance of the detection of models trained on all or all but one sampling year

Fig. S4 Performance of the detection of models trained on all or all but one sampling year, and considering the images properties

Fig. S5 Performance of the detection obtained for all annotation strategies

Fig. S6 Confusion matrix obtained when annotating only one model taxon and Lycopodium

Table S1 Detection metrics obtained using the annotation strategy all-0taxon

Table S2 Performance of the detection of models trained on all or all but one sampling year

Table S3 Proportion of grains left undetected measured separately for each annotated labels, and for each annotation strategy

New Phytologist

Table S4 Confusion between categories and compensation of false negatives (FN) and false positives (FP) confusions

Methods S1 Description of YOLO algorithms

Methods S2 Illustration of the process that eliminates predicted bounding boxes that overlap

Methods S3 ROC curve of the detection, showing the trade-off between precision and recall

Methods S4 Confusion matrix for joint detection and classification and for detection only

Methods S5 Description of the annotation strategies and the associated datasets

Methods S6 Definition of the detection metrics

Methods S7 Variables used to characterize manually annotated grains

Methods S8 Illustration of the categorical variables used to evaluate the causes of undetected grains

Notes S1 Analysis of the performance of the detection from the last vs the best epoch models

Notes S2 Analysis of the performance of the detection for each size of the dataset

New Phytologist

Fig. S1 Performance of the detection method for each sizes of the dataset from the last epoch model, and from the last vs the best epoch models

Detection performance as a function of the size of the training dataset measured as (a) the average ROC curves over 5 cross-validation tests also (b) zoomed for visibility; red crosses represent values for the confidence score threshold of 0.45, and gray colors indicate dataset sizes; (c) F1-scores obtained with the best (brown triangles) or the the last (black circles) epoch models. (d) Recall (blue crosses) and precision (purple squares). Points and bars represent means and 95% confidence intervals over the 5 cross-validation tests, for a confidence score threshold of 0.45.

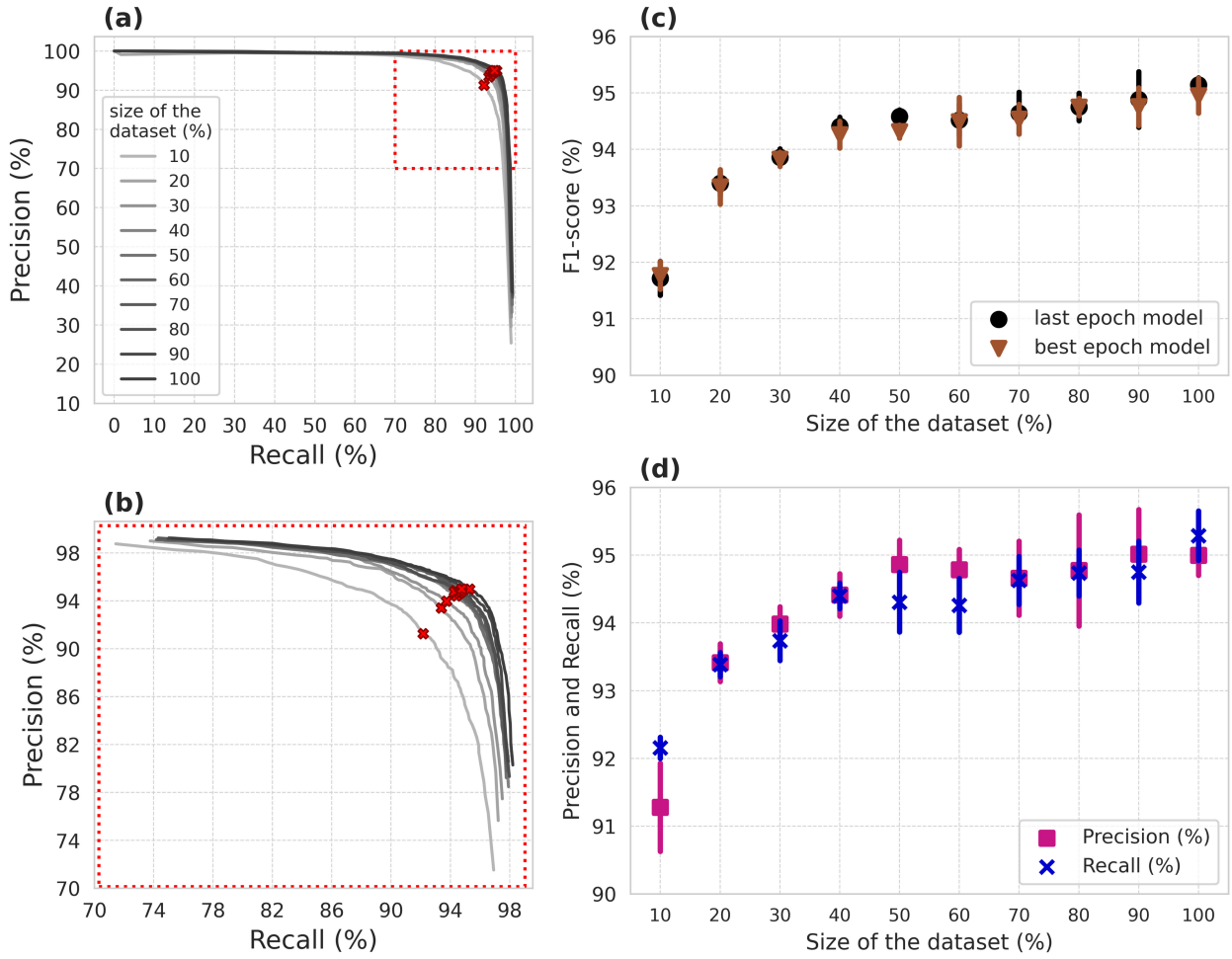


Fig. S2 Distribution of grains in optimal conditions per taxon and category

Percentage of grains in optimal conditions, *i.e.* with good visual quality and not cut on the FOVs edge, among the 12,531 annotated grains for each corresponding category, including the five target taxa, identified but non-target taxa, and not determined grains.

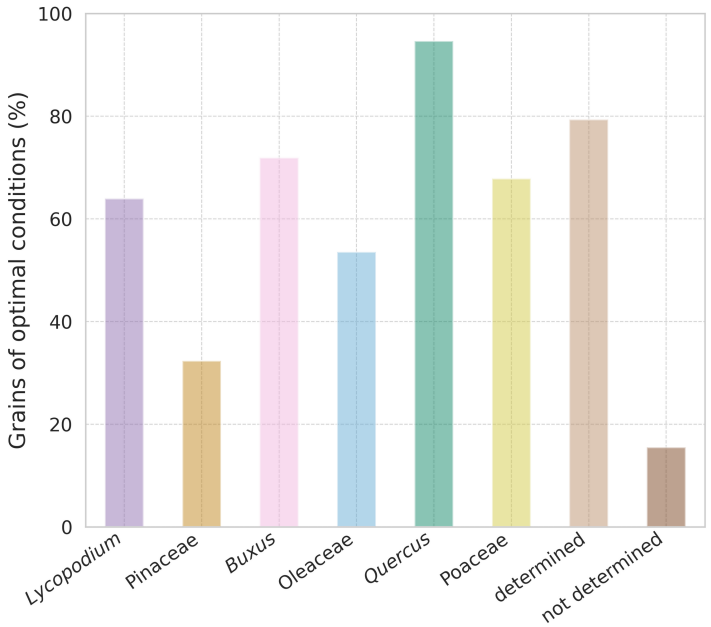


Fig. S3: Performance of the detection of models trained on all or all but one sampling year

(a) ROC curves, (b) also zoomed for visibility, when trained on all (blue) or all but one sampling years: 2019 (lighter gray), 2020 (intermediate gray), or 2021 (darker gray). Curves are averaged over the 5 cross-validation tests, red crosses represent values for the confidence score threshold of 0.45.

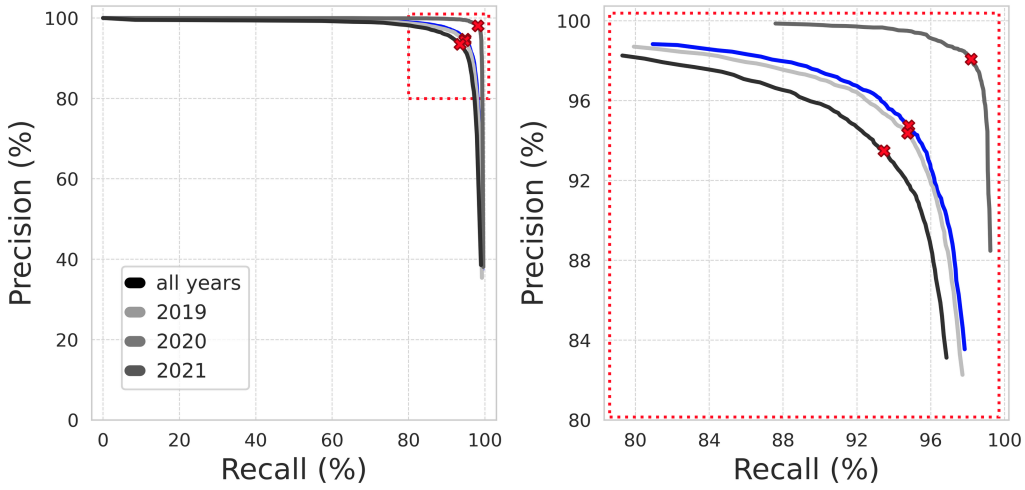


Fig. S4 Performance of the detection of models trained on all or all but one sampling year, and considering the images properties

(a) F1-scores, when trained on all (blue) or all but one sampling years: 2019 (lighter gray), 2020 (intermediate gray), or 2021 (darker gray), as a function of (a) the percentage of grains in optimal conditions (not cut and with good visual quality), and (b) the total number of grains annotated in the test dataset of 820 FOVs, as a proxy for grains density. Points and bars represent means and 95% confidence intervals over the 5 cross-validation tests; note that test sets are the same when extrapolating the models to new sample years explaining the absence of error bars.

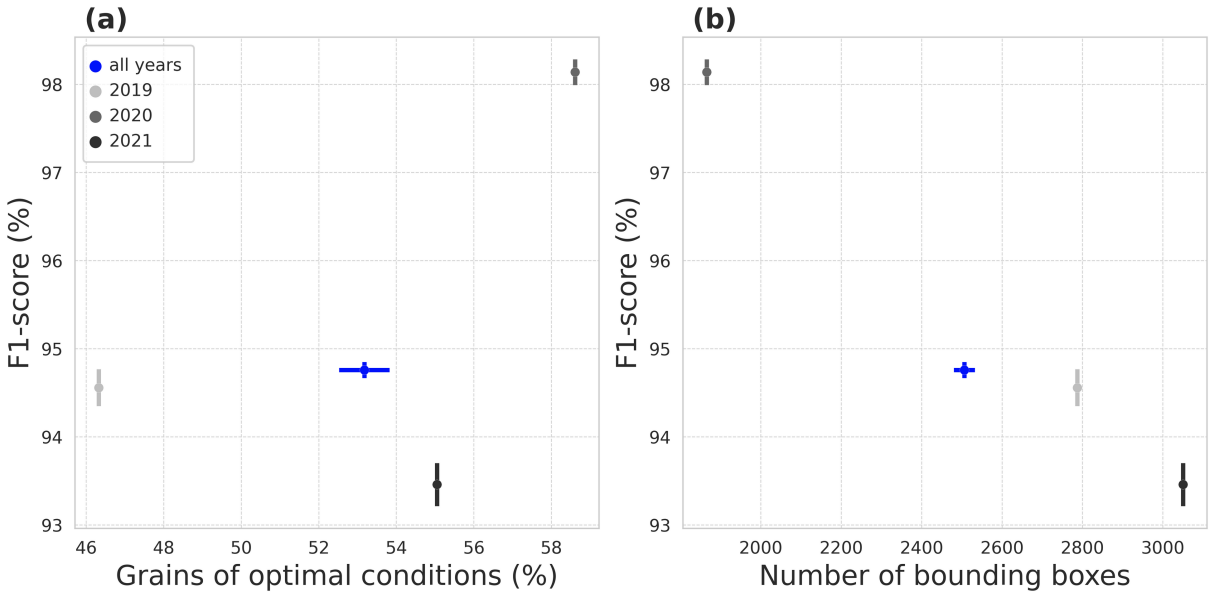


Fig. S5 Performance of the detection obtained for all annotation strategies

Performance of the detection method measured over all annotated grains for each annotation strategy as (a) the ROC curves, with crosses representing values for the confidence score threshold 0.45, (b) the percentages of false positives and false negatives. Circle sizes are proportional to the number of annotated bounding boxes, curves are averaged over the 5 cross-validation tests, and bars represent means and 95% confidence intervals over the 5 cross-validation tests.

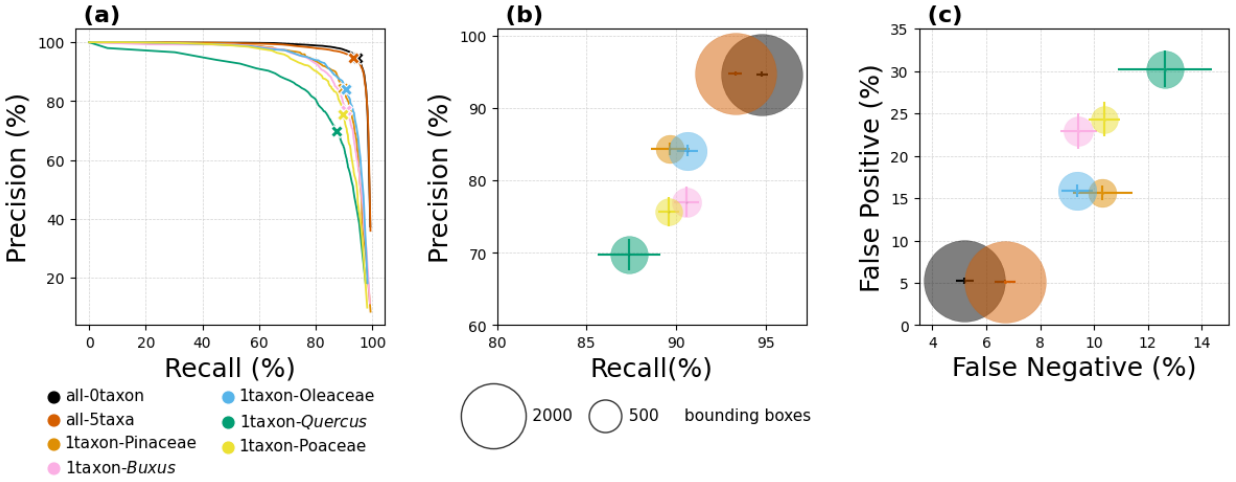
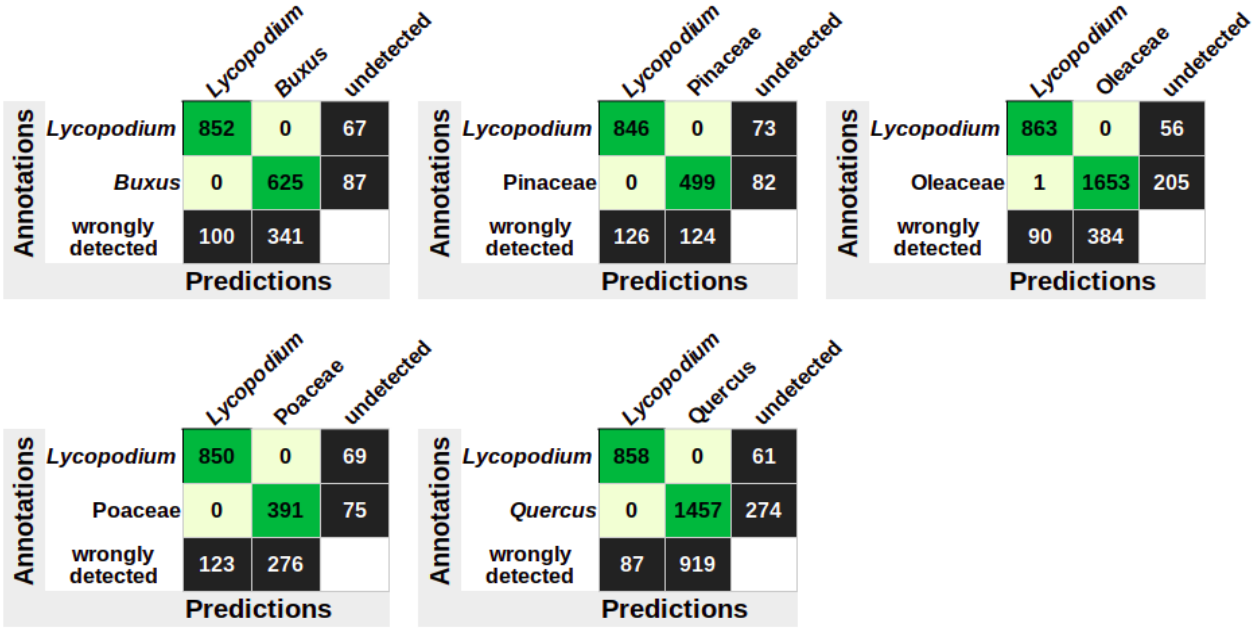


Fig. S6 Confusion matrix obtained when annotating only one model taxon and Lycopodium



New Phytologist

Table S1 Detection metrics obtained using the annotation strategy all-0taxon

Detection metrics (mean ± 2SE over the 5 cross-validated test datasets) obtained using the all-0taxon annotation, measured over all annotated grains, or over each label.

Strategy all-0taxon	# of annotated bounding boxes	False negative (%)	Recall (%)	False positive (%)	Precision (%)	F1-score (%)
All grains	2506 ± 26	5.2 ± 0.33	94.8 ± 0.33	5.3 ± 0.38	94.7 ± 0.38	94.8 ± 0.08
Lycopodium	184 ± 5	5.0 ± 0.25	95.0 ± 0.25	*False positive are all debris falsely detected here*		
pollen	1991 ± 17	4.4 ± 0.42	95.6 ± 0.42			
cut on edge	330 ± 19	10.2 ± 0.98	89.8 ± 0.98			

Table S2 Performance of the detection of models trained on all or all but one sampling year

Performance metrics (means and 2 SE over cross-validation tests) of the detection method when trained on all years or all but one sampling years (2019, 2020, or 2021).

extrapolation	# of annotated bounding box	Percentage of false negative	Percentage of false positive	Precision	Recall	F1-score
all years	2506 ± 26	5.20 ± 0.33	5.26 ± 0.38	94.74 ± 0.38	94.80 ± 0.33	94.77 ± 0.08
2021	3050 ± 0	6.53 ± 0.45	6.51 ± 0.13	93.49 ± 0.13	93.47 ± 0.45	93.48 ± 0.24
2020	1865 ± 0	1.81 ± 0.10	1.92 ± 0.16	98.08 ± 0.16	98.19 ± 0.10	98.14 ± 0.12
2019	2787 ± 0	5.25 ± 0.34	5.62 ± 0.34	94.38 ± 0.34	94.75 ± 0.34	94.56 ± 0.21

Table S3 Proportion of grains left undetected measured separately for each annotated label, and for each annotation strategy

Proportions of grains left undetected (False negative) and of grains correctly detected among all true grains (recall, mean ± 2SE for the 5 cross-validated test datasets), for each label and annotation strategy.

Annotation strategy	Label	# of annotated bounding box	% false negative	Recall
all-0taxon	*Lycopodium*	184 ± 5	5.01 ± 0.25	94.99 ± 0.25
all-0taxon	pollen	1993 ± 17	4.38 ± 0.42	95.62 ± 0.42
all-0taxon	cut on edge	330 ± 19	10.23 ± 0.98	89.77 ± 0.98
all-5taxa	*Lycopodium*	184 ± 5	5.13 ± 0.37	94.87 ± 0.37
all-5taxa	*Buxus*	142 ± 7	1.12 ± 0.47	98.88 ± 0.47
all-5taxa	*Quercus*	346 ± 11	0.28 ± 0.27	99.72 ± 0.27
all-5taxa	Oleaceae	372 ± 19	2.12 ± 0.67	97.88 ± 0.67
all-5taxa	Pinaceae	116 ± 8	10.62 ± 1.69	89.38 ± 1.69
all-5taxa	Poaceae	93 ± 6	1.42 ± 1.33	98.58 ± 1.33
all-5taxa	pollen	923 ± 18	7.29 ± 0.74	92.71 ± 0.74
all-5taxa	cut on edge	330 ± 19	20.14 ± 1.80	79.86 ± 1.80
1taxon-*Buxus*	*Lycopodium*	184 ± 5	7.30 ± 0.92	92.70 ± 0.92
1taxon-*Buxus*	*Buxus*	142 ± 7	12.10 ± 2.0	87.90 ± 2.00

1taxon-Oleaceae	*Lycopodium*	184 ± 5	6.11 ± 1.27	93.89 ± 1.27
1taxon-Oleaceae	Oleaceae	372 ± 19	11.03 ± 0.92	88.97 ± 0.92
1taxon-Pinaceae	*Lycopodium*	184 ± 5	7.99 ± 1.50	92.01 ± 1.50
1taxon-Pinaceae	Pinaceae	116 ± 8	14.02 ± 1.39	85.98 ± 1.40
1taxon-Poaceae	*Lycopodium*	184 ± 5	7.51 ± 0.49	92.49 ± 0.49
1taxon-Poaceae	Poaceae	93 ± 6	15.95 ± 2.18	84.05 ± 2.18
1taxon-*Quercus*	*Lycopodium*	184 ± 5	6.64 ± 1.01	93.36 ± 1.01
1taxon-*Quercus*	*Quercus*	346 ± 11	15.85 ± 2.43	84.15 ± 2.43

Table S4 Confusion between categories and compensation of false negatives (FN) and false positives (FP) confusions

(a) Percentage of detection errors (mean of the summed results from the 5 cross-validations), and compensations between FP and FN of the detection for each label; the labels attributed through classification to the debris falsely detected are considered here.

label	Detection FN per label (%)	Detection FP per label (%)
Buxus	1.12	1.4
Lycopodium	5.11	4.24
Oleaceae	2.1	2.31
Pinaceae	10.67	7.57
Poaceae	1.5	1.07
Quercus	0.29	0.75

(b) Percentage of classification errors and confusion between categories (mean of the summed results from the 5 cross-validations), and compensation between FP and FN of the classification for each category.

label	True	Predicted	Percentage of classification confusion (% of FP and FN) from :					
			FP (+) 5taxa	FN (-) 5taxa	FP (+) pollen	FN (-) pollen	FP (+) barely-visible	FN (-) barely-visible
Buxus	712	923	5.62	0.70	12.08	2.81	15.17	0.00
Lycopodium	919	961	0.11	0.11	0.76	0.44	5.11	0.00
Oleaceae	1859	2047	0.27	2.10	6.78	6.13	11.08	0.00
Pinaceae	581	596	0.00	0.00	1.72	0.69	4.82	0.00
Poaceae	466	599	0.21	0.00	29.18	15.24	14.59	0.00
Quercus	1731	2154	0.12	0.23	26.63	19.01	16.46	0.00

New Phytologist

Methods S1 Description of YOLO algorithms

Several versions of the object detection algorithms of the YOLO family have been published since the first release (Redmon et al., 2016), including the popular YOLOv5 (Jocher, 2020), implemented in this study. YOLO algorithms are single-stage detection algorithms, *i.e.* they process the entire image in a single step, hence the name "You Only Look Once", in contrast to related two-stage algorithms, such as R-CNN and Fast R-CNN networks (Diwan et al., 2022) that use a more complex architecture. YOLO algorithms perform with high speed and accuracy while having high generalization ability (Jiang et al., 2022). YOLOv5 repository (Jocher, 2020) is also associated with detailed implementation guidelines making it user friendly, and has thus become widely used for object-detection in multiple fields. YOLOv5 uses a system of anchors, which are predefined templates of various sizes that serve as reference points and are adjusted by the network through regressions to best fit the actual size of the objects. Images are processed by extracting features at different resolutions, and these feature maps are subsequently segmented into cells. For each cell, YOLO predicts whether an object exists within it, assigning a confidence score for the object presence, an anchor that helps define the size and position of the detected object, and probabilities for each predefined class. YOLOv5 is pretrained on the COCO dataset (Common Objects in COntext) that includes over 200K images of 80 object categories such as bikes or dogs, labeled for object detection (Lin et al., 2015).

References :

Diwan, T., Anirudh, G., & Tembhurne, J. V. (2022). Object detection using YOLO: Challenges, architectural successors, datasets and applications. *Multimedia Tools and Applications*. https://doi.org/10.1007/s11042-022-13644-y

Jiang, P., Ergu, D., Liu, F., Cai, Y., & Ma, B. (2022). A Review of Yolo Algorithm Developments. *Procedia Computer Science*, *199*, 1066–1073. https://doi.org/10.1016/j.procs.2022.01.135

Jocher, G. (2020). *Ultralytics YOLOv5* (7.0) [Computer software]. https://doi.org/10.5281/zenodo.3908559

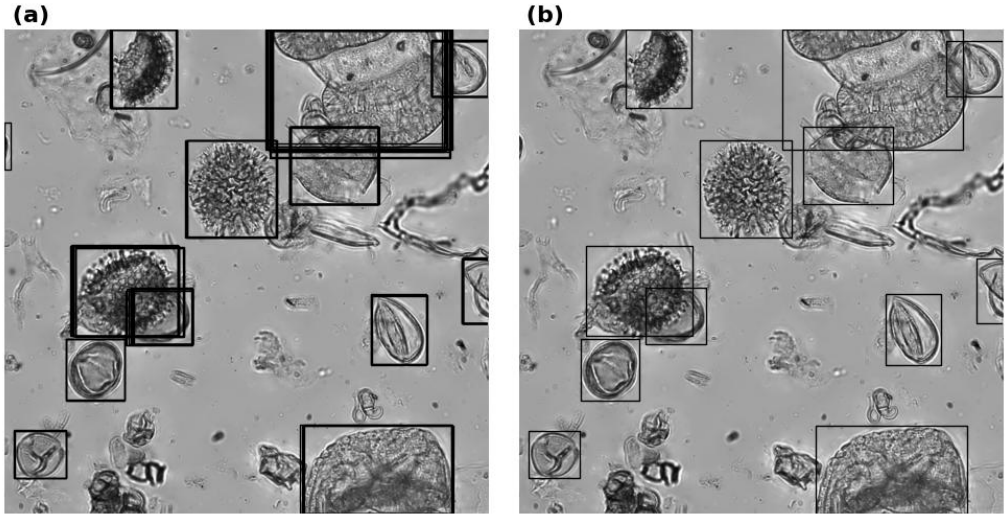
Lin, T.-Y., Maire, M., Belongie, S., Bourdev, L., Girshick, R., Hays, J., Perona, P., Ramanan, D., Zitnick, C. L., & Dollár, P. (2015). *Microsoft COCO: Common Objects in Context* (arXiv:1405.0312). arXiv. https://doi.org/10.48550/arXiv.1405.0312

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You Only Look Once: Unified, Real-Time Object Detection. *arXiv:1506.02640 [Cs]*. http://arxiv.org/abs/1506.02640

New Phytologist

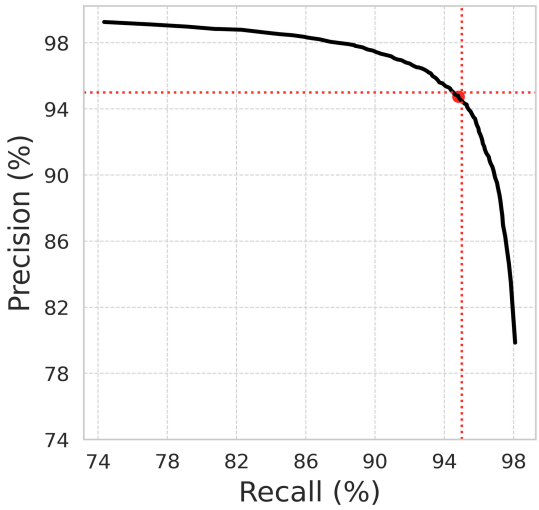
Methods S2 Illustration of the process that eliminates predicted bounding boxes that overlap

FOVs presenting (a) 78 bounding-boxes initially predicted by the YOLOv5 trained model, and (b) the 13 bounding boxes kept after eliminating duplicated bounding-boxes by applying Non-maximum Suppression, with an IoU threhsold of 0.70, and keeping only the bounding-boxes with the highest confidence score.



Methods S3 ROC curve of the detection method, showing the trade-off between precision and recall

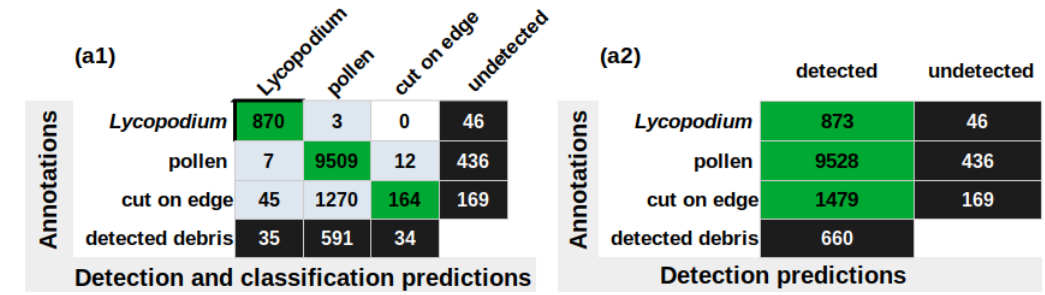
ROC curve (zoomed for visibility) showing the trade-off between recall and precision, generated by changing the threshold for the confidence score of the predicted bounding boxes, and obtained with the annotation strategy all-0taxon (Table 1). The red point represents the value for the chosen confidence score threshold (0.45), providing balanced error types, as indicated by the 0.95:0.95 red dotted lines.

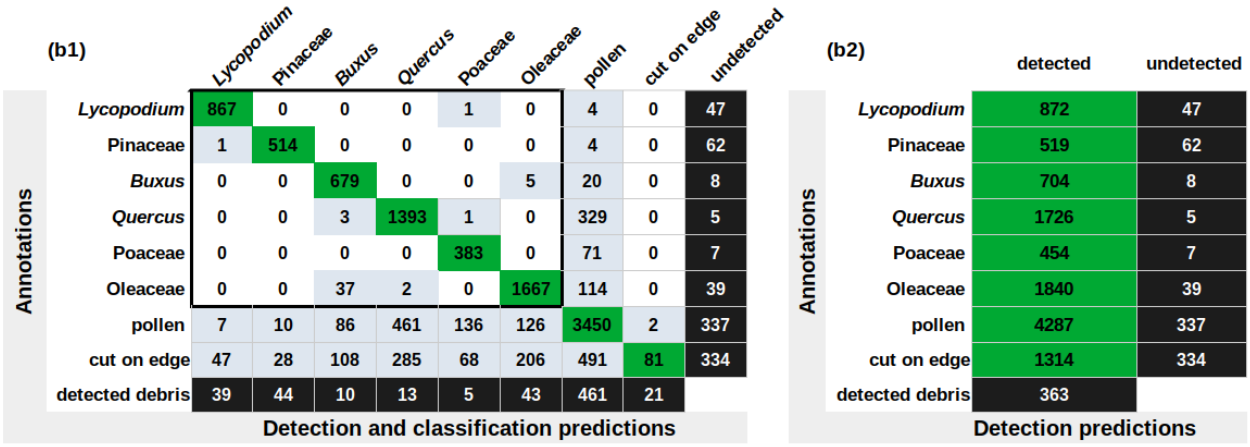


New Phytologist

Methods S4 Confusion matrix for joint detection and classification and for detection only

(a1, b1) Confusion matrix for the joint detection and classification process, showing detection errors (black), classification errors (light blue), and correct joint detection and classification (green); (a2, b2) Confusion matrix only for the detection process, thus showing only detection errors as false negative (undetected grains, black in columns), and false positive (wrongly detected debris, black in lines), and correct detection regardless of the classified labels (green). Values are summed over the 5 cross-validation tests, and illustrated with the annotation strategy all-0taxon (a), and all-5taxa (b).





Methods S5 Definition of the detection metrics

Metrics used to evaluate the performance of the detection method: True Positives (TP) for grains manually tagged and correctly predicted as grains regardless of the label associated during the annotation process; False Negatives (FN) for grains manually tagged but not detected; False Positives (FP) for bounding boxes predicted but not manually tagged, again regardless of the label associated during the automated classification process.

recall = TP/(FN+TP)×100, *i.e.* percentage of correctly detected (predicted) grains in all manually tagged grains

precision = TP/(FP+TP)×100, *i.e.* percentage of correctly detected grains in all detected (predicted) objects

F1-score = 2×(precision×recall)/(precision+recall)×100, *i.e.* harmonic mean of precision and recall

%FN = FN/(FN+TP)×100= 100 - recall, *i.e.* percentage of undetected grains in all annotated grains

%FP = FP/(FP+TP)×100= 100 - precision, *i.e.* percentage of incorrectly detected objects in all detected (predicted) objects

ROC curves (Receiver Operating Characteristics) : precision-recall values measured here for 100 confidence score thresholds ranging from 0.001 to 1

Methods S6 Description of the annotation strategies and the associated datasets

Annotation strategy	**Tagged pollen**	**Label modalities**	**# labels**	**# bounding boxes tagged and labeled**
all-0taxon	all	*Lycopodium*, pollen and grain cut on edge	3	12531
all-5taxa	all	*Lycopodium*, *Buxus*, Pinaceae, Poaceae, Oleaceae, *Quercus*, pollen of other taxa or not determined, and cut on edge grain	8	12531
1taxon-1taxon	1 model taxon	*Lycopodium* and one model taxon (either *Buxus*, Pinaceae, Poaceae, Oleaceae or *Quercus*)	2	919 + [712 or 581 or 466 or 1859 or 1731]

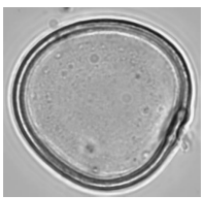
New Phytologist

Methods S7 Variables used to characterize manually annotated grains

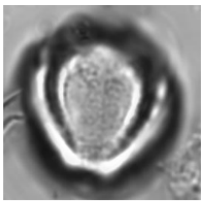
The following variables are used to characterize all manually annotated pollen grains and evaluate the causes of undetected grains.

Variable name	Variable values
identification	**Determined** : the grain taxon can be identified with confidence
	Not determined : the grain taxon cannot be identified with confidence
taxon	*Lycopodium*, *Buxus*, Pinaceae, Poaceae, Oleaceae, *Quercus*
visible section	**Fully visible within the FOV** : less than ¾ of the pollen or *Lycopodium* grain is visible within the FOV because the grain stands partially out of the FOV frame
	Not fully visible within the FOV : more than ¾ of the pollen or *Lycopodium* grain is visible within the FOV because the grain stands fully in the FOV or more than ¾ of its expected surface remains in the FOV frame
image quality	**Poor** : the overall shape and exine texture of the grain is not clear, the grain can be covered by a debris, broken, folded or crumpled, as well as out of focus
	Good : the shape and the exine definition of the grain is clear and visible, it is exempt from the above presented issues
degradation type	**Covered** : the pollen or *Lycopodium* grain is hidden behind a debris, the shape can be visible but not the exine structure or any identification criteria
	Unfocused : the pollen or *Lycopodium* grain is out of focus
	Mixed degradation: **t**he pollen or *Lycopodium* grain combines several types of deterioration, including grains covered, unfocused, broken, folded or crumpled

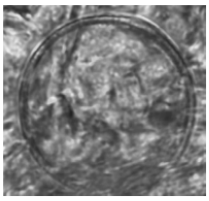
Methods S8 Illustration of the categorical variables used to evaluate the causes of undetected grains



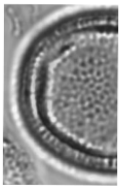
determined
Poaceae
fully in the FOV
good quality



not determined
no taxon
fully in the FOV
poor quality
(unfocused)



not determined
no taxon
fully in the FOV
poor quality
(covered)



determined
Oleaceae
not fully in the FOV
good quality

not determined
no taxon
not fully in the FOV
not qualified

New Phytologist

Notes S1 Analysis of the performance of the detection method from the last vs the best epoch models

We compared detection performance when using the model weights from the last training epoch, and from the best training epoch, both evaluated on the validation dataset. Detection performances, measured as the F1-scores (Fig. S5c), were very similar suggesting that the validation step, and thus the validation dataset, did not help improve results, and could be discarded to save annotation time. In the case the validation step is to be discarded, the number of training epochs has to be optimized to avoid over-fitting. For instance a previous study using YOLOv5 on reference pollen to jointly detect and classify three Betulaceae taxa used 500 epochs and found, based on the validation dataset, no improvement for the last 100 epochs (Kubera et al., 2022). We here used results obtained with the last epoch models, as a conservative strategy, with weights consistently saved after the last 150th epoch.

Reference : Kubera, E., Kubik-Komar, A., Kurasinski, P., Piotrowska-Weryszko, K., & Skrzypiec, M. (2022). Detection and Recognition of Pollen Grains in Multilabel Microscopic Images. Sensors, 22(7), 2690. https://doi.org/10.3390/s22072690

Notes S2 Analysis of the performance of the detection for each size of the dataset

We evaluated the effect of the size of the dataset used to train YOLOv5 on the detection performance of all *Lycopodium* spores and pollen grains, regardless of their taxon (all-0taxon, Table 1). As expected, all metrics of performance (Fig. S5) increased with the number of FOVs (field of views) used to train the algorithm, and the F1-score reached 95.1 ± 0.11% for the maximal size tested here (2,946 training FOVs). With this size, 4.7 ± 0.33% of pollen and *Lycopodium* grains were left undetected (FN) and 5.0% ± 0.26 of debris were falsely detected (FP). Despite variation among cross-validations, the detection performance tended to plateau at about 60% of the dataset, which corresponds to about 1,770 training F0Vs. Training the model for 10% of the dataset with 294 training FOVs already allowed a F1-score as high as 91.7 ± 0.27% (Fig. S5c) and lasted 9 times less than it did for the full dataset with 2,946 training FOVs (20 min compared to 3 hours), and required to annotate 937 compared to 8,981 bounding boxes.

We recommend to determine the number of images required to reduce annotation time while keeping optimal performance, which will be affected by the diversity and abundance of pollen taxa and debris in a given study.

Chapter 3

Variation in pollen rains in the Mediterranean area with automated pollen analysis

Betty Gimenez, Céline Devaux, Sébastien Joannin, Doris Barboni, Sourish Das, Pranabendu Misra, Shubhahish Chauhan, Nathalie Combourieu-Nebout, Laurent Bouby, Sandrine Canal, Sarah Ivorra, Bertrand Limier, Jérôme Pasquet, Odile Peyron

3.1 Introduction

3.1.1 Why monitor pollen rains

Monitoring pollen rains through space and time allows to assess current states and changes in the vegetation and its responses to environmental or anthropogenic disturbances (e.g. van der Knaap et al., 2010; Hattestrand, 2013; Fernandez-Gonzalez et al., 2020; Verstraeten et al., 2023). It also provides modern analogues for paleo-ecological studies helping understand how pollen signals reflect vegetation composition (e.g. Pidek et al., 2010; Filipova-Marinova et al., 2010; Sjögren et al., 2010). Pollen rains have been monitored extensively over the last decades, through airborne pollen concentrations measured from volumetric traps such as Hirst-type traps, or pollen accumulation rates (PAR) measured from gravimetric traps such as Tauber-type traps. Their study have found rapid and specific responses of pollen seasons and pollen emissions to factors including temperature and precipitations (Galán et al., 2016), land use changes (Fernandez-Gonzalez et al., 2020, Algarra et al., 2019), or even CO2 (Ziello et al., 2012). A trend towards earlier and longer pollen seasons combined with higher pollen emissions in response to increased temperatures have been showed in the US (Anderegg et al., 2021), in the Northern hemisphere (Ziska et al., 2019), in Europe (e.g. Bruffaerts et al., 2018; Gehrig and Clot, 2021), and in the Mediterranean area (e.g. Ghitarrini et al., 2017; Fernandez-Gonzalez et al., 2020; Cristofolini et al., 2024). For instance in the Western Mediterranean, a 30-years time series found an advance of the pollen seasons by 5.4 days per 10 years in Northern Italy (Cristofolini et al., 2024) and similar trends have been found for key Mediterranean taxa such as *Quercus* (Fernandez-Gonzalez et al., 2020, Recio et al., 2018), *Olea*, *Pinus* or Cupressaceae (Galera et al., 2018; García-Mozo et al., 2016; Damialis et al., 2011). Altogether, these results are consistent with research on plant phenologies, which have shown advances in flowering times in responses to climate changes throughout the 20th century (Peñuelas et al., 2004; Parmesan, 2006; Gordo and Sanz, 2010; Ellwood et al., 2013).

Mediterranean ecosystems are under increasing climatic and anthropogenic pressures (Giorgi and Lionello, 2008; Stocker et al., 2014), and display strong sensitivity to these disturbances (Newbold et al., 2020; Ali, 2022). Increasing droughts and temperatures in these water-limited ecosystems impact plant communities, the functioning of which largely depends on these factors and for which many species are already reaching their physiological limits (Quezel and Medail, 2003). There is consequently an urgent need to

better understand and anticipate the responses of Mediterranean plant communities to these disturbances. Pollen-data generated from monitoring has the potential to offer valuable insights into these environmental effects by informing on plant communities compositions and phenologies, and detecting rapid ecological changes (Fernandez-Gonzalez et al., 2020; Gómez-Casero et al., 2007; Tormo-Molina et al., 2010; Algarra et al., 2019).
In that context, a long-term international project named POlarise (IRN CNRS, INSU) was developed at the scale of the Mediterranean basin to monitor pollen rains from key Mediterranean taxa. Several monitoring programs are involved in this project, including those focusing on olive trees (Morocco, Boullayali et al., 2024), oak trees (Sicily), and the domesticated grapevine *Vitis* (Pantelleria Island). A monitoring program is also conducted in Southern France since 2019 (Pollimed project, funded by OSU OREME) and focus on *Quercus ilex* (Puechabon experimental station), and in the area of the Pic Saint Loup (Herault, France), on *Vitis*. The study area near the Pic Saint Loup harbors a naturally occurring population of wild grapevine (*Vitis vinifera* subsp. *sylvestris*, later on referred as *Vitis sylvestris*), along with the domesticated grapevine (*Vitis vinifera* subsp. *vinifera* later on referred as *Vitis vinifera*) grown in nearby commercial vineyards. *Vitis sylvestris* is poorly documented while it is an emblematic Mediterranean taxa, present since the Pleistocene, it has become endangered over the last centuries in the Western Mediterranean due to disease but also habitat loss and fragmentation. As a heliophilous liana, *Vitis sylvestris* is strongly dependent of its accompanying vegetation, the presence of which is thus critical for maintaining their relic populations in the Western Mediterranean area. The presence of *Vitis sylvestris* in the study area near the Pic Saint Loup is thus intrinsically tied to its natural Mediterranean habitat. Its monitoring thus also allows to monitor its associated floral assemblages, i.e. several other key Mediterranean taxa. The ultimate goal of this study is to analyze the pollen data collected so far by the monitoring project conducted in the Pic Saint Loup area, and assess the temporal and spatial structures of the vegetation community from the last five years, associated to either *Vitis sylvestris* or *Vitis vinifera*.

3.1.2 Making pollen monitoring efficient and replicable with automated analysis

Monitoring programs are often conducted for several years with weekly, monthly or annual resolutions (Gehrig and Clot, 2021), and thus rely on the analyses of many pollen samples. Still mainly done manually, the process of pollen analysis is time-consuming, labor-intensive and requires skilled specialists while being subject to observer effects. Pollen analysis consequently represents the most limiting process in palynology and to monitor pollen rains. In that context, a first objective in this study is to make pollen analysis rapid, reliable, and replicable by developing an automated pollen analysis pipeline to provide accurate pollen counts for several taxa from the samples collected in the Pollimed monitoring program near the Pic Saint Loup.
Research to automate pollen analysis has been extensive over the last few years, but manual palynology remains standard for environmental-related pollen studies. Automated methods have yet to reach the stage of being used for routine application, i.e. for generating pollen data from new samples collected in the field (e.g. moss pollsters, gravimetric traps, lake sediments), not previously seen when developing the method. Regarding such application conditions, methods are still to their development stage (Punyasena et al., 2022; Theuerkauf et al., 2023; von Allmen et al., 2024; Durand et al., 2024). In a previous work, we developed the detection step, i.e. the first step of an

automated pollen analysis pipeline designed for routine application on environmental samples (Gimenez et al., 2024). So far, our method includes an automated process to get microscopy images from environmental pollen slides that can contain many pollen and non-pollen particles, and to locate, or detect, all pollen grains occurring in these images. Here, we develop the second and final step - classification - and generate a fully automated pollen analysis pipeline. The objective is threefold : (1) build a classification method that can process all pollen and particles generated by the previous automated detection, (2) integrate all automated steps into a full pipeline, and (3) evaluate the pipeline under conditions that are representative of a routine application when monitoring vegetation from pollen rains, i.e. when applied on new environmental pollen samples.
The task of pollen classification has been extensively studied over the last few years, but predominantly by allergists on reference or airborne samples (e.g. Sevillano and Aznarte, 2018; Schaefer et al., 2021; Boldeanu et al., 2021; Zhao et al., 2022; Brdar et al., 2023). Automated classification for environmental-related studies has been investigated for distinct disciplines, including pollination ecology (Olsson et al., 2021), ecological airborne pollen monitoring (Punyasena et al., 2022), and paleoecology (Bourel et al., 2020, Theuerkauf et al., 2023; von Allmen et al., 2024; Durand et al., 2024). Yet, these studies remain scarce, and conducted under conditions that are not always representative of those encountered during routine application, for instance by filtering out debris or poor-quality images generated by the previous detection step. Many technical and scientific challenges remain to achieve efficient automated classification for the environmental monitoring of pollen rains. First, such environmental pollen samples, especially those that cover long-time series such as annual traps, include a wide range a pollen taxa and can include unexpected and rare taxa. Second, such samples may contain large quantities of non-pollen particles of multiple types such as plant or insect debris, with some sharing morphological similarities with pollen grains, and which may have to be processed by the classification task due to errors introduced by the detection step. Third, pollen grains can be damaged or distorted, and in the images, they can occur covered by other particles, cut on the images margins, or blurred if photographed out of focus. These conditions make the classification step a challenging task (Diaz-Lopez et al., 2015; Bourel et al., 2020; Barnes et al., 2023), particularly for samples collected in the Mediterranean region due to its high taxonomic diversity.
Once developed, the automated pollen analysis pipeline is applied on pollen samples from the Pic Saint Loup vegetation monitoring to investigate the potential temporal and spatial changes in the composition of 15 key Mediterranean taxa over the last 5 years.

3.2 Material and Methods

3.2.1 Pollen rain monitoring and environmental samples

3.2.1.1 Study area and sampling sites

The study area is located around the Pic Saint Loup massif, near Montpellier in Southern France (Fig. 1). This area is located in the meso-Mediterranean bioclimatic region, characterized by predominant Mediterranean climatic conditions, and experiences summer droughts with intense precipitations events during autumn, known as Cévenol episodes. The Pic Saint Loup massif reaches an elevation of 658 meters with a crest extending on more than 4 kilometers along an east-west direction and with a very steep northern slope. Winds blow predominantly North, but might be locally modified due to the Pic Saint Loup massif being transverse from this direction.
Seven sampling sites were set up in 2019 within a 1-km radius from the Pic Saint Loup

massif, with each site being in close proximity to either a wild grapevine individual from a relic population growing in the area (This et al., 2006), or in a nearby commercial vineyard (Fig. 1). The overall sampling area is rural and mainly characterized by open and closed scrubland with the presence of typical Mediterranean taxa, including evergreen oaks (*Quercus ilex* and *Q. coccifera*), deciduous oaks (*Quercus pubescens*), and *Buxus*, *Juniperus*, *Cupressus*, *Pistacia* (*lentiscus* and *terebinthus*), *Phillyrea* and *Pinus*. The slopes of the Pic Saint Loup massif are predominantly covered by an evergreen oak forest, which is open by wide stony scree on the Northern steep slope, and transitions to a mixed deciduous and evergreen oak forest on the lowest altitude of the Northern slope. The area also includes vineyards, uncultivated fields, as well as some open Mediterranean meadow, where herbaceous species such as Poaceae, Asteraceae, *Artemisia*, Brassicaceae and *Plantago* are present. *Fraxinus excelsior* and *Fraxinus ornus* are also present at one sampling site, located near an intermittent riparian area. A small fruit tree orchard is grown in close proximity to one of the sampling site, next to the vineyard monitored.

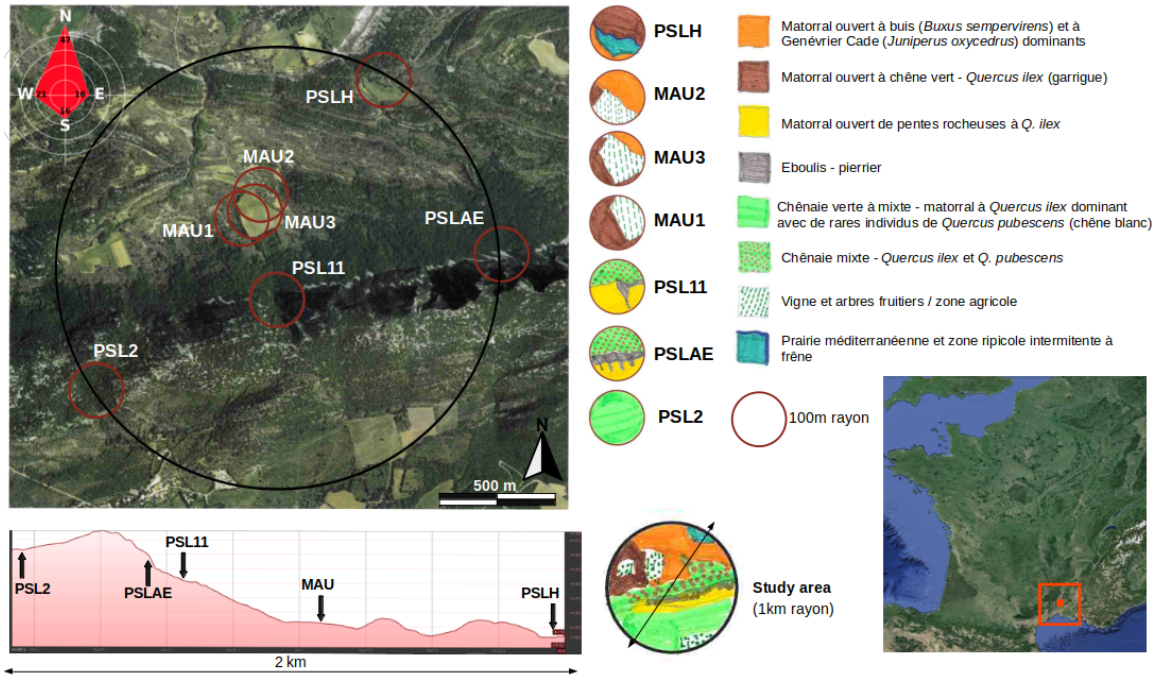


Figure 1: Vegetation cover around each sampling site of the monitoring

Note that, while *Vitis vinifera* is predominantly hermaphroditic, *Vitis sylvestris* is dioecious, with functionally female individuals producing sterile pollen. Sampling sites were thus placed near both wild male (sites PSL2, PSL11, PSLAE) and wild female (site PSLH) individuals, additionally to those placed in the vineyard (sites MAU1 to MAU3, Fig. 1). In these sampling sites, 13 Mediterranean taxa were identified as target taxa for this monitoring, additionally to *Vitis*, based on their ecological significance and their abundance in the study area. These taxa include *Quercus ilex*, *Quercus pubescens*, *Buxus*, *Juniperus*, *Cupressus*, *Pistacia*, *Phillyrea*, *Olea*, and *Pinus*, as key species for both open and closed Mediterranean scrublands. *Poaceae* and *Plantago* were also listed as indicators of more open herbaceous areas, and with *Plantago* potentially reflecting anthropogenic disturbances. *Fraxinus* (*ornus* and *excelsior*) was finally listed as potential indicator for more humid conditions.

Vitis (*sylvestris* and *vinifera*)
Quercus (*ilex* and *pubescens*)
Juniperus and *Cupressus*
Olea, *Phillyrea* and *Fraxinus* (*ornus* and *excelsior*)
Buxus, *Pistacia* and *Pinus*
Poaceae and *Plantago*

Table 1: List of the target taxa in this study

3.2.1.2 Pollen samples

Pollen traps were collected annually in a total of six sampling sites per year, from 2019 to 2023, but with seven locations in total as the sampling site MAU1 was discontinued in 2022 and replaced by the sampling site MAU3. Traps consisted in containers with a 5 cm width opening covered by a grid of 5 mm × 5 mm, inspired from Tauber traps (Tauber, 1965). At each location, the traps were placed in cavities in the soil with the opening around 15 cm above ground (Fig. 2 and 3), at the exception of two traps which were attached to the trunk of the tree supporting the grapevine liana. Glycerol and thyme essential oil were added to the traps before their installation to retain and preserve pollen grains, and avoid fungus growth. The traps were then collected and changed each year in early January starting from 2019, for a total of 28 traps (two traps were damaged or could not be used).

Figure 2: Annual trap located at the sampling location PSL 11

Once pollen samples were collected from the traps, spores of *Lycopodium* were added for calibrating pollen counts (Stockmarr, 1971). The samples were then chemically treated to remove calcium carbonates and silicates, acetolysed for 10 minutes, and finally stored in a few drops of glycerol (see details in chapter I). For each of the 28 pollen samples, one to three fixed slides were mounted using glycerol jelly (see details in chapter I), for a total of 51 slides. Standard mobile slides were also mounted for 13 samples to get manual counts performed by an expert palynologist and provide independent counts on the same samples, but not the same slides.

3.2.2 Individual pollen images : acquisition and detection

The slides were digitized under light microscopy following the process described in chapter II (Gimenez et al., 2024), and using 13 images in the stacks for each field of view (FOV), 8 µm for the z-axis step between these images, 140 µm for the distance between each FOV thus avoiding overlapping images. One to three thousands FOVs per slide were generated, this being adjusted to maximize the number of captured pollen grains in a constrained amount of time. These parameters differed for 16 slides previously acquired, with adjacent

Figure 3: Sampling locations PSL11 (left) and PSL2 (right)

FOVs overlapping by 10 µm and using 11 images in the stacks (still spaced by 8 µm). Each FOV image covered 214 µm × 214 µm and could thus contain multiple pollen and non-pollen particles, such as Non-Pollen Palynomorphs (NPPs) or debris. Pollen grains could also occur in varying states of visibility, for instance if covered by another particle, if out of focus, or when cut on the FOVs margins. The detection method previously developed (Gimenez et al., 2024; chapter II) and based on the object-detection algorithm YOLOv5 (Jocher, 2020) was applied to these FOVs images to generate bounding boxes around pollen and *Lycopodium* grains (detection step, see details in chapter I). The generated bounding boxes were then used to extract images containing each a single detected particle (segmentation step, see details in chapter I). This process generated from 479 to 15,176 bounding boxes, and thus individual images, per slide across the 51 mounted slides, with each individual image containing a single pollen grain or *Lycopodium* spore, in any state, or a falsely detected particle such as a debris, an air-bubble or a light artifact.

3.2.3 Classification of pollen images

To provide accurate counts of the 15 key Mediterranean taxa selected in this study (Table 1) and of *Lycopodium* grains used for calibration, the identification pipeline had thus to be able to recognize these target categories but had also to handle all the other images of non-target pollen taxa and non-pollen particles generated by the automated detection. The first step consisted in implementing a classifier, specifically, a classification Convolution Neural Network (CNN). The ResNet152 algorithm was selected due to the efficiency of the residual network (ResNet) architecture for pollen classification (e.g. Punyasena et al., 2022; Matavulj et al., 2023; Rostami et al., 2023) and the higher performance of ResNet152 compared to other ResNet architectures, including ResNet34, 50, and 101 (He et al., 2015). As any supervised classifier, ResNet152 relies on (1) a training stage during which the model

parameters are fitted to a training dataset of images labeled among a pre-defined number of classes, and (2) on an evaluation stage.

3.2.3.1 Building the dataset for model training

The classifier has to be trained on a specific number of classes, with each class that can represent a single image type (e.g. one pollen taxon) or encompass several image types (e.g. several pollen taxa grouped in one class). Once trained, the classifier can only classify new images among these pre-defined classes from training. Thus, any image type encountered during subsequent monitoring and falling outside these pre-defined classes (e.g. an unexpected pollen taxon) will systematically be misclassified and generate a false positive for one of the training classes. To address this issue, our strategy was to build a training dataset including a maximum number of the image types expected when applying the pipeline to the annual pollen samples. This diversity in the image types was considered in term of (1) pollen taxa, with both abundant and rare taxa, (2) non-pollen particles, with NPPs, debris, or air bubbles, and (3) image visual qualities and pollen preservation states, with pollen grains obscured by debris, cut from the FOVs margin, blurry or damaged, and without necessarily clear identification criteria. For each image type, we also aim to include as many images as possible, with an objective of 1300 images per class, because deep learning models are the most efficient on extensive datasets.

We first built an environmental dataset by compiling images from the annual pollen traps, and using images from 16 out of the 51 slides (i.e. those scanned first). Few monthly samples collected at the same sites but outside this study were also considered to supplement pollen classes with too few images. Pollen grains that could be visually identified in the images were selected and sorted by taxon (Reille and Pons, 1990; Punt et al., 2007). Images of pollen grains that could not be identified because they appeared heavily blurry or because they were obscured by particles were also added into two separate classes, named respectively 'indeterminable pollen blurry' and 'indeterminable pollen covered'. A final class was created for non-pollen objects, combining debris, NPPs, air bubbles and light artifacts, and named 'non-pollen'. To accurately capture the grapevine pollen signal in our study area, separate classes were created for each of the two grapevine pollen morphotypes: (1) the trizonocolporate pollen morphotype, from domesticated hermaphroditic and wild male individuals, referred as *Vitis* fertile, and (2) the inaperturate pollen morphotype from wild functionally female individuals, usually less known from palynologists (Mercuri et al., 2021), referred as *Vitis* sterile.

Despite efforts, most taxa in this environmental dataset had too few images, either because the taxon was scarce in the samples (e.g. only 7 *Juglans* pollen grain), or because its visual identification was made impossible in many cases by the presence of other morphologically similar taxa (e.g. *Corylus* or *Betula* occurring in polar view in the image). To increase the number of images for each taxon and include taxa that were expected in the study area but were not found in the slides used (e.g. *Populus* and *Ilex aquifolium*), images from reference samples of pollen collected from known plant individuals were added. Reference slides from native species to the study area (Hérault, France) or from species of the same genus or family (to ensure representation of the pollen morphotype) were selected from the collection of palynology of the ISEM laboratory, and few slides were also mounted from fresh pollen collected nearby the study area (as part of the Pollimed and the Pollumine projects, https://oreme.org/observation/). The reference slides were processed with the same image acquisition, detection and segmentation

pipeline previously described (section 2.2). Reference images were then sorted by taxon following the level of the manual identifications in the environmental dataset (Reille and Pons, 1990; Punt et al., 2007). Adding reference images allowed to add many more taxa compared to the 15 target Mediterranean taxa of this study. These additional taxa were included to ensure that the method would be able to handle all pollen taxa that may be encountered in routine, even those occurring scarcely (e.g. reference pollen from *Tsuga* were added, as one grain was found in the environmental samples, probably from a fence of a local residence).

The environmental dataset (11 945 images, corresponding to 4% of the total number of particles detected and segmented from the 51 slides) and the reference dataset (63 343 images) were merged. When the combined number of images for a class exceeded the threshold of 1300, the class size was reduced by removing images with a random selection, but removing reference images first. In contrast, some taxa still had too few images and were thus unsuitable to form a standalone training class. To address this issue while keeping the representation of the corresponding pollen taxa in the dataset, a class "other" was created, grouping together taxa with fewer than 17 images each (corresponding to the 10% smallest classes). The final dataset included 56 101 images (21.3% and 78.7% of environmental and reference images respectively) and 83 classes (Fig. 4). These 83 classes included 78 distinct pollen taxa, a heterogeneous class labeled 'other' that grouped several under-represented pollen taxa, two classes of indeterminable pollen (either blurry or covered), one class for non-pollen particles and one class for *Lycopodium*. Each class contained from 33 to 1300 images with an average of 616±109 (±2SE) images (Fig. 4). Note that although we only considered 15 target Mediterranean taxa for the monitoring study, the classes for non-target taxa were kept separate in the training dataset when possible (sufficient number of images). This was done to optimize their representation in the dataset and improve their identification, in view of limiting false positives they may otherwise generate for the key taxa when encountered during routine application.

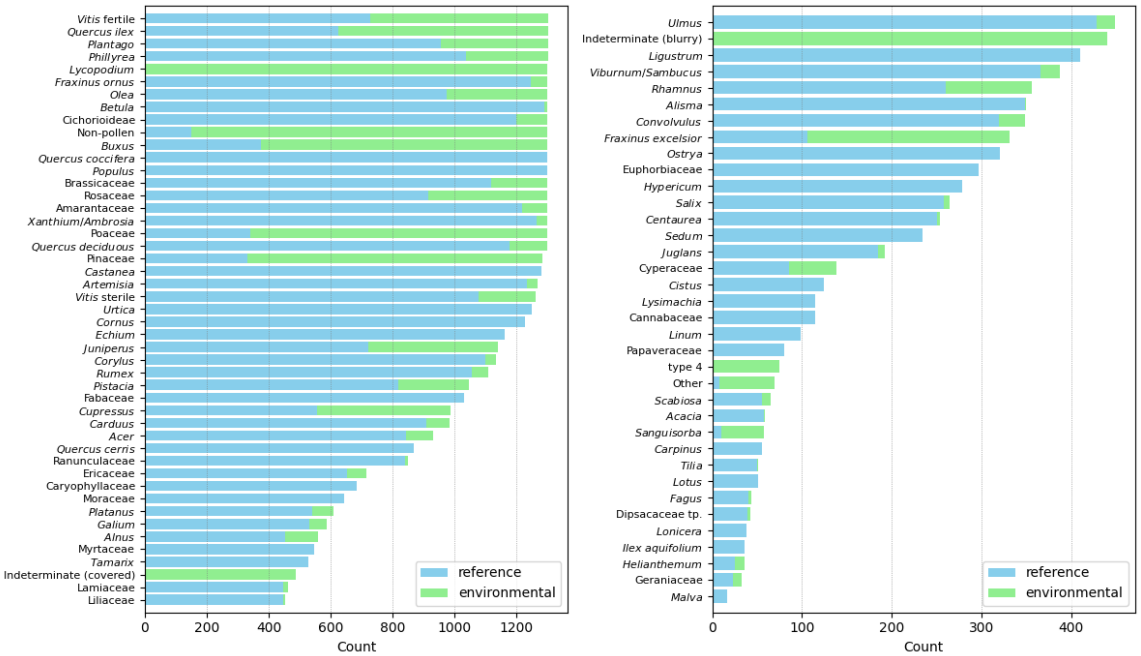


Figure 4: Number of environmental and reference images for each class used to train and evaluate the classification models

3.2.3.2 Classifier training

The classification algorithm processes square images of a standard size and thus transforms images to this format before implementation. To prevent distorting pollen grains during this process, especially those cut on the FOVs image margins (generating very thin images), images were padded to squares before implementation, by adding black pixels, either vertically or horizontally, and equally to both sides of the shorter dimension. The models were cross-validated 10 times on the 83 classes, using 80% for training and 10% for validation, while the remaining 10% were set aside for independent evaluation of the models on images never seen during training (see details in chapter I). Environmental and reference images from each class were equally distributed among the three subsets. Each training was set for 150 epochs, using an early-stopping callback of 30 epochs, and an adaptive learning rate (see details in chapter I). During model training, the standard suite of data augmentation techniques was applied to the training set, including rotations, flips, shearing, changes in the focus, and saturation levels. The models were run on a graphic card from the Chennai Mathematical Institute.
This process generated 10 trained models, set for the identification of 83 classes. The images were reserved only once in the test subset. Thus, they could be used for the performance evaluation of only one trained model, while they were used either for the training or the validation stage of the nine other cross-validated models.

3.2.3.3 Classifier performance evaluation

Each of the 10 cross-validated models were applied to their respective test set (one test set per model), generating one prediction for each image of the full dataset. The classification performance was primarily evaluated on images from the environmental dataset, and not the reference dataset, as these images were representative of the data, and the models were intended to be applied to the annual pollen samples. Predictions were considered for all the 83 classes separately, and also by merging the classes from non-target pollen taxa with the class of 'pollen other'. Two strategies were then followed to evaluate the performance and to determine the one to use in the final pipeline for generating pollen data from new environmental pollen samples. First, we assigned to each image the class with the highest score (top-1), and compared it to the class manually labeled (ground truth) to determine the following statistics: true positive (TP) if the top-1 predicted class and the ground truth matched, and otherwise false negative (FN) for the ground truth class and false positive (FP) for the top-1 predicted class (see chapter I). Second, we evaluated the top-4 predictions by assigning to each image the four classes predicted with the highest scores. The choice of considering specifically the top four predictions was made by evaluating performances using an increasing number of top predictions, and by selecting the lowest possible rank that maximized performances (see details in chapter I, and in Supporting Information Methods M1). This balance aimed to increase the chance of capturing the correct class (ground truth) among the selected predictions while limiting that number of considered top predicted classes. When using the top-4 strategy, a TP was determined if the ground truth class was among these four predictions, otherwise a FN was determined for the ground truth class, while no FP were determined with this approach. The TP, FP and FN statistics were first used to measure the microaccuracy, i.e. the performance of the classification measured over all test images regardless of the classes, and defined as the percentage of correctly classified images (TP) among all images. These statistics were then used to measure three performance metrics for each of the classes : (1) the recall, defined as the percentage of correctly classified images of a given class among all true images from that class, (2) the precision, defined as

the percentage of correctly classified images of a given class among all images predicted to that class, and (3) the F1-score, i.e. the harmonic mean between precision and recall. Note that these three metrics could not be reliably measured on environmental images for 23 classes that had insufficient number of images (< 50 images) and for 26 classes that had no environmental images (only references). Finally, confusion matrices were also generated to get detailed information on misclassification (see chapter I).

3.2.4 Configuring an identification pipeline to get pollen counts for environmental samples

The detection and segmentation pipeline applied to all slides mounted from the annual traps generated many images of many types. Some of these images contained particles cut on the FOV margins and with only a very thin portion visible that often lacked identification criteria. To avoid generating false predictions from these barely visible particles, all images from bounding boxes touching the FOV margin and smaller than 10 µm were eliminated (Fig. 5). This threshold did not remove the smaller pollen taxa, as it is lower than their size and also accounts for the small gap between the grains external edge and the bounding box edge. Note that for the 16 slides digitized first and with overlapping FOVs, and to avoid counting twice particles located at the margins of the FOVs, bounding boxes along the top or left margins of their original FOV images were all removed, regardless of their size. All other images were then processed by the classification models. To increase the robustness and performance of the identification of the remaining images, we considered the top-4 predictions of each model, and combined these predictions from all 10 cross-validated models, as done in ensemble modeling approaches (Mohammed and Kora, 2023; see details in chapter I). This process resulted in a distribution of 40 predictions for each image, for which each class (out of the 83), could occur 0 to 10 times (Fig. 5). The prediction assigned to the image was the class with at least 9 occurrences among the 40 predictions. If two classes had the same number of occurrences, the class with the highest overall rank across the top-4 ones from each model was selected. If no class reached at least 9 occurrences, the predictions were considered unreliable. In that case, if the class 'non-pollen' was not among the 40 predictions, we assumed it was a pollen grain and labeled the image as 'doubtful pollen', otherwise the image was labeled as 'doubtful particle' (see details in Supporting Information Methods M4; Fig. 5).

3.2.5 Analysis of pollen rains by site and year

The total pollen counts for each annual sample corresponded to the sum of pollen counts over the 78 pollen taxa, 'pollen other', indeterminable pollen (covered and blurry) and doubtful pollen. Pollen counts were then calibrated for each slide to which they belonged using the counts of *Lycopodium* spores, thus providing an estimate of the annual pollen accumulation rates (PAR) per site and year. Annual PAR were calculated similarly for each of the 15 target taxa (Table 1). The percentages of each taxon in the PAR (pollen assemblage) were calculated by merging pollen from all other taxa into a category 'other pollen taxa', and without accounting for pollen images labeled as doubtful, indeterminable covered or indeterminable blurry, as these classes can include both target and non-target pollen taxa. When several slides were scanned for a given sample, the calibrated counts were averaged over the slides. Before analyzing pollen rains from the monitoring, and to check the reliability of our method, the automated pollen counts for 13 annual samples

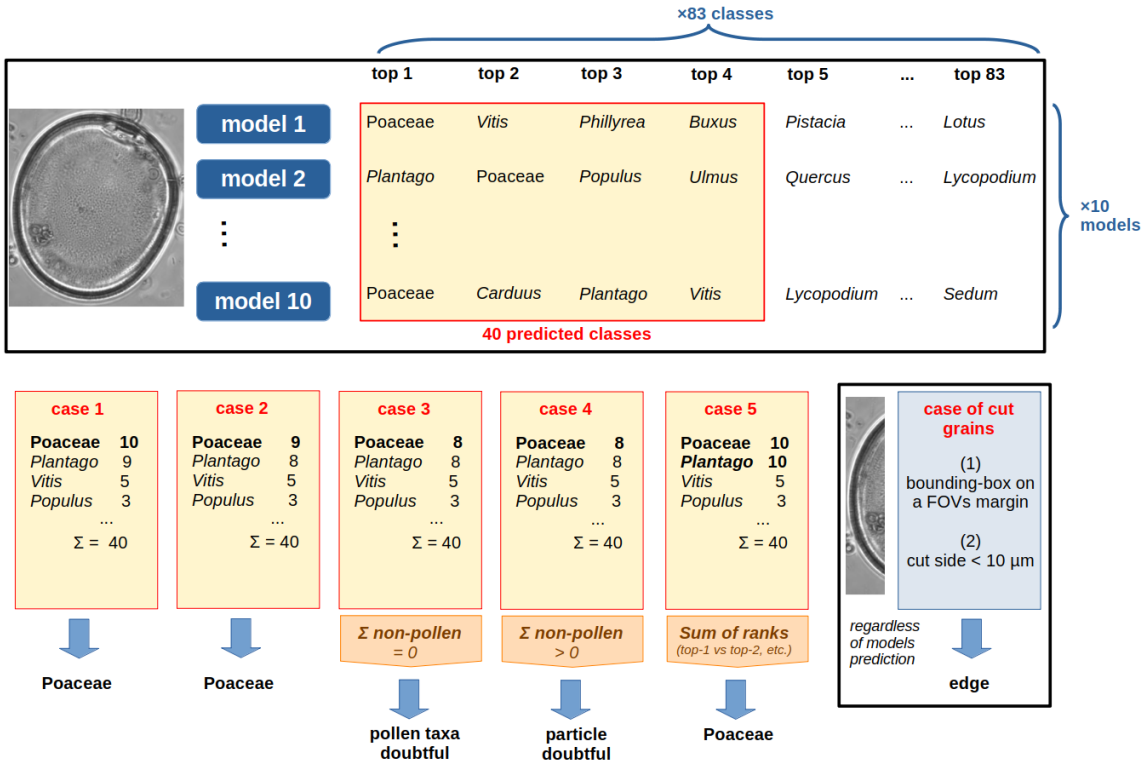


Figure 5: Graphical diagram illustrating the strategy adopted to combine the predictions from the 10 models into a single class in the final pollen analysis.

were compared to the manual pollen counts performed by an expert palynologist, using the same samples but distinct slides (due to distinct protocols for each approach).
Annual PAR were analyzed both regardless of taxon or for each target taxon, for each site and year, for each site over years, to account for expected temporal variations in pollen releases (Nielsen et al., 2010; Haselhorst et al., 2020). PAR were finally investigated at the scale of the whole sampling area over the studied period, i.e. over all years and sites. The composition of the pollen assemblages was analyzed with correspondence analysis at the site level, and per site and year, when accounting for the 15 target pollen taxa only, and also accounting the other identified taxa grouped into a class 'other pollen taxa'.

3.3 Results on the classification method and analysis of pollen rains

3.3.1 Performance of the classification models on the test sets

Results are presented as the mean and standard errors over the 10 cross-validations (mean ± SE), and by merging the classes for the non-target pollen taxa with the class 'other pollen'.

Results obtained on the test data using the top-1 approach were first evaluated. Despite the lower representation of the environmental images in model training, the overall classification accuracy of taxa in the environmental images reached 92.0±0.252% (microaccuracy, Table 2). As expected, F1-scores for easily identifiable and well represented taxa in the environmental dataset were high, including for Pinaceae

(99.3±0.291%), *Buxus* (97.6±0.496%), Poaceae (96.9±0.650%) and for *Lycopodium* (99.1±0.291%, Table 2). The other target taxa mostly reached F1-scores higher than 90%, namely for *Plantago*, *Olea*, the fertile morphotype of *Vitis*, *Pistacia* and *Fraxinus excelsior* (Table 2).

Lower performances achieved for the other target taxa were mostly due to misclassification with morphologically similar taxa from the same family (Fig. 6). It was the case for *Juniperus* and *Cupressus* (F1-scores of 78.8±1.224 and 81.0±1.390 respectively), from the Cupressaceae family, which were misclassified with each other, and for *Phillyrea* (82.0±0.200), which was misclassified with other Oleaceae taxa and especially *Fraxinus* (Fig. 6). These misclassifications between similar taxa partly balanced each other, but also led to some over or underestimation among predictions in the test data, indicated by significant differences between recall (related to FN) and precision (related to FP). For instance for *Phillyrea*, recall higher than precision (by 8%) indicates that this taxa was over-estimated in predictions on the test data, which was mostly at the expense of *Fraxinus ornus*, and which was under-estimated as shown by its lower recall than precision (by 23%), note however that metrics for this taxon were only measured on 5.4±0.163 images per test set (Fig. 6 and Supporting Information Fig. S8). In contrast, despite the lower F1-scores for *Juniperus* and *Cupressus* due to their misclassification (∼80%), their respective recall and precision were of similar range (respectively 1% and 2% differences), suggesting that misidentification between the two taxa balanced each other and thus did not affect their predicted distributions in the test data.

Some misclassifications were more surprising, for instance, 17% of *Quercus deciduous* were misidentified into Rosaceae (Supporting Information Fig. S1), contributing to the under-estimation for this taxa in the predictions (precision higher than recall by 7%). *Quercus* (both *ilex* and *deciduous*) and *Vitis* (both morphotypes) also tended to be misclassified between themselves (Fig. 6 and Supporting Information Fig. S1). *Quercus* and *Vitis* fertile have three colpi and share a similar texture, potentially explaining their misclassification, which contributed to an overestimation of *Quercus ilex* in the predicted test data (4% differences between recall and precision). Comparing precision and recall also informed that *Vitis* sterile was over-estimated in the predictions on the test set (recall higher than precision by 8%), due to false positives from distinct taxa, including Poaceae and *Quercus ilex* (Fig. 6). Additionally, the similar reticulated exine of Brassicaceae also led to misclassification with *Olea* (Supporting Information Fig. S1 and S2), which despite its high F1-score (94.1±0.567), was slightly over-estimated in the test data predictions (6% differences between precision and recall).

Unfocused grains were very well recognized (F1-score of 96.8±0.533%). In contrast, the class of covered pollen grains reached a lower F1-score (84.7±1.100%) due to numerous misclassifications with the class of non-pollen (5.75% of FN and 3.45% of FP, Supporting Information Fig. S1 and S2), and most probably because pollen was frequently covered by non-pollen particles. The class of non-pollen particles reached an F1-score of 93.3±1.303% with similar precision and recall values, and aside from the class of covered pollen, it generated few errors for the other pollen categories (Supporting Information Fig. S5).

Finally, comparing precision and recall indicated that predictions on the test data overestimated *Phillyrea*, *Olea*, *Vitis* sterile and *Quercus ilex*, under-estimated *Fraxinus ornus* and *Quercus deciduous*, and provided similar distribution as the true distribution for the 9 other target pollen taxa, for *Lycopodium*, and for the classes of indeterminable pollen and non-pollen (similar precision and recall).

As expected, when considering the top-4 predictions, recall rates increased significantly compared to those obtained with the top-1 predictions and reached on overall micro-accuracy of 98.88±0.075% (Supporting Information Table M1). The top-4 approach

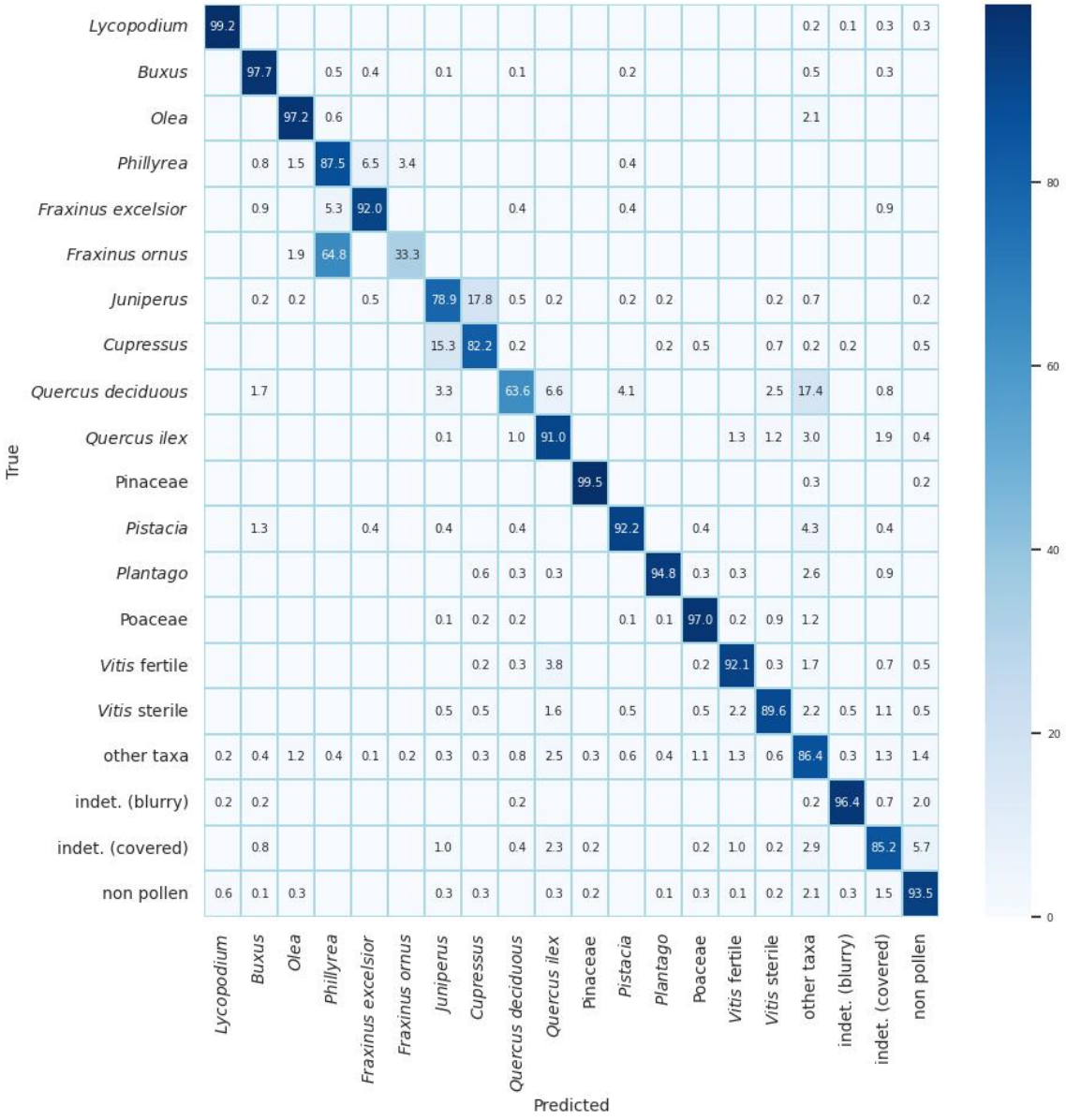


Figure 6: Confusion matrix with the key target taxa of the study. All classes for non-target pollen taxa are merged in the class "other taxa".

allowed to capture the correct class for 99% to 100% of the images for 16 out of the 20 target classes, and recall rates higher than 97.78% for the other four target classes, including for *Quercus deciduous*, *Vitis* fertile, *Vitis* sterile and the class of indeterminable pollen covered (Table 2). These results indicate that for very few numbers of images, the correct class was not among the four classes ranked by the model prediction with the highest scores. For all other predictions in the test data, the correct class for an image was always among the top highest scoring classes from the model prediction on that image. Differences in recall rates between the top-1 and top-4 approach between classes indicates that for some classes (e.g. *Lycopodium* with a top-1 recall rate already of 99.154%), the correct prediction often ranks first, while for some classes (e.g. *Fraxinus ornus* with a top-1 recall of 32.33% and a top-4 recall of 100.0%), the correct predictions can be more frequently ranked in forth position (Table 2). The top-4 approach thus allows to get the correct class for taxa that are often misclassified among closely related taxa, for instance for Oleaceae or Cupressaceae, as the training classes from the taxa family are

likely to rank among the top predictions, the fact that top-1 recall rates are high, nonetheless suggests that the correct taxa most frequently ranks first.

Table 2: Top-1 predictions performance metrics and top-4 recall rates for the 20 key classes on environmental images (mean ±SE over 10 cross validations)

class	**nb env**	**recall**	**precision**	**f1 score**	**top-4 recall**
Lycopodium	130.0±0.0	99.154±0.314	99.081±1.2	99.116±0.291	100.0±0.0
Buxus	92.6±0.221	97.735±0.348	97.547±2.3	97.636±0.496	99.6797±0.2275
Cupressus	43.2±0.133	82.188±0.967	80.594±9.0	80.967±1.39	99.3129±0.3499
Juniperus	42.1±0.1	78.87±1.059	79.596±8.9	78.812±1.224	99.2913±0.3609
Fraxinus excelsior	22.5±0.224	92.011±0.389	88.966±2.6	90.398±0.4	100.0±0.0
Fraxinus ornus	5.4±0.163	32.333±0.371	55.833±1.3	39.786±0.335	100.0±0.0
Olea	32.6±0.267	97.206±0.277	91.337±3.1	94.088±0.567	100.0±0.0
Phillyrea	26.3±0.153	87.436±0.396	79.467±6.1	83.072±0.722	99.6154±0.3846
Pinaceae	95.4±0.221	99.474±0.224	99.168±0.8	99.318±0.291	99.8947±0.1053
Pistacia	23.0±0.0	92.174±0.49	90.535±2.3	91.14±0.448	99.1304±0.5797
Plantago	34.5±0.167	94.782±0.593	96.545±1.2	95.559±0.291	99.1176±0.6278
Poaceae	95.6±0.221	96.965±0.547	96.907±3.0	96.919±0.65	99.4748±0.2351
Quercus deciduous	12.1±0.1	63.526±0.562	71.015±3.6	65.753±0.872	98.3333±1.6667
Quercus ilex	67.7±0.396	90.95±1.038	86.811±9.5	88.724±1.046	99.7037±0.1975
Vitis fertile	57.2±0.133	92.148±0.719	91.975±4.7	91.972±0.746	98.9534±0.5936
Vitis sterile	18.3±0.213	89.585±0.277	81.803±4.0	85.074±0.989	97.7778±1.2284
other pollen taxa	187.1±0.795	86.447±1.881	91.834±14.5	89.014±1.47	99.4691±0.2612
Pollen blurry	44.0±0.0	96.364±0.581	97.353±1.2	96.794±0.533	99.3182±0.3472
Pollen covered	48.7±0.213	85.232±0.646	84.509±7.8	84.736±1.104	98.7794±0.5402
Non pollen	115.1±0.314	93.494±1.302	93.321±7.9	93.323±1.303	99.6537±0.1911

Performances for the other non-target pollen taxa can be found in Supporting Information (Table S2 and Fig. S4 and S3). Some of these classes were well recognized, reaching F1-scores higher than 90% (Cichorioideae, Brassicaceae, Cyperaceae, *Carduus*), although metrics could not be determined for many taxa due to insufficient class sizes. Overall the microaccuracy reached 90.99±0.212% with the top-1 approach when considering all 83 classes separately (Supporting Information Table M1). As expected, when considering both reference and environmental images, performance was higher and reached an overall accuracy of 96.4±0.097% (microaccuracy), with F1-scores exceeding 95% for 53 classes and never below 70%, except for the heterogeneous class 'other' compiling several pollen taxa with very few examples each (Supporting Information Fig. S6 and Table S2). Again, classes with small number of images (less than ∼150 in total) and of taxa sharing morphological features with others, especially among families, achieved the lowest performances (Supporting Information Fig. S7).
When evaluating the top-4 predictions considering all 83 classes separately and on the environmental images only, recall rates exceeded 99% for 19 classes (Supporting Information Fig. M3, numbers 0 to 18), and exceeded 95% for 10 additional classes (Supporting Information Fig. M3, numbers 19 to 28, and Table S3). The two classes that reached accuracies below 90% had very few environmental images (e.g. respectively 5.2±0.13 and 6.2±0.66 for *Rumex* and other) making this metric not very reliable (Supporting Information Table S3) while metrics were not measured for the other classes due to insufficient number of environmental images (less than 5 testing images per cross validation).

3.3.1.1 Comparison of manual and automated counts

The full pipeline was then applied to all the images obtained from the 51 slides (as described in section 2.4) to generate calibrated pollen counts for each slide and then each combination of site and year. The automated counts for 13 of these samples were compared to counts obtained manually, on the same samples but distinct slides (due to distinct protocols for each approach). Pollen counts estimated regardless of pollen taxon, i.e. total annual PAR, correlated between the manual and automated approach (Fig. 7), at the exception for the sample PSL2, and to a smaller degree to PSL11, of 2019. For PSL2-2019, only 3 *Lycopodium* grains were counted manually, making the calibration unreliable and probably explaining the large discrepancies observed with the automated counts (Fig. 7). The same issue occurred for PSL11-2019, with only 6 *Lycopodium* grains counted manually. The number of *Lycopodium* tablets added to the samples in 2019 was insufficient, explaining these low *Lycopodium* counts for this monitoring year, although, more *Lycopodium* grains were systematically counted with the automated method. The protocol was adapted in subsequent monitoring years, allowing to obtain more *Lycopodium* grains with both approaches, and thus improving the quality of the calibration and allowing to compare more accurately calibrated counts between methods.

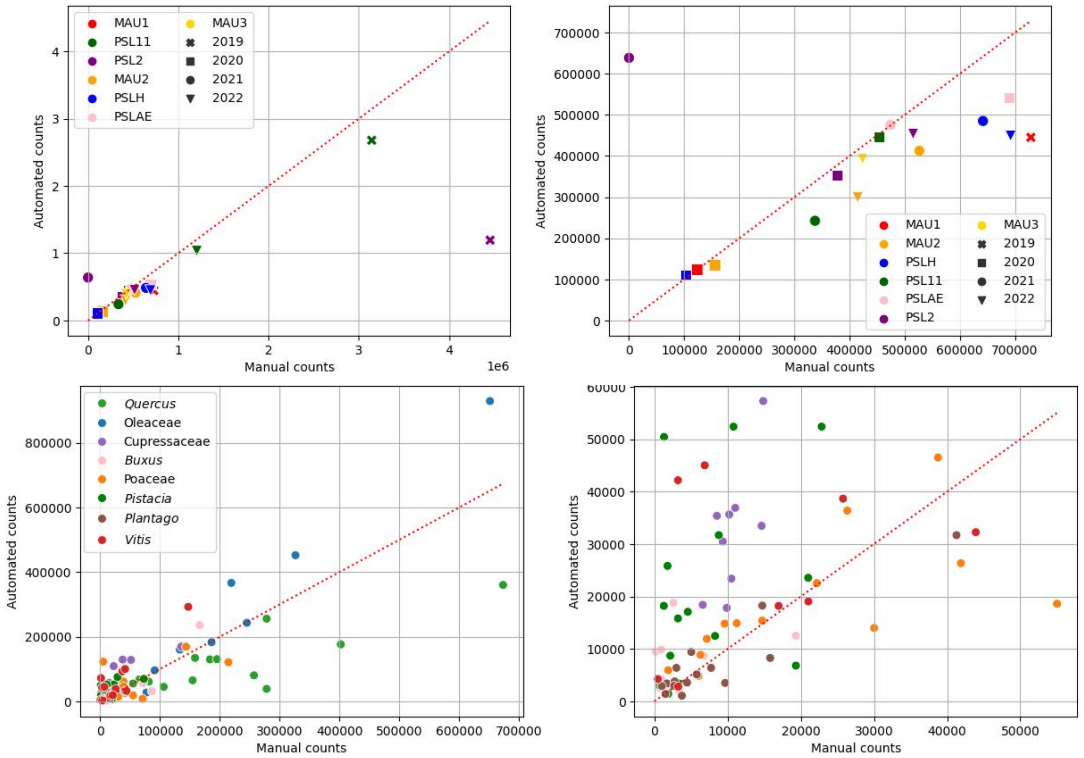


Figure 7: Comparison of manual and automated counts obtained for each sample, i.e. per site and year of collection, and measured (a, top) for all pollen grains regardless of taxa, and (b, bottom) for target taxa separately as the genus or family level. The figures on the right side present a zoomed view of the plots.

Comparing pollen counts per taxa between methods highlighted large differences among taxa from the same family, as expected from the evaluation of performances based on the test data (Supporting Information Fig. S8). For instance for the Oleaceae family, distinct patterns were obtained between the counting methods. The abundances of *Fraxinus* and *Phillyrea* differed between the samples based on the manual counts, while they tended to be more similar across samples based on the automated counts. When comparing pollen counts between both approaches using the overall Oleaceae pollen counts (sums for *Olea*, *Phillyrea*,

and *Fraxinus*), results compared better, confirming that discrepancies observed were mostly related to misclassification inside the Oleaceae family (Supporting Information Fig. S8). These results suggest that the misclassification between these taxa previously observed with the automated approach, especially between *Phillyrea* and *Fraxinus ornus* (to around 12% based on the test set, Fig. 6) tended to level out their respective abundances in the final automated pipeline. In contrast the palynologist expert counting manually tended to count predominantly one or the other taxa in each sample. Similar trends was observed when comparing the counts for *Juniperus* compared to counts for Cupressaceae in general (i.e. when summing counts with *Cupressus*), or for *Quercus ilex* compared to *Quercus* in general (i.e. summed with *Quercus deciduous*). For *Quercus ilex*, counts were nonetheless quite similar between both methods, even at the species level, but counts were underestimated in the automated compared to the manual method (Supporting Information Fig. S8).

3.3.2 Analysis of the pollen assemblages

3.3.2.1 Description of the Mediterranean anemophilous flora

Considered over all 5 monitored years and all 7 sampling sites (Fig. 1), the 15 target taxa represented around 76.7% of the identified pollen grains in the traps. More precisely, most abundant taxa corresponded to *Quercus* (15.9%), *Buxus* (6.8%), *Phillyrea* (7.9%, Fig. 8. a) and Cupressaceae (9.5%). Grapevine pollen was also significant (7.7%), as expected from the experimental design located near grapevine individuals, but its abundance was still relatively high compared to its previously mentioned low emissions (Turner and Brown, 2004). Finally, Poaceae and *Plantago* were also abundant (7.5% including both taxa), more characteristic of open areas, and of more anthropized area for Plantago.

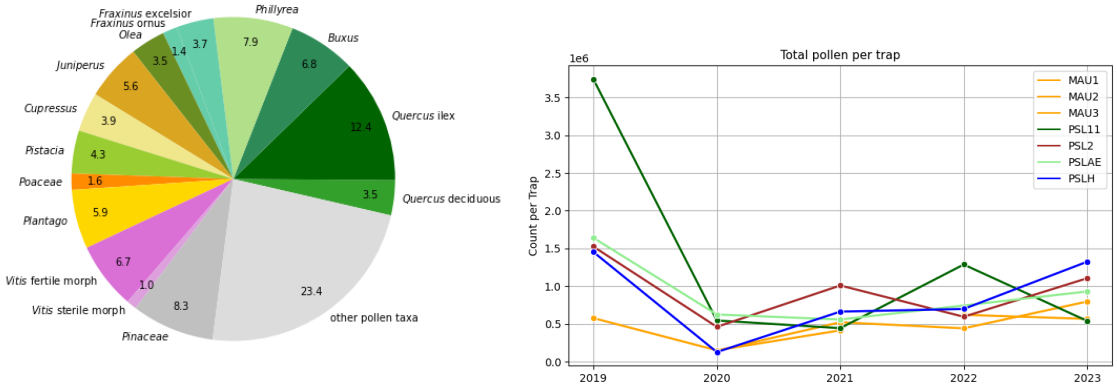


Figure 8: Pollen counts (a) for each of the 15 target taxa with all other determined taxa merged into the class 'other', measured over the whole monitoring sampling sites and years, and (b) for all pollen grains regardless of taxa, collected for each sampling site and year.

Considering pollen counts at the scale of each sampling site, the composition of pollen from the 15 key target taxa displayed associations that aligned with the main vegetation types identified around the sampling sites, including close scrubland, open scrubland, Mediterranean meadow, and vineyard (Fig. 1). The analysis of correspondence distinguished the sampling sites located in the vineyard (MAU1, 2, 3) from the other sites (PSL2, PSLAE, PSL11 and PSLH, Fig. 1 and 9).

These results align with the taxa composition (associations and abundances) in the pollen counts from each sampling site. A first association of pollen taxa was identified for the

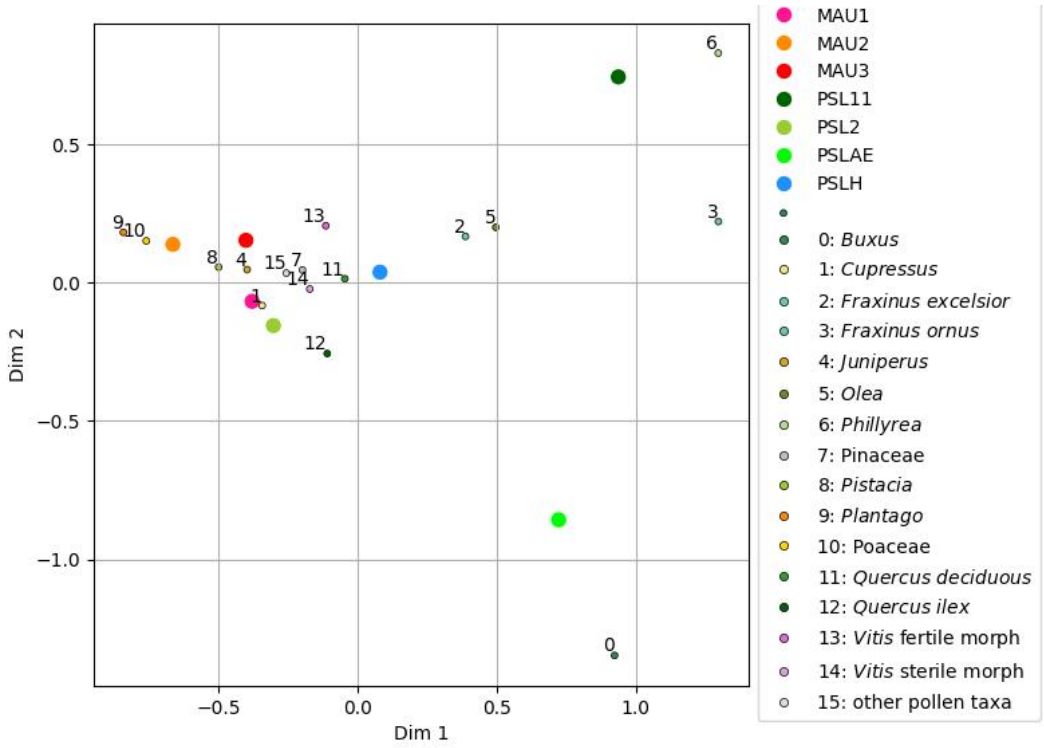


Figure 9: Analysis of correspondences per site including all target taxa and all classes of non-target pollen taxa grouped together into a class 'other pollen taxa'.

sampling sites MAU1, MAU2 and MAU3, located in the vineyard, and which collected a large proportion of Cupressaceae and herbs, and smaller proportions of pollen from trees. In the vineyard, *Plantago* was also largely present along with Poaceae, the presence of these annual or bi-annual herbaceous plants may occur among the grapevine rows, which are worked mechanically while not too intensively as planned by the farm. A second association of pollen taxa was typical of Mediterranean habitats (not cultivated), i.e. nearby the wild grapevine individuals, namely PSL2, PSLAE, PSL11 and PSLH. These sampling sites displayed larger proportions of pollen from trees and shrubs, including pollen from *Quercus*, *Phillyrea*, *Fraxinus* and *Buxus*. Nonetheless, these sampling sites displayed distinct dominant taxa, with a more pronounced presence of *Fraxinus* at PSLH, of *Phillyrea* at PSL11, of *Buxus* at PSLAE and of *Quercus ilex* at PSL2. These dominant taxa are consistent with the vegetation surrounding each sampling area. Indeed, the sampling site PSLH, with large amounts of *Fraxinus excelsior* pollen collected in the traps, is located near an intermittent riparian area; the large presence of *Buxus* at PSLAE also fits with field observations. The abundance of *Quercus ilex* at PSL2 is also consistent with the location of this trap, attached to a *Quercus ilex* trunk and in an area fully occupied by the oak forest. Finally, the traps located at the sampling site PSL11 collected large amounts of Oleaceae pollen grains, mainly *Phillyrea*; again this composition fits field observations of several *Phillyrea* shrubs in that sampling site. The smaller abundance of *Quercus* pollen in both PSLAE and PSL11 fits with the position of the traps, on the Pic Saint Loup massif, but at the edge of the forest opened by large scree slopes.
Pistacia and Poaceae were mostly abundant in the sampling sites characterized by a more open vegetation, including near the vineyard (MAU1, 2, 3) and the sampling site (PSLH), which is located on an intermittent stream separating an open scrubland and a Mediterranean meadow.

Vitis pollen was collected in all sampling sites. In the vineyard, much higher amounts of *Vitis* pollen was collected in the traps placed between the rows of grapevine (17.3%, MAU3) compared to those placed on the sides of the vineyard (2.6% in MAU2 and 4.4% at MAU1), demonstrating that grapevine pollen does not disperse much. Additionally, more pollen was collected in the trap located in the South-West side of the vineyard (MAU1) compared to the North-East side (MAU2), which also fits with the wind direction, which is mainly from the North side (Fig. 1). As expected, a significant amount of the *Vitis* sterile pollen grain was collected at the sampling site PSLH (inaperturate morphotype, 2.6%), i.e. near a functionally female individual of *Vitis sylvestris*, which predominantly produces this morphotype, compared to the other sampling sites. In the PSLH sampling site, the fertile *Vitis* pollen was also identified to similar abundances to that of the sterile morphotype (2.3%, Fig. 10). These grains are produced by male wild individuals and from the cultivated grapevine and might result from the dispersion from distant grapevine individuals, it is indeed expected as flowers from this functionally female individual are effectively fertilized each year. Surprisingly, the sterile *Vitis* pollen was not only identified in PSLH but also in other sampling sites, although in small abundances (<1%, Fig. 10); these counts could be partly due to false positives from the automated classification.

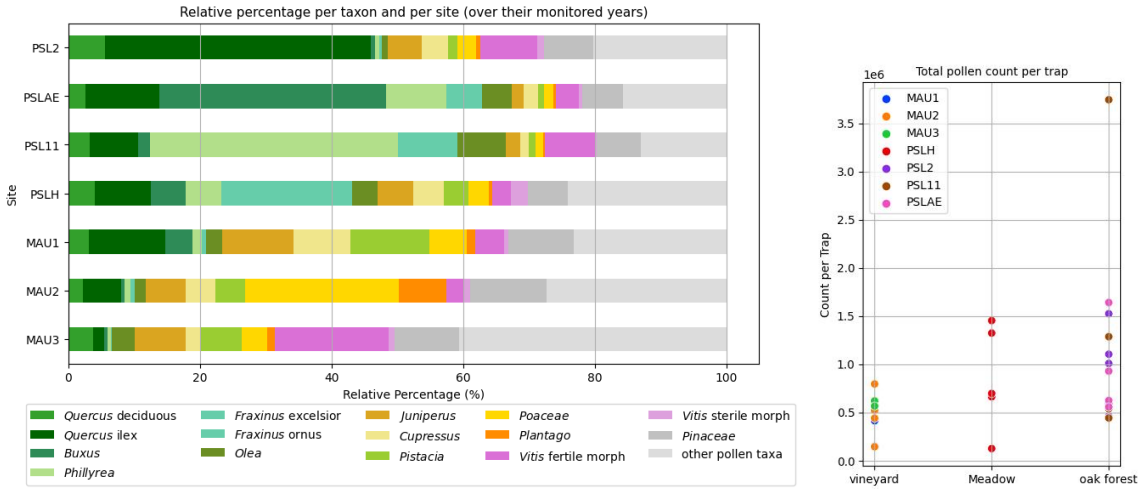


Figure 10: (a, left) Pollen composition of the main target taxa collected at each monitoring site and averaged over the monitored years; pollen grains identified to belong to a pollen taxa but not from a target pollen taxa are grouped into the category 'other pollen taxa'. (b, right) Pollen counts determined regardless of taxa over the whole study area and monitoring period.

The total annual pollen counts measured regardless of taxa and for each sampling sites were consistent with the pollen compositions previously observed. The amounts of collected pollen grains were higher in the locations dominated by trees compared to sampling sites with a more open vegetation, especially compared to the vineyard (Fig. 10.b). The traps placed at the edge of the meadow and the open scrubland (PSLH, Fig. 1) collected an intermediary amount of pollen over the years compared to the two other vegetation types.

3.3.2.2 Differences in pollen compositions and abundances throughout the 5 years of monitoring

Inter-annual differences in the pollen compositions and pollen counts were then evaluated for each sample, i.e. for each sampling location and year (Fig. 11; Supporting Information Fig. S9). The sampling sites characterized by a closed scrubland vegetation type (PSL11, PSLAE, PSL2), displayed much more inter-annual variation across years compared to the sampling sites from the vineyard (MAU1, MAU2, MAU3) and the open vegetation of PSLH (open scrubland and Mediterranean meadow, Fig. 11). The correspondence analysis indicated shifts in composition for PSL2, along a direction associated to *Quercus ilex*, suggesting an enrichment or decrease in this taxa. Shifts were also observed for PSL11 and PSLAE, and seemed to be associated to changes in abundances of *Buxus* and *Phillyrea*.
The counts for each of these taxa across sampling sites and monitoring years indicated that the shifts observed for PSLAE were mainly related to a large and fast decrease in *Buxus* abundances, which disappeared at that sampling location in three years (Supporting Information Fig. S9). Additionally, an increase in *Quercus* pollen was observed in all sampling sites in 2021 (except in PSL11), including for the sampling sites displaying a more open vegetation (MAU1, 2, 3 and PSLH; Supporting Information Fig. S9).

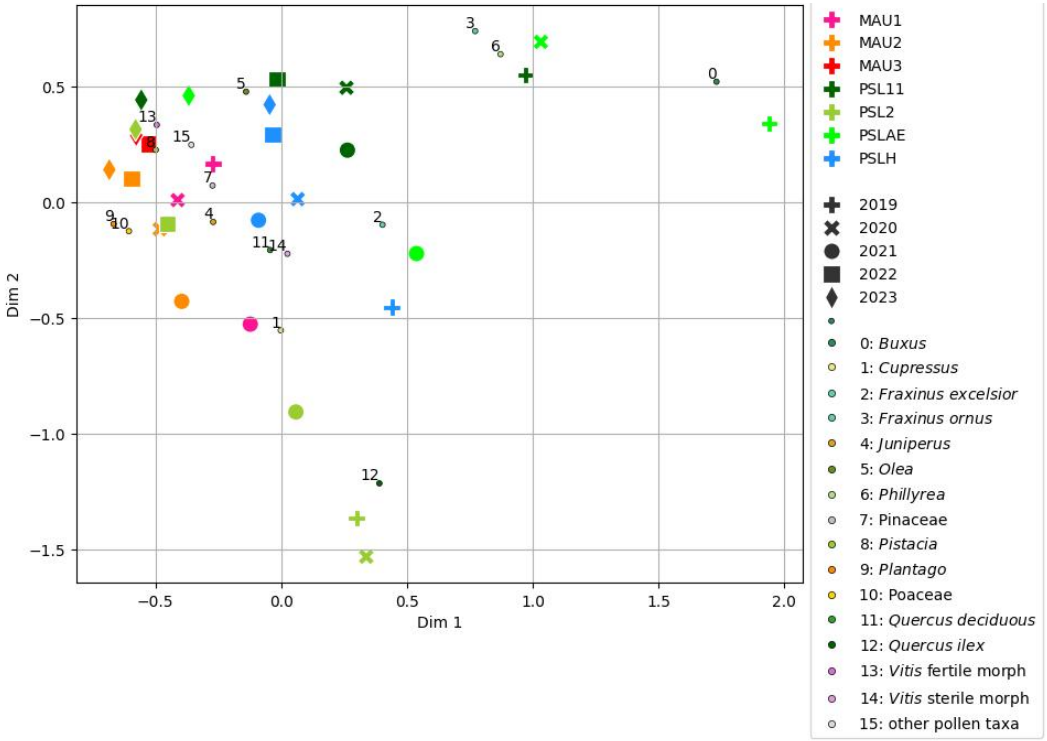


Figure 11: Analysis of correspondences per trap sample, i.e. per sampling location and year, and including all target taxa and all classes of non-target pollen taxa grouped together into a class 'other pollen taxa'.

The amount of pollen collected for each of the target taxa, tended to be much lower in 2020 compared to the other monitored years for the sampling sites PSL2 but also for MAU1, MAU2 (MAU3 not yet monitored), and PSLH, although their relative taxa associations tended to remain similar (Supporting Information Fig. S9). The total amount of pollen grains measured regardless of taxa, and collected in the traps in each sampling sites displayed an overall slight increasing trend from 2020 to 2023, but the

changes remained small (Fig. 8.b). Much larger amounts of pollen grains were collected in the samples from 2019, especially for PSL11, but it remained in the range of pollen counts observed in subsequent years for MAU1 (no data for MAU2, 3 in 2019) and PSLH (Fig. 8.b, Supporting Information Fig. S9).

3.4 Discussion

This study aimed to assess temporal and spatial changes of key Mediterranean pollen taxa. We developed an automated pollen analysis pipeline and applied it to 28 annual pollen samples collected to monitor pollen rains from Mediterranean anemophilous plants in Southern France. The automated pollen analysis pipeline includes a previously developed detection method (chapter III - Gimenez et al., 2024), completed by an identification method, presented in this study. This identification method relies on (1) a dataset of labeled pollen images compiled from pollen samples from the vegetation monitoring and from a collection to increase the number of images per taxon, and include taxon expected in the routine application, (2) the classification algorithm ResNet152, (3) an ensemble modeling strategy to aggregate predictions from cross-validated models.

3.4.1 Discussion on the automated identification method

Discussion related to the properties of our classification dataset

Training the models with images from the environmental samples proved an efficient strategy, as demonstrated by the highest performances reached for the classes that had more than 60% of images from the environmental samples (Poaceae, *Pinus*, *Buxus*, and *Lycopodium*). Including reference images helped getting good performances for a few taxa for which the number of images from the environmental samples were low (e.g. Amarantaceae and Ericaceae reached accuracies above 89% with only 81 and 63 environmental images), but not for all taxon: *Quercus deciduous* and Rosaceae reached F1-scores of 65.8% and 84.7% respectively, despite having classes with 1300 images. The aspect of the slides from the collection for these two taxa was of poor quality. Additionally, the class for *Quercus deciduous* was built using six slides displaying heterogeneous visual aspects. The consequent poor quality and discrepancies between images in the final classes could have contributed to classification errors, suggesting that adding reference images can generate confusion and thus misclassification. Note that the performance of the models could not be evaluated on environmental images for the many classes that had no or few environmental images, and despite large proportions of added reference images (32 classes had less than 50 environmental images and 26 included only references). Overall, achieving lower performance on environmental images from models trained largely on reference images for a class is consistent with a previous study in melissopalynology (Tsiknakis et al., 2022). In that study, models trained on a reference dataset achieved poor performances when applied to real-world honey samples, and despite high accuracy on the reference images (Tsiknakis et al., 2022). Selecting only reference images that display similar properties to images expected in the real-world application (e.g. in pollen transparency or overall shape), even at the expense of fewer images, could minimize the contribution of reference images to errors, and optimize their use. Note finally that although having large class sizes can improve classification performance, it is not always necessary, depending on the singularity of the pollen morphotype. For instance the class Cyperaceae only had 138 images (including 52

environmental images), but still reached 91±0.2% accuracy, likely due to its specificity in both texture and morphology (Supporting Information Table S2, Fig. S4 and S3).

Many more classes were used for training compared to the number of target taxa (83 vs 15). Such training dataset was built to prevent non-target or unknown particles to being assigned to target pollen taxa (i.e. false positives) during subsequent routine application, and because models can only predict classes used during training. The strategy of creating additional classes for debris and unknown particles (e.g. Battiato et al., 2020b, Boldeanu et al., 2021, Schaefer et al., 2021; Oteros et al., 2020; Crouzy et al., 2022), and for 'other' pollen taxa (e.g. Theuerkauf et al., 2023; Barnes et al., 2023) is now common in studies designed for routine application. In our study, the class of non-pollen proved effective in removing most of these particles previously and falsely detected in our pipeline, as indicated by its high recall values (93.5% and 99.6% with the top-1 and top-4 approaches respectively), and low misclassification with the target taxa (Fig. 6). This class of non-pollen particles was mostly misclassified with the class of indeterminable pollen 'covered', with 5.75% of covered pollen lost to the class of non-pollen (FN) and 3.45% of non-pollen particles incorrectly added to this pollen category (FP), based on the top-1 evaluation (Supporting Information Fig. S1 for FN, and Fig. S2 for FP). These errors are explained because when covered, pollen was usually covered by non-pollen particles in the images, but they might have led to under-estimation of pollen counts during the application. The classes for non-target pollen taxa and indeterminable pollen also generated misclassification errors for the target taxa, but errors did not exceed 1% in most cases (Fig. 6). Larger errors involved the classes *Quercus* misclassified with Rosaceae as previously described, and *Olea*, for which Brassicaceae pollen grains contributed to 5% of its identifications (Supporting Information Fig. S2). Overall, using classes for non-target taxa and indeterminable pollen grains (i.e. without visible identification criteria) can generate classification errors, but still contribute to alleviate errors for the target taxa in routine application. Additionally, using these classes allows to estimate errors that can occur in routine, and would otherwise go unnoticed if they were excluded from the model evaluation. A thorough evaluation of the trade-off between misidentifying a non-target taxon encountered during routine application as a target taxon, and of the opposite errors would still provide a better understanding of the most effective strategy. Alternatively, training the models exclusively on target taxa and using thresholds in the predicted scores to remove non-target particles such as debris has been previously used (Crouzy et al., 2016; Crouzy et al., 2022). Such approach nonetheless requires a calibration step to change the predicted scores to probabilities of correct predictions, as it is not initially the case for the ResNet algorithm (Guo et al., 2017).

Predictions were also affected by the class imbalance in the classification dataset, especially by the class imbalance among environmental images, and regardless of the total class sizes with the added references (Fig. 4). Classes that were underrepresented in the environmental dataset tended to have lower recall (i.e. they were often misclassified into other classes) but higher precision (i.e. other classes were rarely misclassified into them). For example, *Platanus* (70 environmental images, among 609) and *Rhamnus* (96 environmental images among 365) achieved precision of 98.33% and 90.5% respectively, but recall rates below 77% (Supporting Information Table S1). For the target taxa, the amounts of misclassification cases occurring among families with morphologically similar pollen were also drawn by the class sizes in the environmental dataset. For instance, *Phillyrea* tended to be over-estimated at the expense of *Fraxinus ornus*, reflecting their class size differences. In contrast, the numerous misclassifications between *Cupressus* and

Juniperus, which had similar class sizes, balanced each other and led to similar predicted and true distributions in the test data (Fig. 6, Table 2). The class imbalance in the environmental dataset was representative of the natural distribution of the pollen taxa in the annual samples from the study area. Taxa that were scarce in the annual samples were thus poorly represented in the environmental dataset. Our analysis focused on key Mediterranean taxa, which were abundant taxa in the environmental dataset (except for *Fraxinus ornus* and *Vitis* sterile), so that their performances were still expected high. To enhance the performance of classes poorly represented in the classification datasets, post-hoc calibrations can be conducted by adjusting classification confidence scores for instance with re-weighting (Olsson et al., 2021; Punyasena et al., 2022). However, these approaches still aim for one specific distribution (whether balanced or not), which might not always reflect taxa distributions encountered during routine application, as in this study with sampling sites located in distinct vegetation types. This issue of distinct distributions between the training/testing data and the real-world application data is inherent to deep learning approaches. In our study, the class size distribution in the training data may have partly homogenized the pollen counts between samples from distinct sampling locations or years during application, as models were driven by the same class distribution. The automated pipeline still generated distinct pollen taxa abundances between samples collected in sites with distinct vegetation types, suggesting this issue did not prevent from capturing large differences in the samples contents. However, it might have prevented from capturing more detailed information in the samples compositions, especially for taxa that can be abundant in one sampling site and absent in another. For instance, many *Phillyrea* grains were counted in the sampling site near the intermittent riparian area (PSLH), where *Fraxinus* was dominant and *Phillyrea* expected rare or absent. This could be explained by the misclassification errors between both taxa and because the abundance of *Phillyrea* in the training dataset might have led the models to predict it more frequently. Also, *Vitis* sterile pollen was expected in the sampling site with the wild female individual, in which 2.6% was counted (PSLH, Fig. 10), but its presence in the other sampling sites was unexpected while it was counted in small proportions by the automated pipeline (<1%). These counts could include false positives driven by the large class size for *Vitis* sterile in the training dataset. Techniques referred as 'domain adaptation' are being developed to handle this issue of differences between training/testing data and application data, e.g. for medical imaging (Guan and Liu, 2022; Farahani et al., 2021; Kumari and Singh, 2024). Investigating these techniques could improve the developed pollen classification in this study, for instance, to obtain a method that can better comprehend small differences in the pollen compositions between the sampling sites and years.

Finally, our dataset still remained small compared to recommendations suggested for automated aero-palynology (Tummon, Bruffaerts, et al., 2024): 1300 images per class compared to a minimum suggested of 5000 pollen images per classes, and even 10 000 images for classes difficult to identify. Nonetheless, using transfer learning based on the ImageNet dataset partly alleviated this issue and allowed to achieve high performances for the selected target taxa.

Discussion related to the integrated identification pipeline

Several ensemble modeling strategies exist (Mohammed and Kora, 2023). The average voting (or soft voting) consists in averaging the scores obtained from several models, and assigning to an image the class with the highest average score. In contrast, the max voting strategy, or hard voting (Kim et al., 2003; Mohammed and Kora, 2023), used in

this study, relies on the occurrence of the classes predicted by each of the models, and assigns to an image the class with the highest occurrence. Both the average voting and max voting strategies were compared in a previous study, combining predictions from models with distinct architectures, including Inception v3 (Szegedy et al., 2015), Inception–ResNet (Szegedy et al., 2017), Xception (Chollet, 2015) and ResNet (He et al., 2015). This study showed similar performances with 97.5% accuracy achieved with both strategies applied to a reference pollen dataset (Tsiknakis et al., 2022). Here, we used a max voting strategy complemented with the top-4 approach and the use of thresholds in the occurrence of the classes predicted by 10 cross-validated ResNet152 models. The max voting strategy was selected as it allowed to use thresholds to account for uncertainty, shown by the absence of the 'most likely' class among the top-4 predictions of some models, and assign the corresponding images to a 'doubtful' category. In a preliminary test, we used a fractional assignment for each of the top-1 predictions from each of the 10 models, but this approach generated false positives for some taxa, based on our expectations for the samples locations. The max voting approach combined with a threshold to only keep predictions that occurred in many models allowed to prevent partially these errors. Classes with low recall rates with the top-4 approach, mainly including classes under-represented in the environmental dataset (Fig. 2), might be predicted as 'doubtful' more frequently, further enhancing their under-estimation. Nonetheless, this issue was considered preferable to indicating a false presence of a taxa in a sampling location, and may have not affected the counts for the target taxa considered in this study, as they are well represented in the environmental dataset. Alternatively, to account for uncertainty in predictions, approaches with an architectural pipeline relying on several CNNs and decision trees have also been investigated to differentiate classes that are predicted with confidence from other predictions (Bourel et al., 2020; Durand et al., 2024).

The sizes and ratio (oblong or round) of pollen grains are major criteria for pollen identification (Reille and Pons, 1990; Punt et al., 2007), but classification CNNs are implemented on images transformed to squares of standard sizes, thus distorting the content in rectangular images and eliminating the image size factor. Few studies have addressed this issue, for instance a study maintained the relative size of the grains and their ratio by integrating the pollen images keeping their relative size in large black images (Sevillano et al., 2020). However this approach is constrained by the size of the largest pollen grain, and therefore may lead to low image resolution for pollen taxa of small sizes. In our study, to prevent the classification algorithm to distort the image contents when re-framing them to squares, a black padding was added, either vertically or horizontally, to generate square images. This approach prevented distortion in the grains, especially for cut grains that occur in very thin images, and also accounted for the ratio of the grains by preserving it. The size parameter was however not considered due to the model resizing process (to 224 pixels × 224 pixels for each image, regardless of its initial size). Using the size parameter as a numeric meta-data into our identification pipeline could prevent few misclassification cases, e.g. ∼1% misclassification between *Buxus* (35 µm) and *Phillyrea* (20 µm). The size numeric value could be used as a separate channel in the CNN, as a multi-channel classification. It could also be used through post-processing filters to remove unlikely predictions or resolve ambiguities. For instance when processing the 40 predicted classes, if two classes are predicted with equal frequencies but have distinct expected sizes.

3.4.2 Relative abundances over the whole study area

The automated pollen analysis generated pollen counts which were consistent with the distinct main types of vegetation around the sampling sites (Fig. 1; Fig. 8.a). *Quercus* represented 15.9% of the total pollen collected, in agreement with values estimated from a Mediterranean oak forest in a study conducted in Northern Spain (14%, Fernandez-Gonzalez et al., 2020).

The amount of *Vitis* pollen found in the samples was relatively high compared to expectations (c. 7%), as it was previously mentioned to be very low in pollen assemblages, even when collected locally, based on soil measurements (Turner and Brown, 2004). A difference was observed between the amount of pollen collected at the center of the vineyard compared to the sides (17% at the center compared to 3% and 4% at the sides), suggesting that grapevine pollen has limited dispersal, as found by others (Turner and Brown, 2004), and thus that it is very difficult to get its signal event. The amount of the grapevine pollen collected very locally and in a *in situ* context, i.e. in the middle of the vineyard thus surrounded by many *Vitis vinifera* individuals, or directly under a single *Vitis sylvestris* individual growing on a tree, was also much higher in the case of the cultivated compared to the wild grapevine (17% compared to less than 8.7%, Fig. 10). Additionally, the overall ratio between the sterile and the fertile morphotypes of the grapevine pollen presented a very distinct pattern near a wild female individual (similar amounts between both morphotypes of around 2.5% each), compared to near hermaphroditic cultivated and wild male individuals with scarce sterile pollen (<1%) and abundant sterile pollen (17% in the middle of the vineyard and >3.4% near wild male individuals). These differences, both in the total amount of fertile grapevine pollen, and of the ratio between both morphotypes, could be used as a baseline to characterize the local presence of a wild grapevine population or of a vineyard. These results are also consistent with a previous study suggesting that the presence of the inaperturate grapevine pollen morphotype could be an indicator of the local presence of a wild population in an archaeological context (Mercuri et al., 2021). The results from this pollen rain monitoring thus complement this suggestion with modern data. Gravimetric traps were used in the protocol from this monitoring. These samples collect more debris compared to volumetric traps such as Hirst-type traps that are more frequently used in recent studies from vegetation monitoring. Nonetheless, gravimetric traps have been showed to provide similar results to those from volumetric traps (Boullayali et al., 2024), and also reproduce better conditions for pollen sedimentation processes, encountered for fossil cores in paleoecology (Pardoe et al., 2010; Latorre et al., 2008; Volkova and Severova, 2017). As previously mentioned, the sterile grapevine pollen morphotype could thus help the interpretation of data in paleo-ecology, although it is rarely known, or considered by palynologists (Mercuri et al., 2021).

The pollen rain monitoring studied here only covers 5 years and has yet to compile long-time series of pollen data, as required for analyzing general trends in pollen rains and relate them to environmental parameters or disturbances (Lindenmayer and Likens, 2010; Cristofolini et al., 2024). Conducting this study during a longer time period, will provide a better insight on inter-annual trends. Some changes could nonetheless already be detected. Pollen of *Buxus* disappeared in 3 years in the sampling site PSLAE, most probably because of damages from *Cydalima perspectalis*, and as observed in the field. Increases of *Quercus ilex* pollen in some years, especially in 2019 but also 2021, could also be indicators of the period for masting, or of indicators of increased temperature or drought (Fernández-Martínez et al., 2012). Additionally, this study showed more

significant inter-annual changes in the sampling sites surrounding by the oak forest, compared to the others with a more open vegetation type.

Finally, note that this data will also be complemented by samples collected in the same locations with a monthly and a weekly resolution. This data will provide refined pollen counts, and thus information on phenologies. From a methodological point of view, they will also allow to compare pollen signals collected from small or large temporal resolutions, for instance when aggregating monthly results over the years to compare with the pollen counts generated from the annual traps.

Supporting Information (SI)

Methods S1: Methodology and results for the selection of the best value for the top-i approach

Principle of the top-i approach

In this approach, referred as top-i, we considered several highest scoring classes generated by a model applied to a given image, i.e. the list of i classes predicted by the model with the highest scores. In that approach, a prediction was considered a true positive if the ground truth class was among these top-i predictions (chapter I). Performance metrics were measured several times with this approach and using i of increasing value. This evaluation aimed to determine an optimal i value allowing to capture the correct class among the selected predictions while minimizing the number of predicted classes to keep for an image. This selection was done considering the objective of getting an automated method that would remain efficient under routine application conditions. This approach is standard when evaluating performances of deep learning algorithms (e.g. He et al., 2015) and informs on the likelihood of finding the correct class depending on the number of highest predictions considered.

Results and decision rules

Performances were evaluated considering a prediction as correct if the true class was among the i highest-scoring classes, measured with i of increasing value. For this evaluation, only the environmental images were considered. Accuracy was first measured over all the environmental images regardless of taxa (microaccuracy), which allowed to also include images from classes with small numbers of environmental images. The accuracy increased by 7.48% when considering the top-4 compared to the top-1 predictions (99.51% vs 92.04% accuracy), i.e. the top-4 predictions captured the correct class for 99.51% of the images, considered regardless of their classes. This increase in performance started to plateau after the top-4 predictions, with an increase by 0.56% between the top-3 and the top-4 predictions compared to an increase by only 0.18% between the top-4 and the top-5 predictions (Supporting Information Methods Fig. M1, Table M1). Based on this evaluation, keeping the top-4 predictions was considered as an optimal compromise to capture the correct class among the highest-scoring classes predicted for each image while limiting the number of predictions to keep, and thus to be processed by the automated pipeline.

This top-4 selection evaluated on the test set images, allowed to capture the correct class for 97.78% to 100% of the images for all 15 target pollen taxa, including *Quercus* (*ilex* and *deciduous*), *Buxus*, *Fraxinus* (*excelsior* and *ornus*), *Olea*, *Phillyrea*, *Vitis* (sterile and fertile morphotypes), *Juniperus*, *Cupressus*, Poaceae, *Pistacia*, *Plantago* and Pinaceae, as well as *Lycopodium*, covered and blurry pollen grains, non-pollen particles and the class grouping all pollen taxa not belonging to the target taxa (Supporting Information Methods Fig. M2, Table 2). The top-4 predictions were selected for the identification pipeline.

Evolution of performance metrics using the top-i approach with i of increasing values when measured using all 83 classes are also presented, when including only environmental

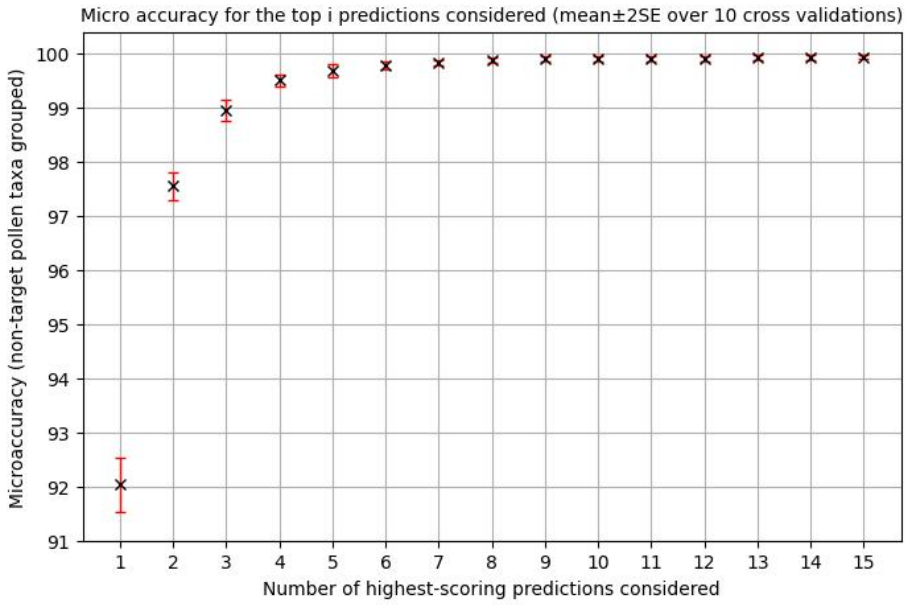


Figure M1: Performances per highest-scoring predictions, as the microaccuracy, and measured over the environmental images regardless of classes (mean±SE over 10 cross-validations)

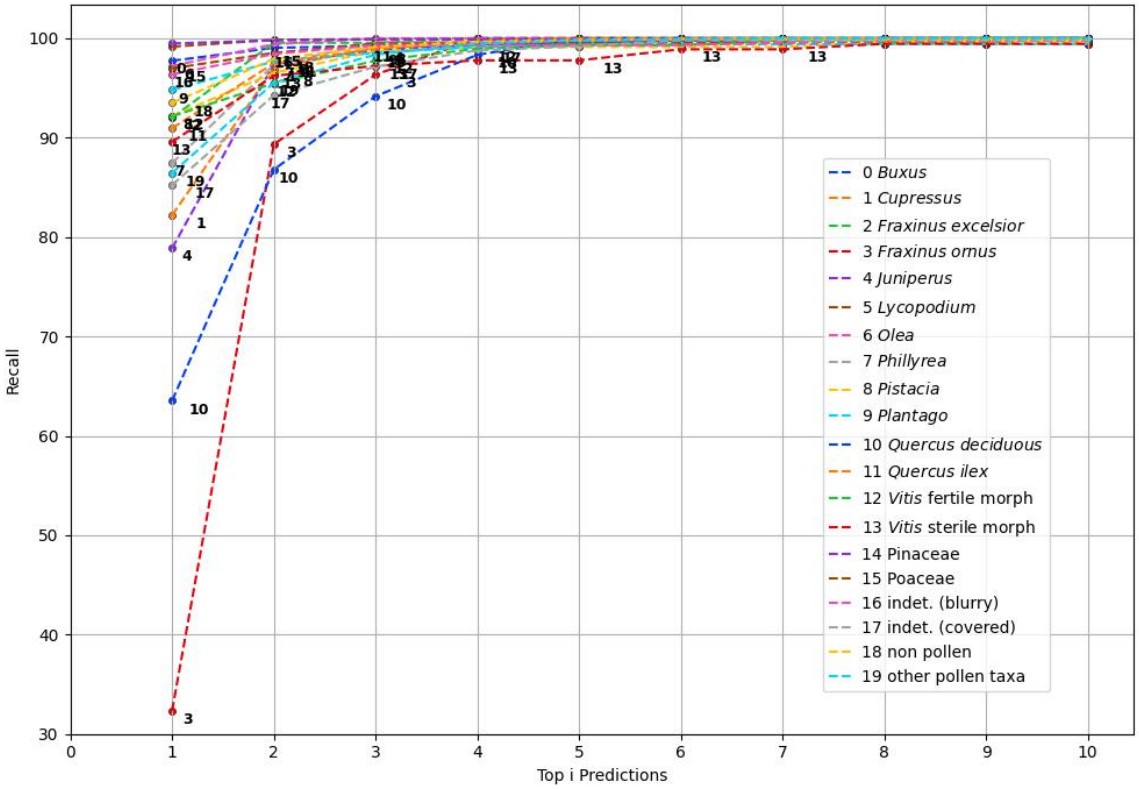


Figure M2: Performances per highest-scoring predictions, as the recall rates measured for each of the 20 main classes separately. Predictions for non-target pollen taxa are grouped into the class 'other pollen taxa'.

images and measuring the metrics over all images regardless of classes (Table M1), or for each class specifically (Fig. M3).

Table M1: Performances per highest-scoring predictions (top-i approach), as the microaccuracy measured over all environmental images and considering all 83 classes when defining a correct or false prediction (mean ±SE over 10 cross-validations).

top-i predictions	**number of images**	**microaccuracy**
1	1193.4±1.3266	90.9915±0.2115
2	1193.4±1.3266	96.7403±0.1325
3	1193.4±1.3266	98.2151±0.0695
4	1193.4±1.3266	98.8772±0.075
5	1193.4±1.3266	99.1284±0.0763
6	1193.4±1.3266	99.363±0.0699
7	1193.4±1.3266	99.4805±0.0671
8	1193.4±1.3266	99.5978±0.0596
9	1193.4±1.3266	99.6564±0.0522
10	1193.4±1.3266	99.6983±0.0454
11	1193.4±1.3266	99.6983±0.0454
12	1193.4±1.3266	99.7318±0.041
13	1193.4±1.3266	99.7905±0.0358
14	1193.4±1.3266	99.8325±0.0353
15	1193.4±1.3266	99.8492±0.0349

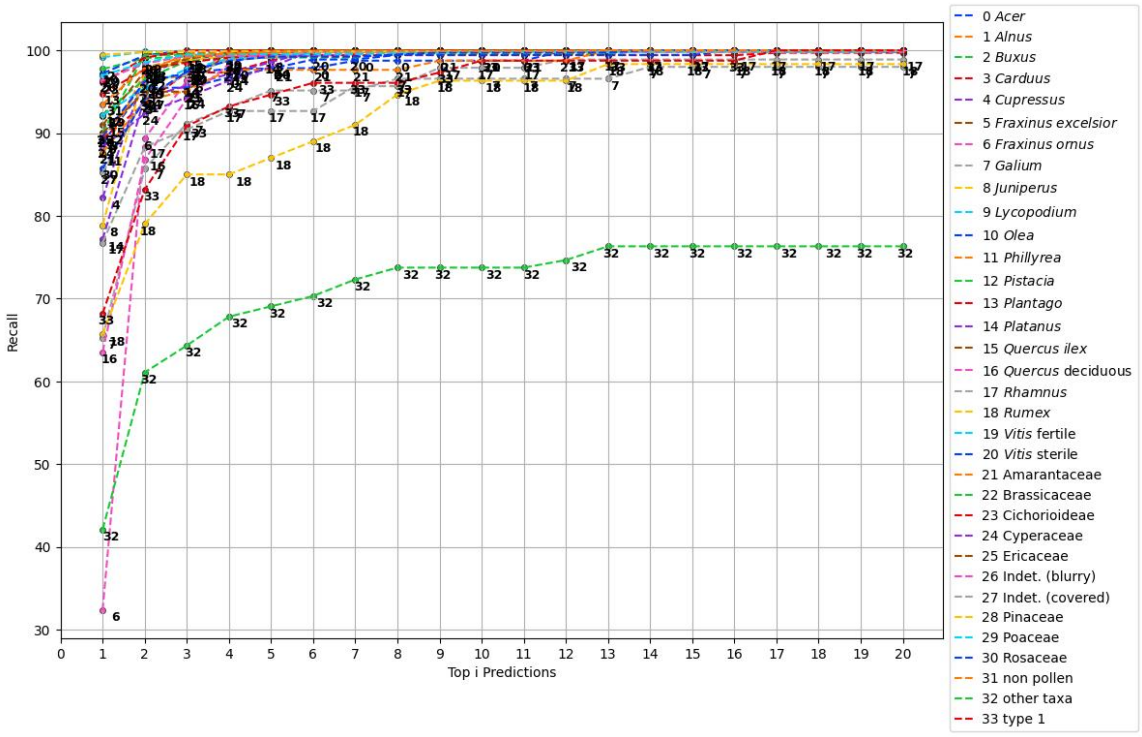


Figure M3: Performances per highest-scoring predictions (top-i approach), as the recall rates measured over all environmental images, per class separately, and considering all 83 classes when defining a correct or false prediction (mean ±SE over 10 cross-validations). The reference number for classes reaching recall rates higher than 99% are not displayed to avoid excessive overlap. Classes with fewer than 5 images per cross-validated test set are excluded from the plot, due to insufficient numbers to correctly measure the metric. The classes in the legend are arranged from highest to lowest recall rates obtained with the top-4 predictions.

Methods S2 : Processing 40 classes (10 models × top-4 predictions) predicted for an image and assigning one final class

Principle of the method

When considering the routine application on new annual trap samples collected for the vegetation monitoring, 10 trained models were available and tailored for the identification in this context. Applying all 10 models with the top-4 approach on each new image generated a distribution of 40 predictions per image, where each class could be predicted up to 10 times. The objective was to design a strategy to aggregate these 40 predictions obtained for an image and assign to the image the most likely prediction.

To determine this strategy, the 40 predictions were first generated for all the images obtained from the detection and segmentation pipeline applied to all the 51 scanned samples from the vegetation monitoring of this study. The distribution of these 40 predicted classes obtained for each new image (10 models × top-4 classes) were then analyzed considering all images, at the exception of those used in the classification dataset or determined as 'edges' based on their dimension (see sections 2.3.3 and 2.3.4), and totaling 284 270 new images.

These distributions were analyzed by considering (a, Fig. M4.a) the number of distinct classes predicted for each image; (b, Fig. M4.b) the frequency of each class, which indicated how similar the top-i predictions were across models; and (c, Fig. M4.c) the frequency of the most frequent class, showing whether models consistently agreed on a single class (if it was predicted 10 times). This analysis helped to determine the best strategy to combine predictions and generate the most likely prediction while filtering out unreliable, or uncertain, predictions. All 83 classes were kept in this step when considering the frequency of each class among the predictions.

Results and decision rules

The distribution of the 40 predicted classes for each new image was evaluated on 284 270 new images. On average, there were 9.25±0.004 distinct classes predicted among the 40 predictions (Fig. M4.a), with each class appearing 4.32±0.002 times (Fig. M4.b). The most frequent classes out of the 40 predictions and regardless of their score ranks were predicted 10 times for 83.7% of the images. This result indicates that for most images, all 10 models agreed on at least one class, even if it was not predicted as the top-1 class in all models. This value reached 94.0% when also considering images for which the most frequent class was predicted 9 times (Fig. M4.c).

Based on these results (Fig. M4), we considered for an image classes that were predicted, or validated, by 10 or by 9 models out of 10 models. This threshold of 9 was determined to maximize the number of images for which a class would be assigned to, and as it was considered possible that one model out of 10 might miss the correct class. For example, *Vitis* sterile achieved 97.7% accuracy with the top-4 approach, meaning that for some images (2.3% in the test set), the correct class was not among the top-4 predictions of a model. Keeping also predictions generated by only 9 out of the 10 models thus allowed greater flexibility for images in this case.

Consequently, in the design of the identification pipeline, a predicted class was kept for an image if it had been predicted 9 or 10 times among the 40 classes. Sometimes, two classes were predicted in these same frequencies, the most likely class was then selected based on their respective sum of ranks, i.e. those ranking the highest among the top-4 predictions

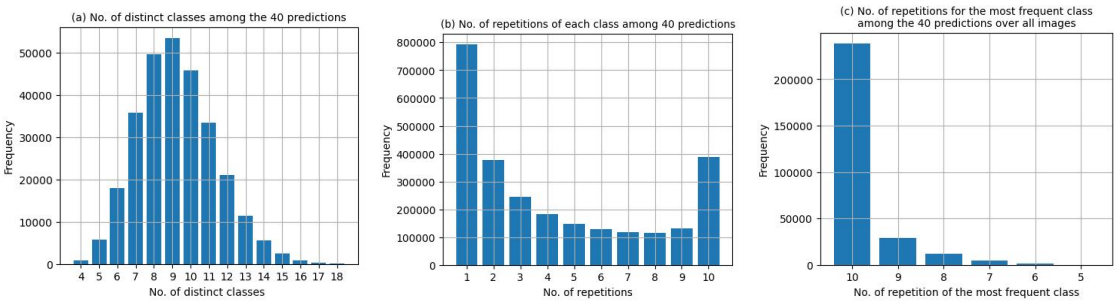


Figure M4: Distribution of the predictions generated for the images when considering 40 classes per image, as (a, left) the number of distinct classes generated among the 40 classes for each image; (b, middle) the number of repetition for each of these classes; (c, right) the number of repetition for the class predicted with the highest occurrence among the 40 classes for each image.

from all models (Fig. 5). When an image did not have any classes predicted at least 9 times among the 40 predicted classes, the image was labeled as 'doubtful'. In that case, if the 40 classes only included classes of pollen taxa (absence of predictions for Non-Pollen), the image was labeled as 'doubtful pollen taxa', otherwise it was labeled as 'doubtful particle' (Fig. 5).

Figure S1: Confusion matrix with recall rates

Confusion matrix showing the recall rates, i.e. percentages of grains correctly classified for each true class, over the number of images for the true class. Data are obtained from the top-1 predictions on the environmental images from the test sets.

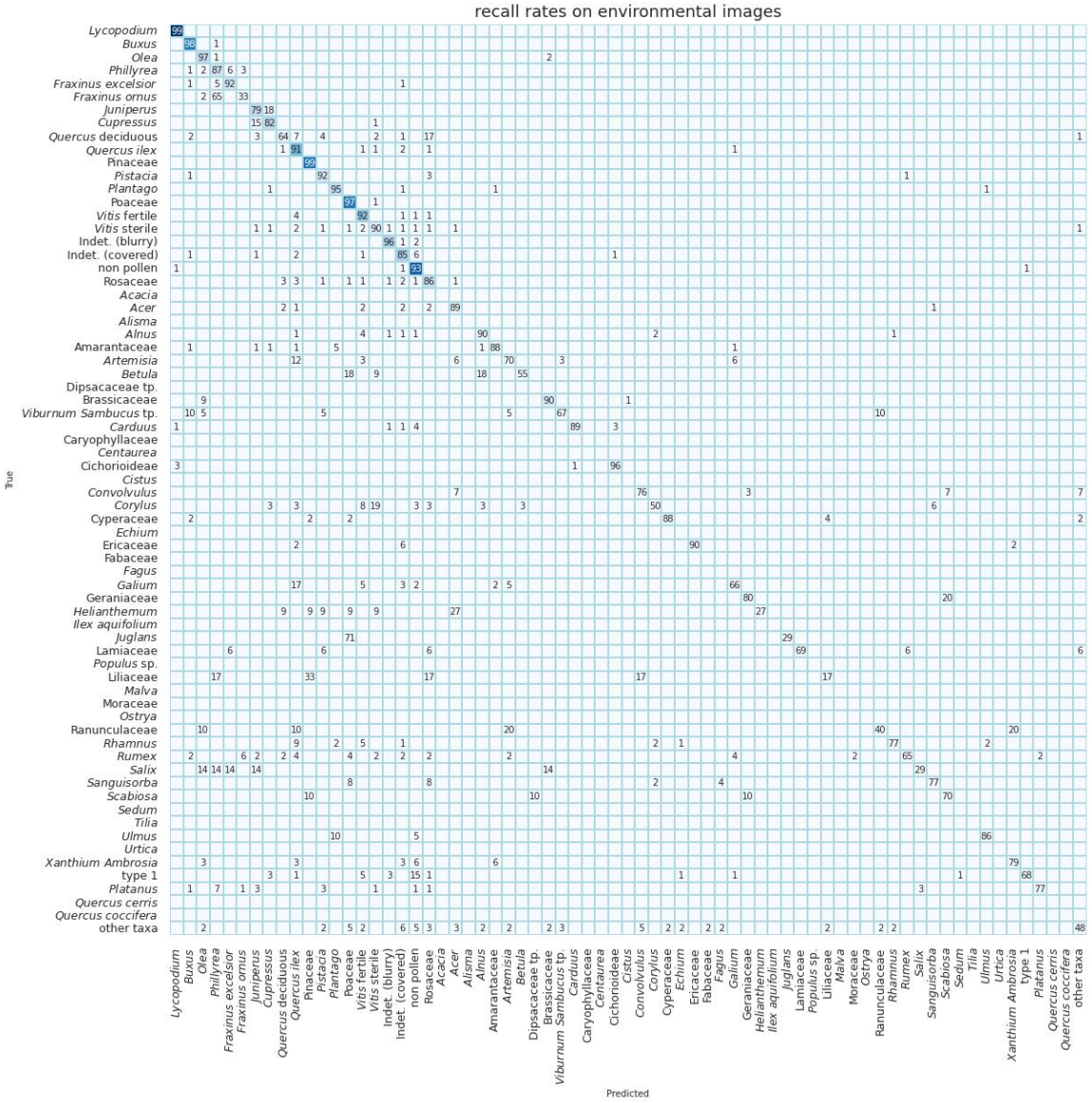


Figure S1

Figure S2: Confusion matrix with precision rates

Confusion matrix showing the precision rates, i.e. percentages of grains correctly classified for each true class, over the number of predicted images for the class. Data are obtained from the top-1 predictions on the environmental images from the test sets.

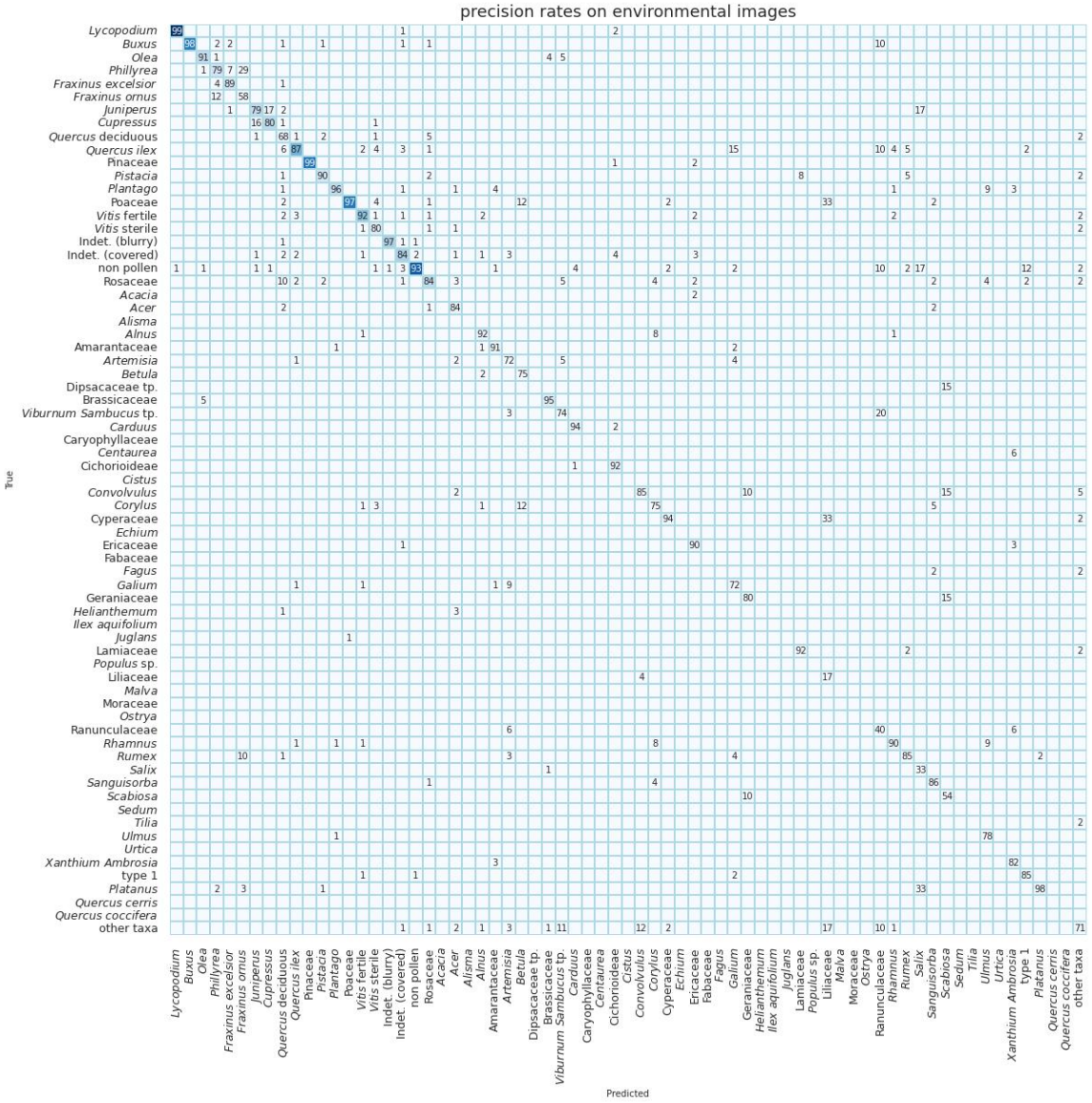


Figure S2

Figure S3: F1-scores evaluated on the environmental images

F1-scores, i.e. harmonic mean between precision and recall values, from the classification models applied on environmental images from the test sets, and as a function of the number of images in the test set (10% of the images used during model training). All classes used for training and including images from the environmental dataset are displayed. Note that metrics measured for classes with fewer than 50 images (i.e. around 5 testing images per cross-validation, below the red dotted line), might not be accurate due to insufficient numbers of images. Classes in the the legend are ordered from highest to lowest F1-scores.

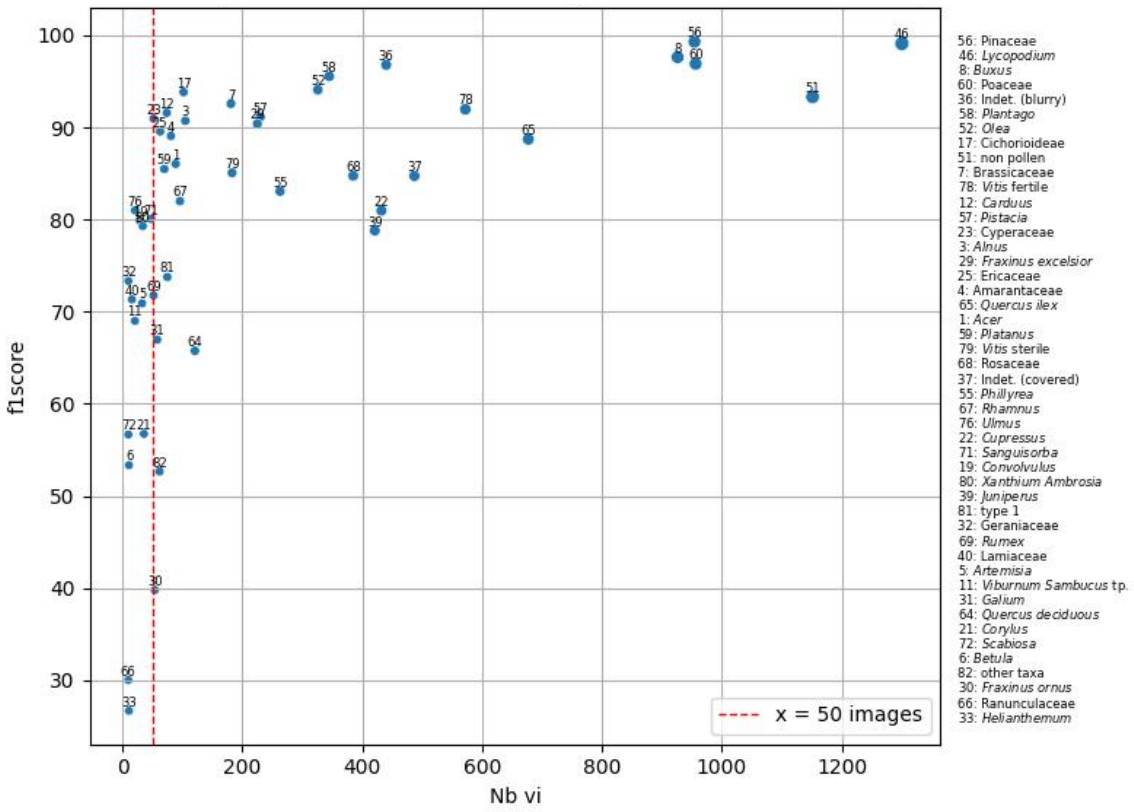


Figure S3

Figure S4 : Precision-Recall for environmental images from the test sets

This figure illustrates the precision-recall performance of the classification models applied on environmental images from the test sets.

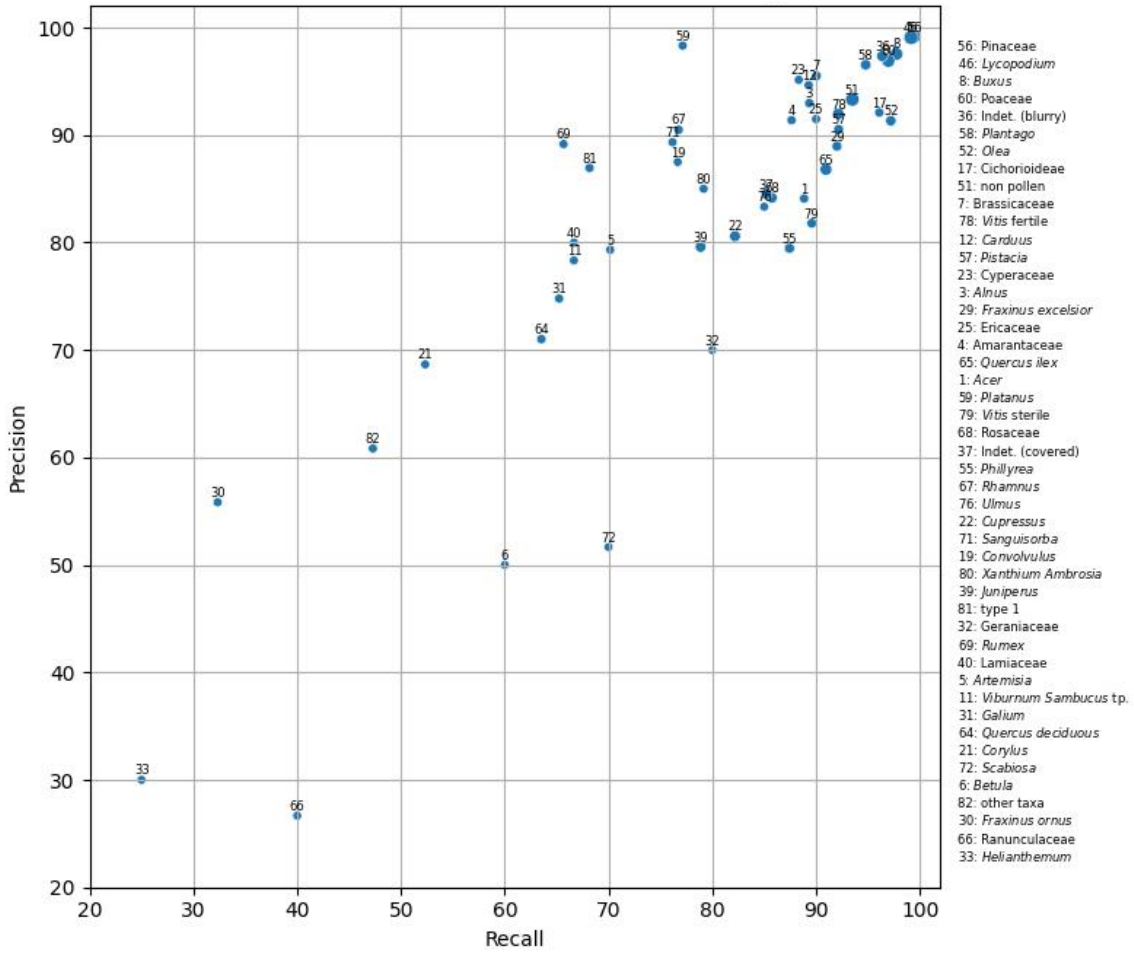


Figure S4

Figure S5: Summarized confusion matrix with a grouped category for all identified pollen grains

Confusion matrix obtained from the models applied on the environmental images from the test sets. All images identified to belong to a given pollen taxa are grouped into one class 'all pollen', and the other classes are kept, i.e. *Lycopodium*, pollen indeterminable blurry, pollen indeterminable covered and non pollen particles.

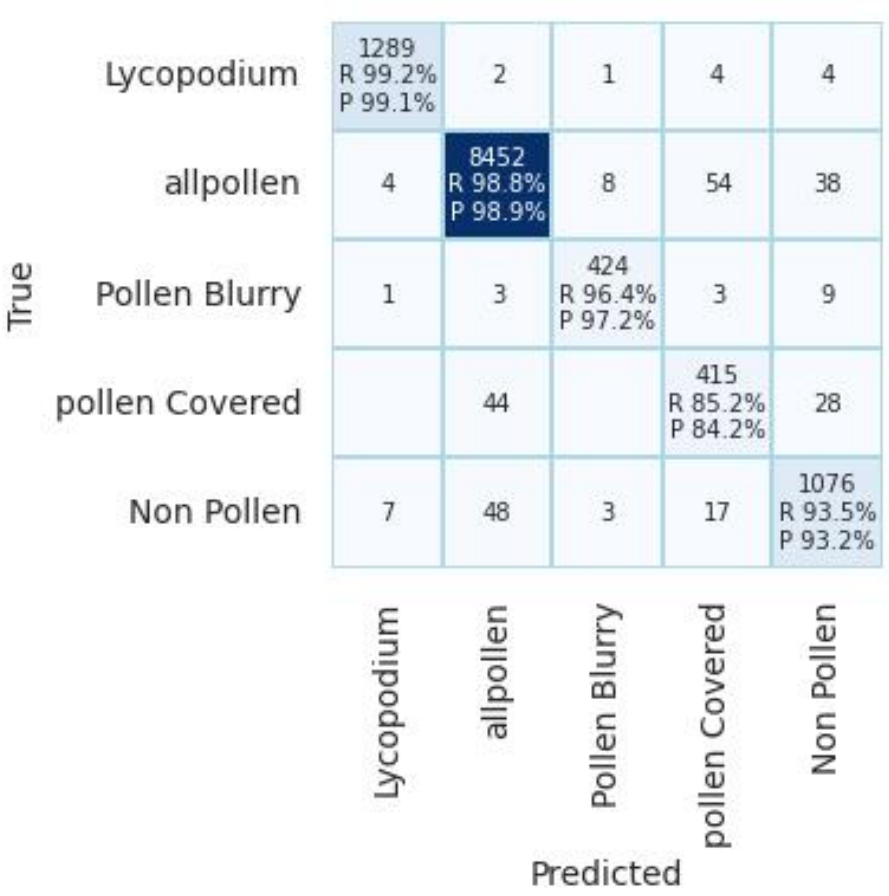


Figure S5

Figure S6 : F1-scores evaluated on the environmental and reference images

F1-scores, i.e. harmonic mean between precision and recall values, from the classification models applied on both environmental and reference images from the test sets, and as a function of the number of images in the test set (10% of the images used during model training). All 83 classes used for training are displayed. Classes in the the legend are ordered from highest to lowest F1-scores. Note that metrics measured for classes with fewer than 50 images (i.e. around 5 testing images per cross-validation, below the red dotted line), might not be accurate due to insufficient numbers of images.

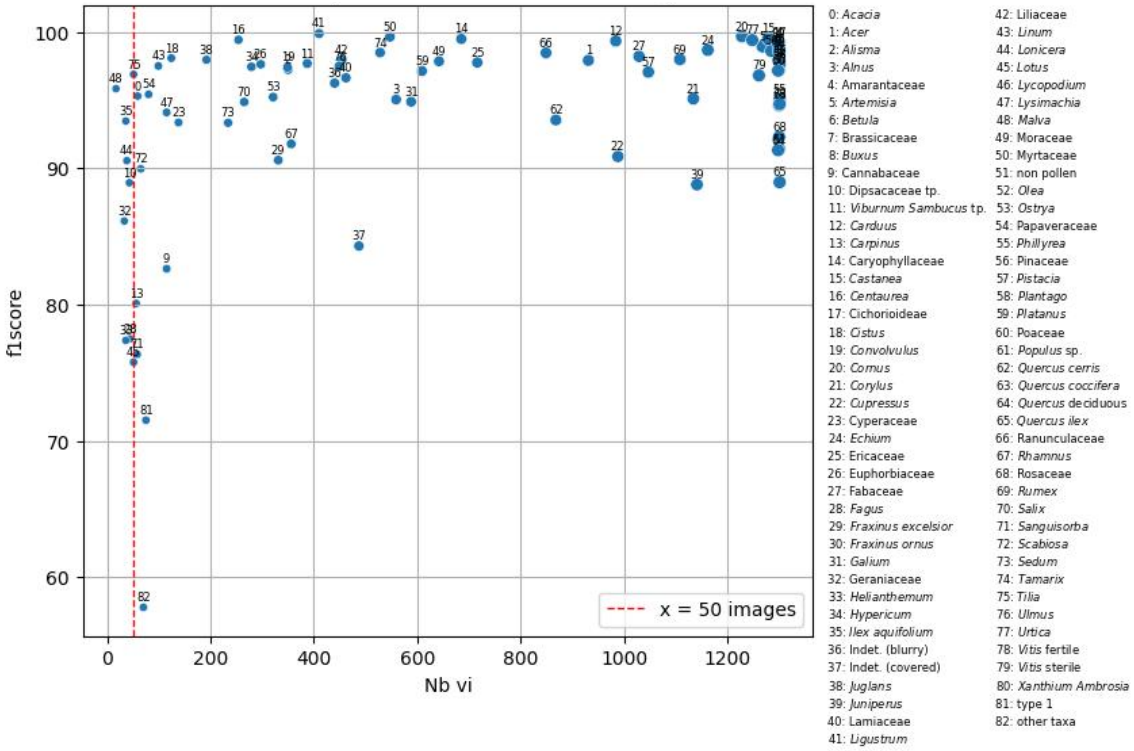


Figure S6

Figure S7: Confusion matrix with precision rates on environmental and reference images

Confusion matrix showing the precision rates, i.e. percentages of grains correctly classified for each true class, over the number of predicted images for the class. Data are obtained from the top-1 predictions on both environmental and reference images from the test sets.

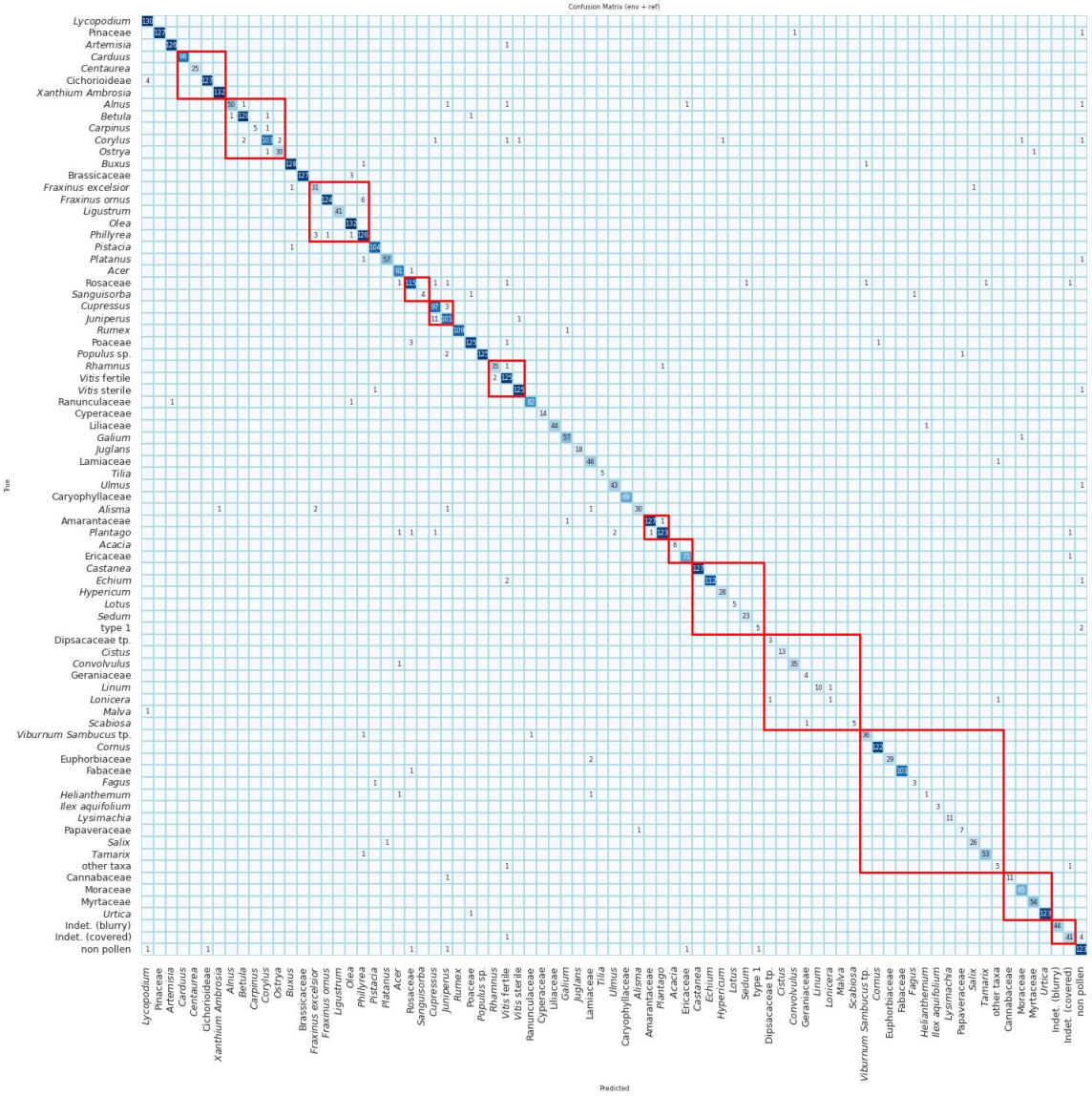


Figure S7

Figure S8 : Comparing manual and automated counts per taxa

Comparison of manual and automated counts per taxa. Taxa belonging to the same genus and misclassified with each other by the automated method are grouped together (Oleaceae includes *Olea*, *Fraxinus*, *Phillyrea*, Cupressaceae includes *Juniperus* and *Cupressus*).

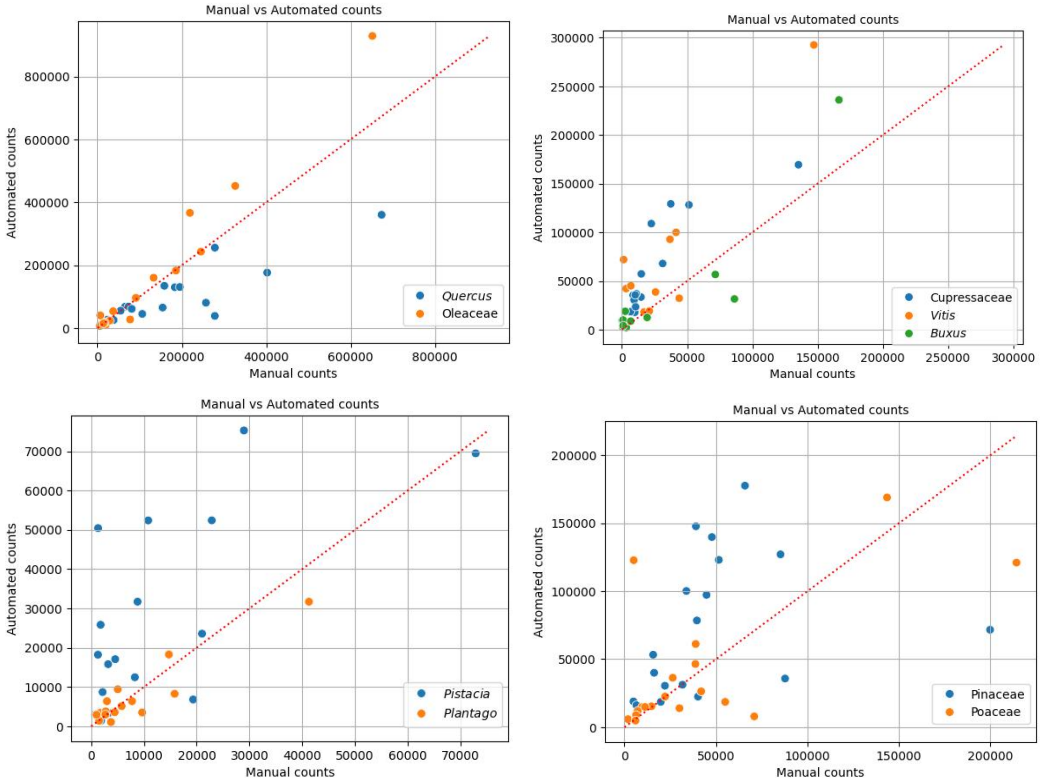


Figure S8

Figure S9 : Counts for each of the target taxa for each sample, i.e. for each sampling location and year

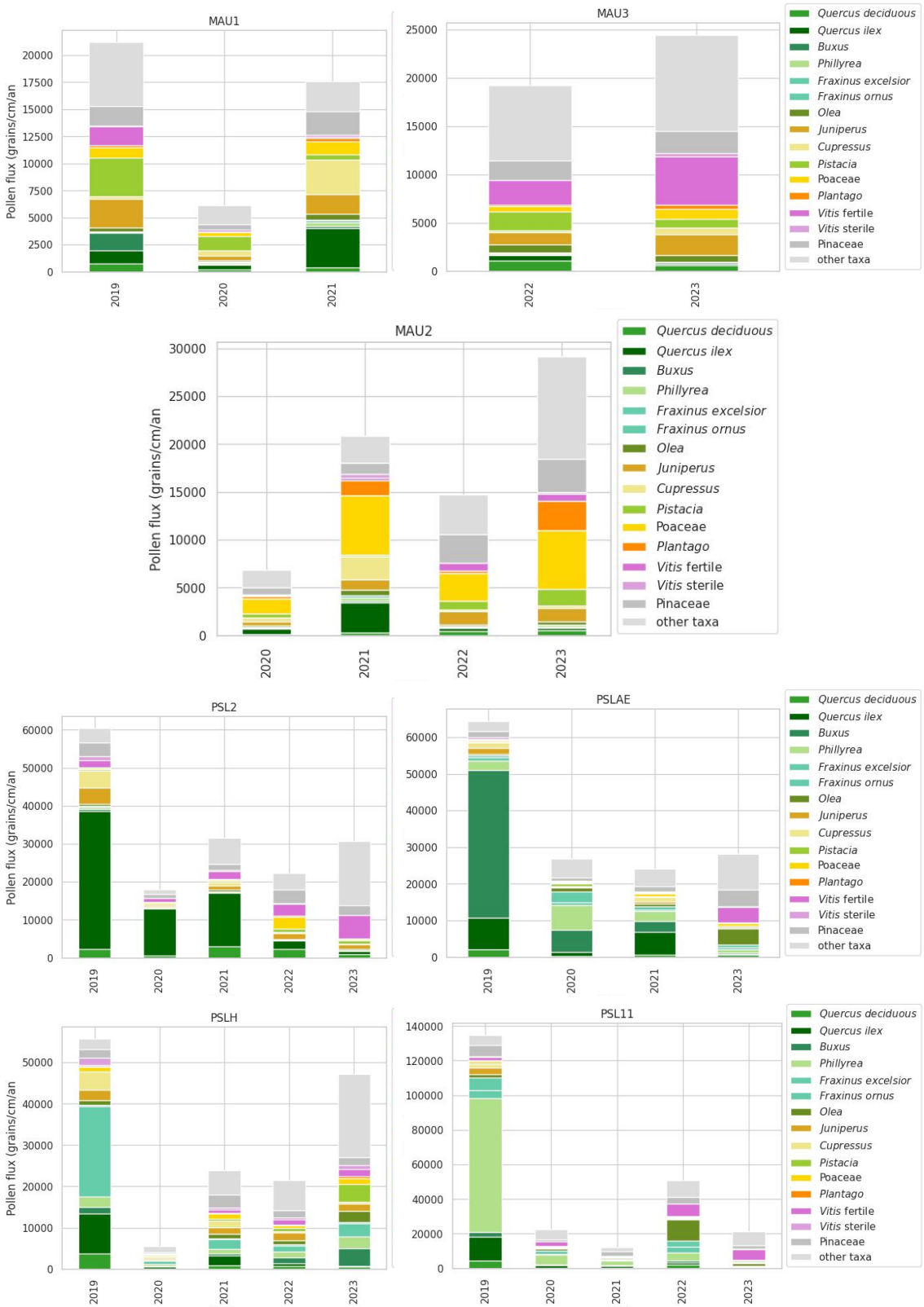


Figure S9

Table S1 : Metrics from the top-1 predictions obtained on environmental images for all 83 classes

Performance of the classification measured for al 83 classes separately, as the recall, precision and f1-scores, and only on the environmental images from the test sets. Values represent mean ±SE over 10 cross-validations. Classes are ordered from highest to lowest f1-scores. Many classes had an insufficient number of images for measuring performance metrics accurately. "Indet." refers to indeterminable pollen.

Table S1

Class	nb images	recall	precision	f1-scores
Pinaceae	95.4±0.2211	99.4737±0.2236	99.1677±0.8	99.3182±0.2906
Lycopodium	130.0±0.0	99.1538±0.3145	99.081±1.2	99.1155±0.2906
Buxus	92.6±0.2211	97.7348±0.348	97.5465±2.3	97.6356±0.4955
Poaceae	95.6±0.2211	96.9649±0.5467	96.9068±3.0	96.9193±0.6498
Indet. (blurry)	44.0±0.0	96.3636±0.5812	97.3527±1.2	96.7937±0.5333
Plantago	34.5±0.1667	94.7815±0.5925	96.5453±1.2	95.5593±0.2906
Olea	32.6±0.2667	97.2064±0.2769	91.3375±3.1	94.0882±0.5667
Cichorioideae	10.2±0.1333	96.0909±0.1633	92.1061±0.9	93.8398±0.2769
Non pollen	115.1±0.3145	93.4938±1.3017	93.3209±7.9	93.3226±1.3034
Brassicaceae	18.1±0.1	90.0292±0.2494	95.5154±0.8	92.5926±0.2494
Vitis fertile	57.2±0.1333	92.1476±0.7188	91.975±4.7	91.9721±0.7461
Carduus	7.4±0.1633	89.2857±0.2494	94.6429±0.4	91.5925±0.2211
Pistacia	23.0±0.0	92.1739±0.4899	90.5352±2.3	91.1397±0.4485
Cyperaceae	5.2±0.1333	88.3333±0.2211	95.1429±0.3	90.9701±0.2134
Alnus	10.5±0.2236	89.3485±0.2769	92.9786±0.8	90.7307±0.2494
Fraxinus excelsior	22.5±0.2236	92.0109±0.3887	88.966±2.6	90.398±0.4
Ericaceae	6.3±0.1528	90.0±0.3055	91.4881±0.6	89.5598±0.2211
Amarantaceae	8.1±0.1	87.6389±0.2981	91.3889±0.7	89.0836±0.2134
Quercus ilex	67.7±0.3958	90.9497±1.0376	86.8106±9.5	88.7236±1.0462
Acer	8.9±0.3145	88.8485±0.2582	84.0985±1.5	86.0549±0.2687
Platanus	7.0±0.0	77.1429±0.3712	98.3333±0.1	85.4872±0.1
Vitis sterile	18.3±0.2134	89.5848±0.2769	81.8032±4.0	85.0741±0.9888
Rosaceae	38.5±0.2236	85.7463±0.5627	84.1721±6.4	84.7492±0.8589
Indet. (covered)	48.7±0.2134	85.2323±0.6464	84.5087±7.8	84.7356±1.1035
Phillyrea	26.3±0.1528	87.4359±0.3958	79.4675±6.1	83.072±0.7219
Rhamnus	9.6±0.2211	76.7576±0.5121	90.4957±0.8	82.0022±0.2
Cupressus	43.2±0.1333	82.1882±0.9667	80.5937±9.0	80.9674±1.3904
Sanguisorba	4.8±0.3266	76.1667±0.3145	89.3333±0.6	80.0714±0.3055
Xanthium / Ambrosia tp	3.4±0.2211	79.1667±0.2603	85.0±0.6	79.2857±0.3055
Juniperus	42.1±0.1	78.8704±1.0588	79.5965±8.9	78.8121±1.2243
Ulmus	2.1±0.1	85.0±0.1528	83.3333±0.5	81.0±0.2236
Convolvulus	2.9±0.3145	76.6667±0.1528	87.5±0.4	79.881±0.2211
type 1	7.5±0.2236	68.1746±0.5617	86.9444±0.9	73.7647±0.348
Geraniaceae	1.0±0.0	80.0±0.1333	70.0±0.2	73.3333±0.1333
Rumex	5.2±0.1333	65.6667±0.4163	89.1667±0.6	71.7677±0.2667
Lamiaceae	1.6±0.2211	66.6667±0.1667	80.0±0.1	71.3333±0.1
Artemisia	3.3±0.2134	70.1667±0.2108	79.3333±0.9	70.9048±0.3145
Viburnum / Sambucus tp	2.1±0.1	66.6667±0.2134	78.3333±0.5	69.0±0.2236
Gallium	5.8±0.2906	65.2381±0.4714	74.7857±1.5	66.9607±0.3727

Continued on next page

Table S1 – continued from previous page

Class	nb images	recall	precision	f1-scores
Quercus deciduous	12.1±0.1	63.5256±0.5617	71.0145±3.6	65.7531±0.8718
Corylus	3.6±0.2211	52.3333±0.4422	68.6667±0.6	56.7381±0.2667
Scabiosa	1.0±0.0	70.0±0.1528	51.6667±0.6	56.6667±0.2667
Betula	1.1±0.1	60.0±0.2236	50.0±0.2	53.3333±0.1333
other	6.2±0.6633	47.2911±0.6289	60.8333±1.2	52.6517±0.2906
Fraxinus ornus	5.4±0.1633	32.3333±0.3712	55.8333±1.3	39.7857±0.335
Ranunculaceae	1.0±0.0	40.0±0.1633	26.6667±0.6	30.0±0.2667
Helianthemum	1.1±0.1	25.0±0.1333	30.0±0.1	26.6667±0.1
Juglans	0.7±0.2603	20.0±0.2687	20.0±0.0	20.0±0.0
Salix	0.7±0.1528	20.0±0.1667	20.0±0.4	20.0±0.2211
Dipsacaceae tp	0.4±0.1633	10.0±0.1528	10.0±0.1	10.0±0.1
Centaurea	0.4±0.2211	10.0±0.2134	10.0±0.0	10.0±0.0
Echium	0.1±0.1	10.0±0.0	10.0±0.3	10.0±0.1528
Fagus	0.4±0.1633	10.0±0.1528	10.0±0.3	10.0±0.1528
Liliaceae	0.6±0.3055	10.0±0.3073	10.0±0.5	10.0±0.3416
Acacia	0.1±0.1	0.0±0.1	0.0±0.0	0.0±0.0
Alisma	0.1±0.1	0.0±0.1	0.0±0.0	0.0±0.0
Cannabaceae	0.0±0.0	0.0±0.0	0.0±0.0	0.0±0.0
Carpinus	0.0±0.0	0.0±0.0	0.0±0.0	0.0±0.0
Caryophyllaceae	0.0±0.0	0.0±0.0	0.0±0.1	0.0±0.1
Castanaea	0.0±0.0	0.0±0.0	0.0±0.0	0.0±0.0
Cistus	0.0±0.0	0.0±0.0	0.0±0.2	0.0±0.1333
Cornus	0.0±0.0	0.0±0.0	0.0±0.0	0.0±0.0
Euphorbiaceae	0.0±0.0	0.0±0.0	0.0±0.0	0.0±0.0
Fabaceae	0.0±0.0	0.0±0.0	0.0±0.5	0.0±0.1667
Hypericum	0.0±0.0	0.0±0.0	0.0±0.0	0.0±0.0
Ilex aquifolium	0.0±0.0	0.0±0.0	0.0±0.1	0.0±0.1
Ligustrum	0.0±0.0	0.0±0.0	0.0±0.0	0.0±0.0
Linum	0.0±0.0	0.0±0.0	0.0±0.0	0.0±0.0
Lonicera	0.0±0.0	0.0±0.0	0.0±0.0	0.0±0.0
Lotus	0.0±0.0	0.0±0.0	0.0±0.0	0.0±0.0
Lysimachia	0.0±0.0	0.0±0.0	0.0±0.0	0.0±0.0
Malva	0.0±0.0	0.0±0.0	0.0±0.1	0.0±0.1
Moraceae	0.0±0.0	0.0±0.0	0.0±0.1	0.0±0.1
Myrtaceae	0.0±0.0	0.0±0.0	0.0±0.0	0.0±0.0
Ostrya	0.0±0.0	0.0±0.0	0.0±0.1	0.0±0.1
Papaveraceae	0.0±0.0	0.0±0.0	0.0±0.0	0.0±0.0
Populus	0.0±0.0	0.0±0.0	0.0±0.1	0.0±0.1
Quercus cerris	0.0±0.0	0.0±0.0	0.0±0.2	0.0±0.1333
Quercus coccifera	0.0±0.0	0.0±0.0	0.0±0.1	0.0±0.1
Sedum	0.0±0.0	0.0±0.0	0.0±0.1	0.0±0.1
Tamarix	0.0±0.0	0.0±0.0	0.0±0.0	0.0±0.0
Tilia	0.1±0.1	0.0±0.1	0.0±0.0	0.0±0.0
Urtica	0.0±0.0	0.0±0.0	0.0±0.3	0.0±0.2134

Table S2 : Metrics from the top-1 predictions obtained on both environmental and reference images for all 83 classes

Performance of the classification measured for all 83 classes separately, as the recall, precision and f1-scores, and on both the environmental and reference images from the test sets. Values represent mean ±SE over 10 cross-validations. Classes are ordered from highest to lowest f1-scores. "Indet." refers to indeterminable pollen.

Table S2

Class	nb images	recall	precision	f1-scores
Acacia	5.9±0.3145	91.417±91.4167	100.0±2.9346	95.293±100.0
Acer	93.1±0.3145	98.393±98.3927	97.453±0.3631	97.914±97.4532
Alisma	35.0±0.2108	96.293±96.2932	98.264±1.3508	97.232±98.2639
Alnus	55.9±0.2333	94.448±94.4477	95.695±1.0129	95.037±95.6949
Amarantaceae	129.9±0.3145	98.998±98.9976	99.382±0.3053	99.187±99.3822
Artemisia	126.8±0.3266	98.975±98.9751	98.912±0.1679	98.939±98.9118
Betula	130.0±0.3651	97.996±97.9957	98.553±0.4192	98.264±98.5534
Brassicaceae	130.0±0.2108	98.231±98.2307	98.776±0.3815	98.496±98.776
Buxus	130.0±0.2108	98.154±98.1544	97.715±0.2345	97.932±97.7149
Cannabaceae	11.5±0.1667	76.439±76.4394	92.253±5.2212	82.621±92.2525
Dipsacaceae tp	4.3±0.2603	86.333±86.3333	93.5±3.8152	88.932±93.5
Viburnum / Sambucus tp	38.7±0.2134	97.928±97.9281	97.483±0.6536	97.684±97.4831
Carduus	98.4±0.1633	99.188±99.1878	99.493±0.2533	99.339±99.4929
Carpinus	5.6±0.1633	74.0±74.0	91.214±5.4388	80.077±91.2143
Caryophyllaceae	68.5±0.2236	99.271±99.271	99.712±0.243	99.488±99.7122
Castanaea	127.9±0.2769	99.922±99.9225	99.077±0.0775	99.497±99.0769
Centaurea	25.4±0.2211	98.875±98.8746	100.0±0.7958	99.419±100.0
Cichorioideae	130.0±0.2582	99.386±99.3858	99.086±0.2969	99.232±99.0856
Cistus	12.4±0.1633	99.167±99.1667	97.183±0.8333	98.081±97.1832
Convolvulus	34.9±0.3145	96.855±96.8548	98.007±0.5143	97.41±98.0075
Cornus	122.8±0.2	99.837±99.8374	99.517±0.1626	99.675±99.5174
Corylus	113.4±0.3055	94.984±94.9842	95.28±0.8517	95.098±95.2798
Cupressus	98.8±0.2494	91.585±91.5851	90.347±1.1755	90.859±90.3467
Cyperaceae	13.8±0.3266	92.864±92.8636	94.042±1.8304	93.356±94.0415
Echium	116.2±0.1333	98.968±98.9677	98.409±0.4214	98.671±98.4086
Ericaceae	71.6±0.2211	97.897±97.8971	97.641±0.6706	97.757±97.641
Euphorbiaceae	29.7±0.2134	97.01±97.01	98.351±1.1531	97.626±98.3512
Fabaceae	102.9±0.2333	98.35±98.3504	98.068±0.4555	98.205±98.0682
Fagus	4.4±0.1633	75.0±75.0	86.833±7.3409	77.48±86.8333
Fraxinus excelsior	33.1±0.3786	90.594±90.5936	90.836±1.7003	90.589±90.836
Fraxinus ornus	130.0±0.2108	96.844±96.8437	97.605±0.3536	97.218±97.6049
Gallium	58.8±0.2906	94.222±94.2216	95.585±1.0177	94.874±95.5845
Geraniaceae	3.3±0.2134	83.333±83.3333	92.167±5.5556	86.127±92.1667
Helianthemum	3.6±0.3055	71.667±71.6667	86.333±5.9835	77.365±86.3333
Hypericum	27.9±0.2769	97.817±97.8175	97.208±1.4553	97.442±97.2085
Ilex aquifolium	3.6±0.2667	93.333±93.3333	95.5±4.4444	93.46±95.5
Indet. (blurry)	44.0±0.0	96.364±96.3636	96.246±1.3209	96.242±96.2459
Indet. (covered)	48.7±0.2134	85.232±85.2323	83.641±1.3086	84.293±83.6411
Juglans	19.2±0.4422	97.086±97.0859	98.947±1.2642	97.957±98.9474
Juniperus	114.1±0.1	88.609±88.6087	89.118±0.975	88.797±89.1179

Continued on next page

Table S2 – continued from previous page

Class	nb images	recall	precision	f1-scores
Lamiaceae	46.2±0.2	96.11±96.1099	97.233±0.7691	96.634±97.2332
Ligustrum	41.0±0.0	100.0±100.0	99.762±0.0	99.88±99.7619
Liliaceae	45.4±0.4522	97.647±97.6474	98.307±0.9617	97.949±98.3065
Linum	9.9±0.2333	97.98±97.9798	97.333±1.3552	97.5±97.3333
Lonicera	3.8±0.2	90.833±90.8333	91.333±6.855	90.551±91.3333
Lotus	5.1±0.1	68.667±68.6667	89.31±6.744	75.765±89.3095
Lycopodium	130.0±0.0	99.154±99.1538	98.411±0.2419	98.776±98.4108
Lysimachia	11.5±0.2236	91.492±91.4918	97.424±2.5038	94.094±97.4242
Malva	1.7±0.335	80.0±80.0	93.75±13.3333	95.833±93.75
Moraceae	64.2±0.1333	98.916±98.9159	96.867±0.6096	97.847±96.8666
Myrtaceae	54.7±0.2603	99.818±99.8182	99.464±0.1818	99.638±99.4642
Non pollen	130.0±0.2108	91.773±91.7732	91.181±1.0312	91.412±91.1807
Olea	130.0±0.3651	98.689±98.6893	97.074±0.3052	97.868±97.074
Ostrya	32.1±0.1	95.938±95.9375	94.667±1.2369	95.221±94.6668
Papaveraceae	8.0±0.0	93.75±93.75	97.321±2.7951	95.417±97.3214
Phillyrea	130.1±0.2769	96.238±96.2377	93.953±0.5266	95.064±93.9528
Pinaceae	128.5±0.3073	98.758±98.7584	98.466±0.2622	98.605±98.4659
Pistacia	104.7±0.2603	97.232±97.2315	96.952±0.5959	97.055±96.9518
Plantago	130.1±0.3786	97.315±97.3152	98.003±0.5262	97.648±98.0032
Platanus	60.9±0.3145	95.56±95.5595	98.859±1.205	97.13±98.8593
Poaceae	129.8±0.2	97.38±97.3804	96.955±0.4168	97.159±96.9548
Populus	130.0±0.0	99.077±99.0769	98.123±0.3939	98.585±98.1231
Quercus cerris	86.8±0.2494	94.815±94.8153	92.382±0.7562	93.537±92.3817
Quercus coccifera	130.0±0.0	96.0±96.0	93.294±0.5936	94.613±93.2939
Quercus deciduous	129.8±0.2906	90.213±90.2131	92.517±0.372	91.322±92.517
Quercus ilex	130.1±0.526	88.684±88.6844	89.378±1.2696	88.974±89.3782
Ranunculaceae	84.9±0.2769	98.008±98.008	98.969±0.7763	98.461±98.9687
Rhamnus	35.6±0.2211	90.999±90.9991	92.814±1.2475	91.783±92.8145
Rosaceae	130.0±0.3333	92.386±92.3859	92.347±0.9742	92.29±92.3465
Rumex	110.8±0.3266	97.381±97.3815	98.65±0.8057	97.986±98.6501
Salix	26.5±0.3073	94.753±94.7525	95.179±1.4951	94.859±95.1791
Sanguisorba	5.8±0.3266	70.81±70.8095	88.81±7.5326	76.344±88.8095
Scabiosa	6.5±0.1667	90.714±90.7143	90.119±2.5459	89.952±90.119
Sedum	23.4±0.3055	91.373±91.3726	95.866±2.3515	93.332±95.8664
Tamarix	52.8±0.2494	98.675±98.6753	98.319±0.4072	98.487±98.3187
Tilia	5.1±0.1	94.667±94.6667	100.0±3.6918	96.889±100.0
Ulmus	44.9±0.3145	98.652±98.6515	96.396±0.3671	97.487±96.3962
Urtica	124.8±0.2494	99.598±99.5981	99.211±0.1794	99.401±99.2112
Vitis fertile	130.1±0.3145	94.39±94.3897	95.114±0.8226	94.715±95.114
Vitis sterile	126.1±0.2769	97.855±97.8552	95.842±0.269	96.808±95.8424
Xanthium / Ambrosia tp	129.8±0.2906	99.23±99.2296	99.083±0.1986	99.154±99.0833
type 1	7.5±0.2236	68.175±68.1746	81.667±7.2717	71.512±81.6667
other	7.0±0.5164	51.053±51.053	74.143±7.7484	57.779±74.1429

Table S3: Recall rates from the top-4 predictions with all 83 classes on environmental images

Recall rates obtained with the top-4 predictions for environmental images on all 83 classes separately (mean ±SE over 10 cross-validations). Only classes with more than 50 images are displayed (i.e. 5 images per cross-validated test set). "Indet." refers to indeterminable pollen.

Table S3

Class	**nb images**	**Recall rates**
Fraxinus excelsior	22.5±0.2236	100.0±0.0
Fraxinus ornus	5.4±0.1633	100.0±0.0
Cichorioideae	10.2±0.1333	100.0±0.0
Ericaceae	6.3±0.1528	100.0±0.0
Olea	32.6±0.2667	100.0±0.0
Lycopodium	130.0±0.0	100.0±0.0
Pinaceae	95.4±0.2211	99.8947±0.1053
Quercus ilex	67.7±0.3958	99.7037±0.1975
Buxus	92.6±0.2211	99.6797±0.2275
Non pollen	115.1±0.3145	99.6537±0.1911
Phillyrea	26.3±0.1528	99.6154±0.3846
Poaceae	95.6±0.2211	99.4748±0.2351
Brassicaceae	18.1±0.1	99.4444±0.5556
Indet. (blurry)	44.0±0.0	99.3182±0.3472
Cupressus	43.2±0.1333	99.3129±0.3499
Juniperus	42.1±0.1	99.2913±0.3609
Pistacia	23.0±0.0	99.1304±0.5797
Plantago	34.5±0.1667	99.1176±0.6278
Alnus	10.5±0.2236	99.0909±0.9091
Vitis fertile	57.2±0.1333	98.9534±0.5936
Indet. (covered)	48.7±0.2134	98.7794±0.5402
Rosaceae	38.5±0.2236	98.7176±0.7881
Quercus deciduous	12.1±0.1	98.3333±1.6667
Vitis sterile	18.3±0.2134	97.7778±1.2284
Amarantaceae	8.1±0.1	97.6389±1.5775
Carduus	7.4±0.1633	97.3214±1.7907
Platanus	7.0±0.0	97.1429±1.9048
Acer	8.9±0.3145	96.7298±1.6846
Cyperaceae	5.2±0.1333	96.3333±2.457
type 1	7.5±0.2236	93.2143±2.9378
Gallium	5.8±0.2906	93.1429±2.8635
Rhamnus	9.6±0.2211	92.6667±2.7477
Rumex	5.2±0.1333	85.0±4.6415
other	6.2±0.6633	67.8108±11.8176

Chapter 4

Differentiating pollen from the wild and the domesticated grapevine using automated image analysis

Betty Gimenez, Céline Devaux, Sébastien Joannin, Doris Barboni, Sourish Das, Pranabendu Misra, Anuraj Kashyap, Laurent Bouby, Maud Tenaillon, Odile Peyron

4.1 Introduction

4.1.1 Distinguishing the wild and domesticated grapevine from pollen analysis: a scientific and methodological challenge

The grapevine is an emblematic species in the Mediterranean area (Guiot et al., 2023) that has shaped landscapes and social interactions since the Neolithic (McGovern et al., 2017; Naqinezhad et al., 2018). It is an ancestral nutritive resource for human populations, with fruits already consumed by paleolithic hunter-gatherers (Zohary et al., 2012). The grapevine has also been a major crop for the last few millennia, driving intense trade and exchanges among Mediterranean people (Terral et al., 2010).
The cultivated grapevine consumed today, *Vitis vinifera subsp. vinifera*, later on *Vitis vinifera*, results from domestication from its wild ancestor, *Vitis vinifera subsp. sylvestris*, or *sativa*, later on *Vitis sylvestris* (Levadoux, 1956; This et al., 2006; McGovern et al., 2017; Dong et al., 2023). The natural range of *Vitis sylvestris* goes from Turkestan to the Atlantic coast of Europe and in North Africa (This et al., 2001; Zecca et al., 2012; Naqinezhad et al., 2018). However over the last centuries, *Vitis sylvestris* has become scarce in the Western Mediterranean basin due to habitat loss and disease (This et al., 2006), and is now considered as endangered in this region. The grapevine is the only species of its genus endemic to Europe (Zecca et al., 2012; Naqinezhad et al., 2018), and two major *Vitis sylvestris* ecotypes have been identified based on genetic differences, an eastern ecotype (accessions in Western Asia and the Caucasus) and a western ecotype (accessions in Central Europe and the Iberian Peninsula), which would have split in the Pleistocene around 400,000 to 200,000 years ago (Dong et al., 2023). Domestication of the grapevine would origin from the Caucasian area, thus from the eastern ecotype, around 11,000 years ago during the Neolithic (Zohary et al., 2012; McGovern et al., 2017; Dong et al., 2023). Cultivated forms along with wine making practices then dispersed during the following millennia, first to Northern Africa and then reaching Iberia and Southern France around 3000 years ago (Terral et al., 2010; McGovern, 2013; Iriarte-Chiapusso et al., 2017).

Vitis sylvestris and *Vitis vinifera* were recognized as two monophyletic subspecies in the early 2000's from morphological differences (leaves, fruits, pips, etc.) and molecular differences (Grassi et al., 2008). The development and diffusion of wine production across

the Mediterranean area is well documented (This et al., 2006; Bouby et al., 2013; Dong et al., 2023; Bernigaud et al., 2024). Yet, it remains difficult to determine in many archaeological contexts whether wine was produced from wild individuals harvested in their natural habitat, from cultivated forms grown in vineyards, or from the cultivation of wild individuals (Arroyo-García et al., 2006). In archaeological records, leaves, fruits, or seeds from grapevine individuals allow to determine the cultivated forms (Terral et al., 2010). These remains are however not always found in such archaeological contexts. In contrast, pollen is more efficiently and ubiquitously preserved. Using pollen data from fossil cores, the presence of grapevine can be tracked through time during the last 11,700 years BP (Holocene period) in the Mediterranean region (e.g. Mercuri et al., 2002; J. Allen et al., 2002; Drescher-Schneider et al., 2007; Joannin et al., 2012; Sadori, 2018). The wild grapevine was likely widespread in the Western Mediterranean area during the Holocene, and thus, while viticulture expended in the area. However, palynology has never been used to assess grapevine history while it has been tested for other emblematic Mediterranean taxa such as *Olea europaea* (Langgut et al., 2019), *Juglans* (Pollegioni et al., 2017) or *Castanea* (Krebs et al., 2019). Few grapevine pollen grains are found in fossil assemblages, which complicates interpretation of the signal obtained. Moreover, from a strictly palynological point of view, it remains impossible to determine whether grapevine pollen grains originate from wild populations or cultivated vineyards. Characterizing more precisely the presence of the grapevine in the Western Mediterranean by discriminating the pollen signature from wild and domesticated grapevines, or from wild populations and cultivated vineyards, is thus relevant for both archaeology and paleoecology, but it remains a challenge.

The wild grapevine is predominantly dioecious, with separate male and female individuals, and with occasional hermaphroditism, while the cultivated grapevine is predominantly hermaphroditic, with few functionally female varieties (e.g. Piccolit cultivar, Cargnello et al., 1980). Two pollen morphotypes can be produced by grapevine individuals (Fig. 1), a fertile morphotype is produced by functionally male and hermaphroditic individuals, and a sterile morphotype is produced by functionally female individuals presenting aborted anthers (Lombardo et al., 1978; Caporali et al., 2003; Preiner et al., 2012). The morphologies of both pollen morphotypes are very distinct, the fertile morphotype is tricolporate (or trizonocolporate), isopolar and radially symmetrical with a foveolate-perforate exine ornamentation (İnceoğlu et al., 2000; Luksic et al., 2022), while the sterile morphotype is inaperturate, with a uniformly thickened wall (Caporali et al., 2003). Additionally, a thicker exine has been shown on the sterile pollen morphotype produced by wild functionally female individuals compared to that of female cultivars (Mercuri et al., 2021), while the fertile and the sterile morphotypes were larger and less variables on domesticated compared to wild individuals (Gallardo et al., 2009). Finally, more (by 50%) fertile pollen grains were estimated in hermaphroditic flowers of domesticated individuals than in male flowers of wild individuals (Gallardo et al., 2009).

Based on these elements, two signals could help discriminate the presence of wild or cultivated subspecies using pollen: (1) morphological differences, and (2) abundance differences between the subspecies in both fertile or sterile pollen morphotypes. In this study, we present a preliminary investigation, as a proof of concept, of differences between the two subspecies. We will address the following questions :
(1) Is there differences in the morphologies between each morphotype produced by the wild or the domesticated grapevine? Can these potential differences be captured in light microscopy (LM) images and be used to differentiate pollen grains produced by a wild or

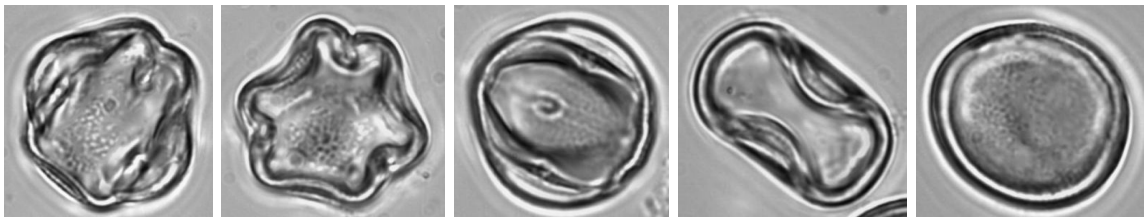


Figure 1: Examples of grapevine pollen images with the tricolporate fertile morphotype (images 1, 2, 3) and with the inaperturate sterile morphotype (images 4 and 5).

a domesticated individual?
(2) Is there differences in the abundance of pollen grains produced by the two subspecies?

4.1.2 AI-based approaches to address our two questions

Differentiating similar pollen morphotypes to develop a taxonomy is a basic principle of palynology that primarily relies on differences in pollen size, exine ornamentation, type and number of apertures. Attempting to reach higher taxonomic ranks, especially for certain taxa of interest, is a common effort in palynology. It has predominantly been carried out using morphometric measurements of pollen grains observed under light or electronic microscopy, such as the equatorial diameter (E), the polar axis (P), their ratio E/P, the exine thickness or the annulus diameter, then processed using multivariate statistical analyses (Portillo et al., 2020). Such approach was for instance conducted to differentiate species of Brassicaceae, usually identified to the family level, to differentiate rice pollen from other Poaceae species (Yang et al., 2018), wild from domesticated Poaceae taxa (Tweddle et al., 2005; Joly et al., 2007), or to differentiate species of *Pinus* using for instance the CART model (Desprat et al., 2015). Over the last decade, artificial intelligence has received increasing interest to distinguish morphologically close pollen taxa and improve taxonomy. Machine learning classifiers differentiated two *Picea* pollen taxa, reaching more than 93% accuracy on fossil and modern samples (Punyasena et al., 2012). More recently, classification Convolutional Neural Networks (CNNs) were also successfully tested in paleoecological contexts to improve the identification of fossil specimens of Amarantaceae, where 6 types were differentiated to 99.3% (Bourel et al., 2020), of legume tribe Amherstieae (Romero et al., 2020), and to infer the phylogenetic placement of extinct specimens of Podocarpus, using confocal microscopy (Adaïmé et al., 2024). Image analysis using artificial intelligence has thus the potential to overcome long-standing limitations of manual palynology, by providing faster and more efficient methods to distinguish pollen taxa with close morphologies. Here, we compare abundance and size of pollen grains produced from 14 wild and 14 domesticated grapevine individuals grown in the ampelographic collection of reference of Vassal-Montpellier, and using light microscopy images. We also use a classification CNN to discriminate between pollen grains from both subspecies. Discriminating between the fertile tricolporate and the sterile inaperturate grapevine morphotypes using a CNN was previously tested, and achieved good results (chapter III). Here, we investigate the possibility of discriminating between pollen grains of the same morphotype, but from the two subspecies. For the CNN classification, we will only focus on the fertile tricolporate pollen morphotype, which is produced by both wild and domesticated individuals. We hypothesize that the tricolporate pollen grains from the wild and domesticated grapevines, although very similar to the human eye, display small discriminant features that may be captured by a CNN for their identification.

4.2 Material and Methods

4.2.1 Pollen samples

We used pollen collected on grapevine reference individuals from 28 accessions grown in the grapevine biological resource center of Vassal-Montpellier (INRAE, SupAgro), with 14 individuals from the wild subspecies and 14 from the domesticated subspecies (Supporting Information Fig. S1). In 2022, during the blooming period of the grapevine (16-20, May 2022), 15 closed flowers were collected on each of the 28 selected individuals, and pooled per individual. Flower stamens were cut off and stored in the laboratory in a closed container. Each sample (i.e. 15 flowers from one individual) were then processed as follows. Tablets of *Lycopodium* spores were added and samples were chemically treated, using KOH to dissolve the organic matter from the stamens, and acetolyzed for 6 minutes. One fixed slide per sample was mounted with glycerol jelly, under 16 mm × 16 mm cover slides (see details in chapter I).

4.2.2 Image acquisition and detection of the grains in the images

Slides were scanned using the automated image acquisition pipeline described previously (Gimenez et al., 2024; chapter II). For each photographed field of view (FOV), we used 13 images spaced 8 µm apart along the z-axis, 256 FOVs of 214 µm × 214 µm were photographed per slide, each separated by 110 µm to avoid their overlap. Note that the focal range acquired for each FOV (104 µm) was much larger than a grapevine pollen (20-25 µm) to capture sharp images of grains standing at distinct depths. Each pile of 13 were then stacked into a single composite 2D image for each FOV. To obtain individual images of all pollen grains occurring in the FOV images, we used the detection method previously developed (Gimenez et al., 2024; chapter II) and based on the object-detection algorithm YOLOv5 (Jocher, 2020). The method generates bounding boxes that enclose each detected particle identified as either pollen, *Lycopodium*, or 'edge' for pollen or *Lycopodium* grains cut on the FOVs margins. This process generated 465 to 3124 bounding boxes per slide.

The contents of the bounding boxes labeled as 'pollen' were then extracted from their FOV images, i.e. segmented, to generate individual images for each detected grapevine pollen grain (see details on segmentation in chapter I). First, the segmentation method was applied to the stacked 2D FOV images. We refer to the single stacked image obtained for one pollen grain, as 2D-stacked image. Second, the segmentation method was applied to each of the 13 images from the pile acquired at distinct focal depths for each FOV. Grapevine pollen grains were sharp in only some images of each pile, and appeared heavily blurred in the others; we thus selected only four consecutive images that had the highest sum of a sharpness index (calculated using the Laplacian variance for each image). We refer to the pile of 4 images obtained for one pollen grain, as 3D-stack images.

As expected from their typology, each sample contained pollen grains from only one morphotype, either the fertile tricolporate or the sterile inaperturate (no pollen grain with the opposite morphotype in a given sample was observed in the dataset). Among domesticated individuals, 12 samples contained the tricolporate fertile morphotype, and 2 the inaperturate sterile morphotype, while among wild individuals, 10 contained the tricolporate fertile morphotype, and 4 the inaperturate sterile morphotype.

4.2.3 Estimating and comparing the size and production of pollen grains from wild and domesticated grapevine individuals

First, we compared pollen grains from both subspecies using the size as a standard morphometric measurement, leveraging the developed AI-based detection method to generate large quantities of individual grapevine images and automatically extract measurements. The size of the grapevine pollen grains from the 28 individuals were estimated from the dimension of detected bounding boxes that were labeled as pollen. Bounding boxes that touched the margin of their FOV were discarded to ensure the grains were not cut off. Similarly, to avoid accounting for grains that overlapped with each other, all bounding boxes with a positive IoU were discarded. The size of a pollen grain was estimated from the surface area of its bounding box.

Then, the amount of pollen grains produced in the flowers from individuals of the wild and domesticated subspecies were compared. To do so, the number of grapevine pollen for 100 *Lycopodium* grains were estimated for each slide (one slide mounted per sample), and considering that pollen from 15 flowers along with the same amount of *Lycopodium* spores were consistently added in each sample. For this measurement, we used the number of bounding boxes detected and labeled as either pollen or *Lycopodium*. Bounding boxes labeled as 'edge' were not considered due to the lack of the particle identity, and bounding boxes touching the bottom and left sides of the FOV, regardless of their labels and surface, were also discarded for calibration.

4.2.4 Compiling the grapevine pollen images for the classification task

A previous study focusing on the classification of several Mediterranean taxa showed that the tricolporate and the inaperturate grapevine morphotypes can be efficiently differentiated using a CNN (<2% mis-identification, chapter III). Our objective was not to repeat this study, but to assess whether pollen grains from the same morphotype (tricolporate) but produced by either the wild or the domesticated grapevine could be differentiated using a CNN. The subsequent analysis was only conducted with samples containing the tricolporate morphotype, using pollen samples from 12 domesticated and 10 wild individuals.

To build the classification dataset, all detected and segmented pollen images were first manually checked to remove images in which grains were damaged, covered by debris, overlapping with each other, or cut off, as well as incorrectly detected or labeled objects, such as dust or *Lycopodium* grains; although these cases were scarce, they can affect the performance of a deep learning model. Two classification datasets were obtained, one from the 2D-stacked pollen images, and one from the 3D-stack pollen images, referred as '1pollen/1image' and '1pollen/4images', respectively (Fig. 2). Both datasets contained the exact same pollen grains, namely from 301 to 1638 pollen grains for each of the 22 samples.

4.2.5 Experimental design for training and testing the classification algorithm

The classification dataset included more pollen grains from domesticated individuals. To balance the two subspecies in the dataset, we randomly sampled a maximum of 550 images per domesticated individual. The dataset included 6594 pollen images from domesticated grapevines (from 12 individuals), and 6277 pollen images from wild grapevines (from 10 individuals). The models were always trained using 2 classes, 'wild' and 'domesticated', corresponding to the two subspecies. Each dataset were split into three subsets for training, validation, and testing, and models were cross-validated by successively interchanging the images in each subset (see details in chapter I). Following these steps, two designs were determined, differing in the way the dataset is divided and

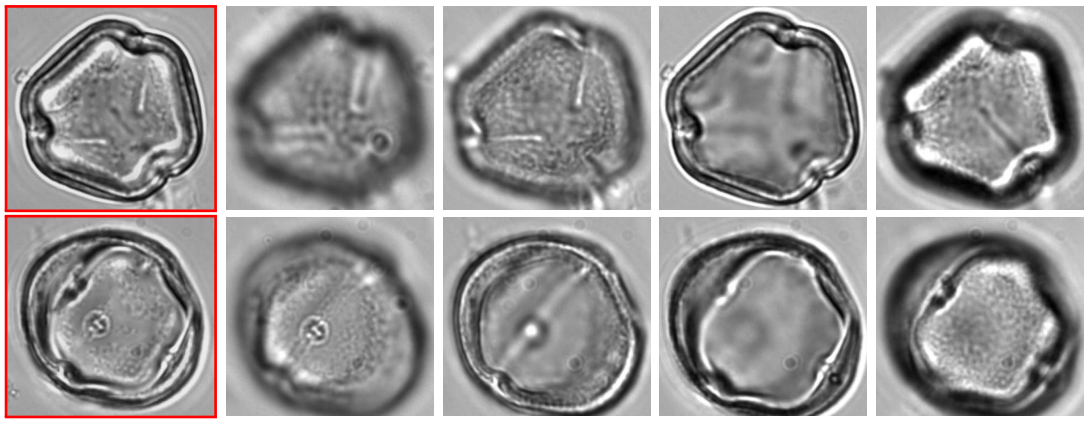


Figure 2: Illustration of two pollen grains from the '1pollen/1image' dataset (left images with red borders), and the '1pollen/4images' dataset (4 images on the right)

cross-validated. First (design A), we split the dataset forcing equal contribution of each individual to each of the three subsets. The dataset was prepared for a five-fold cross-validation, thus with 60% for training, 20% for validation and 20% for testing. In this design, successful classification might rely on different characteristics determined between both subspecies, but also on characteristics of individuals themselves. To avoid this latter caveat, we implemented a second design (design B) based on the leave-one-out strategy. Models were trained on images from all samples but four, including from 2 wild and 2 domesticated individuals kept for independent testing. This approach aimed to assess if the models could accurately identify and learn characteristics specific to each subspecies, and not individuals. For this design, the individuals were randomly paired without replication, forming 6 pairs from the 12 domesticated individuals and 6 pairs from the 10 wild individuals (the last pair combined two individuals not previously paired). The models were cross-validated six times, while successively inter-changing the pairs kept for testing. For each cross-validation, images from samples not used for testing, were split into 80% for training and 20% for validation, randomly, but for each sample separately (see details in chapter I, Supporting Information Fig. S1, Table 1).

Table 1: Description of the main strategies investigated to train and evaluate the models. In the design A, pollen from all individuals are equally represented in all three subsets; in the design B, pollen from 4 grapevine individuals are reserved for testing. 'train-val-test' refers to the distribution of the images into the subsets used for training, validation or testing of the models. 'cval' refers to the number of cross-validated models trained in the corresponding strategy.

design	**dataset**	**train-val-test**	**cval**	**model evaluation**
A	1pollen/1image	60%-20%-20%	5	top-1 predictions
A	1pollen/4images	60%-20%-20%	5	top-1 predictions from the sharpest image of the pile
B	1pollen/1image	test on 2 ind./class + 80%-20%	6	top-1 predictions
B	1pollen/4images	test on 2 ind./class + 80%-20%	6	top-1 predictions from the sharpest image of the pile

Both designs were implemented with the datasets '1pollen/1image' and '1pollen/4images'. For the dataset '1pollen/4images', each of the 4 images from each pollen grain were fed separately to the model and without specifying they belonged to the same pollen grain. For each model training, the algorithm ResNet152 (He et al., 2015) was trained from scratch for 120 epochs, using images resized to 128 $\times$ 128. For the evaluation of performances of each trained model, the predictions on test images were compared to the ground truth to define a True Positive (TP) for a prediction matching with the ground truth class i, otherwise a

False Negative (FN) for class i and a False Positive (FP) for class j. From these statistics, recall, precision and f1-scores were then calculated for both classes (see chapter I). For the models trained on the dataset '1pollen/4images', evaluation were conducted with two approaches. First, we used as prediction for a pollen grain the prediction generated for the sharpest image from the pile of four images. Second, we averaged the scores predicted for the four images, and considered for prediction the class with the highest average score.

4.3 Results

4.3.1 Size and production of pollen grains from wild and domesticated grapevine individuals

The sizes of the grains were larger for wild individuals compared to domesticated individuals (Fig. 3), with an Anova p-value of 6% (Supporting Information Fig. S2). Bounding boxes were respectively of 96 578±930 (mean±2SE) pixels and 90 512±985 pixels for pollen collected from wild and domesticated individuals.

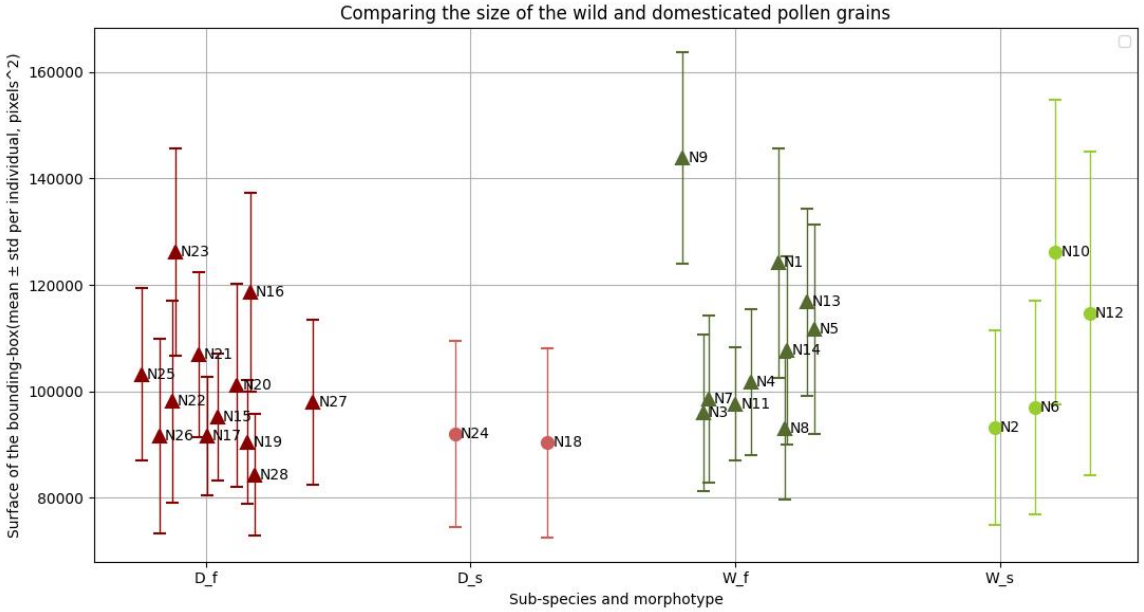


Figure 3: Size of the pollen grains collected from each grapevine individual and estimated from the surface of the detected bounding boxes framing the grains (length × height, in pixels). Values for wild individuals are presented in green, and values from domesticated individuals are presented in red. Triangles represent the tricolporate fertile morphotype, and circles represent the inaperturate sterile morphotype.

The amount of pollen grains produced from flowers of both grapevine subspecies was estimated from calibrated amounts of pollen grains counted for 100 *Lycopodium* spores : we obtained 647.73±106.4 (mean±2SE) pollen grains for domesticated individuals, compared to 309.46±53.08 pollen grains for wild individuals (Fig. 4).

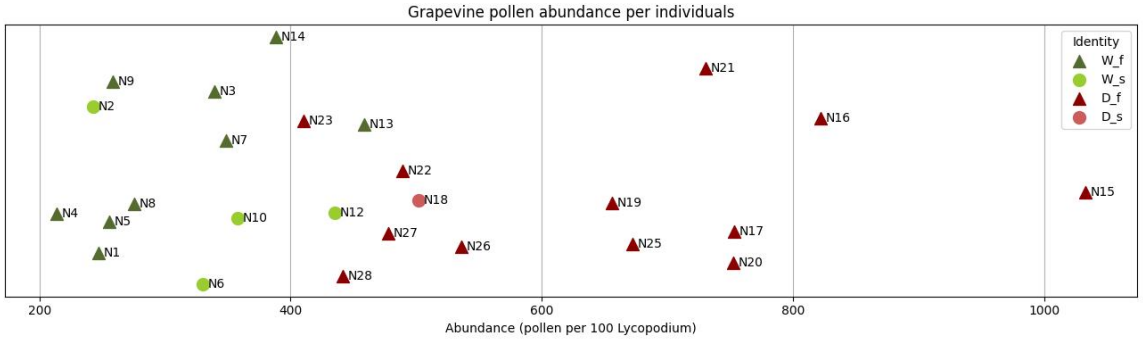


Figure 4: Abundance of pollen grains estimated for 100 *Lycopodium* grains, and from samples including pollen from 15 flowers for each individual. 'W_f' and 'W_s' refer to wild individuals with respectively the fertile or the sterile pollen morphotypes, 'D_f' and 'D_s' refer to domesticated individuals with respectively the fertile or the sterile morphotype.

4.3.2 Using a CNN to differentiate the fertile pollen from wild and domesticated grapevine individuals

4.3.2.1 Experimental design A

When implemented on the dataset '1pollen/1image', the models trained with design A predicted most pollen grains into a single class (i.e. subspecies), and regardless of the true subspecies from which the pollen grains belonged to. Four models out of five predicted most images as 'wild', including pollen images from both wild and domesticated individuals, while the opposite pattern was generated with the last model (e.g. Fig. 5, Supporting Information Fig. S3, Table S2). This pattern suggests that all correct predictions are likely artificial. More cross-validation should be run to establish whether the class predicted most frequently by a model is actually random.

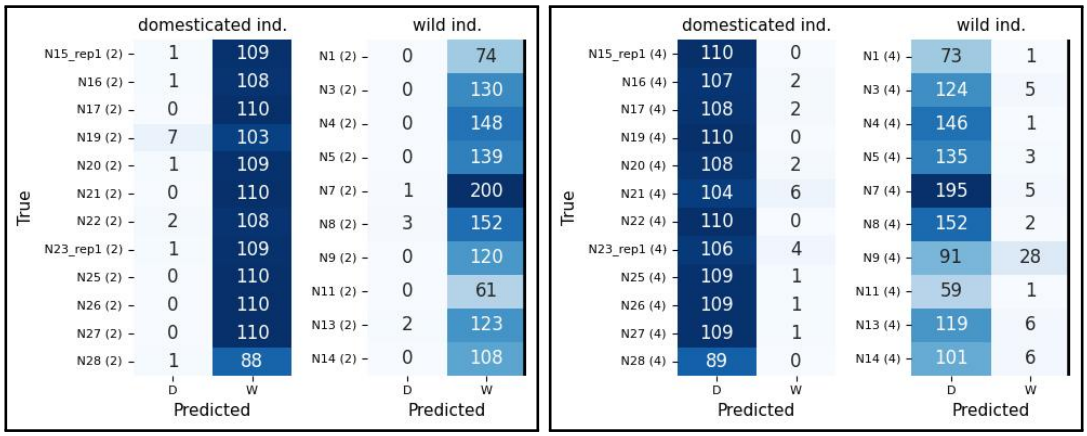


Figure 5: Confusion matrix illustrating the main pattern generated using the design A and the dataset '1pollen/1image', with (a, left panel) the patterns observed in 4 of the models (e.g. of the cross-validation no. 2), and (b, right panel) the opposite pattern generated by the last model (cross-validation no. 4). The number in brackets refers to the number of the cross-validation. 'D' refers to the predictions into the class 'domesticated', and 'W' to the predictions into the class 'wild', 'ind.' refers to 'individuals'.

The same pattern occurred for one of the five models trained with the design A on the '1pollen/4images' dataset, and when considering only predictions for the sharpest image

of the pile of 4 (Supporting Information Fig. S4, cross validation no. 5). In contrast, the other four models predicted both classes, but again predictions seemed to be attributed regardless of the true classes, and rather randomly (e.g. Fig. 6, Supporting Information Fig. S4, Table S2).

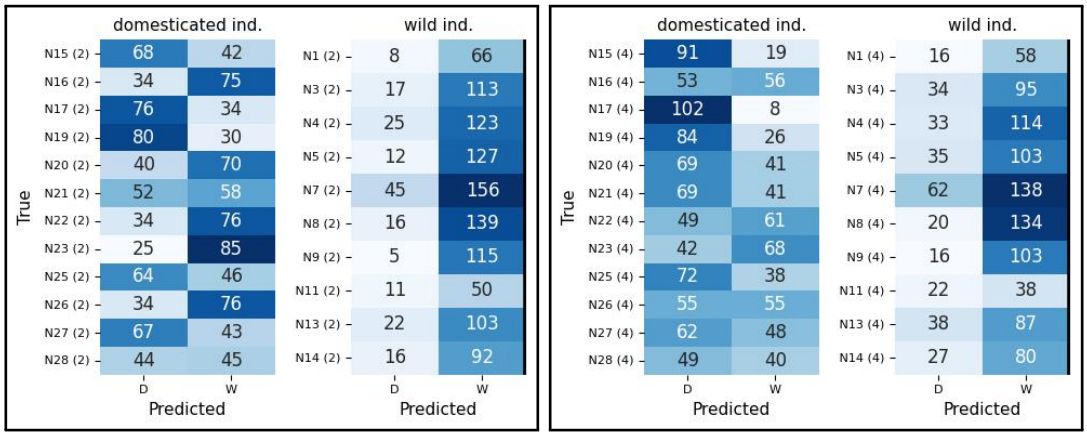


Figure 6: Confusion matrix illustrating the main patterns generated using the design A, the dataset '1pollen/4images' and evaluated on the sharpest image, with as example (a, left panel) the cross-validation no. 2, and (b, right panel) the cross-validation no. 4. The number in brackets refers to the number of the cross-validation. 'D' refers to the predictions into the class 'domesticated', and 'W' to the predictions to the class 'wild', 'ind.' refers to 'individuals'.

4.3.2.2 Experimental design B

When implemented on the dataset '1pollen/1image', the design B also tended to generate predictions at random and not related to the true classes. All but one models predicted most grains into one single class, either wild or domesticated, and regardless of the true class (Fig. 7.a). The last model predicted images among both classes, but again, regardless of the true classes (Fig. 7.a, cross validation no. 2, Supporting Information Table S2).
In this design, images from two wild individuals were used in the test sets of two different cross-validations, from the additional wild pair generated to match the 6 domesticated pairs. Images from these individuals were classified with distinct patterns depending on the cross-validated models, for instance images from the wild individual N8 were classified mostly as domesticated in the cross-validation no. 3 and as wild in the cross-validation no. 6, while pollen from the wild individual N11 were distributed across both classes randomly (cross-validation no. 2) or mostly as wild (cross-validation no. 6, Fig. 7.a). This result suggests that predictions were largely pre-determined by the trained model, and thus the training images, and was not actually related to the test images themselves.

Applying this design to the '1pollen/4images' dataset also produced a similar random pattern, with predictions on the sharpest images being assigned to both classes in a distribution that appeared unrelated to the true classes (Fig. 7.b, Supporting Information Table S2).

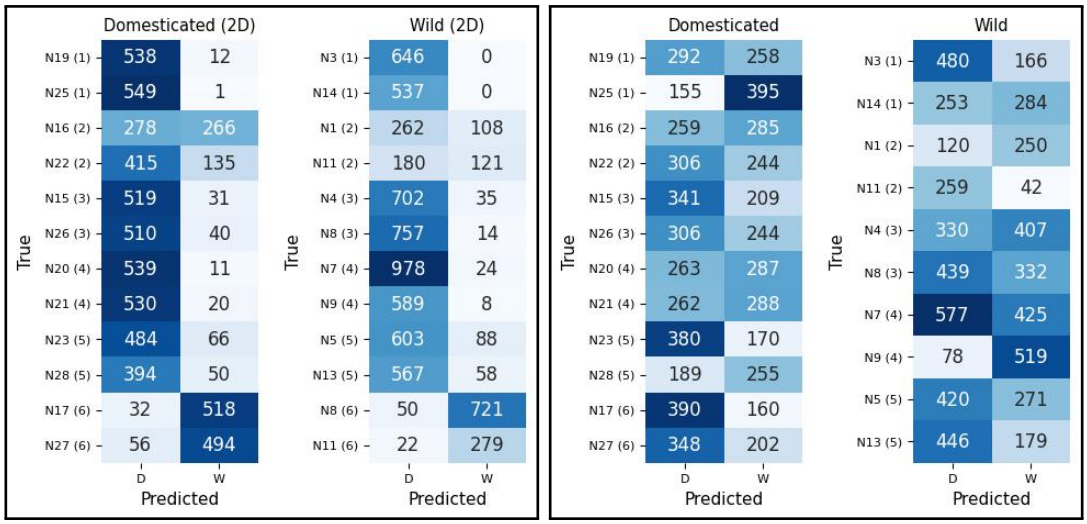


Figure 7: Confusion matrix generated using the design B and including on all cross-validations, as each sample is tested by one model, (a, left panel) when implemented on the dataset '1pollen/1image', and (b, right panel) when implemented on the dataset '1pollen/4images' and evaluated on the sharpest image. The number in brackets refers to the no. of the cross-validation. 'D' refers to the predictions into the class 'domesticated', and 'W' to the predictions to the class 'wild', 'ind.' refers to 'individuals'.

4.4 Discussion

4.4.1 Distinguishing similar pollen morphotypes from wild and domesticated grapevine individuals

Regardless of the design investigated and of the dataset used, the classification models generated predictions that seemed unrelated to the actual true class of the images, but followed specific patterns. Two prediction patterns were observed, either (1) the models predicted only one class predominantly, and that for pollen images from both wild and domesticated individuals, or (2) models predicted images into both classes, but once again, in distribution that were unrelated to the grapevine subspecies from which the pollen grains had been collected. The first pattern mainly occurred when training the models on the dataset '1pollen/1image' while the second pattern was predominant for models trained on the '1pollen/4images' dataset and evaluated on the sharpest image.

We initially hypothesized that the stacking process might result in the loss of discriminant details for classification and that using the pile of images to train the model instead might alleviate this issue and potentially allow more efficient classification. Using a pile of 4 images per pollen grain without specifying they belong to the same grain did not improve the performance of the models. Nonetheless, predictions displayed distinct patterns suggesting that the stacking process modifies images in a way that impacts how they are processed by the CNN. We cannot rule out that the 8 µm step taken between each image of the stacks at distinct focal depths is too large to capture potential discriminant features between both subspecies. We used a 8 µm step to remain consistent with previous acquisitions done with this pipeline on other samples and allow possible bridges between our datasets (see detail chapter I). For instance, we initially aimed to apply the models trained on reference grapevine pollen images from the collection of Vassal-Montpellier to grapevine pollen images obtained from environmental samples, i.e. with grapevine individuals growing in their natural habitats (chapter III). Reducing this 8 µm step, and generating more images per pollen grain might improve the resolution of potentially discriminant features, such as the

exine ornamentation, which may have not be captured with our resolution.

Using our dataset as it is, several approaches were nonetheless identified to improve the robustness of the classification method and to further investigate the possibility of differentiating both subspecies using light microscopy pollen images. Modifying the classification algorithm to compute jointly the 4 images of each pollen grain as a single input, and instead of 4 distinct inputs, may provide a more comprehensive information on the pollen grains, and improve the performance of the models. A 3D classification could be tested with our dataset, for instance relying on multi-layer models, or by extracting local stable points from the images (as in Wang et al., 2021).
The algorithm could also be modified to include complementary morphometric measurements, such as the size parameter, as a meta-data processed in a separate channel to the images. The standard CNN used in this study processed images resized to squares, although size is one major discriminant criteria in palynology. The size of pollen grains were estimated to be slightly smaller on domesticated compared to wild individuals (Fig. 3), suggesting that this parameter might improve the discrimination power of the models. Re-framing images to squares during model implementation also distorted pollen grains in rectangular images, and thus prevented the use of potential differences in the overall elongation of the grains (E/P, pollen grains prolate, sub-prolate, etc.) for the classification. Adding a padding to change the images to squares before model implementation (as done in chapter III) to avoid this distortion could thus improve results, especially as the elongation of the grains has been observed to differ between grapevine pollen samples (İnceoğlu et al., 2000), and subspecies although not significantly (Gallardo et al., 2009). Thus, morphometric measurements such as the E/P ratio, but also pollen roundness or exine thickness, could be extracted from the images, using available tools such as the softwares ImageJ, or Particule Trieur (Marchant et al., 2020a), and used as complementary inputs for image classification using the CNN. Altogether, these approaches could increase the chance of capturing details that may be relevant for the discrimination between pollen from both grapevine subspecies, and may improve further the robustness of the method.

Imaging the samples with more advanced microscopy techniques to provide images of higher resolution and comprising more detailed features might also allow to discriminate pollen from the two grapevine subspecies. Scanning Electronic Microscopy (SEM) for instance can capture much more details on surface features of the exine compared to light microscopy (Jones and Bryant, 2007; Taia, 2022), and this technique was favored in many studies that described grapevine pollen grains morphology, especially for cultivated varieties, which have been more extensively investigated compared to the wild grapevine (İnceoğlu et al., 2000; Perveen and Qaiser, 2008; Gallardo et al., 2009; Jovanovic-Cvetkovic et al., 2016; Mercuri et al., 2021), or for other Vitaceae species (Cartaxo-Pinto et al., 2017). Confocal Scanning Laser Microscopy (CSLM) can also provide high levels of details on pollen grain features, from external but also internal surfaces, allowing for 3D reconstruction of the grains, and may be an adequate solution for our addressed question. Over the last few years, confocal microscopy has been increasingly used in palynology in studies aiming to improve taxonomy (e.g. Gavrilova et al., 2018; Trivedi et al., 2022; Quamar et al., 2022), including of Mediterranean taxa with auto-fluorescence imaging (Castro López et al., 2010), and combined with machine learning classifiers (Romero et al., 2020; Adaïmé et al., 2024; Collevatti et al., 2024).

In this study, the use of a standard CNN on light microscopy images failed to differentiate between pollen grains from the two subspecies, which was explained by the similarity in their morphologies. It is possible that the fertile pollen morphotype from the two subspecies lack discriminant features altogether, potentially due to a too short evolution

period combined with admixtures and introgression. The domestication of the grapevine has been dated to 11 000 years ago in the Caucasian area, where an eastern grapevine ecotype has been identified (Dong et al., 2023). Cultivated forms then dispersed, reaching France and Iberia around 3000 years ago (McGovern, 2013; Myles et al., 2011), where it already occurred naturally granted to suitable habitats, and where a western grapevine ecotype has been identified (Dong et al., 2023). A study mentioned that flowering times of wild individuals differ from that of the cultivated forms reducing the possibility of gene flow through pollen (This et al., 2006), and low levels of pollen-mediated gene immigration from domesticated to wild compartments were identified (Di Vecchi-Staraz et al., 2009). Nonetheless, a recent genomic study found two major introgression events, which were related to the arrival of the cultivated forms to Western Europe (Dong et al., 2023). For instance, introgression of domesticated to wild grapevines in Western Europe would have contributed to ∼25.0 to 30.0% of the ancestry from some Western Europe wine grape, and many cultivars ('Lambrusco' cultivars) share ancestries from the wild western ecotype (Dong et al., 2023). Such genetic similarity between the wild and cultivated compartments thus explain the morphological similarities between their pollen grains, and the potential lack of distinct features in the studied fertile pollen morphotype.

4.4.2 Alternative signals to discriminate wild from domesticated grapevine

The production of pollen grains, measured from 15 flowers per individual, was higher in domesticated compared to wild individuals. Production of sterile inaperturate pollen in the 4 corresponding female wild individuals was similar to that of the other 10 male wild individuals with fertile tricolporate pollen (Fig. 3). This result was surprising since the inaperturate pollen from the female individuals is produced by aborted anthers which are not fully developed (Caporali et al., 2003), and a previous study identified a lower production (Gallardo et al., 2009). Comparing the amount of pollen grains produced by wild or domesticated individuals, and the ratio between the inaperturate and tricolporate morphotypes obtained in wild populations compared to cultivated vineyards could be an alternative approach to discriminate the presence of the wild from the domesticated grapevine. Changes in the ratio between the two morphotypes is a signal for human practices, selecting individuals producing more and heavier fruits (Miller, 2008), and indirectly female individuals producing more sterile pollen, or hermaphroditic individuals producing more fertile pollen. Evaluating changes in the proportion between the two morphotypes through time may thus bring complementary information to infer stages of human selection or cultivation of the grapevine. The presence of the inaperturate morphotype has also been mentioned as a potential tool to assess the local presence of wild individuals in an archaeological context (Mercuri et al., 2021). In environmental samples, the amount of grapevine pollen found is very low, such as in fossil pollen assemblages from peat or lakes, or soil samples (relative abundance around 0.2%), and a minimum of a thousand total land pollen grains from soil samples has been recommended to have a good estimate of the presence of the grapevine (Turner and Brown, 2004). Therefore, the presence of *Vitis* in environmental samples is considered as strong evidence for on-site presence of a vineyard (Turner and Brown, 2004). An approach relying on the ratio between the two pollen morphotypes would thus require counting large amounts of pollen grains for each sample. Our automated pollen analysis pipeline, for acquiring and identifying pollen images from environmental samples, can help investigate this approach more efficiently.

4.4.3 Perspective for this study : grapevine image datasets from the Pic St Loup

Several independent datasets collected in this PhD include pollen from wild and domesticated grapevine individuals, and could be used to improve the developed method aiming at differentiating pollen grains from both grapevine subspecies. The first dataset includes images of grapevine pollen collected from annual passive traps near 5 distinct wild individuals and in a vineyard (chapter III). A second dataset includes pollen collected from flowers of wild grapevine individuals from the Pic Saint Loup population, sampled during their blooming period of 2022 (May-June), and including pollen from 21 additional grapevine reference individuals. Finally, throughout a previous collaboration with a research team from Italy, Anna Maria Mercuri gave us slides mounted from reference pollen samples from wild and domesticated grapevine individuals. In a recent study from this research team, morphometric measurements were presented for some of these pollen samples (Clò et al., 2023). The slides provided were not scanned but are available at the ISEM, and could be analyzed for additional investigations using AI algorithms. Due to a lack of classification success (and a lack of time), the models trained on the reference images from the Vassal-Montpellier collection were not applied to these additional pollen grains. Instead, their images could be used to complement the training dataset and thus generate more robustly trained models, by relying on more images and from more contexts. All images could also be acquired using a smaller step between each images of the pile acquired for a field of view, for instance with a 1 µm step instead of 8 µm as done in this study. Augmenting that way the training dataset would allow to better assess the possibility of discriminating between the pollen from wild and domesticated grapevine individuals. Additionally, these datasets could also allow to address new questions, such as differences between pollen grown in distinct environments, or collected in distinct years.

Supporting Information (SI)

Fig. S1 : Distribution of the test images for each cross-validated models

For each cross-validation (referred as 'cval' in the figure), the dataset was split into three subsets, one for training, one for validation and one for testing. The distribution of pollen images from each grapevine individual reserved for independent testing for each cross-validation are presented. The top figure represent the experimental design A, the bottom figure represent the experimental design B. 'W' and 'D' refer respectively to wild and domesticated individuals.

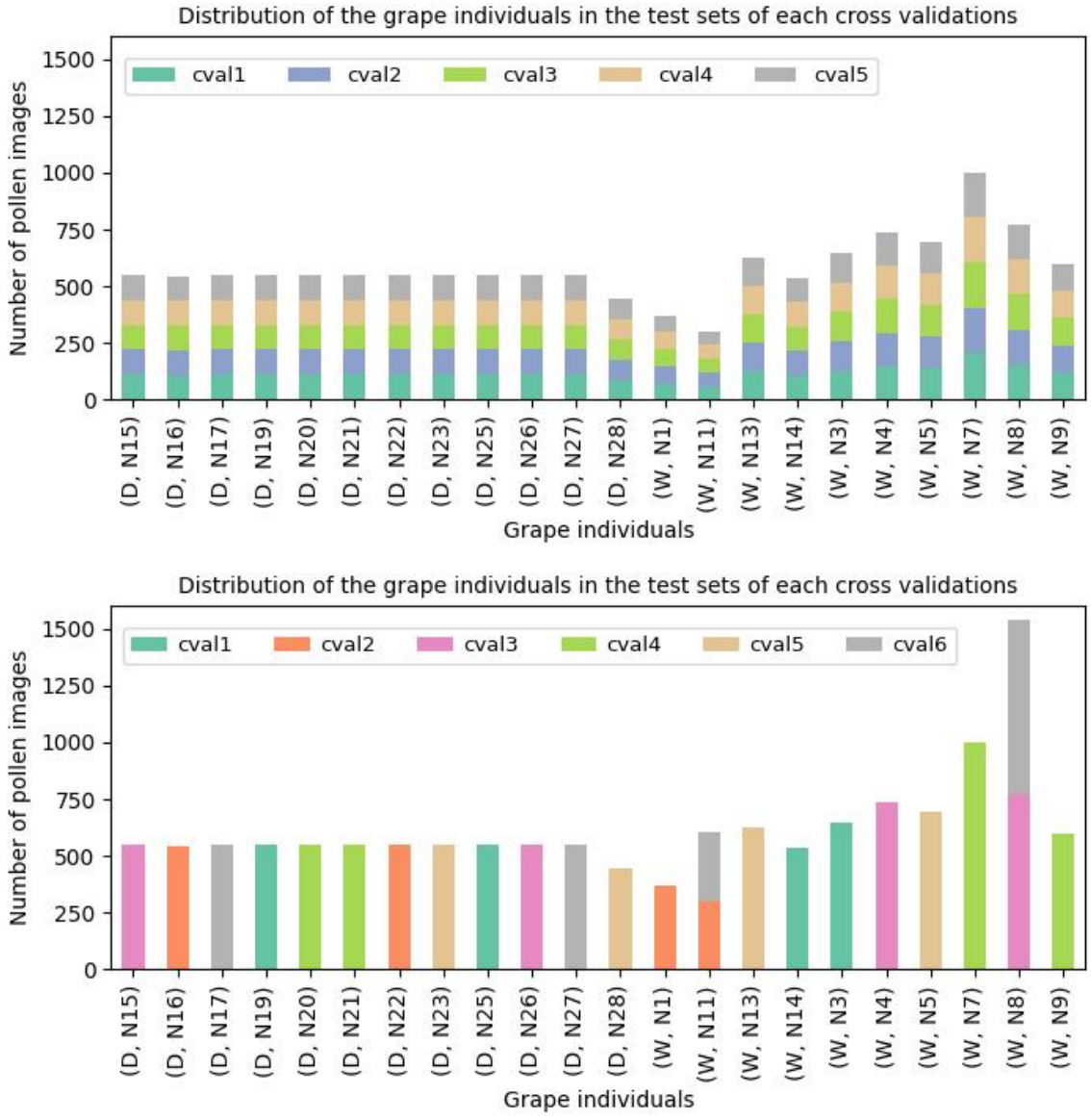


Figure S1

Fig S2 : Anova on the size of the pollen grains

Result from the analysis of variance (ANOVA) conducted on the size of the pollen grains produced by wild and domesticated individuals.

```
> mod1 <- lmer(data=df_size, surface ~ subsp + (1|individual))
> summary(mod1)
Linear mixed model fit by REML ['lmerMod']
Formula: surface ~ subsp + (1 | individual)
   Data: df_size

REML criterion at convergence: 389783.1

Scaled residuals:
    Min      1Q  Median      3Q     Max
-6.2418 -0.6492 -0.1005  0.5942 10.3962

Random effects:
 Groups     Name        Variance  Std.Dev.
 individual (Intercept) 136786535 11696
 Residual               325442042 18040
Number of obs: 17366, groups:  individual, 28

Fixed effects:
            Estimate Std. Error t value
(Intercept)    86694       3250  26.679
subspwild       8231       4442   1.853

Correlation of Fixed Effects:
          (Intr)
subspwild -0.732
> Anova(mod1)
Analysis of Deviance Table (Type II Wald chisquare tests)

Response: surface
       Chisq Df Pr(>Chisq)
subsp 3.4335  1    0.06389 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure S2

Fig. S3 : Confusion matrix obtained from the cross-validated models trained with the design A on the dataset '1pollen/1image'

Confusion matrix obtained from the cross-validated models trained with the design A, on the dataset '1pollen/1image', and using only two classes : 'wild' or 'domesticated'. Each cross-validation is indicated by the number in brackets; 'ind.' refers to individuals; 'W' refers to wild; 'D' refers to domesticated.

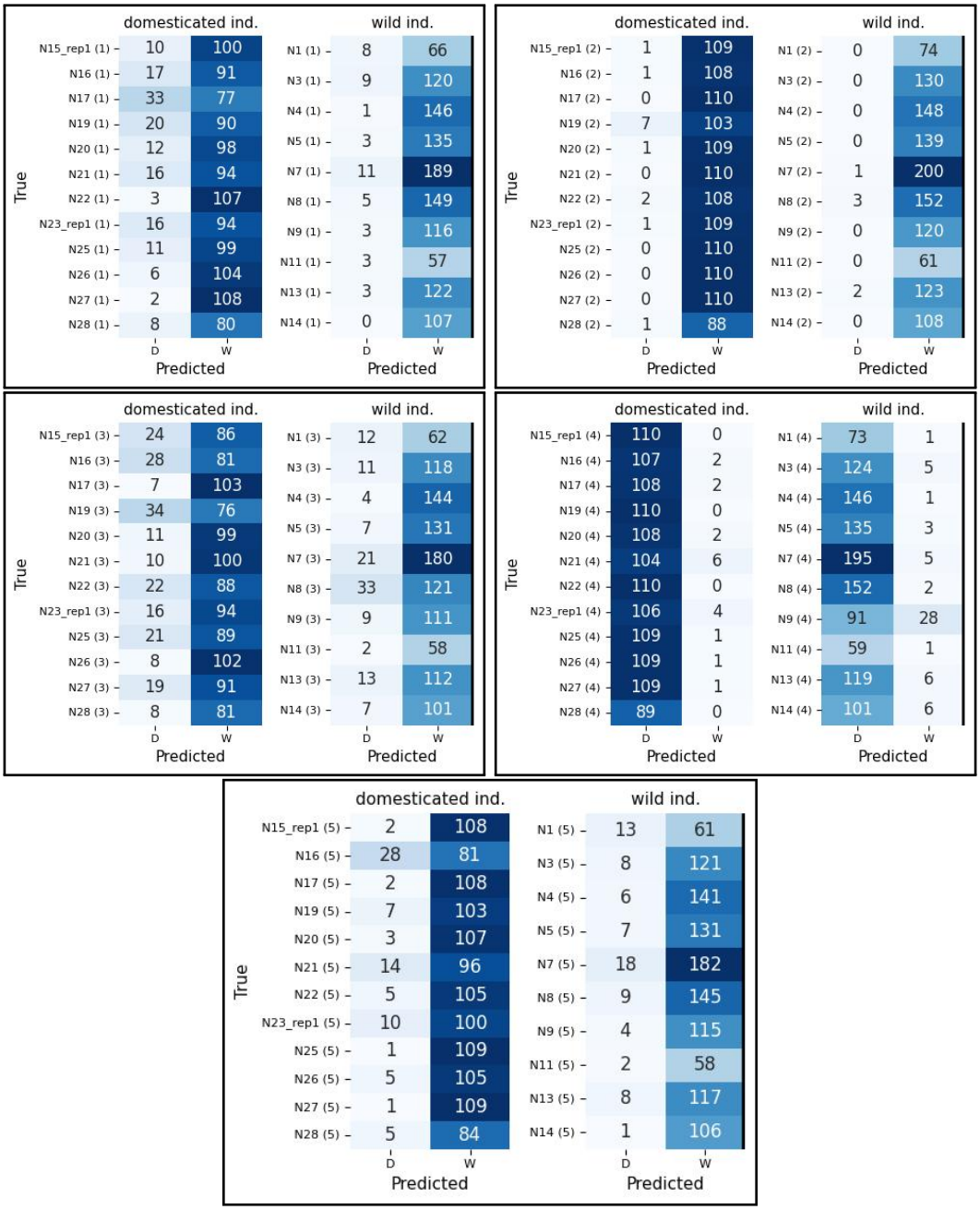


Figure S3

Fig. S4 : Confusion matrix obtained from the cross-validated models trained with the design A on the dataset '1pollen/4image'

Confusion matrix obtained from the cross-validated models trained with the design A, on the dataset '1pollen/4images' but tested only on the top1 sharpest image from the pile of 4, and using only two classes : 'wild' or 'domesticated'. Each cross-validation is indicated by the number in brackets; 'ind.' refers to individuals; 'W' refers to wild; 'D' refers to domesticated.

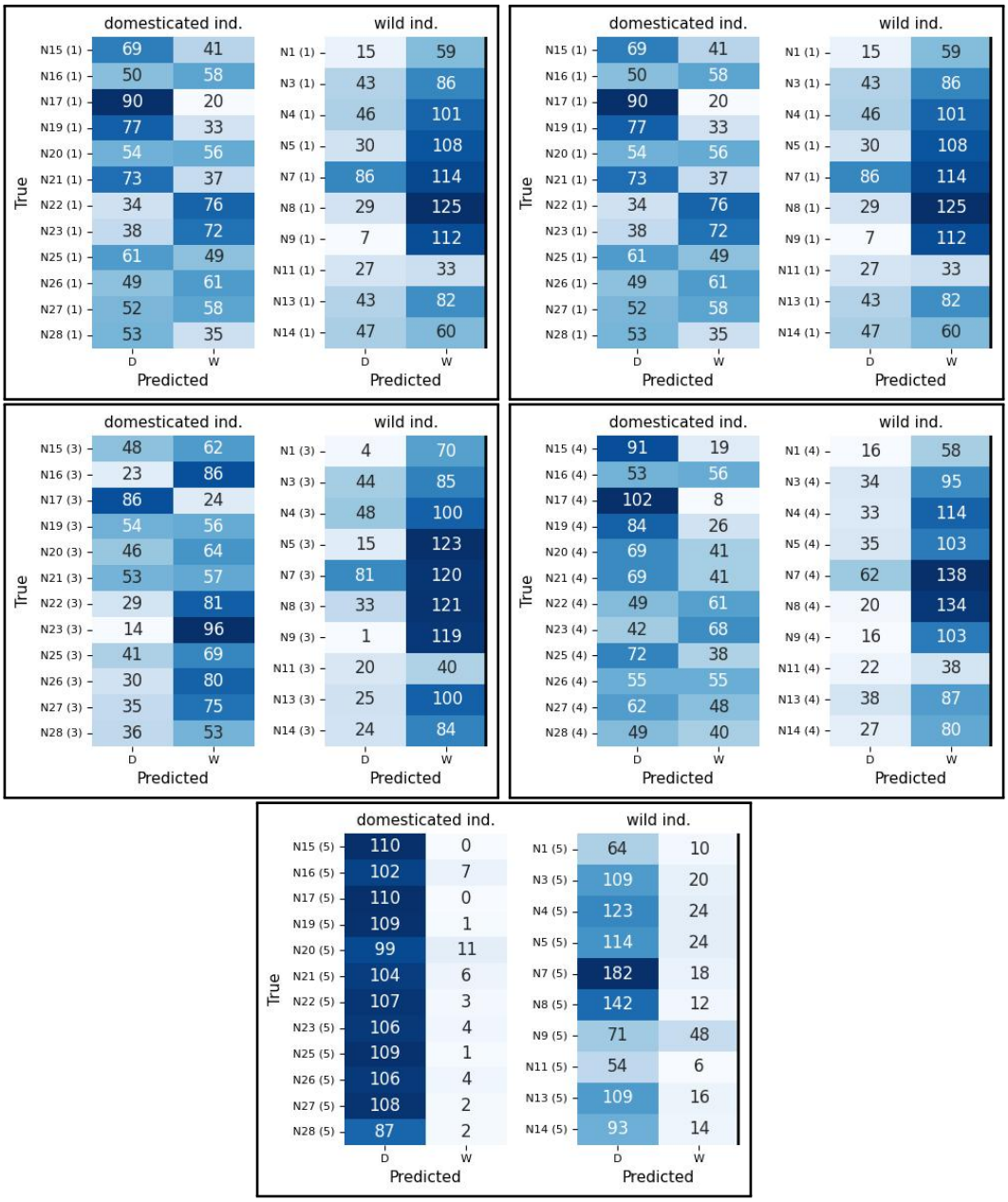


Figure S4

Table S1 : Grapevine individuals sampled from the ampelographic collection of Vassal-Montpellier

Information on all grapevine individuals from which the reference pollen samples were collected, including their subspecies, accession and produced pollen morphotype. 'morph.' refers to morphotype.

Table S1

Id no.	subsp.	Pollen morph.	Var. code	Var. no.	Intro. no.	Accession no.	Pheno location	DNA code
N1	wild	tricolporate	na	na	Vitis silvestris Dirmstein	8500Mtp38	Vass2R08s436-440	GrW1
N2	wild	inaperturate	na	na	Vitis silvestris Ketsch 27	8500Mtp34	Vass2R08s406-410	GrW2
N3	wild	tricolporate	na	na	Lambrusque Chadenac	0Mtp1819	Vass7R09s426-430	GrW5
N4	wild	tricolporate	na	na	Vitis silvestris Mannheim Nr 2	8500Mtp144	Vass7R04s446-450	GrW6
N5	wild	tricolporate	na	na	Lambrusque Ustaritz 3	8500Mtp276	Vass7R12s136-140	GrW8
N6	wild	inaperturate	na	na	V. silvestris Orroa desa ide 1	8500Mtp435	Vass7R06s016-020	GrW9
N7	wild	tricolporate	na	na	V. silvestris Monte Arudo 3	8500Mtp440	Vass7R06s041-045	GrW10
N8	wild	tricolporate	na	na	V. silvestris Valle grande 5	8500Mtp443	Vass7R06s056-060	GrW11
N9	wild	tricolporate	na	na	V. sylvestris Portugal 110104 (T)	8500Mtp412	Vass7R07s001-005	GrW13
N10	wild	inaperturate	na	na	V. sylvestris Portugal 110207 (3)	8500Mtp417	Vass7R07s021-025	GrW15
N11	wild	tricolporate	na	na	V. sylvestris Portugal 110304 (3)	8500Mtp419	Vass7R07s031-035	GrW16
N12	wild	inaperturate	na	na	Vitis vinifera subsp. sylvestris O 1-6 Asturias	8500Mtp369	Vass7R07s121-125	GrW18
N13	wild	tricolporate	na	na	Vitis vinifera subsp. sylvestris J 2-3 Jaén	8500Mtp378	Vass7R07s166-170	GrW19
N14	wild	tricolporate	na	na	Vitis vinifera subsp. sylvestris SS 5-4 Guipuzcoa	8500Mtp384	Vass7R07s196-200	GrW20
N15	domesticated	tricolporate	285	Grüner Veltliner	Velteliner vert	0Mtp1263	Vass3R05s166-170	GrD1
N16	domesticated	tricolporate	1602	Wildbacher blau	Wildbacher blau	1602Mtp1	Vass4R11s346-350	GrD2
N17	domesticated	tricolporate	1840	Amigne	Amigne	1840Mtp1	Vass4R02s301-305	GrD3
N18	domesticated	inaperturate	3477	Wildrebe von Bauer	Wildrebe	0Mtp1220	Vass4R10s366-370	GrD4
N19	domesticated	tricolporate	1446	Maturana blanca	Maturana	1446Mtp1	Vass6R09s031-035	GrD5
N20	domesticated	tricolporate	157	Corbeau	Corbeau	0Mtp1151	Vass4R11s576-580	GrD8
N21	domesticated	tricolporate	424	Lauzet	Lauzet	424Mtp2	Vass3R12s126-130	GrD9
N22	domesticated	tricolporate	1216	Mancin	Mancin	1216Mtp1	Vass3R10s026-030	GrD10
N23	domesticated	tricolporate	328	Petit Verdot	Petit Verdot	0Mtp1451	Vass2R02s341-345	GrD11
N24	domesticated	inaperturate	2045	Lambrusco di Sorbara	Lambrusco di Sorbara	2045Mtp1	Vass6R04s116-120	GrD12
N25	domesticated	tricolporate	1353	Colorino	Colorino	1353Mtp2	Vass6R05s106-110	GrD14
N26	domesticated	tricolporate	2047	Lambrusco Maestri	Lambrusco Maestri	2047Mtp3	Vass6R04s111-115	GrD16
N27	domesticated	tricolporate	3380	Lameiro	Lameiro	3380Mtp2	Vass5R09s431-435	GrD18
N28	domesticated	tricolporate	1488	Tinto Cao	Tinto Cao	0Mtp358	Vass4R11s526-530	GrD19

Table S2 : Classification metrics obtained from the trained CNNs on pollen from wild and domesticated grapevine individuals

Metrics obtained from the classification model ResNet152 trained on two classes : wild and domesticated, and using the design A (all individuals are equally represented in all subsets) and the design B (leave-1-out strategy, i.e. all pollen from 4 individuals are reserved for testing). Metrics are presented per cross-validation and averaged over all cross-validations, and respectively over the full test sets or for each of the two classes separately. '2D' refers to models trained and evaluated on the dataset '1pollen/1image'. '3D-top1' refers to models trained on the dataset '1pollen/4images' using the pile of 4 images and evaluated on the sharpest image of the pile.

Table S2

design	data	images	metric	cval1	cval2	cval3	cval4	cval5	cval6	mean	stderr
A	2D	all	**accuracy**	53.39	49.59	52.68	52.41	49.39	na	**51.49**	0.746
A	2D	wild	recall	96.33	99.52	90.53	4.63	93.93	na	76.99	16.234
A	2D	wild	precision	51.38	49.43	51.08	75.32	49.21	na	55.28	4.498
A	2D	wild	f1score	67.02	66.05	65.31	8.72	64.58	na	54.34	10.206
A	2D	domesticated	recall	11.88	1.08	16.02	98.54	6.39	na	26.78	16.202
A	2D	domesticated	precision	77.0	70.0	63.61	51.7	52.2	na	62.9	4.426
A	2D	domesticated	f1score	20.59	2.12	25.6	67.82	11.39	na	25.5	10.12
A	3D-top1	all	**accuracy**	61.99	66.51	57.03	68.48	56.8	na	**62.16**	2.136
A	3D-top1	wild	recall	70.23	85.96	76.53	75.82	15.32	na	64.77	11.286
A	3D-top1	wild	precision	59.62	61.45	54.5	65.47	82.4	na	64.69	4.263
A	3D-top1	wild	f1score	64.49	71.67	63.67	70.27	25.84	na	59.19	7.586
A	3D-top1	domesticated	recall	54.01	47.61	38.14	61.4	96.84	na	59.6	9.002
A	3D-top1	domesticated	precision	65.24	77.74	62.66	72.45	54.23	na	66.46	3.626
A	3D-top1	domesticated	f1score	59.1	59.05	47.41	66.47	69.52	na	60.31	3.419
B	2D	all	**accuracy**	47.61	52.24	41.33	40.79	44.33	50.09	**46.07**	1.747
B	2D	wild	recall	0.0	34.13	3.25	2.0	11.09	93.28	23.96	13.499
B	2D	wild	precision	0.0	36.35	40.83	50.79	55.73	49.7	38.9	7.573
B	2D	wild	f1score	0.0	35.2	6.02	3.85	18.5	64.85	21.4	9.264
B	2D	domesticated	recall	98.82	63.35	93.55	97.18	88.33	8.0	74.87	13.129
B	2D	domesticated	precision	47.89	61.06	41.36	40.55	42.87	55.0	48.12	3.097
B	2D	domesticated	f1score	64.51	62.18	57.36	57.23	57.73	13.97	52.16	7.062
B	3D-top1	all	**accuracy**	39.29	48.56	53.14	54.43	44.11	67.09	**51.1**	3.597
B	3D-top1	wild	recall	38.04	43.52	49.01	59.04	34.19	0.0	37.3	7.545
B	3D-top1	wild	precision	40.8	35.57	62.0	62.15	51.43	0.0	41.99	8.663
B	3D-top1	wild	f1score	39.37	39.14	54.74	60.55	41.08	0.0	39.15	7.884
B	3D-top1	domesticated	recall	40.64	51.65	58.82	47.73	57.24	67.09	53.86	3.45
B	3D-top1	domesticated	precision	37.88	59.85	45.69	44.49	39.65	100.0	54.59	8.777
B	3D-top1	domesticated	f1score	39.21	55.45	51.43	46.05	46.85	80.3	53.22	5.349

Table S3 : Complementary classification metrics obtained from the trained CNNs on pollen from wild and domesticated grapevine individuals

Complementary metrics (not presented in the main document) obtained from the classification model ResNet152 trained on the dataset '1pollen/4images' (i.e. a pile of 4 images for each pollen grain), using two classes : wild and domesticated, and with the design A (all individuals are equally represented in all subsets) and the design B (leave-1-out strategy, i.e. all pollen from 4 grapevine individuals are reserved for testing). Metrics are presented per cross-validation and averaged over all cross-validations, and respectively over the full test sets or for each of the two classes separately. '3D-4' refers to models trained and evaluated on all images, including all 4 images of the piles associated to each pollen grain, all images from each pile are processed separately. '3Dsumscores' refers to the same process as '3D-4' but the model scores predicted for each of the 4 images of each pile are aggregated to generate a unique prediction per pollen grain. '3D-top2' refers to models evaluated only one the second sharpest image of the pile, '3D-top3' on the third sharpest image, and '3D-top' on the fourth sharpest image.

Table S3

design	data	images	metric	cval1	cval2	cval3	cval4	cval5	cval6	mean	stderr
A	3D-4	all	**accuracy**	61.03	63.11	57.38	67.78	54.47	na	**60.75**	2.058
A	3D-4	wild	recall	80.55	80.04	83.75	73.34	9.62	na	65.46	12.578
A	3D-4	wild	precision	57.38	59.32	54.34	65.32	80.6	na	63.39	4.169
A	3D-4	wild	f1score	67.02	68.14	65.91	69.1	17.18	na	57.47	9.021
A	3D-4	domesticated	recall	42.17	46.67	31.84	62.4	97.77	na	56.17	10.291
A	3D-4	domesticated	precision	69.16	70.64	66.92	70.8	52.84	na	66.07	3.023
A	3D-4	domesticated	f1score	52.39	56.21	43.15	66.34	68.6	na	57.34	4.172
A	3Dsumscores	all	**accuracy**	63.75	67.25	59.22	73.03	52.61	na	**63.17**	3.108
A	3Dsumscores	wild	recall	90.98	89.14	91.17	81.32	3.83	na	71.29	15.17
A	3Dsumscores	wild	precision	58.43	61.59	55.18	69.18	92.31	na	67.34	5.957
A	3Dsumscores	wild	f1score	71.16	72.85	68.75	74.76	7.36	na	58.97	11.576
A	3Dsumscores	domesticated	recall	37.42	45.99	28.27	65.02	99.69	na	55.28	11.317
A	3Dsumscores	domesticated	precision	81.1	81.34	76.78	78.29	51.78	na	73.86	4.996
A	3Dsumscores	domesticated	f1score	51.21	58.76	41.33	71.04	68.16	na	58.1	4.89
A	3D-top2	all	accuracy	60.81	63.03	58.0	68.13	55.19	na	**61.03**	1.977
A	3D-top3	all	accuracy	61.55	61.43	56.83	67.74	53.9	na	**60.29**	2.109
A	3D-top4	all	accuracy	59.79	61.47	57.65	66.76	51.98	na	**59.53**	2.16
B	3D-4	all	**accuracy**	42.79	50.18	44.65	55.44	47.85	57.62	**49.76**	2.189
B	3D-4	wild	recall	54.92	34.17	30.69	67.31	35.75	46.48	44.89	5.286
B	3D-4	wild	precision	45.68	34.38	53.75	61.28	56.7	58.96	51.79	3.763
B	3D-4	wild	f1score	49.88	34.27	39.07	64.15	43.85	51.98	47.2	3.951
B	3D-4	domesticated	recall	29.75	60.01	63.8	38.18	63.86	68.48	54.01	5.956
B	3D-4	domesticated	precision	38.03	59.78	40.17	44.55	42.88	56.76	47.03	3.369
B	3D-4	domesticated	f1score	33.38	59.89	49.3	41.12	51.31	62.07	49.51	4.08
B	3Dsumscores	all	**accuracy**	42.79	50.18	44.65	55.44	47.85	57.62	**49.76**	2.189
B	3Dsumscores	wild	recall	54.92	34.17	30.69	67.31	35.75	46.48	44.89	5.286
B	3Dsumscores	wild	precision	45.68	34.38	53.75	61.28	56.7	58.96	51.79	3.763
B	3Dsumscores	wild	f1score	49.88	34.27	39.07	64.15	43.85	51.98	47.2	3.951
B	3Dsumscores	domesticated	recall	29.75	60.01	63.8	38.18	63.86	68.48	54.01	5.956
B	3Dsumscores	domesticated	precision	38.03	59.78	40.17	44.55	42.88	56.76	47.03	3.369
B	3Dsumscores	domesticated	f1score	33.38	59.89	49.3	41.12	51.31	62.07	49.51	4.08
B	3D-top2	all	accuracy	40.69	57.17	56.78	57.95	46.06	62.71	**53.56**	3.11
B	3D-top3	all	accuracy	43.93	51.61	42.64	58.98	50.61	71.64	**53.24**	4.014
B	3D-top4	all	accuracy	47.26	59.97	49.43	50.39	50.61	56.81	**52.41**	1.821

Chapter 5

Discussion and perspectives

In this thesis, automated pollen analysis was investigated using artificial intelligence (AI) to generate pollen data and (1) assess spatial and temporal changes in key Mediterranean plant communities, and (2) investigate potential differences in the pollen signature of wild and domesticated grapevines. In the frame of the Pollimed Project (OSU OREME), a first study conducted in Southern France focused on pollen rain monitorings of (1) oak dominated and open scrublands, (2) a relic population of wild grapevine and a commercial vineyard (domesticated grapevine). The monitoring, based on pollen traps, was still recent (2019-....) but intended to be conducted in the long term in Southern France but also in other mediterranean countries (POlarise project). Here, we aimed to provide preliminary results and to set the foundation for its onward deployment.

5.1 Discussing the methodological developments made in the thesis

5.1.1 Main objectives and initial challenges for methodological developments

The vegetation monitoring is expected to generate large amounts of samples. Additionally, extensive amounts of pollen grains per sample had to be analyzed, first in the annual traps, to better represent the grapevine that is scarce in pollen samples, and second in the reference samples to better account for the variability of each sampled reference individual. Finally, pollen from the wild and domesticated grapevines are very similar, so that new methods had to be investigated to test for possible discrimination, compared to traditional manual observations and descriptions. Artificial intelligence (AI), and namely convolutional neural networks (CNN), was thus investigated to develop automated pollen analysis methods for Mediterranean vegetation. The pollen samples displayed two very distinct properties, targeting two distinct but complementary goals :

-(1) in order to monitor key anemophilous Mediterranean plants in Southern France, we wanted to get pollen counts per taxa from environmental samples collected from annual passive traps, thus generating challenging images for automated analysis due to uncontrolled taxonomic pollen diversity, damaged grains, non-pollen particles, noisy images with pollen lacking visibility ;

-(2) to discriminate morphologically very similar pollen morphotypes, we needed to process reference pollen samples of wild and domesticated grapevine individuals, thus with samples containing only one pollen taxon and almost no debris, but for which a more detailed analysis of the grains has to be conducted.

These two methodological approaches : get pollen counts per taxon from environmental samples and describe pollen morphotypes are complementary, and major foundations of palynology.

When this PhD started, in 2021, developed methods were not yet ready to handle environmental samples and generate pollen data usable to conduct research in environmental-related disciplines, i.e. beyond methodological developments. Several challenges had yet to be solved : the few studies addressing the issue of damaged pollen grains had been conducted under controlled conditions (e.g. on fossil pollen, Bourel et al., 2020); methods to handle non-pollen particles such as debris in the samples had only been addressed by allergists (Crouzy et al., 2022) and thus from volumetric traps collecting few debris compared to typical sample in ecology or paleo-ecology such as passive traps, moss pollsters or fossil cores; the uncontrolled taxa diversity occurring in such samples was also not previously addressed; the first detection step of automated pollen analysis remained largely overlooked and still conducted with standard machine learning approaches that performs poorly on the noisy images obtained from environmental samples (Diaz-Lopez et al., 2015). Altogether, aside from allergology, conditions considered to train and, most importantly, evaluate the models were explored under theoretical contexts, with image datasets not reflecting conditions encountered when studying pollen samples for research in environmental-related disciplines. Consequently, providing automated pollen data to address questions in ecology or paleoecology beyond method developments was still a challenge. Automated methods were thus not ready to process full environmental samples, without intervening manually to filter and clean the image datasets. Considering the biological objectives for this PhD, the method developments had thus to address these challenges related to routine application to the environmental samples of the vegetation monitoring.

5.1.2 An automated pollen analysis pipeline to monitor Mediterranean vegetation

Considering the passive pollen traps from the vegetation monitoring, a fully automated pollen analysis pipeline was developed, and included automated steps for image acquisition (Chapter I), detection, i.e. locating pollen grains in the images (Gimenez et al., 2024 in Chapter II), and identification, i.e. classifying all automatically detected particles and generating pollen counts for target taxa (Chapter III).

5.1.2.1 Methodological developments and main results

The detection step was the object of the first study in this thesis (Gimenez et al., 2024 in Chapter II), developments were guided by two main statements drawn from the literature. First, developments on detection were significantly late compared to those for classification, and still relied on methods that performed poorly on noisy images typical for environmental samples; yet, errors generated at the detection stage would propagate in a fully automated pipeline. Second, developments were rarely presented for easy implementation by other researchers, especially palynologists usually with no expertise in machine learning, and thus limiting a broader adoption of the developments made and of automated palynology. The objectives in the first conducted study were thus two-folds : to address the challenging step of detection on noisy images from environmental samples, and to make automated pollen analysis more accessible for those interested in pollen-based studies. In that frame, a user-friendly, open-source algorithm with online general application guidelines was selected (YOLOv5, guidelines at github.com/ultralytics/yolov5, Jocher, 2020) and several strategies in using this algorithm were investigated, for instance in building the training dataset to retrieve distinct types of pollen data. The method proved very effective to detect pollen, even in noisy images, solving previous limitations when encountering clumped, covered or blurry pollen grains

occurring with sometimes very similar non-pollen particles. This method displayed similar detection errors between taxa, except for grains with rare morphologies left undetected slightly more frequently (e.g. *Lycopodium*, but <1%), and thus did not introduce major biases in the detected pollen assemblage. Although the objective with that algorithm was initially to automate both detection and classification, the image dataset composed of full microscope field of views was inadequate to compile classes for most pollen taxa, due to their scarcity in the images and to many grains with uncertain identification. Using the algorithm as a detector only, and without labeling pollen grains with their taxa, while conducting the classification separately, was found to be the most efficient strategy with that type of images.

The second development of the thesis aimed to obtain automated pollen counts for 15 key Mediterranean taxa, typical of open and close scrublands (e.g. *Quercus*, *Juniperus*, *Pistacia*, *Phillyrea*) and of *Vitis*, (Chapter III). The identification method had to deal with an uncontrolled taxa diversity, some non-pollen particles incorrectly detected, and pollen images lacking identification criteria (e.g. when unfocused or covered by debris). The adopted strategy involved (a) a training dataset composed of images identified from the vegetation monitoring trap samples and complemented by references, to represent a maximum of particle types expected for the routine application, (b) several cross-validated classification models (ResNet152, He et al., 2015) combined with an ensemble modeling approach, and (c) complementary filters, such as thresholds in prediction occurrences, to account for uncertainty in predictions. A second challenge was to evaluate how the method would perform when applied under routine conditions, i.e. how automated pollen counts generated by the entirely automated pipeline applied on new monitoring samples reflected the true pollen content from the samples, including outside of training/testing data. Results show that the most important classes, including the 15 target taxa, *Lycopodium*, as well as non-pollen particles, were classified with high performances with only few exceptions. Taxa from the same family were misclassified with each other due to their similar morphological features: *Phillyrea* and *Fraxinus* (Oleaceae), and *Juniperus* and *Cupressus* (Cupressaceae). Note that *Lycopodium* spores, which identification is crucial for calibrating the samples, and thus comparing pollen counts per taxon across sampling locations and years, reached one of the highest classification performance, likely granted to its distinctive features. The number of images per class used in training also affected predictions evaluated on the test sets, by predicting less frequently some pollen taxa with small classes, and even compared to their actual smaller distribution in testing data. Finally, the addition of images from reference samples contributed to improve the classification performances for several classes. However it also generated confusion in few cases (e.g. Rosaceae and *Quercus*) probably caused by differences in the image aspects (linked to chemical treatment, slide mounting protocol, or time since preparation, and resulting in distinct contrasts, transparency or distortion in the pollen images).

5.1.2.2 Quality of the pollen data automatically generated by the fully automated pipeline

The pipeline combined four automated steps, each with their own advantages and limitations, including : (a) image acquisition with an automated microscope, (b) detection with an object-detection CNN, (c) classification with a classification CNN, and (d) integration of classifiers to generate calibrated pollen counts per samples. This full pipeline was applied to all images obtained from the vegetation monitoring samples available at the time (2019-2023). No ground truth were available for all the new identifications during this application. To assess the quality of the automated pollen counts, our results were thus complemented by both field observations and manual counts

done by a palynologist. Automated counts displayed very distinctive patterns in the pollen assemblages between the sampling locations, and reflected the main vegetation types observed on site (e.g. Mediterranean meadow, open scrubland, vineyard, mixed oak forest, open scrubland on rocky slopes, Fig. 1 from Chapter III), and the local presence of some taxa (e.g. *Fraxinus* near the intermittent watercourse, or local *Buxus* shrubs). Automated pollen counts were also consistent with manual counts for several abundant taxa. Large discrepancies were however observed. For *Phillyrea* and *Fraxinus*, most of the signal obtained from automated counts remained in line with field observations, i.e. more *Fraxinus* pollen was identified at the intermittent riparian area where *Fraxinus* trees do occur (sampling site PSLH), while more *Phillyrea* pollen was identified in the sampling site PSL11, where *Phillyrea* shrubs were observed near the pollen traps location. However, their abundances were probably over-estimated for one taxon and under-estimated for the other at each location where one of the two occurs. This is a main limitation of the developed method, as even if displaying very similar pollen, both taxa have largely distinct ecological implications, one being characteristic of xerophilous scrubland, while the other being indicative of more humid conditions. The occurrence of false positives was also another issue of the method, as it prevented for determining with certainty the presence of a taxon counted with a small abundance due to the possibility of a false signal, as it was the case for the sterile morphotype of *Vitis* in sampling locations far from a female grapevine individual (where $< 1\%$ pollen was found in these samples), or for *Buxus* identified in the riparian area. These potential errors might have also been driven by the class distribution in the training data, which is another limitation from this method, relying on a standard supervised classification algorithm, such as ResNet. As suggested by the evaluation of the classification models on their respective test sets, predictions tended to be drawn by the class distribution in the training data, and thus by a unique class distribution, with always similar amounts of *Fraxinus* and *Phillyrea*, and with large amounts of *Vitis* and *Buxus*. In this study, we combined several classification models and used a threshold of occurrences in the predictions to account for uncertain predictions; this filter might have not be sufficient to remove all false positives, although it is not excluded that the identified pollen taxa do occur in small abundances in the processed samples.

To conclude, general trends in the pollen assemblage from automated pollen data were relevant. Automated pollen data informed well on the general abundances of the target pollen taxa, on the main vegetation types occurring near the sampling locations, and on the large and marked changes occurring between the sampling sites and years (e.g. disappearance of *Buxus* in three years due to the Pyrale invasion). The automated method remained limited to provide more detailed information on the pollen contents in each sample.

5.1.3 Automated methods to analyze reference pollen samples from the grapevine

The third study of the thesis (Chapter IV) focused on the wild (*Vitis sylvestris*) and domesticated (*Vitis vinifera*) grapevines. Here, automated image analysis was investigated to (1) quantify pollen production in grapevine flowers, (2) extract morphological measurement, i.e. estimate pollen grains sizes, and (3) attempt to discriminate very similar pollen morphotypes with a classification CNN. Our pipeline was adjusted to focus only on the pollen of *Vitis* and new classification models had to be trained. The previously developed detection method was already set for the identification of pollen versus *Lycopodium* (Chapter II, Gimenez et al., 2024). Thus, directly applying this method to the reference images already allowed to retrieve several information. First,

it allowed to quantify the amount of pollen grains in each sample and thus to estimate the pollen production per flower of both subspecies. Second, it allowed to estimate the size of pollen grains using as proxy the dimensions of the detected bounding boxes. Third, it allowed to build a new classification dataset of many individual grapevine pollen images. The direct application of that method was very effective in that context: it was fast, easy to implement, did not require any new developments, and generated pollen data that could already be used to address the biological objectives of the study, i.e. compare and identify potential differences between the two grapevine subspecies from a palynological perspective. Differences in pollen sizes and productions were identified, but the classification of the images with the ResNet152 algorithm was not effective. One possible explanation for this result is the lack of any morphological differences between pollen grains produced by *Vitis sylvestris* or *Vitis vinifera*. However, it could also be related to the employed method, relying on light microscopy (LM) images and on a standard supervised CNN, that might be ineffective to capture very subtle differences between pollen morphotypes. This method, especially considering the imaging technique, was indeed relatively limited compared to state of the art research conducting this type of analysis. Studies describing pollen morphotypes are often based on scanning electronic microscopy (SEM), as it provides much more visible details compared to LM, for instance on the exine textures (e.g. folveolate, rugulate) that was showed to differ between pollen samples for *Vitis* or between some species of Vitaceae (İnceoğlu et al., 2000; Marasali et al., 2005; Gallardo et al., 2009; Cartaxo-Pinto et al., 2017; Mercuri et al., 2021), and although no study has yet systematically compared pollen from *Vitis sylvestris* and *Vitis vinifera*. Additionally, most related studies investigating AI to differentiate morphologically close pollen morphotypes, are increasingly conducted using super-resolution and three-dimensional imaging, such as confocal laser scanning microscopy. Confocal microscopy has been used to improve taxonomy of fossil specimens, or for phylogeny placement of species (Punyasena et al., 2012; Kong et al., 2016; Romero et al., 2020; Adaïmé et al., 2024; Collevatti et al., 2024), and is becoming a reference technique for this type application. These imaging approaches have great potential for our objective of discriminating the pollen grains produced by wild and domesticated grapevines.

Finally, this study also illustrates another and complementary use of the detection method developed in the first study of this thesis (Gimenez et al., 2024 in Chapter II). This method can be easily implemented by others and can already provide relevant pollen data for research such as counting pollen in samples, for plant productivity assessments, or to extract large amounts of pollen images on which further morphometric measurements can be conducted.

5.1.4 Development of the automated image acquisition system

The developments made during my PhD relied on a standard optic microscope Leica with a ×63 objective under oil immersion, equipped with an automated stage and camera. From these tools, the acquisition step consisted in scanning multiple separate field of views (FOV) per slide, and by acquiring for each FOV a pile of images at distinct focal depths, then stacked into one image per FOV. This process generated grey-level images, with high magnification (FOVs of 214µm × 214µm), which could contain several pollen and non-pollen particles, and displayed a good resolution (2048 pixels × 2048 pixels per FOV). The high resolution, achieved with the used objective, allowed a good visibility of the identification criteria of the grains, which facilitated the process of building the training datasets, especially for the classification. Nonetheless, as the size of pollen grains

ranged around 20-50 µm, or even more (e.g. Pinaceae pollen that can exceed 100 µm), a large amount of pollen grains were truncated on the margins of the FOV images, which was an challenge for their identification and on how to deal with them. The high magnification also required significantly more images (both FOV and focus levels per FOV) for scanning a given slide surface compared to a ×40 objective for instance. The acquisition process was consequently time consuming (around 24 hours to scan a full slide), and required hundreds of Giga to store the generated images, which is a significant constrain for effective deployment of the method. Besides, the resolution of the images is largely reduced when implemented in the CNNs, in this study : to 640 pixels for each side for the detection (full FOV images), and to 224 pixels for each side for the classification (cropped pollen images). Consequently, the high quality image resolution was not exploited by the models. The method was kept throughout the thesis for consistency. However, using a lower magnification, for instance with a ×40 objective, might not affect the models performance while it can make the image acquisition process significantly more efficient and can reduce the amounts of truncated pollen grains.

Many other imaging techniques have also been investigated for automated palynology and several display great potential for the task. Whole slide imaging techniques with slide scanners (e.g. Nanozoomer, Punyasena et al., 2022) generate one single image covering the full slide, thus avoiding truncating pollen grains, and in very short amounts of time (<1 hour per slide). Flow cytometry avoids the step of detection and the issue of covered pollen grains, by processing each particle separately, and proved efficient in palynology (Dunker et al., 2021), including samples from lake sediments (Barnes et al., 2023). Scanning electronic microscopy (C. Li et al., 2023), or confocal microscopy (Adaïmé et al., 2024) can capture much more morphological features in the pollen grains, and are thus adequate for studies comparing close morphotypes. Imaging techniques also extend to elastic light scattering microscopy (Daunys et al., 2024; Brdar et al., 2023) or digital holography images (Sauvageat et al., 2020), which were also tested on airborne pollen samples.

The imaging system used in this thesis had the advantage of relying on a standard material (optic microscope) that is common in any palynology laboratory, at the difference of its automated stage. It is thus more accessible compared to these other imaging techniques, both in term of cost and ease of use. Relying on simple and common tools as done in this work might have more potential to make automated palynology accessible and utilizable to a broader part of the community.

5.2 Recent and ongoing advances in automated palynology, and crucial role of palynologists

5.2.1 The state of the art has recently changed towards real-world application

When the thesis started in 2021, there was no development designed to fully automate pollen analysis for environmental samples (from traps, moss pollsters or lake sediments), and to address questions for environmental-related research. These conditions, referred here as routine application conditions, were identified as the ultimate goal of automated palynology (by K. A. Holt and Bennett, 2014). In that frame, several challenges were initially identified, related to the samples content (many taxa, unknown taxa, debris, etc.), and to the objective of getting a fully automated method (i.e. not relying on any intermediary manual intervention to filter the images and cleaning the dataset). Since 2021, the interest of the community has largely switched towards that objective of developing methods for routine application, and the state of the art has radically changed over the last three years. Studies have now been conducted on fossil core samples (Barnes

et al., 2023; Theuerkauf et al., 2023; Durand et al., 2024; von Allmen et al., 2024), vegetation monitoring samples (Punyasena et al., 2022, this thesis), and have been extensive in aeropalynology with most studies conducted in the discipline (Grant-Jacob and Mills, 2022; Levetin et al., 2023; Tummon, Adams-Groom, et al., 2024). Consequently, the challenges previously identified have been increasingly addressed and many issues were solved, as listed in the four points bellow :

(1) For detection, this step was previously overlooked and the used standard machine learning algorithms were inadequate for noisy images as encountered for environmental samples. In the 2020's, deep learning algorithms designed for this task (e.g. R-CNN, YOLO, RetinaNet) were tested with success for environmental pollen samples as they can handle overlapping objects, pollen of multiple morphologies, or non-pollen objects (Theuerkauf et al., 2023; von Allmen et al., 2024; Gimenez et al., 2024 - Chapter II). These algorithms are thus getting increasingly adopted, in aeropalynology (Cao et al., 2020a; Jin et al., 2023) but also in paleoecology to process samples from peat or lake sediments (Theuerkauf et al., 2023; von Allmen et al., 2024) or in ecology to study modern pollen rains (Punyasena et al., 2022; Gimenez et al., 2024 - Chapter II; Zhang et al., 2024; Chapter IV), and major limitations related to this stage can be considered solved.

(2) For classification, methods were not developed to deal with the diversity of particle types and of image qualities encountered in routine for environmental samples. Many recent studies have now been designed with the objective of identifying key pollen taxa under such conditions, i.e. among non-target taxa, unknown or indeterminable pollen grains, and non-pollen particles such as spores and debris. The most frequent solution has been to generate class specifically for the undesired classes (e.g. non pollen), which was tested on airborne samples (Zhao et al., 2022), but also in paleoecology (Theuerkauf et al., 2023; Durand et al., 2024), and for pollen traps (Chapter III). Other innovative approaches have also been investigated, such as designing models that can identify new pollen taxa to higher taxonomic ranks based on morphological criteria from the known taxa (Barnes et al., 2023). A large number of issues have thus be solved for this task, but several remain considering the objective of generating full and accurate pollen counts per taxon from environmental samples.

(3) These previous advances have now lead researchers to start investigating automatically generated pollen data for their research beyond method developments. In allergology, automated methods notably implemented in commercialized automated devices are already used to monitor airborne allergen pollen (Buters et al., 2024, Tummon, Adams-Groom, et al., 2024). In paleoecology, an automated method was used for the placement of new fossil Podocarpidites specimens (Adaïmé et al., 2024). In this thesis, pollen data from the vegetation monitoring were assessed to provide preliminary results on spatial and temporal changes in the pollen assemblages (Chapter III).

(4) A final challenge initially identified was the limited accessibility of developments and of user-friendly methods for researchers interested in conducting pollen-based studies. Over the last three years, making automated pollen analysis more easily accessible has been an objective that has become widespread. Methods are now often made available, and online tools or practical studies have been proposed. For instance, the Tofsi-POST online interface allows to build a dataset, train the models and run them on new data (Theuerkauf et al., 2023), and few methods simple to implement were presented for pollen detection and to quantifying pollen grains in samples (Gimenez et al., 2024 - Chapter II; Zhang et al., 2024; Chapter IV).

Overall, the objectives pursued in this thesis aligned with ongoing developments and research in the field of automated pollen analysis. The thesis thus contributed to address

key and current challenges of the community, and participated to the ongoing transition from methodological developments, to real-world applications, especially for environmental-related disciplines.

5.2.2 Remaining challenges for automated palynology applied to real-world environmental-related research

Despite extensive advancements in automated palynology towards methods designed for environmental samples, some challenges remain, especially for the classification. A major identified challenge arising at that stage of developments is related to the capacity of generalization of the methods beyond the training and testing data, such as trap samples from a new collection year, moss pollsters from a new location, or new levels in a fossil sequence. So far, supervised CNNs have been the most widespread investigated methods. They are very efficient when applied to data that remains close to the training data, but can lose significant performance if the application data differs, regarding both (1) the quality of the images related to distinct chemical treatments, slides mounts, preservation conditions, microscope acquisition and resolution for instance, and (2) the class distributions related to distinct pollen assemblage from one environmental sample to another.

The generalization capacity of the developed method was investigated in the first study on detection (Gimenez et al., 2024 - Chapter II) by comparing performances when applying the models to images from distinct sampling year in the vegetation monitoring. In that context, performances remained similar, indicating a good generalization ability. This result could be explained by particularly efficient and strong data-augmentation associated to the YOLOv5 algorithm, for instance with the mosaic data loader functionality that combines portions of several images. The method also remained very effective when applied to the reference images from the grapevine, which was expected as reference images are simpler to process (Chapter IV). The identification method developed for the vegetation monitoring was more limited by the generalization ability on the classification CNNs, as suggested by the sometimes ineffective, or even confusing, addition of reference pollen images from the collection of palynology of ISEM. Similar results with an inefficient use of a reference image dataset for the identification of pollen on honey samples was previously showed (Tsiknakis et al., 2022). These results were explained by differences in the images in terms of contrast, transparency, grain distortions, and caused by differences in sample properties and slide processing, even when the images displayed the same pollen taxon. Using more extensive data-augmentation transformations, to train the models on a wider range of contrasts, transparency, blur could alleviate this issue and optimize the utilization of reference images by the models, for instance with the python Albumentation library. Artificially generating images of pollen covered by debris or unfocused, from pollen images with an initial good quality and accurate taxon identification, could also improve the method compared to using classes for indeterminable grains covered and blurry as done in this study (Chapter III). Overall, efficient solutions to optimize the training datasets, but mainly, to allow an optimal use of reference images from collection is crucial for automated palynology, considering the extensive resource and huge potential that constitute the palynology collections compiled worldwide by numerous people and in many laboratories.

Another remaining challenge was related to the class distribution in the training and testing data compared to the application data. As observed in the Chapter III, predictions were likely drawn by the distribution of the classes in the training dataset. This issue was notably related in our dataset to large class imbalance, with taxa under-represented in the training dataset that were under-predicted in the test set. The

issue of the class imbalance has previously been addressed with post-hoc calibration techniques, such as re-weighting (Olsson et al., 2021; Punyasena et al., 2022), or temperature scaling (Guo et al., 2017), which consists in readjusting the predicted scores to fit with probabilities of correct prediction. Nonetheless, these procedures still rely one single class distribution, while the pollen assemblage composition might significantly differ from one sample to another. This might lead to homogenize predictions among the processed samples, and thus reduce the potential dissimilarity in pollen assemblages. Here, in the Chapter III, for a taxon well represented in the dataset, its potential over-predictions might have generated false signals of presence in sampling locations where it might not occur. To address this issue of differences between training and testing data compared to application data, solutions are being developed, including semi-supervised or unsupervised algorithms especially for medical imaging (Guan and Liu, 2022; Farahani et al., 2021; Kumari and Singh, 2024). These techniques could be used to improve automated palynology applied to environmental samples.

5.2.3 Automated methods in palynology can support but not replace palynologists, which expertise remains crucial

So far, studies on automated palynology based on environmental samples relied on custom made image datasets compiled by manually annotating and identifying the pollen images generated from the studied environmental samples, to be representative of the application data (Punyasena et al., 2022; Theuerkauf et al., 2023; Barnes et al., 2023; von Allmen et al., 2024; Durand et al., 2024; Gimenez et al., 2024 in Chapter II; Chapter III). Methods are not yet ready to directly use images from reference samples, for which manual identification is not necessary (Tsiknakis et al., 2022; Chapter III). Consequently, the expertise of palynologists is crucial for automated methods, to constitute large and representative training datasets on which the CNNs rely. In that context, automated pollen analysis could be seen as a tool to fasten the process for pollen-based studies that have to process large number of samples, for instance for long fossil cores or long-term monitoring. For a given study, a training dataset could be compiled from a selected number of samples, or using all samples but only a small number of images to annotate per sample, the trained models could then be used to automate the analysis on the remaining images. Similarly, automated methods could also be applied to obtain a general idea of the samples contents to indicate major trends of some key taxa, or for instance for a fossil core, to point towards more interesting levels to analyze manually or to sample with a higher resolution.

In the future, it could be expected that an extensive and wide contribution of palynologists to generate large amounts of annotated datasets could allow to constitute one single but really extensive dataset allowing models to be efficient on all types of samples and contexts, for instance as achieved with the ImageNet dataset, that contains over 1.4 million images from 1000 object classes (Deng et al., 2009a). Making datasets open-source is becoming more and more frequent in palynology, and several pollen datasets are now available online. For instance, POLLEN23E (Gonçalves et al., 2016), and POLLEN73S (Astolfi et al., 2020) are classification datasets of optic microscope images respectively with 23 and 73 pollen types from the Brazilian Savannah, POLLEN13K is a detection and classification dataset from airborne samples (Battiato et al., 2020a), and POLLEN20L_det is a detection and classification dataset from reference samples from 13 temperate allergen taxa from Russia (Khanzhina et al., 2022). The detection datasets from Mediterranean taxa used in the first study of this thesis were also made open-source (Gimenez et al., 2024, at doi: 10.5281/zenodo.11126431.) and we

plan to do the same with the classification dataset used in Chapter III. Many other studies indicate their datasets can be shared under request. The platform OMERO (Open Microscopy Environment, openmicroscopy.org) might help in the process by allowing collaborative annotations and shared datasets.

Finally, the knowledge of palynologists on their region, time-period and ecosystems of expertise extends to information ranging well beyond the identification of pollen morphotypes, and this knowledge cannot be replaced by automated approaches. Altogether, the future automated methods in palynology most likely will be to help palynologists in their research. It will provide rapidly large amounts of pollen data and allowing an overall estimation trends, or to provide complementary information such as quantification of pollen fluxes of pollen production, or for instance for phylogeny placement of extinct species. To conduct more refined analysis of the pollen assemblage, automatically generated pollen data might still need to be reviewed or even sometimes revised based on human expertise to ensure its quality.

5.3 Discussing the application of the methods to address our ultimate biological questions

5.3.1 Monitoring vegetation from pollen rains and using automated data

The automated methods developed during this thesis have allowed to identify spatial and temporal changes in pollen rains from the Mediterranean area; it has also showed differences in the pollen signatures from the wild and domesticated grapevines *in situ* (wild population or vineyard in Southern France), and from the Vassal collection. Considering the vegetation monitoring, pollen rains from 15 target Mediterranean taxa collected throughout five years with annual passive traps were analyzed. Pollen rains allowed to differentiate the main type of vegetation occurring in each sampling location, based on the relative abundance of each target taxon, with more tree or shrub pollen (e.g. *Quercus*, *Phillyrea*) in the oak forest areas compared to more herbaceous pollen (Poaceae, *Plantago*) in the vineyard with a more open vegetation. Dominant pollen taxa also differed in the pollen assemblages from each sampling sites, as observed in the field, and including indicator taxa such as *Plantago* which correctly reflected more anthropized areas, and *Fraxinus* for more humid conditions. Few temporal changes could also be identified, such as the large decrease of *Buxus* pollen due to the Pyrale invasion, which aligned with field observations, or the increase in *Quercus* pollen in specific years that were shared across all sampling locations. The pollen data generated so far thus already revealed some interesting patterns, indicating that the automated method can already support research by generating pollen data that is relevant to study and monitor vegetation, including large general trends on the main surrounding vegetation type, dominant species, and drastic changes such as the disappearance of a taxon. For instance, the automated method could already identify changes in *Quercus* pollen between years in the whole study area. Such data could inform on disturbances that affected the vegetation in each year, as suggested by several studies that found rapid responses in *Quercus* pollen rains to disturbances, and mentioned it as a bio-indicator (Fernandez-Gonzalez et al., 2020; Gómez-Casero et al., 2007; Tormo-Molina et al., 2010; Algarra et al., 2019).

Nonetheless, the interpretation of the pollen data generated with the automated method still faced some limitations. First, few pollen taxa were largely misclassified due to the presence of similar pollen morphotypes (e.g. *Fraxinus* and *Phillyrea*). Thus, it is highly likely that their respective counts were either under or over estimated each time at least

one of the two occurred in a sample. These errors potentially lead to high error margins that remain difficult to quantify with precision in the application data since there is no ground truth for these identifications. Errors in the automated classification also prevented to identify with certainty the presence of taxa in small abundance in a sample due to the possibility of a false signal. Consequently, the automatically generated data is still limited to identify small but relevant changes in the vegetation.

Finally, note that the monitoring is still recent and upcoming data will allow to provide more detailed analysis on global trends and assess their link with climatic conditions. The monitoring is also conducted with a monthly and a weekly (during the grapevine pollination period) resolution, using similar passive traps placed in the same sampling sites. Applying the automated method to these samples should not require any further developments, due to their similar contents, with even less debris. These samples will allow to address new questions, especially regarding the phenology of the monitored species and inter-annual changes in flowering times. Comparing the sum over the 12 monthly samples with the corresponding annual sample could also be interesting on a methodological point of view. Finally, a smaller temporal resolution can help to solve the issue of identifying the presence of taxa identified in low abundance in the annual traps. If they are actually present, they should occur in larger abundance in monthly or weekly samples collected during their blooming period. This data could thus be combined with the annual resolution to improve the quality of the automated data. Finally, note that so far only a maximum of three slides were mounted for each annual sample, but much more material was collected throughout the year. Mounting more slides and counting more pollen grains per sample could be a tool to improve the identification for rare taxa, and more generally the whole pollen assemblage.

5.3.2 Comparisons of pollen signature of the wild and domesticated grapevines in situ and from collection based on automated data

The monitoring also allowed to describe the pollen signal associated of the grapevine in a vineyard (*Vitis vinifera*), and from a wild population with male and female individuals naturally occurring (*Vitis sylvestris*), i.e. *in situ*. Differences were observed in the pollen rains associated to both subspecies in their usual context of occurrence. First, and as expected, differences were identified in the accompanying vegetation of both subspecies. Wild grapevine individuals were associated to a larger abundance of typical Mediterranean taxa, with more pollen from trees and shrubs, such as *Quercus ilex* and *Phillyrea*, while the domesticated grapevine were associated to a larger abundance of taxa from open an anthropized vegetation, such as *Plantago* and Poaceae. The presence of the grapevine pollen itself also differed between the sampling sites, and displayed three distinct patterns between the sampling location : in the middle of the vineyard with the largest amount of collected fertile grapevine pollen, near wild male individuals with a more variable amount of fertile pollen, and near the wild female individual with a large amount of sterile grapevine pollen along with a similar amount of fertile pollen. The larger quantity of grapevine pollen found in the middle of the vineyard compared to below the wild male individual can be both explained by the larger production of pollen found in flowers from *Vitis vinifera* compared to *Vitis sylvestris* (Chapter IV), and of course, by the much larger number of individuals in a vineyard. A much lower amount of grapevine pollen was also found in the edge compared to in the middle of the vineyard, which is in line with the very small dispersion distance of this species, compared to other taxa (Turner and Brown, 2004).
Considering these patterns as indicators to discriminate between the presence of the wild or the domesticated subspecies is limited here by the number of replications in this study, with only one female individual and only one vineyard. Yet, the presence of a large

abundance of sterile pollen near the female individual aligns with a previous study, considering an archaeological context, suggesting the presence of that morphotype could be an indicator of local presence of wild grapevine (Mercuri et al., 2021). Note however, that some old indigenous grapevine varieties are functionally female and produce exclusively the inaperturate pollen morphotype, for instance the Picolit cultivar (Cargnello et al., 1980), and as found in two of the domesticated reference individuals from the ampelographic collection of Vassal-Montpellier (Chapter IV). A more detailed understanding of the grapevine pollen in both wild and cultivated contexts is thus required to use these approaches to discriminate between the presence of both subspecies. Nonetheless, these results suggest that such method could be achieved, by relying on the ratio between the two morphotypes, and the relative abundance of the fertile pollen among the full pollen assemblage and its accompanying vegetation. To consolidate these results, a complementary study could consists in adding other locations, near female individuals, maybe vineyards growing other varieties, or in another wild population. The same annual traps could be collected in many more locations in Southern France and other Mediterranean sites, but for only two to three years for instance, which could already be very informative. Again the automated method could allow to process a large amount of pollen grains, as required to represent the grapevine pollen. In that case, adjustments in the automated classification could be done considering this specific objective of identifying the grapevine pollen in the samples. A semi-automated method could also be used, to rely on the already developed method, for instance by keeping all images in which the grapevine was predicted with a certain level of probability (relatively low to avoid missing any grapevine pollen), and then by manually checking these images to remove all non grapevine pollen. This approach would still considerably reduce the workload as the human operator would only have to look at a much smaller amount of images.

Such information has a great potential for paleoecology research to reconstruct past landscapes and study practices from past Mediterranean societies. Here, we used annual traps, informing on pollen accumulation rates. These measurements were showed to provide comparable results with lake sediments (Hattestrand, 2013), and are relevant to help Quaternary scientists (Gerasimidis et al., 2006; Gaillard et al., 2008; Abraham et al., 2021), confirming the possibility of using the information from these modern traps to interpret fossil records.

Beyond the analysis of the grapevine pollen signature in its context 'in situ', we also investigated the possibility of using the image of individual grapevine pollen grains to discriminate between the two subspecies (Chapter IV), by comparing pollen size and attempting classification with a CNN. The classification method did not discriminate the fertile pollen grains produced by both subspecies. This could result from an absence of morphological differences between these pollen grains, potentially due to an insufficient number of generations since the two subspecies diverged. Domestication started during the Neolithic period (Dong et al., 2023; Grassi et al., 2008), however vegetative propagation is very common. Also, it could also be explained by genetic flows that occurred between both compartments throughout the domestication process, notably through major introgression events (Arroyo-García et al., 2006; Dong et al., 2023). Nonetheless, the identified differences in the size of the grains from both subspecies indicates that the two pollen morphotypes display differences. The classification results could thus also be related to methodological limitations, with an inadequate method to capture and differentiate subtle features, as previously mentioned, and for instance as the size parameter was not used in the classification process.

5.3.3 Other available data and potential applications for the automated method

Additionally to the study of annual trap samples, a weekly resolution in the vegetation monitoring was conducted during the blooming period of the grapevine. The pollen data obtained from these samples has great potential to assess flowering phenologies of the wild and domesticated grapevine as well as their inter-annual changes, which can have great impacts, especially for the wild subspecies. Preliminary results based on manual counts showed an advance of more than one week on wild individuals (N. Combourieu-Nebout, pers. comm.). Such changes in flowering phenologies can have drastic consequences for the wild grapevine. First, it could decorrelate its phenology with that of its pollinators, since this subspecies is partly insect pollinated, mostly by beetles and bees (Zito et al., 2018). Second, it could modify gene flows with the domesticated subspecies if they display distinct phenological response to disturbances. A previous study found low levels of gene flow from domesticated to wild individuals, from parentage tests of pips from wild individuals and including from female individuals from the Pic Saint Loup population (Di Vecchi-Staraz et al., 2009). Another study, based on molecular data, found that wild individuals were not related to the nearby or past cultivated varieties of domesticated grapevines (This et al., 2001). Flowering times between both subspecies would differ (pers. comm. mentioned in This et al., 2006), explaining these results. Distinct responses in their flowering times to upcoming climatic disturbances could modify these gene flows between both compartments. As previously mentioned, the application of the automated method should be direct on these samples, and some adjustments can also be made to improve the accuracy of the identification for the grapevine.

Additionally, the automated method could be also used to study in more details the pollen morphotypes produced by each individuals. In all pollen grains observed from the reference samples, only one morphotype was observed, either the fertile or the sterile. Nonetheless, it is likely that in few cases the other morphotype is produced, as suggested by a previous study which found a small proportion of seedlings collected on female wild individuals, including from the Pic St Loup, being issued from self-fertilization (<1%, Di Vecchi-Staraz et al., 2009). The classification of both morphotypes has proved efficient (Chapter III), when included in a dataset with many more classes, so it is likely very effective with just two classes. Using a classification CNN to quantify the amount of sterile and fertile pollen from reference samples could thus be tested on the grapevine image datasets generated during this PhD. Images from the ampelographic collection of Vassal-Montpellier could be included but also images from reference samples collected from wild individuals from the population of Pic Saint Loup. A total of 30 wild individuals from the wild population of the Pic Saint Loup were sampled, chemically processed, and photographed in view of compiling a classification dataset. Initially, these images were generated to evaluate the classification models trained on the reference collection from Vassal-Montpellier (Chapter IV), and as an intermediate step before applying it to the grapevine pollen collected from the annual traps. Finally, these steps were not conducted due to the ineffective method in discriminating the fertile pollen produced by both subspecies from the Vassal collection. The pollen images acquired could be used instead to search for the presence of the 'other' morphotype in each sample, either the sterile or the fertile depending on the sampled individual.

Overall, the method developments and the results obtained here regarding the pollen signature of the wild and domesticated grapevines offer a proof of concept for the possibility of developing a pollen-based automated method that could differentiate the presence of the wild or the domesticated grapevine. Many more studies and questions could be conducted and addressed after this work.

5.3.4 Implications and future research for automated palynology

After more than 20 years of developments and significant progress over the last few years, an increasing number of automated methods can now be used to help pollen-based research in environmental-related disciplines. This level where automated method can finally be applied to the real-world has already been achieved in aeropalynology for allergology, with many monitoring conducted automatically (Buters et al., 2024; Tummon, Adams-Groom, et al., 2024). Methods are also already used in the agroindustry to help anticipate time of harvests based on flowering phenologies (Tummon, Adams-Groom, et al., 2024). For environmental-related sciences, automated methods can already provide some relevant information. For instance, to quantify pollen from flowers, which has application in ecology, or in evolution (Zhang et al., 2024; as done in Chapter IV). This thesis also showed that few species can also be monitored through space and time (for some of the target taxa presented in the Chapter III). Numerous developments are being currently conducted in paleoecology with the objective of getting pollen counts per taxon from environmental samples, and especially fossil cores (Theuerkauf et al., 2023, Durand et al., 2024 , von Allmen et al., 2024, Barnes et al., 2023). These studies are promising and let to believe that methods should be able to process more and more taxa in such environmental samples in the near future.

Beyond automating, thus increasing the efficiency of tasks already performed by humans, automated methods can also open new ways of conducted pollen-based research. Automated image analysis has the potential to extract much more information from a pollen image compared to what a human expert could potentially ever do. For instance, automated method have been previously used to place extinct taxon in their phylogenetic tree (Adaïmé et al., 2024). There are lots of potential in the use of AI to compare close morphotypes and improve taxonomy of fossil taxa for instance (Punyasena et al., 2012, Bourel et al., 2020, Kong et al., 2016, Romero et al., 2020). Morphometric parameters such as pollen size, exine thickness, or the variability of these parameters have been related to environmental factors such as temperature (Kurtz and Liverman, 1958, Ejsmond et al., 2015 but Bell et al., 2018 on *Cedrus* pollen) and moisture (Payne, 1981). Pollen morphology and exine patterns have also been correlated to pollination syndrome (Lu et al., 2022). Studying pollen morphology using AI can thus definitely contribute to studies focused on climate reconstructions (as in Griener and Warny 2015), or to study pollen dispersal mechanisms, by processing large amounts of pollen grains from which they can directly extract relevant morphometric parameters (as done in Chapter IV).

So far, research on automated palynology has mostly been conducted by relatively small teams of researchers from different countries with limited interactions. Custom made datasets generated from environmental samples such as fossil cores are not yet often shared online; this limitation is due to reluctance due to the extremely intensive workload that represents constituting such image databases, and potentially because such datasets reflects the samples content, for instance when it is a fossil core, which data might be the object of further studies. A global contribution, from many pollen experts, in building and sharing large database is one of the most relevant approach to improve the accuracy of automated pollen analysis methods. Online platforms are already available and can support interactive annotations and construction of datasets, which can also help improve the quality of the data by cross-validating results across distinct experts. Globally sharing pollen datasets and collaboratively improving their quality will be a corner stone for efficient and globally accessible automated palynology. This largely relies on the will of pollen experts, and will play a large role in future advancements in this discipline.

Chapter 6

Conclusion

Automated pollen analysis has been an objective for a long time and is now switching to its final stage : application and use to real world samples and questions. Several challenges remained when this PhD started and were addressed to develop a method that would allow to investigate our ultimate objectives : assess spatial and temporal changes in pollen rains from a vegetation monitoring, and assess the pollen signature associated to the wild and domesticated grapevines. In that context, two fundamental approaches in palynology were thus investigated using AI : get pollen counts per taxon from environmental samples, and describe pollen morphotypes.

To automate the pollen analysis from the environmental samples, our developments covered distinct steps including the detection, the classification of all detected particles, and the integration of these steps in a fully automated pipeline to get calibrated pollen counts for 15 target Mediterranean taxa. The automated pipeline developed could be used directly to get the pollen counts from the vegetation monitoring. Complementary methods were investigated to analyze and compare reference samples from close pollen morphotypes, namely from the wild and domesticated grapevine. Automated methods allowed to estimate the production of pollen grains in flowers, and the size of pollen grains. The classification CNN classification did not discriminate the light microscopy images of the very close morphotypes of fertile grapevine pollen.

Automatically generated pollen data allowed to identify large but relevant spatial and temporal variations in pollen rains from the vegetation monitoring, identify distinct patterns in the pollen signatures associated to the wild and domesticated grapevines in term of accompanying cortege, identify abundance and ratio between the fertile and sterile pollen grains from the grapevine, and identify difference in the sizes of the grapevine pollen and in its production in flowers. The results are a proof of concept for methods to discriminate their respective pollen signature, from in situ and reference conditions. More work is required but such discrimination method has great potential for paleoecology to reconstruct viticulture.

Overall, the developed methods and the automatically generated pollen data thus allowed to address questions related to environmental sciences. The method developed during the PhD can already generate pollen data for distinct biological purposes and serve to fellow palynologists for some steps of their analysis or to complement their analysis with new type of data, being easy and already available.

Overall automated methods are far from replacing palynologists, but they can already revolutionize many aspects of palynology by supporting their research, by making some processes faster, and open new doors for pollen-based research by doing tasks not conducted by human experts. It can open new ways of doing research in palynology and of using data, solve new questions, facilitate and speed the work, leaving more time to palynologists to analyze the data and improve our understanding of main and crucial processes that palynology is providing to society, such as past vegetation and climate reconstructions, allergen calendars, future of pollination ecology and so on. A global share of the image

datasets will be key in years to come, to reach higher levels of accuracy in the automated methods, and such path, if taken, will rely on the expert knowledge of palynologists across the world.

Bibliography

Abas, S. M., Abdulazeez, A. M., & Zeebaree, D. Q. (2022). A YOLO and convolutional neural network for the detection and classification of leukocytes in leukemia. *Indonesian Journal of Electrical Engineering and Computer Science*, *25*(1), 200–213. https://doi.org/10.11591/ijeecs.v25.i1.pp200-213

Abdullah, Ali, S., Khan, Z., Hussain, A., Athar, A., & Kim, H.-C. (2022). Computer vision based deep learning approach for the detection and classification of algae species using microscopic images. *Water*, *14*(14), 2219. https://doi.org/10.3390/w14142219

Abraham, V., Hicks, S., Svobodova-Svitavska, H., Bozilova, E., Panajiotidis, S., Filipova-Marinova, M., Jensen, C. E., Tonkov, S., Pidek, I. A., Swieta-Musznicka, J., Zimny, M., Kvavadze, E., Filbrandt-Czaja, A., Hattestrand, M., Kilic, N. K., Kosenko, J., Nosova, M., Severova, E., Volkova, O., … Giesecke, T. (2021). Patterns in recent and holocene pollen accumulation rates across europe - the pollen monitoring programme database as a tool for vegetation reconstruction. *Biogeosciences*, *18*(15), 4511–4534. https://doi.org/10.5194/bg-18-4511-2021

Adaïmé, M.-É., Kong, S., & Punyasena, S. W. (2024). Deep learning approaches to the phylogenetic placement of extinct pollen morphotypes. *PNAS Nexus*, *3*(1), pgad419. https://doi.org/10.1093/pnasnexus/pgad419

Aguilera, F., & Ruiz-Valenzuela, L. (2019). A new aerobiological indicator to optimize the prediction of the olive crop yield in intensive farming areas of southern spain. *Agricultural and Forest Meteorology*, *271*, 207–213. https://doi.org/10.1016/j.agrformet.2019.03.004

Algarra, J. A., Carinanos, P., Herrero, J., Delgado-Capel, M., Ramos-Lorente, M. M., & Diaz de la Guardia, C. (2019). Tracking montane mediterranean grasslands: analysis of the effects o snow with other related hydro-meteorological variables and land-use change on pollen emissions. *Science of the Total Environment*, *649*, 889–901. https://doi.org/10.1016/j.scitotenv.2018.08.311

Ali. (2022). Cross-chapter paper 4: mediterranean region. in: climate change 2022: impacts, adaptation and vulnerability. contribution of working group II to the sixth assessment report of the intergovernmental panel on climate change.

Allen, G. (2006). *An automated pollen recognition system : a thesis submitted to massey university, turitea, palmerston north, new zealand in fulfilment of the requirements for the degree of master of engineering* [Thesis]. Massey University. Retrieved December 2, 2022, from https://mro.massey.ac.nz/handle/10179/613

Allen, J., Watts, W., McGee, E., & Huntley, B. (2002). Holocene environmental variability—the record from lago grande di monticchio, italy. *Quaternary International*, *88*(1), 69–80. https://doi.org/10.1016/S1040-6182(01)00074-X

Anderegg, W. R. L., Abatzoglou, J. T., Anderegg, L. D. L., Bielory, L., Kinney, P. L., & Ziska, L. (2021). Anthropogenic climate change is worsening north american pollen seasons. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(7), e2013284118. https://doi.org/10.1073/pnas.2013284118

Anzani, R., Failla, O., Scienza, A., & Campostrini, F. (1990). Wild grapevine (*Vitis vinifera* var. *silvestris*) in italy: distribution, characteristics and germplasm preservation-1989 report. *VITIS - Journal of Grapevine Research*, *29*, 97–97. https://doi.org/10.5073/vitis.1990.29.special-issue.97-112

Arnold, C. (2002). *Ecologie de la vigne sauvage, Vitis vinifera L. ssp. sylvestris (Gmelin) Hegi, dans les forêts alluviales et colluviales d'Europe*. vdf Hochschulverlag AG.

Arroyo-García, R., Ruiz-García, L., Bolling, L., Ocete, R., López, M. A., Arnold, C., Ergul, A., Söylemezo˝lu, G., Uzun, H. I., Cabello, F., Ibáñez, J., Aradhya, M. K., Atanassov, A., Atanassov, I., Balint, S., Cenis, J. L., Costantini, L., Gorislavets, S., Grando, M. S., … Martinez-Zapater, J. M. (2006). Multiple origins of cultivated grapevine (vitis vinifera l. ssp. sativa) based on chloroplast DNA polymorphisms. *Molecular Ecology*, *15*(12), 3707–3714. https://doi.org/10.1111/j.1365-294X.2006.03049.x

Astolfi, G., Gonçalves, A. B., Menezes, G. V., Borges, F. S. B., Astolfi, A. C. M. N., Matsubara, E. T., Alvarez, M., & Pistori, H. (2020). POLLEN73s: an image dataset for pollen grains classification. *Ecological Informatics*, *60*, 101165. https://doi.org/10.1016/j.ecoinf.2020.101165

Barnes, C. M., Power, A. L., Barber, D. G., Tennant, R. K., Jones, R. T., Lee, G. R., Hatton, J., Elliott, A., Zaragoza-Castells, J., Haley, S. M., Summers, H. D., Doan, M., Carpenter, A. E., Rees, P., & Love, J. (2023). Deductive automated pollen classification in environmental samples via exploratory deep learning and imaging flow cytometry. *New Phytologist*, *240*(3), 1305–1326. https://doi.org/10.1111/nph.19186

Barredo, J. I., Mauri, A., Caudullo, G., & Dosio, A. (2018). Assessing shifts of mediterranean and arid climates under RCP4.5 and RCP8.5 climate projections in europe. *Pure and Applied Geophysics*, *175*(11), 3955–3971. https://doi.org/10.1007/s00024-018-1853-6

Battiato, S., Ortis, A., Trenta, F., Ascari, L., Politi, M., & Siniscalco, C. (2020a). Detection and classification of pollen grain microscope images. *2020 Ieee/Cvf Conference on Computer Vision and Pattern Recognition Workshops (cvprw 2020)*, 4220–4227. https://doi.org/10.1109/CVPRW50498.2020.00498

Battiato, S., Ortis, A., Trenta, F., Ascari, L., Politi, M., & Siniscalco, C. (2020b). Pollen13k: a large scale microscope pollen grain image dataset. *2020 Ieee International Conference on Image Processing (icip)*, 2456–2460. Retrieved May 11, 2022, from http://arxiv.org/pdf/2007.04690

Bell, B. A., Bishop, T. H., Fletcher, W. J., Ryan, P., & Ilmen, R. (2018). Cedrus atlantica pollen morphology and investigation of grain size variability using laser diffraction granulometry. *Palynology*, *42*(3), 339–353. https://doi.org/10/gn3ttg

Bernigaud, N., Bondeau, A., Guiot, J., Bouby, L., Bertoncello, F., & Ouriachi, M.-J. (2024). Reconstructing vineyard geography, yields and the profitability of wine production in the roman empire: new insights from spatial analysis, and agent-based and climate-vegetation modelling. Bloomsbury Publishing Plc. https://doi.org/10.5040/9781350346680.ch-015

Blondel, J., & Aronson, J. (1999). *Biology and wildlife of the mediterranean region*. Oxford University Press.

Boldeanu, M., Cucu, H., Burileanu, C., & Marmureanu, L. (2021). Automatic pollen classification using convolutional neural networks. In H. Herencsar (Ed.), *2021 44th International conference on Telecommunications and Signal Processing (TSP)* (pp. 130–133). IEEE. https://doi.org/10.1109/TSP52935.2021.9522626

Boldeanu, M., Gonzalez-Alonso, M., Cucu, H., Burileanu, C., Maya-Manzano, J. M., & Buters, J. T. M. (2022). Automatic pollen classification and segmentation using u-nets and synthetic data. *Ieee Access*, *10*, 73675–73684. https://doi.org/10.1109/ACCESS.2022.3189012

Bouby, L., Figueiral, I., Bouchette, A., Rovira, N., Ivorra, S., Lacombe, T., Pastor, T., Picq, S., Marinval, P., & Terral, J.-F. (2013). Bioarchaeological insights into the process of domestication of grapevine (vitis vinifera l.) during roman times in southern france. *PLOS ONE*, *8*(5), e63195. https://doi.org/10/gnrffv

Boucher, A., Hidalgo, P. J., Thonnat, M., Belmonte, J., Galan, C., Bonton, P., & Tomczak, R. (2002). Development of a semi-automatic system for pollen recognition. *Aerobiologia*, *18*(3), 195–201. https://doi.org/10.1023/A:1021322813565

Boullayali, A., Ater, M., Terral, J.-F., & Bouziane, H. (2024). Comparison of *Olea* pollen sampling between gravimetric and volumetric traps (NW of morocco). *Science of The Total Environment*, *951*, 175663. https://doi.org/10.1016/j.scitotenv.2024.175663

Bourel, B., Marchant, R., de Garidel-Thoron, T., Tetard, M., Barboni, D., Gally, Y., & Beaufort, L. (2020). Automated recognition by multiple convolutional neural networks of modern, fossil, intact and damaged pollen grains. *Computers & Geosciences*, *140*, 104498. https://doi.org/10.1016/j.cageo.2020.104498

Braconnot, P., Leduc, G., Arnaud, F., Barboni, D., Bassinot, F., Blard, P., Boisserie, J.-R., Bourquin, S., Chase, B. M., Eynaud, F., de Garidel-Thoron, T., Jacob, J., Khodri, M., Landais, A., Martinerie, P., Malaizé, B., Turcq, B., Vannière, B., Robin, C., … Bern, U. B. /. U. o. (2024). Livre blanc - paléoclimats et paléoenvironnements. Retrieved August 24, 2024, from https://agris.fao.org/search/en/providers/122439/records/66718aa875507fed3402bf46

Brdar, S., Panić, M., Matavulj, P., Stanković, M., Bartolić, D., & Šikoparija, B. (2023). Explainable AI for unveiling deep learning pollen classification model based on fusion of scattered light

patterns and fluorescence spectroscopy. *Scientific Reports*, *13*(1), 3205. https://doi.org/10.1038/s41598-023-30064-6

Brooks, J., & Shaw, G. (1978). Sporopollenin: a review of its chemistry, palaeochemistry and geochemistry. *Grana*, *17*(2), 91–97. https://doi.org/10.1080/00173137809428858

Brown, J., Qiao, Y., Clark, C., Lomax, S., Rafique, K., & Sukkarieh, S. (2022). Automated aerial animal detection when spatial resolution conditions are varied. *Computers and Electronics in Agriculture*, *193*, 106689. https://doi.org/10.1016/j.compag.2022.106689

Bruffaerts, N., De Smedt, T., Delcloo, A., Simons, K., Hoebeke, L., Verstraeten, C., Van Nieuwenhuyse, A., Packeu, A., & Hendrickx, M. (2018). Comparative long-term trend analysis of daily weather conditions with daily pollen concentrations in brussels, belgium. *International Journal of Biometeorology*, *62*(3), 483–491. https://doi.org/10.1007/s00484-017-1457-3

Buters, J., Antunes, C., Galveias, A., Bergmann, K. C., Thibaudon, M., Galán, C., Schmidt-Weber, C., & Oteros, J. (2018). Pollen and spore monitoring in the world. *Clinical and Translational Allergy*, *8*(1), 9. https://doi.org/10.1186/s13601-018-0197-8

Buters, J., Clot, B., Galán, C., Gehrig, R., Gilge, S., Hentges, F., O'Connor, D., Sikoparija, B., Skjoth, C., Tummon, F., Adams-Groom, B., Antunes, C. M., Bruffaerts, N., Çelenk, S., Crouzy, B., Guillaud, G., Hajkova, L., Seliger, A. K., Oliver, G., ... Stjepanovic, B. (2024). Automatic detection of airborne pollen: an overview. *Aerobiologia*, *40*(1), 13–37. https://doi.org/10.1007/s10453-022-09750-x

Cantos, M., Arroyo-García, R., García, J. L., Lara, M., Morales, R., López, M. Á., Gallardo, A., Ocete, C. A., Rodríguez, Á., Valle, J. M., Vaca, R., González-Maestro, M., Bánáti, H., & Ocete, R. (2017). Current distribution and characterization of the wild grapevine populations in Andalusia (Spain). *Comptes Rendus. Biologies*, *340*(3), 164–177. https://doi.org/10.1016/j.crvi.2017.01.004

Cao, N., Meyer, M., Thiele, L., & Saukh, O. (2020a). Automated pollen detection with an affordable technology. *Proceedings of the International Conference on Embedded Wireless Systems and Networks (EWSN)*, 108–119.

Cao, N., Meyer, M., Thiele, L., & Saukh, O. (2020b). Pollen video library for benchmarking detection, classification, tracking and novelty detection tasks: dataset. *Proceedings of the Third Workshop on Data: Acquisition To Analysis*, 23–25. https://doi.org/10.1145/3419016.3431487

Caporali, E., Spada, A., Marziani, G., Failla, O., & Scienza, A. (2003). The arrest of development of abortive reproductive organs in the unisexual flower of vitis vinifera ssp silvestris. *Sexual Plant Reproduction*, *15*(6), 291–300. https://doi.org/10.1007/s00497-003-0169-5

Cargnello, G., Carraro, L., Lombardo, G., & Gerola, F. (1980). Pollen morphology of picolit grown in different italian regions. *Vitis*, *19*(3), 201–206. Retrieved January 13, 2022, from https://www.webofscience.com/wos/woscc/full-record/WOS:A1980KU43700001

Cartaxo-Pinto, S., Mendonça, C. B. F., Lopes, R. C., & Gonçalves-Esteves, V. (2017). Pollen morphology of species of cissus (vitaceae): an evaluation of ornamentation. *Palynology*, *41*(3), 359–369. https://doi.org/10/gmd5vj

Castañeda-Álvarez, N. P., Khoury, C. K., Achicanoy, H. A., Bernau, V., Dempewolf, H., Eastwood, R. J., Guarino, L., Harker, R. H., Jarvis, A., Maxted, N., Müller, J. V., Ramirez-Villegas, J., Sosa, C. C., Struik, P. C., Vincent, H., & Toll, J. (2016). Global conservation priorities for crop wild relatives. *Nature Plants*, *2*(4), 1–6. https://doi.org/10.1038/nplants.2016.22

Castro López, A. J., Rejón, J. D., Fendri, M., Jiménez-Quesada, M. J., Zafra, A., Jiménez-López, J. C., Rodríguez García, M. I., & Alché Ramírez, J. d. D. (2010, December). *Taxonomical discrimination of pollen grains by using confocal laser scanning microscopy (CLSM) imaging of autofluorescence.* Formatex. Retrieved September 19, 2024, from https://digital.csic.es/handle/10261/32458

Chen, X., & Ju, F. (2022). Automatic classification of pollen grain microscope images using a multi-scale classifier with SRGAN deblurring. *Applied Sciences-Basel*, *12*(14), 7126. https://doi.org/10.3390/app12147126

Chollet. (2015). *Keras* [GitHub]. Retrieved June 20, 2022, from https://github.com/fchollet/keras

Clò, E., Torri, P., Baliva, M., Brusco, A., Marchianò, R., Sgarbi, E., Palli, J., Mercuri, A. M., Piovesan, G., & Florenzano, A. (2023). A multidisciplinary study of wild grapevines in the river crati natural reserve, south italy (calabria): implications in conservation biology

and palaeoecological reconstructions. *Quaternary*, *6*(3), 43. https://doi.org/10.3390/quat6030043

Collevatti, R. G., Castañeda, M., Silva-Caminha, S. A. F., & Jaramillo, C. (2024). Application of confocal laser microscopy for identification of modern and fossil pollen grains, an example in palm mauritiinae. *Review of Palaeobotany and Palynology*, *327*, 105140. https://doi.org/10.1016/j.revpalbo.2024.105140

Connolly, R. M., Jinks, K. I., Herrera, C., & Lopez-Marcano, S. (2022). Fish surveys on the move: adapting automated fish detection and classification frameworks for videos on a remotely operated vehicle in shallow marine waters. *Frontiers in Marine Science*, *9*, 918504. https://doi.org/10.3389/fmars.2022.918504

Corvucci, F., Nobili, L., Melucci, D., & Grillenzoni, F.-V. (2015). The discrimination of honey origin using melissopalynology and raman spectroscopy techniques coupled with multivariate analysis. *Food Chemistry*, *169*, 297–304. https://doi.org/10.1016/j.foodchem.2014.07.122

Cramer, W. (2018). Climate change and interconnected risks to sustainable development in the mediterranean. *Nature Climate Change*, *8*, 972–980. https://doi.org/10.1038/s41558-018-0299-2

Cristofolini, F., Cristofori, A., Corradini, S., & Gottardini, E. (2024). The impact of temperature on increased airborne pollen and earlier onset of the pollen season in trentino, northern italy. *Regional Environmental Change*, *24*(2), 60. https://doi.org/10.1007/s10113-024-02223-6

Crouzy, B., Lieberherr, G., Tummon, F., & Clot, B. (2022). False positives: handling them operationally for automatic pollen monitoring. *Aerobiologia*, *38*(3), 429–432. https://doi.org/10.1007/s10453-022-09757-4

Crouzy, B., Stella, M., Konzelmann, T., Calpini, B., & Clot, B. (2016). All-optical automatic pollen identification: towards an operational system. *Atmospheric Environment*, *140*, 202–212. https://doi.org/10.1016/j.atmosenv.2016.05.062

Diversity and composition of pollen loads carried by pollinators are primarily driven by insect traits, not floral community characteristics. Retrieved August 28, 2024, from https://doi.org/10.1007/s00442-021-04911-0

Cunha, M., Ribeiro, H., & Abreu, I. (2016). Pollen-based predictive modelling of wine production: application to an arid region. *European Journal of Agronomy*, *73*, 42–54. https://doi.org/10.1016/j.eja.2015.10.008

Cuttelod, A., García, N., Malak, D. A., Temple, H., & Katariya, V. (2008). The Mediterranean: a biodiversity hotspot under threat. *IUCN, Gland, Switzerland_ The 2008 Review of The IUCN Red List of Threatened Species*, 17.

Damialis, A., Fotiou, C., Halley, J. M., & Vokou, D. (2011). Effects of environmental factors on pollen production in anemophilous woody species. *Trees*, *25*(2), 253–264. https://doi.org/10.1007/s00468-010-0502-1

Datar, P., Jain, K., & Dhedhi, B. (2018). Detection of birds in the wild using deep learning methods. *2018 4th International Conference for Convergence in Technology (I2CT)*, 1–4. https://doi.org/10.1109/I2CT42659.2018.9057933

Daunys, G., Šukienė, L., Vaitkevičius, L., Valiulis, G., Sofiev, M., & Šaulienė, I. (2024). Comparison of computer vision models in application to pollen classification using light scattering. *Aerobiologia*, *40*(1), 109–121. https://doi.org/10.1007/s10453-022-09769-0

De Linares, C., Delgado, R., Jesus Aira, M., Alcazar, P., Alonso-Perez, S., Boi, M., Carinanos, P., Cuevas, E., Diaz de la Guardia, C., Elvira-Rendueles, B., Fernandez-Gonzalez, D., Galan, C., Montserrat Gutierrez-Bustillo, A., Perez-Badia, R., Javier Rodriguez-Rajo, F., Ruiz-Valenzuela, L., Tormo-Molina, R., del Mar Trigo, M., Valencia-Barrera, R. M., ... Belmonte, J. (2017). Changes in the mediterranean pine forest: pollination patterns and annual trends of airborne pollen. *Aerobiologia*, *33*(3), 375–391. https://doi.org/10.1007/s10453-017-9476-4

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009a). ImageNet: A Large-Scale Hierarchical Image Database. *CVPR09*.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009b). ImageNet: a large-scale hierarchical image database. *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 248–255. https://doi.org/10.1109/CVPR.2009.5206848

Desprat, S., Diaz Fernandez, P. M., Coulon, T., Ezzat, L., Pessarossi-Langlois, J., Gil, L., Morales-Molino, C., & Sanchez Goni, M. F. (2015). Pinus nigra (european black pine) as the dominant species of the last glacial pinewoods in south-western to central iberia:

a morphological study of modern and fossil pollen. *Journal of Biogeography*, *42*(10), 1998–2009. https://doi.org/10/gn8psv

Di Vecchi-Staraz, M., Laucou, V., Bruno, G., Lacombe, T., Gerber, S., Bourse, T., Boselli, M., & This, P. (2009). Low level of pollen-mediated gene flow from cultivated to wild grapevine: consequences for the evolution of the endangered subspecies vitis vinifera l. subsp. silvestris. *Journal of Heredity*, *100*(1), 66–75. https://doi.org/10.1093/jhered/esn084

Diaz-Lopez, E., Rincon, M., Rojo, J., Vaquero, C., Rapp, A., Salmeron-Majadas, S., & Perez-Badia, R. (2015). Localisation of pollen grains in digitised real daily airborne samples. In J. M. F. Vicente, J. R. AlvarezSanchez, F. D. Lopez, F. J. ToledoMoreo, & H. Adeli (Eds.), *Artificial computation in biology and medicine, pt i (iwinac 2015)* (pp. 348–357, Vol. 9107). Springer Int Publishing Ag. https://doi.org/10.1007/978-3-319-18914-7_37

Diwan, T., Anirudh, G., & Tembhurne, J. V. (2022). Object detection using YOLO: challenges, architectural successors, datasets and applications. *Multimedia Tools and Applications*. https://doi.org/10.1007/s11042-022-13644-y

Donders, T., Panagiotopoulos, K., Koutsodendris, A., Bertini, A., Mercuri, A. M., Masi, A., Combourieu-Nebout, N., Joannin, S., Kouli, K., Kousis, I., Peyron, O., Torri, P., Florenzano, A., Francke, A., Wagner, B., & Sadori, L. (2021). 1.36 million years of mediterranean forest refugium dynamics in response to glacial-interglacial cycle strength. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(34), e2026111118. https://doi.org/10.1073/pnas.2026111118|1of9

Dong, Y., Duan, S., Xia, Q., Liang, Z., Dong, X., Margaryan, K., Musayev, M., Goryslavets, S., Zdunić, G., Bert, P.-F., Lacombe, T., Maul, E., Nick, P., Bitskinashvili, K., Bisztray, G. D., Drori, E., De Lorenzis, G., Cunha, J., Popescu, C. F., ... Chen, W. (2023). Dual domestications and origin of traits in grapevine evolution. *Science*, *379*(6635), 892–901. https://doi.org/10.1126/science.add8655

Drescher-Schneider, R., de Beaulieu, J.-L., Magny, M., Walter-Simonnet, A.-V., Bossuet, G., Millet, L., Brugiapaglia, E., & Drescher, A. (2007). Vegetation history, climate and human impact over the last 15,000 years at lago dell'accesa (tuscany, central italy). *Vegetation History and Archaeobotany*, *16*(4), 279–299. https://doi.org/10.1007/s00334-006-0089-z

Dunker, S., Motivans, E., Rakosy, D., Boho, D., Mäder, P., Hornick, T., & Knight, T. M. (2021). Pollen analysis using multispectral imaging flow cytometry and deep learning. *New Phytologist*, *229*(1), 593–606. https://doi.org/10.1111/nph.16882

Durand, M., Paillard, J., Ménard, M.-P., Suranyi, T., Grondin, P., & Blarquez, O. (2024). Pollen identification through convolutional neural networks: first application on a full fossil pollen sequence. *PLOS ONE*, *19*(4), e0302424. https://doi.org/10.1371/journal.pone.0302424

Ejsmond, M. J., Ejsmond, A., Banasiak, Ł., Karpińska-Kołaczek, M., Kozłowski, J., & Kołaczek, P. (2015). Large pollen at high temperature: an adaptation to increased competition on the stigma? *Plant Ecology*, *216*(10), 1407–1417. https://doi.org/10/f78ztm

Ellwood, E. R., Temple, S. A., Primack, R. B., Bradley, N. L., & Davis, C. C. (2013). Record-breaking early flowering in the eastern united states. *PLOS ONE*, *8*(1), e53788. https://doi.org/10.1371/journal.pone.0053788

Farahani, A., Voghoei, S., Rasheed, K., & Arabnia, H. R. (2021). A brief review of domain adaptation. In R. Stahlbock, G. M. Weiss, M. Abou-Nasr, C.-Y. Yang, H. R. Arabnia, & L. Deligiannidis (Eds.), *Advances in data science and information engineering* (pp. 877–894). Springer International Publishing. https://doi.org/10.1007/978-3-030-71704-9_65

Fernandez-Gonzalez, M., Gonzalez-Fernandez, E., Ribeiro, H., Abreu, I., & Javier Rodriguez-Rajo, F. (2020). Pollen production of quercus in the north-western iberian peninsula and airborne pollen concentration trends during the last 27 years. *Forests*, *11*(6), 702. https://doi.org/10.3390/f11060702

Fernández-Martínez, M., Belmonte, J., & Maria Espelta, J. (2012). Masting in oaks: disentangling the effect of flowering phenology, airborne pollen load and drought. *Acta Oecologica*, *43*, 51–59. https://doi.org/10.1016/j.actao.2012.05.006

Filipova-Marinova, M. V., Kvavadze, E. V., Connor, S. E., & Sjögren, P. (2010). Estimating absolute pollen productivity for some european tertiary-relict taxa. *Vegetation History and Archaeobotany*, *19*(4), 351–364. https://doi.org/10.1007/s00334-010-0257-z

France, I., Duller, A. W. G., Duller, G. a. T., & Lamb, H. F. (2000). A new approach to automated pollen analysis. *Quaternary Science Reviews*, *19*(6), 537–546. https://doi.org/10.1016/S0277-3791(99)00021-9

Gaillard, M.-J., Sugita, S., Bunting, M. J., Middleton, R., Broström, A., Caseldine, C., Giesecke, T., Hellman, S. E. V., Hicks, S., Hjelle, K., Langdon, C., Nielsen, A.-B., Poska, A., von Stedingk, H., Veski, S., & POLLANDCAL members. (2008). The use of modelling and simulation approach in reconstructing past landscapes from fossil pollen data: a review and results from the POLLANDCAL network. *Vegetation History and Archaeobotany*, *17*(5), 419–443. https://doi.org/10.1007/s00334-008-0169-3

Galán, C., Alcázar, P., Oteros, J., García-Mozo, H., Aira, M. J., Belmonte, J., Diaz de la Guardia, C., Fernández-González, D., Gutierrez-Bustillo, M., Moreno-Grau, S., Pérez-Badía, R., Rodríguez-Rajo, J., Ruiz-Valenzuela, L., Tormo, R., Trigo, M. M., & Domínguez-Vilches, E. (2016). Airborne pollen trends in the iberian peninsula. *Science of The Total Environment*, *550*, 53–59. https://doi.org/10.1016/j.scitotenv.2016.01.069

Galera, M. D., Elvira-Rendueles, B., Moreno, J. M., Negral, L., Ruiz-Abellón, M. C., García-Sánchez, A., & Moreno-Grau, S. (2018). Analysis of airborne *Olea* pollen in cartagena (spain). *Science of The Total Environment*, *622-623*, 436–445. https://doi.org/10.1016/j.scitotenv.2017.11.349

Gallardo, A., Ocete, R., López, M. Á., Lara, M., & Rivera, D. (2009). Assessment of pollen dimorphism in populations of vitis vinifera l. subsp. sylvestris (gmelin) hegi in spain, 4.

Gallardo-Caballero, R., García-Orellana, C. J., García-Manso, A., González-Velasco, H. M., Tormo-Molina, R., & Macías-Macías, M. (2019). Precise pollen grain detection in bright field microscopy using deep learning techniques. *Sensors*, *19*(16), 3583. https://doi.org/10.3390/s19163583

García-Mozo, H., Oteros, J. A., & Galán, C. (2016). Impact of land cover changes and climate on the main airborne pollen types in southern spain. *Science of The Total Environment*, *548-549*, 221–228. https://doi.org/10.1016/j.scitotenv.2016.01.005

Gavrilova, O., Zavialova, N., Tekleva, M., & Karasev, E. (2018). Potential of CLSM in studying some modern and fossil palynological objects. *Journal of Microscopy*, *269*(3), 291–309. https://doi.org/10.1111/jmi.12639

Gehrig, R., & Clot, B. (2021). 50 years of pollen monitoring in basel (switzerland) demonstrate the influence of climate change on airborne pollen. *Frontiers in Allergy*, *2*. https://doi.org/10.3389/falgy.2021.677159

Gerasimidis, A., Panajiotidis, S., Hicks, S., & Athanasiadis, N. (2006). An eight-year record of pollen deposition in the pieria mountains (n. greece) and its significance for interpreting fossil pollen assemblages. *Review of Palaeobotany and Palynology*, *141*(3), 231–243. https://doi.org/10.1016/j.revpalbo.2006.04.004

Geus, A. R. d., Barcelos, C. A., Batista, M. A., & Silva, S. F. d. (2019). Large-scale pollen recognition with deep learning. *2019 27th European Signal Processing Conference (EUSIPCO)*, 1–5. https://doi.org/10.23919/EUSIPCO.2019.8902735

Ghitarrini, S., Emma, T., Veronica, T., & Giuseppe, F. (2017). Climate change: consequences on the pollination of grasses in perugia (central italy). a 33-year-long study. *International Journal of Biometeorology*, *61*(1), 149–158. https://doi.org/10.1007/s00484-016-1198-8

Gimenez, B., Joannin, S., Pasquet, J., Beaufort, L., Gally, Y., de Garidel-Thoron, T., Combourieu-Nebout, N., Bouby, L., Canal, S., Ivorra, S., Limier, B., Terral, J.-F., Devaux, C., & Peyron, O. (2024). A user-friendly method to get automated pollen analysis from environmental samples. *New Phytologist*, *243*(2), 797–810. https://doi.org/10.1111/nph.19857

Giorgi, F. (2006). Climate change hot-spots. *Geophysical Research Letters*, *33*(8), L08707. https://doi.org/10.1029/2006GL025734

Giorgi, F., & Lionello, P. (2008). Climate change projections for the mediterranean region. *Global and Planetary Change*, *63*(2), 90–104. https://doi.org/10.1016/j.gloplacha.2007.09.005

Githumbi, E., Fyfe, R., Gaillard, M.-J., Trondman, A.-K., Mazier, F., Nielsen, B., Poska, A., Sugita, S., Theuerkauf, M., Woodbridge, J., Feurdean, A., Grindean, R., Lebreton, V., & Marquer, L. (2021). 1 european pollen-based REVEALS land-cover reconstructions for the 2 holocene: methodology, mapping and potentials. *Open Access*.

Gómez-Casero, M. T., Galán, C., & Domínguez-Vilches, E. (2007). Flowering phenology of Mediterranean Quercus species in different locations (Córdoba, SW Iberian Peninsula). *Acta Botanica Malacitana*, *32*, 127–146. https://doi.org/10.24310/abm.v32i0.7033

Gonçalves, A. B., Souza, J. S., Silva, G. G. D., Cereda, M. P., Pott, A., Naka, M. H., & Pistori, H. (2016). Feature extraction and machine learning for the classification of brazilian savannah

pollen grains (H. A. Kestler, Ed.). *PLOS ONE*, *11*(6), e0157044. https://doi.org/10.1371/journal.pone.0157044

Gonçalves-Esteves, V., Mezzonato-Pires, A. C., Marinho, E. B., de Souza, R. M. B. S., Esteves, R. L., Cartaxo-Pinto, S., & Mendonça, C. B. F. (2022). The importance of palynology to taxonomy. In M. F. T. Medeiros & B. de Sá Haiad (Eds.), *Aspects of brazilian floristic diversity : from botany to traditional communities* (pp. 119–134). Springer International Publishing. https://doi.org/10.1007/978-3-031-07453-0_5

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. Retrieved August 9, 2024, from https://www.deeplearningbook.org/

Gordo, O., & Sanz, J. J. (2010). Impact of climate change on plant phenology in mediterranean ecosystems. *Global Change Biology*, *16*(3), 1082–1106. https://doi.org/10.1111/j.1365-2486.2009.02084.x

Grant-Jacob, J. A., & Mills, B. (2022). Deep learning in airborne particulate matter sensing: a review. *Journal of Physics Communications*, *6*(12), 122001. https://doi.org/10.1088/2399-6528/aca45e

Grassi, F., De Mattia, F., Zecca, G., Sala, F., & Labra, M. (2008). Historical isolation and quaternary range expansion of divergent lineages in wild grapevine. *Biological Journal of the Linnean Society*, *95*(3), 611–619. https://doi.org/10.1111/j.1095-8312.2008.01081.x

Grindean, R., Nielsen, A. B., Tanţău, I., & Feurdean, A. (2019). Relative pollen productivity estimates in the forest steppe landscape of southeastern romania. *Review of Palaeobotany and Palynology*, *264*, 54–63. https://doi.org/10.1016/j.revpalbo.2019.02.007

Guan, H., & Liu, M. (2022). Domain adaptation for medical image analysis: a survey. *IEEE Transactions on Biomedical Engineering*, *69*(3), 1173–1185. https://doi.org/10.1109/TBME.2021.3117407

Guiot, J., Bernigaud, N., Bondeau, A., Bouby, L., & Cramer, W. (2023). Viticulture extension in response to global climate change drivers – lessons from the past and future projections. *Climate of the Past*, *19*(6), 1219–1244. https://doi.org/10.5194/cp-19-1219-2023

Guo, C., Pleiss, G., Sun, Y., & Weinberger, K. Q. (2017, August 3). On calibration of modern neural networks. Retrieved March 18, 2024, from http://arxiv.org/abs/1706.04599

Haselhorst, D. S., Moreno, J. E., & Punyasena, S. W. (2020). Assessing the influence of vegetation structure and phenological variability on pollen-vegetation relationships using a 15-year neotropical pollen rain record. *Journal of Vegetation Science*, *31*(4), 606–615. https://doi.org/10.1111/jvs.12897

Hattestrand, M. (2013). Eight years of annual pollen monitoring in northern sweden, from the boreal forest to above the birch forest-line. *Grana*, *52*(1), 26–48. https://doi.org/10.1080/00173134.2012.744427

He, K., Zhang, X., Ren, S., & Sun, J. (2015, December 10). Deep residual learning for image recognition. https://doi.org/10.48550/arXiv.1512.03385

Hicks, S., Ammann, B., Latałowa, M., Pardoe, H., & Tinsley, H. (1996, January 1). *European pollen monitoring programme: project description and guidelines*.

Hicks, S., Tinsley, H., Huusko, A., Jensen, C., Hättestrand, M., Gerasimides, A., & Kvavadze, E. (2001). Some comments on spatial variation in arboreal pollen deposition: first records from the pollen monitoring programme (PMP). *Review of Palaeobotany and Palynology*, *117*(1), 183–194. https://doi.org/10.1016/S0034-6667(01)00086-0

Holt, K., Allen, G., Hodgson, R., Marsland, S., & Flenley, J. (2011). Progress towards an automated trainable pollen location and classifier system for use in the palynology laboratory. *Review of Palaeobotany and Palynology*, *167*(3), 175–183. https://doi.org/10.1016/j.revpalbo.2011.08.006

Holt, K. A., & Bennett, K. D. (2014). Principles and methods for automated palynology. *New Phytologist*, *203*(3), 735–742. https://doi.org/10.1111/nph.12848

Huang, G., Liu, Z., van der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. *30th Ieee Conference on Computer Vision and Pattern Recognition (cvpr 2017)*, 2261–2269. https://doi.org/10/gfhw3n

Hussain, A., Barua, B., Osman, A., Abozariba, R., & Asyhari, A. T. (2021). Low latency and non-intrusive accurate object detection in forests. *2021 Ieee Symposium Series on Computational Intelligence (ieee Ssci 2021)*. https://doi.org/10.1109/SSCI50451.2021.9660175

İnceoğlu, Ö., Pinar, N. M., & Dönmez, E. O. (2000). Pollen morphology of wild vitis sylvestris gmelin (vitaceae) özden Inceoğlu, n. münevver Pinar. *Turkish Journal of Botany*, *24*(2), 147–150.

Retrieved January 13, 2022, from https://journals.tubitak.gov.tr/botany/abstract.htm?id=3894

International Union for Conservation of Nature, I. (2016, March 28). *Mediterranean-type ecosystems* [IUCN: commission on ecosystem management]. Retrieved March 9, 2020, from https://www.iucn.org/commissions/commission-ecosystem-management/our-work/cems-specialist-groups/mediterranean-type-ecosystems

Iriarte-Chiapusso, M. J., Ocete-Perez, C. A., Hernandez-Beloqui, B., & Ocete-Rubio, R. (2017). Vitis vinifera in the iberian peninsula: a review. *Plant Biosystems*, *151*(2), 245–257. https://doi.org/10.1080/11263504.2016.1165751

Jackson, S. T., & Kearsley, J. B. (1998). Quantitative representation of local forest composition in forest-floor pollen assemblages. *Journal of Ecology*, *86*(3), 474–490. https://doi.org/10.1046/j.1365-2745.1998.00277.x

Jackson, S. T., & Lyford, M. E. (1999). Pollen dispersal models in quaternary plant ecology: assumptions, parameters, and prescriptions. *The Botanical Review*, *65*(1), 39–75. https://doi.org/10.1007/BF02856557

Jiang, P., Ergu, D., Liu, F., Cai, Y., & Ma, B. (2022). A review of yolo algorithm developments. *Procedia Computer Science*, *199*, 1066–1073. https://doi.org/10.1016/j.procs.2022.01.135

Jin, B., Milling, M., Plaza, M. P., Brunner, J. O., Traidl-Hoffmann, C., Schuller, B. W., & Damialis, A. (2023). Airborne pollen grain detection from partially labelled data utilising semi-supervised learning. *Science of The Total Environment*, *891*, 164295. https://doi.org/10.1016/j.scitotenv.2023.164295

Joannin, S., Brugiapaglia, E., de Beaulieu, J.-L., Bernardo, L., Magny, M., Peyron, O., Goring, S., & Vannière, B. (2012). Pollen-based reconstruction of holocene vegetation and climate in southern italy: the case of lago trifoglietti. *Climate of the Past*, *8*(6), 1973–1996. https://doi.org/10.5194/cp-8-1973-2012

Jocher, G. (2020). *Ultralytics YOLOv5* (Version 7.0). https://doi.org/10.5281/zenodo.3908559

Johnsrud, S., Yang, H., Nayak, A., & Waduge Punyasena, S. (2013). Semi-automated segmentation of pollen grains in microscopic images: a tool for three imaging modes. *Grana*, *52*(3), 181–191. https://doi.org/10.1080/00173134.2013.768291

Joly, C., Barillé, L., Barreau, M., Mancheron, A., & Visset, L. (2007). Grain and annulus diameter as criteria for distinguishing pollen grains of cereals from wild grasses. *Review of Palaeobotany and Palynology*, *146*(1), 221–233. https://doi.org/10.1016/j.revpalbo.2007.04.003

Jones, G. D., & Bryant, V. M. (2007). A comparison of pollen counts: light versus scanning electron microscopy. *Grana*, *46*(1), 20–33. https://doi.org/10.1080/00173130601173897

Jovanovic-Cvetkovic, T., Micic, N., Djuric, G., & Cvetkovic, M. (2016). Pollen morphology and germination of indigenous grapevine cultivars Zilavka and Blatina (vitis vinifera l.) *Agrolife Scientific Journal*, *5*(1), 105–109. Retrieved January 17, 2022, from https://www.webofscience.com/wos/woscc/full-record/WOS:000378272900014

Jubayer, F., Soeb, J. A., Mojumder, A. N., Paul, M. K., Barua, P., Kayshar, S., Akter, S. S., Rahman, M., & Islam, A. (2021). Detection of mold on the food surface using YOLOv5. *Current Research in Food Science*, *4*, 724–728. https://doi.org/10.1016/j.crfs.2021.10.003

Khanzhina, N., Filchenkov, A., Minaeva, N., Novoselova, L., Petukhov, M., Kharisova, I., Pinaeva, J., Zamorin, G., Putin, E., Zamyatina, E., & Shalyto, A. (2022). Combating data incompetence in pollen images detection and classification for pollinosis prevention. *Computers in Biology and Medicine*, *140*, 105064. https://doi.org/10.1016/j.compbiomed.2021.105064

Khanzhina, N., Putin, E., Filchenkov, A., & Zamyatina, E. (2018). Pollen grain recognition using convolutional neural network. *Computational Intelligence*, 6.

Kim, H.-C., Pang, S., Je, H.-M., Kim, D., & Yang Bang, S. (2003). Constructing support vector machine ensemble. *Pattern Recognition*, *36*(12), 2757–2767. https://doi.org/10.1016/S0031-3203(03)00175-4

Klein Goldewijk, K., Beusen, A., van Drecht, G., & de Vos, M. (2011). The HYDE 3.1 spatially explicit database of human-induced global land-use change over the past 12,000 years. *Global Ecology and Biogeography*, *20*(1), 73–86. https://doi.org/10.1111/j.1466-8238.2010.00587.x

Knausgard, K. M., Wiklund, A., Sordalen, T. K., Halvorsen, K. T., Kleiven, A. R., Jiao, L., & Goodwin, M. (2022). Temperate fish detection and classification: a deep learning based approach. *Applied Intelligence*, *52*(6), 6988–7001. https://doi.org/10.1007/s10489-020-02154-9

Kong, S., Punyasena, S., & Fowlkes, C. (2016, May 3). Spatially aware dictionary learning and coding for fossil pollen identification. https://doi.org/10.48550/arXiv.1605.00775

Krebs, P., Pezzatti, G. B., Beffa, G., Tinner, W., & Conedera, M. (2019). Revising the sweet chestnut (*Castanea sativa* mill.) refugia history of the last glacial period with extended pollen and macrofossil evidence. *Quaternary Science Reviews*, *206*, 111–128. https://doi.org/10.1016/j.quascirev.2019.01.002

Krishnadas, P., Chadaga, K., Sampathila, N., Rao, S., Swathi, K. S. K., & Prabhu, S. (2022). Classification of malaria using object detection models. *Informatics-Basel*, *9*(4), 76. https://doi.org/10.3390/informatics9040076

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*, *25*. Retrieved September 14, 2024, from https://proceedings.neurips.cc/paper/2012/hash/c399862d3b9d6b76c8436e924a68c45b-Abstract.html

Kubera, E., Kubik-Komar, A., Kurasinski, P., Piotrowska-Weryszko, K., & Skrzypiec, M. (2022). Detection and recognition of pollen grains in multilabel microscopic images. *Sensors*, *22*(7), 2690. https://doi.org/10.3390/s22072690

Kubera, E., Kubik-Komar, A., Piotrowska-Weryszko, K., & Skrzypiec, M. (2021). Deep learning methods for improving pollen monitoring. *Sensors*, *21*(10), 3526. https://doi.org/10.3390/s21103526

Kumari, S., & Singh, P. (2024). Deep learning for unsupervised domain adaptation in medical imaging: recent advancements and future perspectives. *Computers in Biology and Medicine*, *170*, 107912. https://doi.org/10.1016/j.compbiomed.2023.107912

Kurtz, E. B., & Liverman, J. L. (1958). Some effects of temperature on pollen characters. *Bulletin of the Torrey Botanical Club*, *85*(2), 136–138. https://doi.org/10.2307/2483028

Langgut, D., Cheddadi, R., Carrión, J. S., Cavanagh, M., Colombaroli, D., Eastwood, W. J., Greenberg, R., Litt, T., Mercuri, A. M., Miebach, A., Roberts, C. N., Woldring, H., & Woodbridge, J. (2019). The origin and spread of olive cultivation in the mediterranean basin: the fossil pollen evidence. *The Holocene*, *29*(5), 902–922. https://doi.org/10.1177/0959683619826654

Latorre, F., Romero, E. J., & Mancini, M. V. (2008). Comparative study of different methods for capturing airborne pollen, and effects of vegetation and meteorological variables. *Aerobiologia*, *24*(2), 107–120. https://doi.org/10.1007/s10453-008-9090-6

Lazar, J., Prasad, S., Barboni, D., Das, L., Kumaresan, V., & Anupama, K. (2024). Diversity matters: diet of apis cerana in southeast india includes one consistently occurring and several seasonally available floral sources. *Palynology*, *48*(1), 2255990. https://doi.org/10.1080/01916122.2023.2255990

LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., & Jackel, L. D. (1989). Backpropagation applied to handwritten zip code recognition. *Neural Computation*, *1*(4), 541–551. https://doi.org/10.1162/neco.1989.1.4.541

Levadoux, L. (1956). Les populations sauvages et cultivées de Vitis Vinifera L. *Annales de l' Amélioration des Plantes*, (1), 59–117.

Levetin, E., Pityn, P. J., Ramon, G. D., Pityn, E., Anderson, J., Bielory, L., Dalan, D., Codina, R., Rivera-Mariani, F. E., & Bolanos, B. (2023). Aeroallergen monitoring by the national allergy bureau: a review of the past and a look into the future. *The Journal of Allergy and Clinical Immunology: In Practice*, *11*(5), 1394–1400. https://doi.org/10.1016/j.jaip.2022.11.026

Li, C., Polling, M., Cao, L., Gravendeel, B., & Verbeek, F. J. (2023). Analysis of automatic image classification methods for urticaceae pollen classification. *Neurocomputing*, *522*, 181–193. https://doi.org/10.1016/j.neucom.2022.11.042

Li, F.-S., Phyo, P., Jacobowitz, J., Hong, M., & Weng, J.-K. (2019). The molecular structure of plant sporopollenin. *Nature Plants*, *5*(1), 41–46. https://doi.org/10.1038/s41477-018-0330-7

Li, K., Zhai, L., Pan, H., Shi, Y., Ding, X., & Cui, Y. (2022). Identification of the operating position and orientation of a robotic kiwifruit pollinator. *Biosystems Engineering*, *222*, 29–44. https://doi.org/10.1016/j.biosystemseng.2022.07.014

Lin, T.-Y., Goyal, P., Girshick, R., He, K., & Dollar, P. (2017). Focal loss for dense object detection, 2980–2988. Retrieved October 9, 2023, from https://openaccess.thecvf.com/content_iccv_2017/html/Lin_Focal_Loss_for_ICCV_2017_paper.html

Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., & Zitnick, C. L. (2014). Microsoft COCO: common objects in context. In D. Fleet, T. Pajdla, B. Schiele, &

T. Tuytelaars (Eds.), *Computer vision – ECCV 2014* (pp. 740–755). Springer International Publishing. https://doi.org/10.1007/978-3-319-10602-1_48

Lin, T. (2015). *labelImg: LabelImg is a graphical image annotation tool and label object bounding boxes in images*. HumanSignal. Retrieved October 26, 2023, from https://github.com/HumanSignal/labelImg

Lindenmayer, D. B., & Likens, G. E. (2010). The science and application of ecological monitoring. *Biological Conservation*, *143*(6), 1317–1328. https://doi.org/10.1016/j.biocon.2010.02.013

Linneberg, A., Dam Petersen, K., Hahn-Pedersen, J., Hammerby, E., Serup-Hansen, N., & Boxall, N. (2016). Burden of allergic respiratory disease: a systematic review. *Clinical and Molecular Allergy*, *14*(1), 12. https://doi.org/10.1186/s12948-016-0049-9

Lionello, P., Malanotte-Rizzoli, P., Boscolo, R., Alpert, P., Artale, V., Li, L., Luterbacher, J., May, W., Trigo, R., Tsimplis, M., Ulbrich, U., & Xoplaki, E. (2006, January 1). The mediterranean climate: an overview of the main characteristics and issues. In P. Lionello, P. Malanotte-Rizzoli, & R. Boscolo (Eds.), *Developments in earth and environmental sciences* (pp. 1–26, Vol. 4). Elsevier. https://doi.org/10.1016/S1571-9197(06)80003-0

Lionello, P., & Scarascia, L. (2018). The relation between climate change in the mediterranean region and global warming. *Regional Environmental Change*, *18*(5), 1481–1493. https://doi.org/10.1007/s10113-018-1290-1

Lombardo, G., Cargnello, G., Bassi, M., Gerola, F., & Carraro, L. (1978). Pollen ultrastructure in different vine cultivars with low productivity. *Vitis*, *17*(3), 221–228. Retrieved May 3, 2022, from https://www.webofscience.com/wos/woscc/full-record/WOS:A1978FZ00800001

Lu, X., Ye, X., & Liu, J. (2022). Morphological differences between anemophilous and entomophilous pollen. *Microscopy Research and Technique*, *85*(3), 1056–1064. https://doi.org/10.1002/jemt.23975

Luksic, K., Zdunic, G., Mucalo, A., Marinov, L., Rankovic-Vasic, Z., Ivanovic, J., & Nikolic, D. (2022). Microstructure of croatian wild grapevine (*Vitis vinifera* subsp. *sylvestris* gmel hegi) pollen grains revealed by scanning electron microscopy. *Plants-Basel*, *11*(11), 1479. https://doi.org/10.3390/plants11111479

Mahmood, T., Choi, J., & Park, K. R. (2023). Artificial intelligence-based classification of pollen grains using attention-guided pollen features aggregation network. *Journal of King Saud University - Computer and Information Sciences*, *35*(2), 740–756. https://doi.org/10.1016/j.jksuci.2023.01.013

Marasali, B., Pinar, M., & Büyükkartal, H. N. (2005). Palynological study on the pollen grains of selected turkish grape (vitis vinifera l.) cultivars. *Turkish Journal of Agriculture and Forestry*, *29*(1), 75–81. Retrieved January 13, 2022, from https://journals.tubitak.gov.tr/agriculture/abstract.htm?id=7400

Marchant, R., Tetard, M., Pratiwi, A., Adebayo, M., & de Garidel-Thoron, T. (2020a). Automated analysis of foraminifera fossil records by image classification using a convolutional neural network. *Journal of Micropalaeontology*, *39*(2), 183–202. https://doi.org/10.5194/jm-39-183-2020

Marchant, R., Tetard, M., Pratiwi, A., Adebayo, M., & de Garidel-Thoron, T. (2020b). Automated analysis of foraminifera fossil records by image classification using a convolutional neural network. *Journal of Micropalaeontology*, *39*(2), 183–202. https://doi.org/10.5194/jm-39-183-2020

Marini, S., Bonofiglio, F., Corgnati, L. P., Bordone, A., Schiaparelli, S., & Peirano, A. (2022). Long-term automated visual monitoring of antarctic benthic fauna. *Methods in Ecology and Evolution*, *13*(8), 1746–1764. https://doi.org/10.1111/2041-210X.13898

Matavulj, P., Panić, M., Šikoparija, B., Tešendić, D., Radovanović, M., & Brdar, S. (2023). Advanced CNN architectures for pollen classification: design and comprehensive evaluation. *Applied Artificial Intelligence*, *37*(1), 2157593. https://doi.org/10.1080/08839514.2022.2157593

McGovern, P. (2013, October 31). *Ancient wine: the search for the origins of viniculture*. Princeton University Press. https://doi.org/10.1515/9781400849536

McGovern, P., Jalabadze, M., Batiuk, S., Callahan, M. P., Smith, K. E., Hall, G. R., Kvavadze, E., Maghradze, D., Rusishvili, N., Bouby, L., Failla, O., Cola, G., Mariani, L., Boaretto, E., Bacilieri, R., This, P., Wales, N., & Lordkipanidze, D. (2017). Early neolithic wine of georgia in the south caucasus. *Proceedings of the National Academy of Sciences*, *114*(48), E10309–E10318. https://doi.org/10.1073/pnas.1714728114

Menzel, A., Sparks, T. H., Estrella, N., Koch, E., Aasa, A., Ahas, R., Alm-Kübler, K., Bissolli, P., Braslavská, O., Briede, A., Chmielewski, F. M., Crepinsek, Z., Curnel, Y., Dahl, Å., Defila, C., Donnelly, A., Filella, Y., Jatczak, K., Måge, F., ... Zust, A. (2006). European phenological response to climate change matches the warming pattern. *Global Change Biology*, *12*(10), 1969–1976. https://doi.org/10.1111/j.1365-2486.2006.01193.x

Mercuri, A. M., Accorsi, C. A., & Bandini Mazzanti, M. (2002). The long history of cannabis and its cultivation by the romans in central italy, shown by pollen records from lago albano and lago di nemi. *Vegetation History and Archaeobotany*, *11*(4), 263–276. https://doi.org/10.1007/s003340200039

Mercuri, A. M., Torri, P., Florenzano, A., Clò, E., Mariotti Lippi, M., Sgarbi, E., & Bignami, C. (2021). Sharing the agrarian knowledge with archaeology: first evidence of the dimorphism of vitis pollen from the middle bronze age of n italy (terramara santa rosa di poviglio). *Sustainability*, *13*(4), 2287. https://doi.org/10.3390/su13042287

Miller, N. F. (2008). Sweeter than wine? the use of the grape in early western asia. *Antiquity*, *82*(318), 937–946. https://doi.org/10.1017/S0003598X00097696

Mohammed, A., & Kora, R. (2023). A comprehensive review on ensemble deep learning: opportunities and challenges. *Journal of King Saud University - Computer and Information Sciences*, *35*(2), 757–774. https://doi.org/10.1016/j.jksuci.2023.01.014

Morente-López, J., Lara-Romero, C., Ornosa, C., & Iriondo, J. M. (2018). Phenology drives species interactions and modularity in a plant - flower visitor network. *Scientific Reports*, *8*(1), 9386. https://doi.org/10.1038/s41598-018-27725-2

Myers, N., Mittermeier, R. A., Mittermeier, C. G., da Fonseca, G. A. B., & Kent, J. (2000). Biodiversity hotspots for conservation priorities. *Nature*, *403*(6772), 853–858. https://doi.org/10.1038/35002501

Myles, S., Boyko, A. R., Owens, C. L., Brown, P. J., Grassi, F., Aradhya, M. K., Prins, B., Reynolds, A., Chia, J.-M., Ware, D., Bustamante, C. D., & Buckler, E. S. (2011). Genetic structure and domestication history of the grape. *Proceedings of the National Academy of Sciences*, *108*(9), 3530–3535. https://doi.org/10.1073/pnas.1009363108

Naqinezhad, A., Ramezani, E., Djamali, M., Schnitzler, A., & Arnold, C. (2018). Wild grapevine (vitis vinifera subsp sylvestris) in the hyrcanian relict forests of northern iran: an overview of current taxonomy, ecology and palaeorecords. *Journal of Forestry Research*, *29*(6), 1757–1768. https://doi.org/10.1007/s11676-017-0549-6

Newbold, T., Oppenheimer, P., Etard, A., & Williams, J. J. (2020). Tropical and mediterranean biodiversity is disproportionately sensitive to land-use and climate change. *Nature Ecology & Evolution*, *4*(12), 1630–1638. https://doi.org/10.1038/s41559-020-01303-0

Nielsen, A. B., Møller, P. F., Giesecke, T., Stavngaard, B., Fontana, S. L., & Bradshaw, R. H. W. (2010). The effect of climate conditions on inter-annual flowering variability monitored by pollen traps below the canopy in draved forest, denmark. *Vegetation History and Archaeobotany*, *19*(4), 309–323. https://doi.org/10.1007/s00334-010-0253-3

Olsson, O., Karlsson, M., Persson, A. S., Smith, H. G., Varadarajan, V., Yourstone, J., & Stjernman, M. (2021). Efficient, automated and robust pollen analysis using deep learning (R. Freckleton, Ed.). *Methods in Ecology and Evolution*, *12*(5), 850–862. https://doi.org/10.1111/2041-210X.13575

Oteros, J., Orlandi, F., Garcia-Mozo, H., Aguilera, F., Ben Dhiab, A., Bonofiglio, T., Abichou, M., Ruiz-Valenzuela, L., Mar del Trigo, M., Diaz de la Guardia, C., Dominguez-Vilches, E., Msallem, M., Fornaciari, M., & Galan, C. (2014). Better prediction of mediterranean olive production using pollen-based models. *Agronomy for Sustainable Development*, *34*(3), 685–694. https://doi.org/10.1007/s13593-013-0198-x

Oteros, J., Sofiev, M., Smith, M., Clot, B., Damialis, A., Prank, M., Werchan, M., Wachter, R., Weber, A., Kutzora, S., Heinze, S., Herr, C. E. W., Menzel, A., Bergmann, K.-C., Traidl-Hoffmann, C., Schmidt-Weber, C. B., & Buters, J. T. M. (2019). Building an automatic pollen monitoring network (ePIN): selection of optimal sites by clustering pollen stations. *Science of The Total Environment*, *688*, 1263–1274. https://doi.org/10.1016/j.scitotenv.2019.06.131

Oteros, J., Weber, A., Kutzora, S., Rojo, J., Heinze, S., Herr, C., Gebauer, R., Schmidt-Weber, C. B., & Buters, J. T. M. (2020). An operational robotic pollen monitoring network based on automatic image recognition. *Environmental Research*, *191*, 110031. https://doi.org/10.1016/j.envres.2020.110031

Pardoe, H. S., Giesecke, T., van der Knaap, W. O., Svitavská-Svobodová, H., Kvavadze, E. V., Panajiotidis, S., Gerasimidis, A., Pidek, I. A., Zimny, M., Świeta-Musznicka, J., Latałowa, M., Noryśkiewicz, A. M., Bozilova, E., Tonkov, S., Filipova-Marinova, M. V., van Leeuwen, J. F. N., & Kalniņa, L. (2010). Comparing pollen spectra from modified tauber traps and moss samples: examples from a selection of woodlands across europe. *Vegetation History and Archaeobotany*, *19*(4), 271–283. https://doi.org/10.1007/s00334-010-0258-y

Parmesan, C. (2006). Ecological and evolutionary responses to recent climate change. *Annual Review of Ecology, Evolution, and Systematics*, *37*, 637–669. https://doi.org/10.1146/annurev.ecolsys.37.091305.110100

Parmesan, C., & Yohe, G. (2003). A globally coherent fingerprint of climate change impacts across natural systems. *Nature*, *421*(6918), 37–42. https://doi.org/10.1038/nature01286

Payne, W. W. (1981). Structure and function in angiosperm pollen wall evolution. *Review of Palaeobotany and Palynology*, *35*(1), 39–59. https://doi.org/10.1016/0034-6667(81)90013-0

Peñuelas, J., Filella, I., Zhang, X., Llorens, L., Ogaya, R., Lloret, F., Comas, P., Estiarte, M., & Terradas, J. (2004). Complex spatiotemporal phenological shifts as a response to rainfall changes. *New Phytologist*, *161*(3), 837–846. https://doi.org/10.1111/j.1469-8137.2004.01003.x

Perveen, A., & Qaiser, M. (2008). Pollen flora of pakistan-LVII. vitaceae. *Pakistan Journal of Botany*, *40*(2), 501–506. Retrieved January 13, 2022, from https://www.webofscience.com/wos/woscc/full-record/WOS:000257835400004

Peyron, O., Combourieu-Nebout, N., Brayshaw, D., Goring, S., Andrieu-Ponel, V., Desprat, S., Fletcher, W., Gambin, B., Ioakim, C., Joannin, S., Kotthoff, U., Kouli, K., Montade, V., Pross, J., Sadori, L., & Magny, M. (2017). Precipitation changes in the mediterranean basin during the holocene from terrestrial and marine pollen records: a model–data comparison. *Climate of the Past*, *13*(3), 249–265. https://doi.org/10.5194/cp-13-249-2017

Pidek, I. A., Piotrowska, K., & Kasprzyk, I. (2010). Pollen-vegetation relationships for pine and spruce in southeast poland on the basis of volumetric and tauber trap records. *Grana*, *49*(3), 215–226. https://doi.org/10.1080/00173134.2010.514006

Polade, S. D., Pierce, D. W., Cayan, D. R., Gershunov, A., & Dettinger, M. D. (2015). The key role of dry days in changing regional climate and precipitation regimes. *Scientific Reports*, *4*(1), 4364. https://doi.org/10.1038/srep04364

Pollegioni, P., Woeste, K., Chiocchini, F., Lungo, S. D., Ciolfi, M., Olimpieri, I., Tortolano, V., Clark, J., Hemery, G. E., Mapelli, S., & Malvolti, M. E. (2017). Rethinking the history of common walnut (juglans regia l.) in europe: its origins and human interactions. *PLOS ONE*, *12*(3), e0172541. https://doi.org/10.1371/journal.pone.0172541

Portillo, M., Ball, T. B., Wallace, M., Murphy, C., Perez-Diaz, S., Ruiz-Alonso, M., Aceituno, F. J., & Lopez-Saez, J. A. (2020). Advances in morphometrics in archaeobotany. *Environmental Archaeology*, *25*(2), 246–256. https://doi.org/10/gf5zdx

Pratt, C. (1971). Reproductive anatomy in cultivated grapes - a review. *American Journal of Enology and Viticulture*, *22*(2), 92–109. https://doi.org/10.5344/ajev.1971.22.2.92

Preiner, D., Kontić, J. K., Šimon, S., Marković, Z., Stupić, D., & Maletić, E. (2012). Intravarietal agronomic variability in croatian native vitis vinifera l. cultivar grk with female flower and seedless berries. *American Journal of Enology and Viticulture*, *63*(2), 291–295. https://doi.org/10.5344/ajev.2012.11121

Prentice, I. C., & Webb III, T. (1986). Pollen percentages, tree abundances and the fagerlind effect. *Journal of Quaternary Science*, *1*(1), 35–43. https://doi.org/10.1002/jqs.3390010105

Pruvost, E., Tulet, H., Delon, E., Shokouh, G. S., Montesinos, P., Magnier, B., Jourdan, C., Belaud, E., & Lledde, M. (2022). Invertebrates detection with YOLOv5: towards study of soil organisms using deep learning. *2022 10th European Workshop on Visual Information Processing (euvip)*. https://doi.org/10.1109/EUVIP53989.2022.9922675

Punt, W., Hoen, P., Blackmore, S., Nilsson†, S., & Le Thomas, A. (2007). Glossary of pollen and spore terminology. *Review of Palaeobotany and Palynology*, *143*(1), 1–81. https://doi.org/10.1016/j.revpalbo.2006.06.008

Punyasena, S. W., Haselhorst, D. S., Kong, S., Fowlkes, C. C., & Moreno, J. E. (2022). Automated identification of diverse neotropical pollen samples using convolutional neural networks. *Methods in Ecology and Evolution*, *13*(9), 2049–2064. https://doi.org/10.1111/2041-210X.13917

Punyasena, S. W., Tcheng, D. K., Wesseln, C., & Mueller, P. G. (2012). Classifying black and white spruce pollen using layered machine learning. *New Phytologist*, *196*(3), 937–944. https://doi.org/10.1111/j.1469-8137.2012.04291.x

Quamar, M. F., Singh, P., Garg, A., Tripathi, S., Farooqui, A., Shukla, A. N., & Prasad, N. (2022). Pollen characters and their evolutionary and taxonomic significance: using light and confocal laser scanning microscope to study diverse plant pollen taxa from central india. *Palynology*, *46*(4), 1–13. https://doi.org/10.1080/01916122.2022.2070294

Quezel, P., & Medail, F. (2003). Que faut-il entendre par "forêts méditerranéennes" ? *Forêt Méditerranéenne*, *XXIV*(1), 11–31. Retrieved October 13, 2024, from https://hal.science/hal-03564585

Recio, M., Picornell, A., Trigo, M. M., Gharbi, D., Garcia-Sanchez, J., & Cabezudo, B. (2018). Intensity and temporality of airborne quercus pollen in the southwest mediterranean area: correlation with meteorological and phenoclimatic variables, trends and possible adaptation to climate change. *Agricultural and Forest Meteorology*, *250*, 308–318. https://doi.org/10.1016/j.agrformet.2017.11.028

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: unified, real-time object detection. *arXiv:1506.02640 [cs]*. Retrieved May 11, 2022, from http://arxiv.org/abs/1506.02640

Redmon, J., & Farhadi, A. (2016). YOLO9000: better, faster, stronger. *arXiv:1612.08242 [cs]*. Retrieved May 11, 2022, from http://arxiv.org/abs/1612.08242

Reille, M., & Pons, A. (1990). *Leçons de palynologie et d'analyse pollinique*. CNRS, Editions du.

Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster r-CNN: towards real-time object detection with region proposal networks. *Advances in Neural Information Processing Systems*, *28*. Retrieved May 18, 2022, from https://proceedings.neurips.cc/paper/2015/hash/14bfa6bb14875e45bba028a21ed38046-Abstract.html

Rocha, M., Claro, M., Neto, L., Aires, K., Machado, V., & Veras, R. (2023). Malaria parasites detection and identification using object detectors based on deep neural networks: a wide comparative analysis. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, *11*(3), 351–368. https://doi.org/10.1080/21681163.2022.2111715

Romero, I. C., Kong, S., Fowlkes, C. C., Jaramillo, C., Urban, M. A., Oboh-Ikuenobe, F., D'Apolito, C., & Punyasena, S. W. (2020). Improving the taxonomy of fossil pollen using convolutional neural networks and superresolution microscopy. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(45), 28496–28505. https://doi.org/10.1073/pnas.2007324117

Rostami, M. A., Balmaki, B., Dyer, L. A., Allen, J. M., Sallam, M. F., & Frontalini, F. (2023). Efficient pollen grain classification using pre-trained convolutional neural networks: a comprehensive study. *Journal of Big Data*, *10*(1), 151. https://doi.org/10.1186/s40537-023-00815-3

Sadori, L. (2018). The lateglacial and holocene vegetation and climate history of lago di mezzano (central italy). *Quaternary Science Reviews*, *202*, 30–44. https://doi.org/10.1016/j.quascirev.2018.09.004

Sauvageat, E., Zeder, Y., Auderset, K., Calpini, B., Clot, B., Crouzy, B., Konzelmann, T., Lieberherr, G., Tummon, F., & Vasilatou, K. (2020). Real-time pollen monitoring using digital holography. *Atmospheric Measurement Techniques*, *13*(3), 1539–1550. https://doi.org/10.5194/amt-13-1539-2020

Schaefer, J., Milling, M., Schuller, B. W., Bauer, B., Brunner, J. O., Traidl-Hoffmann, C., & Damialis, A. (2021). Towards automatic airborne pollen monitoring: from commercial devices to operational by mitigating class-imbalance in a deep learning approach. *Science of the Total Environment*, *796*, 148932. https://doi.org/10.1016/j.scitotenv.2021.148932

Schiele, J., Rabe, F., Schmitt, M., Glaser, M., Häring, F., Brunner, J. O., Bauer, B., Schuller, B., Traidl-Hoffmann, C., & Damialis, A. (2019). Automated classification of airborne pollen using neural networks. *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 4474–4478. https://doi.org/10.1109/EMBC.2019.8856910

Sevillano, V., & Aznarte, J. L. (2018). Improving classification of pollen grain images of the POLEN23e dataset through three different applications of deep learning convolutional neural networks (S. Rutherford, Ed.). *PLOS ONE*, *13*(9), e0201807. https://doi.org/10.1371/journal.pone.0201807

Sevillano, V., Holt, K., & Aznarte, J. L. (2020). Precise automatic classification of 46 different pollen types with convolutional neural networks (M. Raza, Ed.). *PLOS ONE*, *15*(6), e0229751. https://doi.org/10.1371/journal.pone.0229751

Simonyan, K., & Zisserman, A. (2015, April 10). Very deep convolutional networks for large-scale image recognition. https://doi.org/10.48550/arXiv.1409.1556

Sjögren, P., Connor, S. E., & van der Knaap, W. O. (2010). The development of composite dispersal functions for estimating absolute pollen productivity in the swiss alps. *Vegetation History and Archaeobotany*, *19*(4), 341–349. https://doi.org/10.1007/s00334-010-0247-1

Sozzi, M., Cantalamessa, S., Cogato, A., Kayad, A., & Marinello, F. (2022). Automatic bunch detection in white grape varieties using YOLOv3, YOLOv4, and YOLOv5 deep learning algorithms. *Agronomy-Basel*, *12*(2), 319. https://doi.org/10.3390/agronomy12020319

Stillman, E. C., & Flenley, J. R. (1996). The needs and prospects for automation in palynology. *Quaternary Science Reviews*, *15*(1), 1–5. https://doi.org/10.1016/0277-3791(95)00076-3

Stocker, T., Qin, D., & Plattner, G.-K. (2014). *Climate change 2013 : the physical science basis working group i contribution to the fifth assessment report of the intergovernmental panel on climate change*. Cambridge University Press.

Stockmarr, J. (1971). Tablets with spores used in absolute pollen analysis. *Pollen et Spores*, (13), 615–621.

Sugita, S. (2007a). Theory of quantitative reconstruction of vegetation i: pollen from large sites REVEALS regional vegetation composition. *The Holocene*, *17*(2), 229–241. https://doi.org/10.1177/0959683607075837

Sugita, S. (2007b). Theory of quantitative reconstruction of vegetation II: all you need is LOVE. *The Holocene*, *17*(2), 243–257. https://doi.org/10.1177/0959683607075838

Sugita, S., Hicks, S., & Sormunen, H. (2010). Absolute pollen productivity and pollen-vegetation relationships in northern finland. *Journal of Quaternary Science*, *25*(5), 724–736. https://doi.org/10.1002/jqs.1349

Sun, L., Xu, Y., Rao, Z., Chen, J., Liu, Z., & Lu, N. (2022). YOLO algorithm for long-term tracking and detection of escherichia coli at different depths of microchannels based on microsphere positioning assistance. *Sensors*, *22*(19), 7454. https://doi.org/10.3390/s22197454

Szegedy, C., Ioffe, S., Vanhoucke, V., & Alemi, A. (2017). Inception-v4, inception-ResNet and the impact of residual connections on learning. *Proceedings of the AAAI Conference on Artificial Intelligence*, *31*(1). https://doi.org/10.1609/aaai.v31i1.11231

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., & Rabinovich, A. (2015). Going deeper with convolutions. *2015 Ieee Conference on Computer Vision and Pattern Recognition (cvpr)*, 1–9. https://doi.org/10/gftjd8

Taia, W. K. (2022). Pollen grain diversity and application in taxonomy and evolution. *Taeckholmia*, *42*(1), 27–42. https://doi.org/10.21608/taec.2022.258213

Tan, M., Chao, W., Cheng, J.-K., Zhou, M., Ma, Y., Jiang, X., Ge, J., Yu, L., & Feng, L. (2022). Animal detection and classification from camera trap images using different mainstream object detection architectures. *Animals*, *12*(15), 1976. https://doi.org/10.3390/ani12151976

Tauber, H. (1965). Differential pollen dispersion and the interpretation of pollen diagrams. with a contribution to the interpretation of the elm fall. *Danmarks Geologiske Undersøgelse II. Række*, *89*, 1–69. https://doi.org/10.34194/raekke2.v89.6880

Terral, J.-F., Tabard, E., Bouby, L., Ivorra, S., Pastor, T., Figueiral, I., Picq, S., Chevance, J.-B., Jung, C., Fabre, L., Tardy, C., Compan, M., Bacilieri, R., Lacombe, T., & This, P. (2010). Evolution and history of grapevine (vitis vinifera) under domestication: new morphometric perspectives to understand seed domestication syndrome and reveal origins of ancient european cultivars. *Annals of Botany*, *105*(3), 443–455. https://doi.org/10/dq2k6v

Tetard, M., Marchant, R., Cortese, G., Gally, Y., de Garidel-Thoron, T., & Beaufort, L. (2020a, July 9). *A new automated radiolarian image acquisition, stacking, processing, segmentation, and identification workflow* (preprint). Proxy Use-Development-Validation/Marine Archives/Cenozoic. https://doi.org/10.5194/cp-2020-76

Tetard, M., Marchant, R., Cortese, G., Gally, Y., de Garidel-Thoron, T., & Beaufort, L. (2020b). Technical note: a new automated radiolarian image acquisition, stacking, processing, segmentation and identification workflow. *Climate of the Past*, *16*(6), 2415–2429. https://doi.org/10.5194/cp-16-2415-2020

Theuerkauf, M., Siradze, N., & Gillert, A. (2023). A trainable object finder, selector and identifier for pollen, spores and other things: a step towards automated pollen recognition in lake sediments. *The Holocene*, 09596836231211876. https://doi.org/10.1177/09596836231211876

This, P., Lacombe, T., & Thomas, M. R. (2006). Historical origins and genetic diversity of wine grapes. *Trends in genetics: TIG*, *22*(9), 511–519. https://doi.org/10.1016/j.tig.2006.07.008

This, P., Roux, C., Parra, P., Siret, R., Bourse, T., Adam, A.-F., Yvon, M., Lacombe, T., David, J., & Boursiquot, J.-M. (2001). Caractérisation de la diversité d'une population de vignes sauvages du Pic Saint-Loup (Hérault) et relations avec le compartiment cultivé. *Genetics Selection Evolution*, *33*(1), S289. https://doi.org/10.1186/BF03500885

Tormo-Molina, R., Gonzalo-Garijo, M. A., Silva-Palacios, I., & Munoz-Rodriguez, A. F. (2010). General trends in airborne pollen production and pollination periods at a mediterranean site (badajoz, southwest spain). *Journal of Investigational Allergology and Clinical Immunology*, *20*(7), 567–574. Retrieved January 17, 2023, from https://www.webofscience.com/wos/woscc/full-record/WOS:000286704300005

Tramblay, Y., Koutroulis, A., Samaniego, L., Vicente-Serrano, S. M., Volaire, F., Boone, A., Le Page, M., Llasat, M. C., Albergel, C., Burak, S., Cailleret, M., Kalin, K. C., Davi, H., Dupuy, J.-L., Greve, P., Grillakis, M., Hanich, L., Jarlan, L., Martin-StPaul, N., … Polcher, J. (2020). Challenges for drought assessment in the mediterranean region under future climate scenarios. *Earth-Science Reviews*, *210*, 103348. https://doi.org/10.1016/j.earscirev.2020.103348

Trivedi, A., Srivastava, A., Farooqui, A., Khan, S., Pokharia, A. K., Ferguson, D. K., & Singh, V. K. (2022). Pollen morphological study in subfamily papilionoideae using confocal laser scanning microscopy. *Journal of Palaeosciences*, *71*(2), 123–142. https://doi.org/10.54991/jop.2022.538

Tsiknakis, N., Savvidaki, E., Kafetzopoulos, S., Manikis, G., Vidakis, N., Marias, K., & Alissandrakis, E. (2021). Segmenting 20 types of pollen grains for the cretan pollen dataset v1 (CPD-1). *Applied Sciences-Basel*, *11*(14), 6657. https://doi.org/10.3390/app11146657

Tsiknakis, N., Savvidaki, E., Manikis, G. C., Gotsiou, P., Remoundou, I., Marias, K., Alissandrakis, E., & Vidakis, N. (2022). Pollen grain classification based on ensemble transfer learning on the cretan pollen dataset. *Plants*, *11*(7), 919. https://doi.org/10.3390/plants11070919

Tummon, F., Adams-Groom, B., Antunes, C. M., Bruffaerts, N., Buters, J., Cariñanos, P., Celenk, S., Choël, M., Clot, B., Cristofori, A., Crouzy, B., Damialis, A., Fernández, A. R., González, D. F., Galán, C., Gedda, B., Gehrig, R., Gonzalez-Alonso, M., Gottardini, E., … de Weger, L. A. (2024). The role of automatic pollen and fungal spore monitoring across major end-user domains. *Aerobiologia*, *40*(1), 57–75. https://doi.org/10.1007/s10453-024-09820-2

Tummon, F., Bruffaerts, N., Celenk, S., Choël, M., Clot, B., Crouzy, B., Galán, C., Gilge, S., Hajkova, L., Mokin, V., O'Connor, D., Rodinkova, V., Sauliene, I., Sikoparija, B., Sofiev, M., Sozinova, O., Tesendic, D., & Vasilatou, K. (2024). Towards standardisation of automatic pollen and fungal spore monitoring: best practises and guidelines. *Aerobiologia*, *40*(1), 39–55. https://doi.org/10.1007/s10453-022-09755-6

Turner, S. D., & Brown, A. G. (2004). Vitis pollen dispersal in and from organic vineyards: i. pollen trap and soil pollen data. *Review of Palaeobotany and Palynology*, *129*(3), 117–132. https://doi.org/10/bp6fp9

Tweddle, J. C., Edwards, K. J., & Fieller, N. R. J. (2005). Multivariate statistical and other approaches for the separation of cereal from wild poaceae pollen using a large holocene dataset. *Vegetation History and Archaeobotany*, *14*(1), 15–30. https://doi.org/10/bh3tp8

van der Knaap, W. O., van Leeuwen, J. F. N., Svitavska-Svobodova, H., Pidek, I. A., Kvavadze, E., Chichinadze, M., Giesecke, T., Kaszewski, B. M., Oberli, F., Kalnina, L., Pardoe, H. S., Tinner, W., & Ammann, B. (2010). Annual pollen traps reveal the complexity of climatic control on pollen productivity in europe and the caucasus. *Vegetation History and Archaeobotany*, *19*(4), 285–307. https://doi.org/10.1007/s00334-010-0250-6

Verstraeten, A., Bruffaerts, N., Cristofolini, F., Vanguelova, E., Neirynck, J., Genouw, G., De Vos, B., Waldner, P., Thimonier, A., Nussbaumer, A., Neumann, M., Benham, S., Rautio, P., Ukonmaanaho, L., Merilä, P., Lindroos, A.-J., Saarto, A., Reiniharju, J., Clarke, N., … Gottardini, E. (2023). Effects of tree pollen on throughfall element fluxes in european forests. *Biogeochemistry*, *165*(3), 311–325. https://doi.org/10.1007/s10533-023-01082-3

Viertel, P., Koenig, M., & Rexilius, J. (2022). *Pollen detection from honey sediments via region-based convolutional neural networks*. Gesellschaft für Informatik e.V. Retrieved March 9, 2022, from http://dl.gi.de/handle/20.500.12116/38415

Volkova, O., & Severova, E. (2017). A comparison of modern pollen rain collected by hirst type and tauber traps in moscow, russia. *Ecological Questions*, *26*, 15–17. https://doi.org/10.12775/EQ.2017.010

Von Der Ohe, W., Persano Oddo, L., Piana, M. L., Morlot, M., & Martin, P. (2004). Harmonized methods of melissopalynology. *Apidologie*, *35*, S18–S25. https://doi.org/10.1051/apido:2004050

von Allmen, R., Brugger, S. O., Schleicher, K. D., Rey, F., Gobet, E., Courtney Mustaphi, C. J., Tinner, W., & Heiri, O. (2024). Method development and application of object detection and classification to quaternary fossil pollen sequences. *Quaternary Science Reviews*, *327*, 108521. https://doi.org/10.1016/j.quascirev.2024.108521

Waithe, D., Brown, J. M., Reglinski, K., Diez-Sevilla, I., Roberts, D., & Eggeling, C. (2020). Object detection networks and augmented reality for cellular detection in fluorescence microscopy. *Journal of Cell Biology*, *219*(10), e201903166. https://doi.org/10.1083/jcb.201903166

Walther, G.-R., Post, E., Convey, P., Menzel, A., Parmesan, C., Beebee, T. J. C., Fromentin, J.-M., Hoegh-Guldberg, O., & Bairlein, F. (2002). Ecological responses to recent climate change. *Nature*, *416*(6879), 389–395. https://doi.org/10.1038/416389a

Wang, Z., Wang, Z., & Wang, L. (2021). Automatic 3d pollen recognition based on convolutional neural network. *Scientific Programming*, *2021*, 5577307. https://doi.org/10.1155/2021/5577307

Yang, S., Zheng, Z., Mao, L., Li, J., & Chen, B. (2018). Pollen morphology of selected crop plants from southern china and testing pollen morphological data in an archaeobotanical study. *Vegetation History and Archaeobotany*, *27*(6), 781–799. https://doi.org/10.1007/s00334-018-0696-5

Zecca, G., Abbott, J. R., Sun, W.-B., Spada, A., Sala, F., & Grassi, F. (2012). The timing and the mode of evolution of wild grapes (vitis). *Molecular Phylogenetics and Evolution*, *62*(2), 736–747. https://doi.org/10/dstqph

Zedda, L., Loddo, A., & Di Ruberto, C. (2022). A deep learning based framework for malaria diagnosis on high variation data set. In S. Sclaroff, C. Distante, M. Leo, G. M. Farinella, & F. Tombari (Eds.), *Image analysis and processing, iciap 2022, pt ii* (pp. 358–370, Vol. 13232). Springer International Publishing Ag. https://doi.org/10.1007/978-3-031-06430-2_30

Zhang, C.-J., Liu, T., Wang, J., Zhai, D., Chen, M., Gao, Y., Yu, J., & Wu, H.-Z. (2024). DeepPollenCount: a swin-transformer-YOLOv5-based deep learning method for pollen counting in various plant species. *Aerobiologia*. https://doi.org/10.1007/s10453-024-09828-8

Zhao, L.-N., Li, J.-Q., Cheng, W.-X., Liu, S.-Q., Gao, Z.-K., Xu, X., Ye, C.-H., & You, H.-L. (2022). Simulation palynologists for pollinosis prevention: a progressive learning of pollen localization and classification for whole slide images. *Biology-Basel*, *11*(12), 1841. https://doi.org/10.3390/biology11121841

Zhuang, J., Li, X., Bagavathiannan, M., Jin, X., Yang, J., Meng, W., Li, T., Li, L., Wang, Y., Chen, Y., & Yu, J. (2022). Evaluation of different deep convolutional neural networks for detection of broadleaf weed seedlings in wheat. *Pest Management Science*, *78*(2), 521–529. https://doi.org/10.1002/ps.6656

Ziello, C., Sparks, T. H., Estrella, N., Belmonte, J., Bergmann, K. C., Bucher, E., Brighetti, M. A., Damialis, A., Detandt, M., Galán, C., Gehrig, R., Grewling, L., Bustillo, A. M. G., Hallsdóttir, M., Kockhans-Bieda, M.-C., Linares, C. D., Myszkowska, D., Pàldy, A., Sánchez, A., … Menzel, A. (2012). Changes to airborne pollen counts across europe. *PLOS ONE*, *7*(4), e34076. https://doi.org/10.1371/journal.pone.0034076

Ziska, L. H., Makra, L., Harry, S. K., Bruffaerts, N., Hendrickx, M., Coates, F., Saarto, A., Thibaudon, M., Oliver, G., Damialis, A., Charalampopoulos, A., Vokou, D., Heiđmarsson, S., Gudjohnsen, E., Bonini, M., Oh, J.-W., Sullivan, K., Ford, L., Brooks, G. D., … Crimmins, A. R. (2019). Temperature-related changes in airborne allergenic pollen abundance and seasonality across the northern hemisphere: a retrospective data analysis. *The Lancet Planetary Health*, *3*(3), e124–e131. https://doi.org/10.1016/S2542-5196(19)30015-4

Zito, P., Serraino, F., Carimi, F., Tavella, F., & Sajeva, M. (2018). Inflorescence-visiting insects of a functionally dioecious wild grapevine (vitis vinifera subsp sylvestris). *Genetic Resources and Crop Evolution*, *65*(5), 1329–1335. https://doi.org/10/gdjjxb

Zohary, D. (1996). The domestication of the grapevine vitis vinifera l. in the near east. In *The origins and ancient history of wine*. Routledge.

Zohary, D., Hopf, M., & Weiss, E. (2012). *Domestication of plants in the old world: the origin and spread of domesticated plants in southwest asia, europe, and the mediterranean basin* (4th ed). Oxford University Press.

Zu, B., Cao, T., Li, Y., Li, J., Ju, F., & Wang, H. (2024). SwinT-SRNet: swin transformer with image super-resolution reconstruction network for pollen images classification. *Engineering Applications of Artificial Intelligence*, *133*, 108041. https://doi.org/10.1016/j.engappai.2024.108041