

2025 Activity Report

RESEARCH CENTRE: Inria Branch at the University of Montpellier

IN PARTNERSHIP WITH: Université de Montpellier

Project-Team

IROKO

Data Driven Environmental Sciences

In collaboration with Laboratoire d'informatique, de robotique et de microélectronique de Montpellier (LIRMM), Institut Montpellierain Alexander Grothendieck (IMAG)



Project-Team IROKO

Creation of the Project-Team: 2024 October 01

Each year, Inria research teams publish an Activity Report presenting their work and results over the reporting period. These reports follow a common structure, with some optional sections depending on the specific team. They typically begin by outlining the overall objectives and research programme, including the main research themes, goals, and methodological approaches. They also describe the application domains targeted by the team, highlighting the scientific or societal contexts in which their work is situated. The reports then present the highlights of the year, covering major scientific achievements, software developments, or teaching contributions. When relevant, they include sections on software, platforms, and open data, detailing the tools developed and how they are shared. A substantial part is dedicated to new results, where scientific contributions are described in detail, often with subsections specifying participants and associated keywords. Finally, the Activity Report addresses funding, contracts, partnerships, and collaborations at various levels, from industrial agreements to international cooperations. It also covers dissemination and teaching activities, such as participation in scientific events, outreach, and supervision. The document concludes with a presentation of scientific production, including major publications and those produced during the year.

Keywords

Computer sciences and digital sciences

- A3.1. – Data
- A3.1.2. – Data management, quering and storage
- A3.1.3. – Distributed data
- A3.1.4. – Uncertain data
- A3.1.5. – Control access, privacy
- A3.1.7. – Open data
- A3.1.8. – Big data (production, storage, transfer)
- A3.1.9. – Database
- A3.1.10. – Heterogeneous data
- A3.1.11. – Structured data
- A3.2. – Knowledge
- A3.2.2. – Knowledge extraction, cleaning
- A3.3. – Data and knowledge analysis
- A3.3.2. – Data mining
- A3.3.3. – Big data analysis
- A3.4. – Machine learning and statistics
- A5.2. – Data visualization
- A5.3. – Image processing and analysis
- A5.3.3. – Pattern recognition
- A9. – Artificial intelligence
- A9.2. – Machine learning
- A9.2.1. – Supervised learning
- A9.2.3. – Reinforcement learning
- A9.2.6. – Neural networks
- A9.2.8. – Deep learning
- A9.12.3. – Content retrieval

Other research topics and application domains

- B1. – Life sciences
- B1.1. – Biology
- B2.6. – Biological and medical imaging
- B3. – Environment and planet
- B3.2. – Climate and meteorology
- B3.3. – Geosciences
- B3.5. – Agronomy
- B3.6. – Ecology
- B3.6.1. – Biodiversity

Contents

Project-Team IROKO	1
1 Team members, visitors, external collaborators	5
2 Overall objectives	7
3 Research program	8
3.1 Big Data and Scalability	9
3.2 Machine Learning with Humans in the Loop	9
3.3 Multiscale & Multimodal Data Analytics	10
4 Application domains	10
5 Social and environmental responsibility	11
5.1 Footprint of research activities	11
5.2 Impact of research results	12
6 Highlights of the year	12
6.1 Awards	12
6.2 Other key achievements	12
7 Latest software developments, platforms, open data	12
7.1 New Features in the PI@ntNet Platform	12
7.2 New platforms	13
7.2.1 GeoPI@ntNet	13
7.2.2 PROMISE	13
7.3 Open data	14
8 New results	14
8.1 Distributed Data and Model Management	14
8.1.1 A Logic-Based Approach for Knowledge Graph Data Integration	14
8.1.2 Federated Learning	15
8.1.3 Distributed Web Infrastructure for Integrated Pest Management	15
8.2 Data Analytics	15
8.2.1 Event Detection in Time Series	15
8.2.2 Scalable Multivariate Anomaly Detection	16
8.2.3 Detecting Anomalies with Any Duration in Climate Time Series	16
8.2.4 Energy Efficient Time Series Anomaly Detection	17
8.2.5 Extending Matrix Profile for Seasonal Anomaly Detection	17
8.3 Machine Learning for Biodiversity and Agroecology	17
8.3.1 Learning Ecological Structure with Large Language Models	17
8.3.2 Scalable Plant Vision Models for Operational Monitoring	18
8.3.3 Conformal Prediction for uncertainty quantification	18
8.3.4 AI-Based Species Distribution Modeling and Mapping	19
8.3.5 Coral Reef Monitoring	19
8.3.6 Evaluation of Species Identification and Prediction Algorithms	20
8.3.7 Importance of fossil pollen data for vegetation species distribution modeling	20
9 Bilateral contracts and grants with industry	21

10 Partnerships and cooperations	21
10.1 International initiatives	21
10.1.1 Associate Teams in the framework of an Inria International Lab or in the framework of an Inria International Program	21
10.1.2 Participation in other International Programs	22
10.2 International research visitors	22
10.2.1 Visits of international scientists	22
10.3 European initiatives	23
10.3.1 Horizon Europe	23
10.4 National initiatives	26
10.4.1 Others	29
10.5 Regional initiatives	30
10.6 Public policy support	30
11 Dissemination	30
11.1 Promoting scientific activities	30
11.1.1 Scientific events: organization	30
11.1.2 Scientific events: selection	30
11.1.3 Journal	31
11.1.4 Invited talks	31
11.1.5 Leadership within the scientific community	31
11.1.6 Scientific expertise	32
11.1.7 Research administration	32
11.2 Teaching - Supervision - Juries - Educational and pedagogical outreach	32
11.2.1 Teaching	32
11.2.2 Supervision	33
11.2.3 Juries	33
11.2.4 Educational and pedagogical outreach	33
11.3 Popularization	34
11.3.1 Specific official responsibilities in science outreach structures	34
11.3.2 Productions (articles, videos, podcasts, serious games, ...)	34
11.3.3 Participation in Live events	34
11.3.4 Others science outreach relevant activities	34
12 Scientific production	34
12.1 Major publications	34
12.2 Publications of the year	35
12.3 Cited publications	41

1 Team members, visitors, external collaborators

Research Scientists

- Florent Massegla [Team leader, INRIA, Senior Researcher, HDR]
- Reza Akbarinia [INRIA, Researcher, HDR]
- Christophe Botella [INRIA, ISFP]
- Benjamin Bourel [INRIA, Researcher]
- Hervé Goëau [CIRAD, Researcher]
- Alexis Joly [INRIA, Senior Researcher, HDR]
- Fabio Andre Machado Porto [INRIA, Senior Researcher, from Apr 2025 until Jun 2025]
- Christophe Pradal [CIRAD, Researcher, from Mar 2025]
- Maxime Ryckewaert [INRIA, Starting Research Position, until Mar 2025]
- Joseph Salmon [UNIV MONTPELLIER, Professor Detachement, HDR]
- Patrick Valduriez [INRIA, Emeritus, HDR]

Faculty Members

- Esther De Castro Pacitti [UNIV MONTPELLIER, Professor, from Sep 2025, HDR]
- François Munoz [UGA, Associate Professor, from Mar 2025]
- Maximilien Servajean [LIRMM, Associate Professor Delegation, from Feb 2025 until Aug 2025]

Post-Doctoral Fellows

- Aimie Berger Dauxere [INRAE]
- Jean-Baptiste Fermanian [UNIV MONTPELLIER, Post-Doctoral Fellow, until Sep 2025]
- Ilyass Moummad [INRIA, Post-Doctoral Fellow, from Mar 2025]
- Lukas Picek [INRIA, Post-Doctoral Fellow, until Oct 2025]
- Rebecca Pontes Salles [INRIA, Post-Doctoral Fellow, until Nov 2025]

PhD Students

- Raphael Benerradi [UNIV MONTPELLIER]
- Matteo Contini [IFREMER, until Oct 2025]
- Guillaume Coulaud [UNIV MONTPELLIER]
- Lo’Ai Gandeel [INRIA, from Jul 2025]
- Sebastien Gigot-Leandri [CNRS]
- Théo Larcher [UNIV MONTPELLIER]
- Cesar Leblanc [INRIA, until Oct 2025]
- Alex Maleknia [UNIV MONTPELLIER, from Nov 2025]
- Giulio Martellucci [INRAE, from Jun 2025]
- Kawtar Zaher [INA, CIFRE]

Technical Staff

- Antoine Affouard [INRIA, Engineer]
- Mathias Chouet [CIRAD]
- Hugo Gresse [INRIA, Engineer, until May 2025]
- Benoit Lange [INRIA, Engineer]
- Pierre Leroy [INRIA, Engineer]
- Thomas Paillot [INRIA, Engineer, from Sep 2025]
- Thomas Paillot [INRIA, until Aug 2025]
- Remi Palard [CIRAD, Engineer, from Nov 2025]
- Remi Palard [CIRAD, until Oct 2025]
- Lukas Picek [INRIA, Engineer, from Nov 2025]
- Rebecca Pontes Salles [INRIA, Engineer, from Dec 2025]
- Julien Thomazo [LIRMM, Engineer, from Jun 2025 until Sep 2025]
- Jozef Ba Tran [INRIA, Engineer, from Nov 2025]
- Axel Vaillant [INRIA, Engineer, until Feb 2025]

Interns and Apprentices

- Bronislav Abadie [UNIV MONTPELLIER, Intern, until Jun 2025]
- Marion Cann [INRIA, Intern, from Sep 2025 until Oct 2025]
- Marion Cann [INRIA, Intern, from Jun 2025 until Aug 2025]
- Raphael Lemarie [INRIA, Intern, from Jun 2025 until Jul 2025]
- Massilya Raked [INRIA, Intern, from Mar 2025 until Jul 2025]

Administrative Assistant

- Anouk Renaud [INRIA, from Dec 2025]

Visiting Scientist

- Diletta Santovito [UNIV BOLOGNE, until Apr 2025]

External Collaborators

- Fabio Andre Machado Porto [LNCC-PETROPOLIS, from Aug 2025]
- Fabio Andre Machado Porto [LNCC-PETROPOLIS, until Mar 2025]
- Jean Marc Sadaillan [INRAE]
- Jozef Ba Tran [UNIV MONTPELLIER, from Mar 2025 until Sep 2025]

2 Overall objectives

Environmental sciences combine various scientific disciplines to understand and address critical environmental issues such as climate change, pollution and biodiversity loss, and to develop sustainable solutions to preserve the planet's ecosystems and resources. Today, the increasing production of observation and experimentation data in environmental sciences requires advanced data science skills and tools to manage, analyze, and interpret large-scale and complex datasets and make sense of it. Data science focuses on extracting insights from data through pattern identification, outcome prediction, and process optimization. It is an interdisciplinary science that relies on well-established research fields such as machine learning, statistics, data mining, and data management, which need to work in synergy.

Iroko ¹ advocates an interdisciplinary scientific approach to address the challenges of environmental sciences by using and improving data science. This approach should have a high impact on both data science, by proposing new solutions and new systems, and on environmental sciences, by contributing to findings applied to real use cases in biodiversity, agriculture and one-health.

The team's research focuses on the intersection of data science and environmental sciences.

Data science is an interdisciplinary field that utilizes scientific methods, processes, algorithms, and systems to extract knowledge and insights from structured and unstructured data. It combines various disciplines, including statistics, computer science, and domain expertise, to tackle complex problems and make data-driven decisions. The ultimate goal of data science is to discover patterns and trends, predict future outcomes, and optimize processes through the analysis of vast amounts of data.

Deep learning, a kind of machine learning, plays a crucial role in data science. It employs artificial neural networks, specifically deep neural networks, which imitate the human brain's structure and functionality. Deep learning algorithms extract knowledge from data using multiple layers of abstraction, allowing them to identify patterns and generate highly accurate predictions. These methods have been effectively utilized in various applications such as image and speech recognition, natural language processing, and recommendation systems.

Data mining is another crucial aspect of data science. It is the process of discovering previously unknown, valid, and potentially useful patterns in large datasets. Data mining techniques include clustering, classification, association rule learning, and anomaly detection, among others. These methods enable data scientists to gain insights and identify trends, relationships, and dependencies within the data, which can be used to inform decision-making initiatives.

Time series analysis is a valuable method in data science, focusing on ordered, often time-stamped data points. Key aspects include comparing different time series using techniques like cross-correlation and dynamic time warping, and detecting anomalies with statistical tests or machine learning algorithms. Pattern recognition in time series analysis aims to find recurring motifs or sub-sequences, helping to discover underlying structures in the data. By identifying patterns and anomalies, data scientists can better understand system dynamics, predict future behavior, and make informed decisions across various domains.

Ultimately, models are central not only in data science but also in environmental sciences. While machine learning models are central in data science, mechanistic models (mathematical, physical and process-based models) allow us to capture at different scales the scientific knowledge of different disciplines (*e.g.*, soil, plant, atmosphere, disease) in order to simulate the behavior of complex systems and to predict their behavior under different scenarios.

Environmental sciences encompass a diverse range of disciplines that focus on understanding the complex relationships between humans and the natural world. By studying the Earth's ecosystems, climate, and resources, environmental scientists address critical issues such as climate change, pollution, habitat loss, and biodiversity conservation. This multidisciplinary field combines knowledge from areas such as biology, chemistry, geology, meteorology, physics, and agronomy to provide a comprehensive understanding of the environment and the challenges it faces. In addition to understanding the Earth's physical processes,

¹The iroko tree is a good emblem for this project, as it is a vital contributor to carbon sequestration, absorbing CO₂ from the atmosphere and helping to mitigate climate change. Its dense wood provides a substantial carbon sink, storing this critical greenhouse gas for decades. Additionally, it enhances biodiversity by offering habitat for various species and contributes to soil conservation.

environmental scientists also investigate the ecological and social dimensions of environmental problems, recognizing that human well-being is intricately linked to the health of ecosystems.

The primary objective of environmental sciences is to develop sustainable solutions to preserve and protect the planet's ecosystems and resources for present and future generations. This includes the conservation of biodiversity, which is essential for maintaining ecosystems stability, resilience, and the provision of valuable ecosystem services. Agronomy, the study of agricultural production and soil management, is another key component of environmental sciences. By optimizing agricultural practices and promoting sustainable land use, agronomists help ensure global food security while minimizing negative environmental impacts.

One Health is an emerging approach in environmental sciences that emphasizes the interconnectedness of human, animal, and environmental health. It recognizes that the health of people, animals, and ecosystems are interdependent, and that collaborative, interdisciplinary efforts are needed to address complex challenges such as zoonotic diseases, antimicrobial resistance, and climate change. Environmental scientists engaging in One Health research collaborate with public health experts, veterinarians, ecologists, and social scientists to develop integrated solutions that promote health and well-being across species and ecosystems.

To achieve their objectives, environmental scientists engage in research, data analysis, and policy development to inform decision-making processes. They collaborate with industries, governments, and communities to promote environmentally responsible practices and policies. This involves conducting environmental impact assessments, developing strategies for climate change adaptation and mitigation, and designing programs for habitat restoration and species conservation. Environmental sciences require interdisciplinary collaboration, critical thinking, and ethical commitment to the well-being of the environment and future generations.

Unsurprisingly, and as envisioned by the fourth paradigm of discovery [84], data production and analysis have become fundamental activities of environmental sciences. The generation of data has increased exponentially, from remote sensing satellites that monitor climate patterns and land use changes, to biodiversity databases that track species distribution and abundance. These vast datasets, provide unprecedented opportunities for understanding and responding to environmental changes. However, this influx of data also presents significant challenges. It requires advanced tools and methodologies to store, manage, and analyze data. Environmental scientists must therefore develop data literacy skills, and there is a growing need for specialists in environmental data science. This calls for combining at least computer science, statistics, and environmental sciences to derive meaningful insights from complex, large-scale datasets.

Objectives:

The objectives of the team are *improving data science and contributing to new findings in environmental sciences*.

We expect our impact to be measured by three main aspects:

- Academic recognition of our contributions. This aspect should be assessed as usual.
- The interdisciplinary extent of our results. This may involve results in ecology, biology, climatology, or any other science in which we collaborate with those scientists. The results obtained through this collaboration, which might not have otherwise been obtained, are significant to us. For example, it may be measuring a change in biodiversity in a region, selecting and improving plant varieties adapted to specific environmental conditions, or identifying climate anomalies in global measurement history.
- Our impact in the real world. We hope that our work will help humanity reduce its environmental footprint and eventually slow down the course of global warming. This could be done by, for example, preserving biodiversity in a particular area, replacing one type of crop with another, or avoiding the overuse of antibiotics in animal agriculture, just to name a few of the ways we are currently working on.

3 Research program

Iroko develops data science methods and systems to support *data-driven environmental sciences*. Our research program is organized around three tightly connected themes: (i) *Big Data and Scalability*, (ii) *Machine Learning with Humans in the Loop*, and (iii) *Multiscale & Multimodal Data Analytics*. Across

these themes, we pursue three cross-cutting objectives: (1) make analyses *reusable and reproducible* through well-engineered data/model/workflow services, (2) make learning *reliable and trustworthy* by explicitly handling bias and uncertainty, and (3) maximize impact through *open science* (software, models, and FAIR data whenever possible).

3.1 Big Data and Scalability

Unified data–model–workflow services. Environmental science pipelines increasingly combine heterogeneous data (e.g., images, omics, epidemiology, climate) with heterogeneous models (statistical, machine learning, mechanistic) and complex workflows. Yet current solutions remain largely domain-specific and ad hoc, making it difficult to connect artifacts, reproduce results, and reuse components across projects. Our goal is to provide *integrated data and model management* together with workflow services that can interoperate with established environments such as Galaxy [71] and OpenAlea [95], as well as distributed execution engines. Building on our LifeSWS initiative [72], we aim to treat all scientific artifacts (datasets, models, metadata, workflow components, intermediate results) as first-class citizens, searchable via catalogs and executable through standardized interfaces. Compared to existing platforms (the closest being CyVerse [83]), our focus is on tighter integration of model life-cycle management, provenance, and caching to support end-to-end scientific investigations.

Scalable time-series analytics at climate scale. Many environmental questions rely on large collections of time series and call for motif discovery, clustering, and anomaly detection. At scale, naive distributed adaptations can be inefficient due to communication and synchronization costs [100]. We will therefore design *distribution-aware* algorithms for time-series analytics, with a particular focus on anomaly detection in large climate datasets using Matrix Profile ideas [101], and on implementations that can leverage modern distributed infrastructures (e.g., Spark) without sacrificing usability for domain partners.

- **Mid-term:** a first operational version of LifeSWS enabling integrated artifact search and workflow execution across heterogeneous environments; a distributed Matrix-Profile-based anomaly detection prototype validated on large environmental time series.
- **Long-term:** a production-grade, scalable service stack (catalog, provenance, caching, scheduling) enabling reproducible analyses across data modalities and models; a toolbox of distributed time-series operators usable in multiple environmental and One Health contexts.

3.2 Machine Learning with Humans in the Loop

Cooperative learning in citizen science. Platforms such as PI@ntNet and iNaturalist continuously improve their identification models from community-produced observations and revisions. This *cooperative learning* loop is powerful but raises new issues: sparse and opportunistic revisions, strong imbalance across taxa/regions/users, and extreme scale (tens of millions of observations; tens of thousands of classes). Standard crowdsourcing inference tools (e.g., Bayesian aggregation) are not directly applicable at this scale [86]. We will develop end-to-end human-in-the-loop models that represent user behavior and its impact on training dynamics, building on our initial contributions [88] and leveraging modern approaches to detect label issues [94]. A key aim is to prevent negative feedback loops and to learn principled user weighting strategies that remain robust under sparsity and imbalance.

Bias-aware species distribution models from opportunistic data. To monitor biodiversity under rapid global change [85], species distribution models (SDMs) increasingly rely on citizen science data, whose scale is unmatched but whose biases are substantial [73, 79]. We will continue to develop statistically grounded bias-correction methods [78, 77] and extend them to modern AI-based SDMs [76] and to Bayesian dynamic SDMs [75]. We will also address the fundamental limitation of presence-only data by inferring absences from visit histories and multi-species information [97], and by modeling observer profiles (persistent users, learners, taxonomic preferences) to better disentangle detectability from ecological signals.

Uncertainty, trust, and interpretability. Reliable downstream decisions require explicit management of predictive uncertainty. We will build on our work on set-valued classification and abstention mechanisms [91, 82] and investigate uncertainty quantification and propagation for structured biodiversity predictions (assemblages, indicators, abundance maps). This includes generic tools such as conformal prediction [96, 81] and scalable Bayesian approximations [70]. Finally, we will strengthen user trust through transparency and interpretability mechanisms, extending prior work on user-facing uncertainty and interactive features in PI@ntNet [90].

- **Mid-term:** new scalable HITL (Human In The Loop) models evaluated on real PI@ntNet-style revision streams; open-source SDM training components with improved bias handling; first results on uncertainty propagation for biodiversity indicators.
- **Long-term:** integration of HITL and uncertainty-aware decision modules into PI@ntNet/GeoPI@ntNet-like services; robust bias-aware dynamic SDM pipelines for long-term monitoring and scenario analysis.

3.3 Multiscale & Multimodal Data Analytics

Multimodal foundation models for biodiversity and agro-ecology monitoring. High-resolution monitoring of biotic components remains challenging despite major advances in sensors and AI [74]. We will study learning strategies that combine smartphone observations, scientific imaging workflows, drones, remote sensing, and environmental covariates (e.g., bioclimatic layers [80]) into multimodal pipelines. Because fully end-to-end multimodal training can be costly, we will emphasize self-supervised and foundation-model approaches, then reuse learned representations for downstream ecological tasks. We will also prioritize interpretability, combining transparency principles with post-hoc explanations (e.g., Shapley-based methods [98]).

Multivariate time series and scalable similarity. Environmental and One Health applications increasingly involve *multivariate* time series, where variables interact across time and scales. We will develop parallel anomaly detection and similarity search methods (including kNN Matrix Profile variants) building on our prior distributed indexing and analytics experience [100, 92]. We will also investigate visualization and representation learning for complex multivariate temporal data to support interactive exploration by domain experts.

Biodiversity trajectories and community structure. Predicting biodiversity trajectories requires models that capture long-term dependencies and integrate heterogeneous historical evidence. We will investigate deep architectures (including transformer-based models) for integrative forecasting [89], and contrast them with interpretable dynamic models grounded in ecological mechanisms (e.g., Bayesian DSDMs [75]). At the community level, we will leverage graph-based approaches to analyze and predict species assemblages via co-occurrence networks, relating network structure to dispersal, filtering, and interactions [93], and validating on vegetation survey datasets [99].

- **Mid-term:** prototypes for multi-species monitoring from complex imagery and environmental covariates; multivariate time-series analytics components; first case studies on forecasting short-term biodiversity trends.
- **Long-term:** reusable multimodal foundation models shared openly; operational toolchains for multiscale forecasting and scenario analysis; new graph-based methods to predict and interpret species assemblages.

4 Application domains

The application domains covered by Iroko focus on the environment, with the specific needs of data-intensive scientific applications, *i.e.*, management and analytics of large amounts of (streaming) data. Since the interaction with scientists is critical to identify and tackle data management problems, we are dealing

primarily with application domains for which Montpellier has an excellent track record, *i.e.*, agronomy, botany and life sciences, with our scientific partners CIRAD, INRAE and IRD.

Let us briefly illustrate some representative examples of scientific applications on which we will work.

- **Monitoring and preservation of plant biodiversity.** In the continuity of Zenith, Iroko is the host team for the **PI@ntNet** citizen science platform. This initiative, piloted by a consortium of four research organisms (Inria, CIRAD, INRAE and IRD), began in 2011 and has become one of the largest citizen science platforms in the world. Its mobile front-end, allowing to identify and share plant observations, is used by more than 20 million users worldwide, of which 15% are professionals in the fields of land management, biodiversity management, education, agriculture, trade and tourism. PI@ntNet is one of the official publisher of the Global Biodiversity Information Facility (GBIF), the world's largest government-funded biodiversity data infrastructure. More than 13 million PI@ntNet observations have been published and have been used in **hundreds** of scientific publications on various themes ranging from conservation, to agro-ecology or to the impact of climate change.
- **Biological data integration and analysis.** Biology and its applications, from medicine to agronomy and ecology, are now producing massive data, which is revolutionizing the way life scientists work. For instance, using plant phenotyping platforms such as HIRROS and PhenoArch at INRAE Montpellier, quantitative genetic methods allow to identify genes involved in phenotypic variation in response to environmental conditions. These methods produce large amounts of data at different time intervals (minutes to months), at different sites and at different scales ranging from small tissue samples to the entire plant until whole plant population. Analyzing such big data creates new challenges for data management and data integration, but also for plant modeling. We will address this application in the context of the French initiative OpenAlea, with CIRAD and INRAE.
- **One Health approach to fight antimicrobial resistance (AMR).** Antimicrobial resistance (AMR) refers to the ability of microorganisms, such as bacteria to resist the effects of antimicrobial drugs that were previously effective in treating infections. It is a growing public health threat that can make infections more difficult and costly to treat, leading to longer hospital stays, and increased mortality rates. A promising approach for fighting AMR would be the One Health approach that recognizes that the health of humans, animals, and the environment are interconnected. However, our ongoing PROMISE project with experts from different health and environmental sectors has revealed that addressing AMR through the One Health approach is a complex and multifaceted issue, which poses significant challenges from the data science point of view, including the following: 1) Heterogeneous data collection and standardization; 2) Multivariate data analysis; 3) Predictive modeling; and 4) Data sharing and access. This application will eventually bring together 21 professional networks and 42 academic partners. Iroko will be central to interdisciplinarity at the interface with data analytics in this application through the PROMISE PPR project led by INSERM.

5 Social and environmental responsibility

5.1 Footprint of research activities

The footprint of IROKO's research activities mainly stems from (i) large-scale computation and storage (e.g., deep learning training on GPUs, large-scale analytics, and data management) and (ii) travel for collaboration and dissemination. In continuity with practices established in the predecessor team, we take several measures to mitigate this footprint.

We promote computing frugality by prioritizing model reuse (transfer learning, warm starts, reuse of pretrained models) and by improving experimental pipelines to avoid unnecessary retraining. When models are deployed, we explore compression and efficiency-oriented architectures to reduce memory and computational requirements.

For widely deployed services such as PI@ntNet, GeoPI@ntNet, and PROMISE, we adopt an eco-design approach by focusing on purposeful, non-addictive functionalities and by optimizing workflows to limit unnecessary computation, storage, and data transfers.

We also favor open and reusable software, models, and FAIR data (Findable, Accessible, Interoperable, and Reusable), which encourages reuse and reproducibility and reduces redundant data collection and

re-computation across projects. Finally, we limit long-distance travel when possible, favor train for domestic trips, and increasingly rely on hybrid or remote meetings.

5.2 Impact of research results

The team aims to produce data science results with direct impact on environmental sciences, One Health, and sustainable practices. In 2025, this impact is materialized through operational platforms and openly shared resources that are already reused beyond the team.

GeoPI@ntNet provides high-resolution (50 × 50 m) plant species distribution maps for more than 15,000 species across Europe, with freely downloadable outputs and biodiversity indicators supporting research, conservation planning, and territorial management.

PI@ntNet continues to act as a large-scale citizen observatory used by more than 20 million users worldwide, including a significant share of professionals. Its open data published on GBIF were used in 292 scientific publications in 2025.

In marine ecology, the Seatizen Atlas dataset (more than 1.6 million underwater and aerial images) supports large-scale training and reuse of AI models for cost-effective coral reef and habitat monitoring.

For agriculture and agroecology, the Deep-Plant-Disease dataset (about 250K images covering 55 crops and 175 diseases) provides a large and diverse benchmark to improve plant disease identification and generalization.

In public health, the PROMISE multi-cloud platform supports One Health surveillance and research on antimicrobial resistance by integrating aggregated data from human, animal, and environmental sectors.

As a longer-term strategy, the team also explores the transfer of its scalable data management and learning techniques to other domains. This is a key motivation behind our participation in initiatives such as the OMICFINDER challenge, which aims to unlock the potential of vast public genomic databases to enable new advances in medicine, ecology, and agriculture.

6 Highlights of the year

6.1 Awards

Prix science ouverte des données de la recherche - Seatizen Atlas: a collaborative dataset of underwater and aerial marine imagery [18]. First author: Matteo Contini (IROKO PhD student). Last author: Alexis Joly (PhD director).

6.2 Other key achievements

- Publication in the journal *Nature Plants*: Learning the syntax of plant assemblages [23]. First author: César Leblanc (IROKO PhD student). Last author: Alexis Joly (PhD director). Most cited *Nature Plants* article during several weeks. Highlighted by *Nature Plants* as “Crystal Ball Time” paper.
- **GeoPI@ntNet**: a new software of the PI@ntNet family dedicated to the high-resolution mapping of plant biodiversity has been released. It has been already used by more than 1K users per month.

7 Latest software developments, platforms, open data

7.1 New Features in the PI@ntNet Platform

Participants: Antoine Affouard, Hugo Gresse, Jean-Christophe Lombardo, Thomas Paillot, Joseph Salmon, Alexis Joly, Józef Tran.

PI@ntNet is a large-scale citizen observatory relying on AI technologies to support plant identification and biodiversity monitoring through mobile and web applications. In 2025, platform developments focused on strengthening PI@ntNet’s role as an operational biodiversity data infrastructure, with particular emphasis

on community-level identification, interoperability, and integration into decision-support workflows, notably within the GUARDEN European project.

A major development effort concerned the consolidation and deployment of community-level plant identification services, extending PI@ntNet beyond individual plant observations. In particular, the platform's workflow for vegetation survey and plot images was strengthened and operationalized, enabling the identification of multiple co-occurring plant species from complex imagery such as quadrats, drone acquisitions, and roadside surveys. These services were deployed and validated in several real-world GUARDEN case studies and integrated into downstream biodiversity monitoring and mapping pipelines.

Significant progress was also made on scalability and interoperability. PI@ntNet services were further integrated with external platforms and decision-support tools through improved APIs, facilitating their use within broader analytical chains combining citizen science data, remote sensing, and predictive modeling. In particular, PI@ntNet identification services were connected to GeoPI@ntNet and the GUARDEN Decision Support Applications, enabling the seamless flow from raw observations to high-resolution biodiversity indicators and maps served through standard web services (e.g. WMS).

In parallel, developments targeted data ingestion and management workflows. New mechanisms were implemented to support the batch import of large collections of plant observations, addressing the needs of institutions and organizations willing to contribute existing datasets to the platform. These features reduce barriers to data sharing and strengthen PI@ntNet's capacity to act as a hub for heterogeneous biodiversity observations.

7.2 New platforms

7.2.1 GeoPI@ntNet

Participants: Lukas Picek, César Leblanc, Benjamin Deneu, Rémi Palard, Thomas Paillet, Christophe Botella, Alexis Joly.

GeoPI@ntNet is a new, large-scale web application developed in the context of the PI@ntNet platform for the exploration, analysis, and dissemination of plant biodiversity information, offering an unprecedented combination of taxonomic coverage, spatial extent, and spatial resolution. The application provides high-resolution distribution maps (50×50 m) for more than 15,000 plant species across the entire European continent, making it one of the most comprehensive operational systems currently available for plant biodiversity mapping at this scale. GeoPI@ntNet relies on state-of-the-art deep learning-based species distribution models that integrate heterogeneous environmental data—such as satellite imagery, climatic variables, land-use information, and topography—with millions of in situ plant observations collected through the PI@ntNet platform. Beyond interactive visualization, the application allows users to explore regions of interest, compute biodiversity indicators (including protected, invasive, and endemic species), and access detailed, spatially explicit reports to support research, conservation planning, and territorial management. A key feature of GeoPI@ntNet is the open availability of its outputs: all species distribution maps are made freely downloadable, fostering transparency, reuse, and integration into external scientific studies, public policies, and operational workflows. By combining continental-scale coverage, fine spatial resolution, and open data dissemination within a single platform, GeoPI@ntNet represents a unique operational contribution to large-scale plant biodiversity monitoring and decision support. The application is already used by more than 1K users per month.

7.2.2 PROMISE

Participants: Reza Akbarinia, Benoit Lange, Florent Masseglia.

The objective of the **PROMISE** (PROfessional coMmunity network on antimicrobial reSistanceE) project [87] is to build a large data warehouse for managing and analyzing antimicrobial resistance (AMR) data. The PROMISE platform, of the same name, is a multi-cloud data management and analytics platform developed

in the context of the PROMISE project to support One Health surveillance and research on antimicrobial resistance. The platform integrates data from the human, animal and environmental sectors. At present, the data handled by the platform are aggregated (no personal data) and largely derived from public sources. PROMISE relies on a modular architecture organized into five independent "bubbles" (diffusion, query, storage, administration and processing) that can be deployed on any cloud. Services are containerized (Docker) and orchestrated with Kubernetes; inter-service communication is performed through REST APIs, and WebSockets are used for notifications.

The diffusion bubble provides both the web user interface and the API entry point, with a React-based viewer and a Quarkus (Java) gateway that routes requests to the relevant services. The administration bubble manages authentication and observability (monitoring and metrics collection). The query bubble normalizes user requests and aggregates results, while the storage bubble isolates raw data on the providers' infrastructures, translates normalized queries into database-specific queries, and returns aggregated outputs. Current storage connectors support PostgreSQL, InfluxDB and MongoDB. The processing bubble orchestrates analytics over aggregated time series and supports correlation modules implemented in Python and connected to an event bus. Finally, an HDS-oriented deployment option is being investigated to enable the use of more sensitive health data while preserving a strict separation between raw data and aggregated outputs.

Contact: Reza Akbarinia

7.3 Open data

Prix science ouverte des données de la recherche - Seatizen Atlas: a collaborative dataset of underwater and aerial marine imagery [18]. Seatizen Atlas is a citizen science dataset made of more than 1.6M underwater and aerial imagery collected in shallow tropical coastal areas by using various low cost platforms operated either by citizens or researchers. Data discovery and access rely on DOI assignment while data interoperability and reuse is ensured by complying with widely used community standards. The open-source data workflow is provided to ease contributions from anyone collecting pictures.

PI@ntNet GBIF data - A new release of PI@ntNet open data has been published on GBIF (the world's largest open data infrastructure for biodiversity). In 2025, this data has been used in **292 scientific publications**.

Deep-Plant-Disease Dataset - We aggregated and published the largest and most diverse dataset ever built for plant disease identification [38]. It comprises about 250K images across 55 crop species, 175 disease classes, and 333 unique crop-disease composition as well as novel text data designed to enhance model generalization in multi crop disease identification.

8 New results

8.1 Distributed Data and Model Management

8.1.1 A Logic-Based Approach for Knowledge Graph Data Integration

Participants: Fabio Porto, Patrick Valduriez.

In the context of the **Dinizia** Inria associated team with Brazil, we started a collaboration with the Boreal Inria team to study the combination of a knowledge graph with rule-based reasoning. In particular, we are interested in leveraging the InteGraal framework developed within the Boreal team, which enables semantic integration and reasoning over heterogeneous data sources. In this context, we proposed Gypscie-KG [55], an ML (Machine Learning) system that combines data integration, rule-based reasoning, and prediction services to provide semantic access to domain knowledge using a knowledge graph. In addition to providing integration of heterogeneous ML data within a knowledge graph, we explore the use of logic-based declarative techniques to enable reasoning and semantic querying over ML data.

8.1.2 Federated Learning

Participants: Patrick Valduriez.

Federated Learning (FL) is a promising distributed machine learning approach that enables collaborative training of a global model using multiple edge devices. The data distributed among the edge devices is highly heterogeneous. Thus, FL faces the challenge of data distribution and heterogeneity, where non-independent and identically distributed (non-IID) data across edge devices may result in a significant accuracy drop. Furthermore, the limited computation and communication capabilities of edge devices increase the likelihood of stragglers, thus leading to slow model convergence. To address this problem, we proposed the FedDHAD FL framework [26], which comes with two novel methods: Dynamic Heterogeneous model aggregation (FedDH) and Adaptive Dropout (FedAD). The combination of these two methods makes FedDHAD significantly outperform state-of-the-art solutions in terms of accuracy (up to 6.7% higher), efficiency (up to 2.02 times faster), and computation cost (up to 15.0% smaller).

8.1.3 Distributed Web Infrastructure for Integrated Pest Management

Participants: Christophe Pradal.

Crop protection and pest management are major economic and environmental concerns throughout Europe. The consultation of decision support systems (DSS) to guide decisions relating to Integrated Pest Management (IPM) is one of the key principles of IPM, reducing the ambiguity around potential risks to crop health. *Pests* in this context include invertebrate pests, weeds and pathogens.

In [63], to facilitate the use of these models, two Application Programming Interfaces (APIs) were designed to access catalog of DSS models and European online weather data sources. While these APIs are integrated into the IPM Decisions Platform (IPM Decisions Platform), they are also open source, allowing other crop protection and farm management software to inspect, download, modify, install, run, and use them.

The scientific platform OpenAlea provides a new service, the IPM Decision Factory, that enables DSS researchers and developers to advance, combine and create DSS interactively into its scientific workflow management system. These workflows are then automatically transformed into web services to be readily integrated into the IPM Decisions platform. This ensures that new DSS have access to required weather data and can be made readily accessible across Europe, for validation and use. OpenAlea.EpyMix [25] is a model describing canopy growth and epidemic dynamics on species mixture that has been integrated into the IPM Decision platform to understand how weather data, provided by the platform, and wheat-based crop mixtures are a promising strategy to improve disease management.

8.2 Data Analytics

8.2.1 Event Detection in Time Series

Participants: Esther Pacitti, Fabio Porto, Rebecca Salles.

Event detection in time series is a basic function in surveillance and monitoring systems and has been extensively explored over the years.

The new book [56] published by Springer and authored by Eduardo Ogasawara (CEFET-RJ, Brazil), Rebecca Salles (Iroko), Fabio Porto (LNCC, Brazil) and Esther Pacitti (Iroko), reflects our productive collaboration with Brazil in the context of the Dinizia associated team. It provides a general taxonomy for event detection according to the specific event types: anomaly detection, change-point, and motif discovery. It discusses state-of-the-art metric evaluations for event detection methods and on online event detection, including the challenges of incremental and adaptive learning.

Anomaly detection methods implicitly define detection criteria, such as deviation measures, filter thresholds, and candidate anomaly selection strategies. Choosing inappropriate criteria results in inaccurate outputs, generating spurious alerts or missing events. Adjusting these criteria is essential for monitoring systems. To address this challenge, we explored the fine-tuning of deviation measures, filter thresholds, and candidate selection strategies [52]. Experimental results show that the proper choice of criteria significantly improves anomaly detection performance, often with greater impact than changing the detection methods.

Concept drift detection (CDD) is the general problem of identifying significant changes in streaming data distribution over time. Current CDD methods face challenges in large-scale, multivariate datasets, where single drift detectors (DD) often fail to capture variable interdependencies. While ensemble drift detectors (EDD) are usually adopted to mitigate the adoption of a single DD, EDD may suffer when detections do not converge. This misalignment can cause voting mechanisms to neglect critical intervals with high detection rates. To address this issue, we proposed a fuzzy ensemble drift detector (FEDD) [44] that integrates unsupervised threshold voting with fuzzy logic to provide time tolerance and reconcile minor temporal misalignments in drift detection. Our evaluation shows that FEDD outperforms existing approaches by improving detection robustness and coverage.

8.2.2 Scalable Multivariate Anomaly Detection

Participants: Reza Akbarinia, Benoit Lange, Florent Masegla, Esther Pacitti, Rebecca Salles.

The continuous monitoring of dynamic processes generates vast amounts of streaming multivariate time series data. Detecting anomalies within them is crucial for real-time identification of significant events, such as environmental phenomena, security breaches, or system failures, which can critically impact sensitive applications. Despite significant advances in univariate time series anomaly detection, scalable and efficient solutions for online detection in multivariate streams remain underexplored. This challenge becomes increasingly prominent with the growing volume and complexity of multivariate time series data in streaming scenarios.

In [33], we provide the first structured survey primarily focused on scalable and online anomaly detection techniques for multivariate time series, offering a comprehensive taxonomy. Additionally, we introduce the Online Distributed Outlier Detection (2OD) methodology, a novel well-defined and repeatable process designed to benchmark the online and distributed execution of anomaly detection methods. Experimental results with both synthetic and real-world datasets, covering up to hundreds of millions of observations, demonstrate that a distributed approach can enable centralized algorithms to achieve significant computational efficiency gains, averaging tens and reaching up to hundreds in speedup, without compromising detection accuracy.

8.2.3 Detecting Anomalies with Any Duration in Climate Time Series

Participants: Reza Akbarinia, Guillaume Coulaud, Florent Masegla.

Detecting abnormal climate events across temporal and spatial scales is crucial to the understanding of local and regional climate trends. Existing methods often depend on prior knowledge about the timing, location, or duration of such events, limiting their versatility. In [15], we propose ClimBurst, an approach to detect climate bursts (unusually high or low values of climate variables) without prior assumptions about their temporal duration. ClimBurst offers the ability to: (a) identify climate bursts of any duration within the time series of single locations, (b) link climate bursts across neighboring locations, and (c) analyze the spatio-temporal propagation of these anomalies. Applying ClimBurst to sea surface temperature data from the Mediterranean Sea (1960–2021) shows some detected hot bursts and anomalies coincide in time with known severe marine heatwaves. ClimBurst also shows how detected hot (cold) bursts are spatio-temporally connected and these connected bursts have increased (decreased) in duration, intensity, spatial extent and frequency historically.

In [40], we propose a demonstration of ClimBurst allowing users to interact directly with our system to see both a summary showing the presence/absence of bursts over a user-specified year and spatial range. The demonstration will also allow users to perform time-travel queries to see how bursts propagate over space and time.

8.2.4 Energy Efficient Time Series Anomaly Detection

Participants: Reza Akbarinia, Benoit Lange, Florent Masegla, Esther Pacitti, Rebecca Salles.

Traditionally, choosing an anomaly detection method for a given application is mainly driven by detection accuracy and runtime. However, with the rapid evolution of hardware and connected devices, massive amounts of time series data are produced, and the real-time analysis of such time series brings new demands not only for accurate and scalable solutions, but also for energy consumption management. In this scenario, any improvement in energy efficiency can have a considerable impact on both the environmental footprint and the monetary expenses. In [53], we address the problem of benchmarking time series anomaly detection methods based on the trade-off between accuracy, runtime, and energy consumption. We introduce a new metric for evaluating relative energy efficiency performance, called `saveUp`, and provide a novel methodology, inspired by skyline queries, for benchmarking methods based on a more comprehensive set of metrics, including peak power usage and total energy consumption. Experimental results based on large datasets show that our methodology is useful for selecting the methods that provide the best performance with the lowest energy impacts. Moreover, results indicate that speedup and `saveUp` are not always directly correlated as believed a priori, and sometimes it is best to "take it slow" in favor of green applications.

8.2.5 Extending Matrix Profile for Seasonal Anomaly Detection

Participants: Reza Akbarinia, Guillaume Coulaud, Florent Masegla.

Seasonal time series analysis is fundamental in domains such as climate science, where detecting and understanding anomalies, patterns, and data changes are essential. The classical Matrix Profile approach does not consider the data's seasonality, failing to detect seasonal anomalies and patterns. In [60], we propose the Interval Matrix Profile (IMP), a novel extension of the Matrix Profile specifically designed for analyzing periodic and seasonal time series data. The Interval Matrix Profile enables flexible interval-based comparisons across seasons, allowing the detection of anomalies that conventional approaches miss. We further propose the constrained k Nearest Neighbor Interval Matrix Profile, designed to identify anomalies that may appear across multiple periods, a common characteristic of abnormal climate events and extreme weather phenomena. Our approach leverages a scalable block-based algorithm that achieves significant performance gains through caching, vectorization, and parallelism. Additionally, we introduce a novel methodology to detect the first or last occurrence of a pattern, enabling the discovery of pattern emergence or disappearance within seasonal time series. The algorithms are demonstrated in case studies on temperature climate time series. They effectively capture seasonal anomalies and find pattern disappearance. Our results illustrate that the IMP consistently outperforms the classical Matrix Profile both in the accuracy of seasonal anomaly detection and in computational efficiency.

8.3 Machine Learning for Biodiversity and Agroecology

8.3.1 Learning Ecological Structure with Large Language Models

Participants: César Leblanc, Hervé Goëau, Maximilien Servajean, Alexis Joly, Diego Marcos, Pierre Bonnet.

This research axis explores how large language models (LLMs) can be adapted to capture and exploit structured ecological knowledge, with a focus on plant communities and functional traits. By transferring ideas from natural language processing to ecology, these works investigate how latent structure in species assemblages and unstructured textual resources can be leveraged to improve biodiversity understanding and modeling.

In [23], the team introduces an approach inspired by language modeling to learn the “syntax” of plant assemblages, treating abundance-ordered species lists as ecological sequences. Trained on more than 10,000 European plant species, the model captures latent associations shaped by environmental constraints, dispersal processes, and species interactions. The learned representations can be fine-tuned for multiple downstream tasks, including predicting missing species in assemblages and classifying habitat types, where the method consistently outperforms co-occurrence-based models, expert systems, and standard neural networks. This work demonstrates how sequence-based modeling provides a powerful and flexible framework for representing plant community structure.

Complementing this community-level perspective, [27] focuses on species-level functional information and addresses the challenge of assembling large trait databases. Leveraging the information extraction capabilities of large language models, this work proposes a fully automatic pipeline to extract plant morphological traits from unstructured online textual descriptions. The approach successfully reconstructs expert-curated species–trait matrices with high accuracy, showing that LLMs can transform heterogeneous textual resources into structured ecological knowledge at scale, albeit with current limitations linked to data availability.

Together, these contributions illustrate the potential of large language models to bridge different levels of ecological organization, from individual traits to species assemblages, and to open new avenues for scalable, data-driven biodiversity modeling, mapping, and conservation science.

8.3.2 Scalable Plant Vision Models for Operational Monitoring

Participants: Hervé Goëau, Vincent Espitalier, Alexis Joly, Pierre Bonnet.

This research axis investigates how large-scale plant vision models can be designed and adapted for operational monitoring tasks, with a strong emphasis on scalability, robustness, and reduced annotation requirements in real-world conditions.

In [20], the team addresses the early detection of invasive alien plant species along roadsides, a major vector for biological invasions. Rather than relying on object detection or segmentation pipelines that require extensive manual annotation, this work evaluates the reuse of a global plant identification model trained on citizen science data. Using a vision transformer from the PI@ntNet platform, the study compares multi-label classification and tiling-based strategies applied to high-resolution roadside imagery. The results show that the tiling approach achieves strong detection performance even without task-specific fine-tuning, demonstrating the potential of large pretrained models for large-scale invasive species monitoring at low cost.

From a methodological perspective, [16] contributes to this axis by proposing PlantAIM, a hybrid vision architecture that combines global attention mechanisms with local feature extraction. By fusing transformer-based and convolutional representations, the model improves robustness and generalization in challenging plant visual recognition settings, including limited training data and heterogeneous environments. These architectural insights directly support the development of scalable plant vision systems capable of reliable deployment in operational monitoring scenarios.

8.3.3 Conformal Prediction for uncertainty quantification

Participants: Joseph Salmon, Jean-Baptiste Fermanian.

Deep neural networks in computer vision produce overconfident predictions without statistical guarantees, making uncertainty calibration essential. Conformal prediction provides distribution-free guarantees but

struggles in the long-tailed, highly unbalanced settings typical of large citizen science platforms, where many classes are rare. Recent work highlights both theoretical limitations and possible adaptations, including transductive and grouped conformal approaches [42, 61].

Handling ambiguity further requires integrating domain knowledge and leveraging multiple observations of the same instance to better separate aleatoric from epistemic uncertainty. Recent conformal approaches extend classification to multi-input settings by aggregating conformal p-values across observations, reducing prediction set size while preserving class-conditional coverage. Such aggregation frameworks are particularly well suited to citizen science applications, where multiple images per instance are available, and naturally support refined decision rules and rejection for uncertain predictions [41].

8.3.4 AI-Based Species Distribution Modeling and Mapping

Participants: Christophe Botella, Alexis Joly, Théo Larcher, César Leblanc, Diego Marcos, François Munoz, Rémi Palard, Lukáš Pícek, Maximilien Servajean, Dennis Shasha, Benjamin Bourel.

This research axis focuses on advancing species distribution modeling (SDMs) through deep learning and multimodal data integration, with the goal of overcoming key limitations of classical approaches, including limited training data, the absence of biotic interactions, and insufficient spatial resolution for biodiversity mapping.

A first line of work investigates how deep learning can extend SDMs beyond presence-only prediction. In [14], the team demonstrates that convolutional neural network–based SDMs can effectively model species abundance by exploiting transfer learning from large presence-only datasets. This strategy significantly improves abundance predictions, particularly for rare species and locally rare occurrences, and leads to clear performance gains over classical SDMs.

A complementary direction explores the explicit integration of biotic structure into SDMs. In [35], a cascading prediction framework is proposed in which common and dominant plant species are first predicted from environmental variables, and these predictions are then used to inform the distribution of less common species. By leveraging species co-occurrence patterns and competitive hierarchies, this approach improves prediction accuracy at fine spatial resolutions, especially in species-rich environments.

In parallel, [46] presents a large-scale, multimodal deep-SDM pipeline for very-high-resolution biodiversity mapping across Europe. Based on the integration of remote sensing data, climate time series, and species occurrence records at 50×50 m resolution, this work produces continental-scale species distribution maps, biodiversity indicators, and habitat maps. The approach enables joint modeling of interspecies dependencies and large-scale inference from heterogeneous data sources, supporting operational biodiversity monitoring at unprecedented spatial detail.

8.3.5 Coral Reef Monitoring

Participants: Matteo Contini, Sylvain Bonhommeau, Victor Illien, Sylvain Poulain, Serge Bernard, Julien Barde, Alexis Joly.

This research axis, conducted in close collaboration with Ifremer and IRD, develops scalable AI-based methods for coral reef monitoring by combining citizen-driven data collection, multi-scale imaging, and deep learning. The overarching objective is to enable accurate, fine-grained ecological assessment over large reef areas while relying on low-cost and operational data acquisition.

The Seatizen Atlas [18] provides the data backbone of this effort, bringing together more than 1.6 million underwater and aerial images collected in shallow tropical environments by citizens and researchers using diverse platforms. The dataset captures the strong variability inherent to real-world marine imagery and is distributed through an open, standards-compliant workflow, enabling large-scale training and reuse of AI models for marine biodiversity mapping.

Building on this resource, the collaboration explores how fine-scale ecological information extracted from underwater imagery can be transferred to broader spatial scales. In [17], a multi-scale learning

framework propagates detailed coral and habitat classifications from underwater images to drone-based aerial imagery through knowledge distillation. This approach is further extended in [39], which introduces a weakly supervised semantic segmentation method that combines underwater-derived supervision, spatial interpolation, and self-distillation to minimize annotation effort. Together, these contributions demonstrate how multi-scale deep learning and weak supervision can support cost-effective, high-resolution coral reef monitoring at scale.

8.3.6 Evaluation of Species Identification and Prediction Algorithms

Participants: Alexis Joly, Lukáš Pícek, Hervé Goëau, Christophe Botella, Diego Marcos, César Leblanc, Théo Larcher.

This research axis focuses on the large-scale, rigorous evaluation of species identification and prediction algorithms, with the objective of characterizing state-of-the-art performance under realistic conditions and identifying key methodological challenges for biodiversity-oriented AI systems.

A central activity in this area is the organization of the LifeCLEF evaluation campaign [45, 57], which continues to attract hundreds of research teams and data scientists worldwide. The 2025 edition featured five complementary, data-driven tasks covering a wide range of ecological modalities and problem settings: AnimalCLEF for open-set individual animal re-identification, BirdCLEF+ for species recognition in complex acoustic soundscapes, FungiCLEF for few-shot classification of rare species, GeoLifeCLEF [51] for plant species distribution prediction from multimodal environmental data, and PlantCLEF [49] for identifying multiple co-occurring plant species in vegetation-plot imagery. Together, these benchmarks provide a unique and controlled view of current capabilities and limitations in species-level AI.

A key insight emerging across tasks is the persistent impact of domain shift, particularly when training and test data differ in geography, sensing modality, or species composition. While baseline models offered strong starting points, the most effective solutions relied on large-scale pretraining, self-supervised and semi-supervised learning, and multimodal data fusion. The results of BirdCLEF+ highlighted the potential of unlabeled audio data through contrastive learning, whereas GeoLifeCLEF exposed the difficulty of generalizing even with high-resolution, multimodal inputs. Similarly, FungiCLEF and PlantCLEF confirmed that few-shot and weakly supervised scenarios remain challenging, despite progress enabled by vision transformers, prototype-based methods, and metadata-aware pipelines. Overall, multimodality consistently emerged as a key driver of robustness and performance, alongside growing interest in efficient and deployable architectures.

Complementing these benchmarking efforts, [29] explores species identification in a heritage biodiversity context, focusing on herbarium specimens. This study compares hyperspectral leaf reflectance measurements with RGB image-based identification using Pl@ntNet, showing that spectral approaches can achieve high species-level accuracy from relatively small datasets, even in the absence of reproductive structures. The results highlight the complementarity of spectral and vision-based methods and point to practical solutions for reducing taxonomic knowledge gaps in large digitized collections.

8.3.7 Importance of fossil pollen data for vegetation species distribution modeling

Participants: Benjamin Bourel, Christophe Botella.

Given the current acceleration of climate change, anticipating future responses in plant biodiversity is a major scientific and societal challenge. We propose a resolutely innovative approach to improve the predictability of European vegetation dynamics, based on a long-term perspective covering the last 20,000 years. By combining, for the first time on a European scale, more than 72,000 harmonized fossil pollen records, high-resolution paleoclimate simulations and indicators of anthropogenic pressure, we aim to unravel the respective roles of climate and human activities in past ecosystem transformations. We are tackling a major conceptual barrier in ecology and paleoecology: the validity of the principle of actualism and the underestimation of plant species' climatic niches due to niche truncation.

The preliminary results obtained in 2025 during a Master's internship carried out by Marion Cann and supervised by Benjamin Bourel and Christophe Botella, researchers who will supervise this postdoctoral project, are very encouraging. The coupling of LegacyPollen 1.0 data with the paleoclimate simulations highlighted that the climatic hypervolume occupied by *Olea* in Europe increased by 25% when taking into account past data (data in Europe since the Last Glacial Maximum), in addition to present data. The integration of fossil data therefore makes it possible to identify plant communities with no modern analogues and to reconstruct more realistic fundamental niches for key taxa in European ecosystems. These methodological advances pave the way for more robust species distribution models, capable of improving biodiversity projections in the face of future climate change.

9 Bilateral contracts and grants with industry

Participants: Antoine Affouard, Jean-Christophe Lombardo, Hugo Gresse, Alexis Joly.

- CIFRE contract with INA (Institut National de l'Audiovisuel): PhD of Kawtar Zaher.
- **PI@ntNet API for developers:** 32 companies have signed up for paid use of the service (110K euros in revenue in 2025).

10 Partnerships and cooperations

10.1 International initiatives

10.1.1 Associate Teams in the framework of an Inria International Lab or in the framework of an Inria International Program

Dinizia

Title: Data Science for the Natural Environment

Duration: 2025-2027

Coordinator: Esther Pacitti (Iroko) and Eduardo Ogasawara (CEFET-RJ, Rio de Janeiro, Brazil)

Partners:

- CEFET-RJ, Rio de Janeiro, RJ
- Fiocruz, Rio de Janeiro, RJ
- LNCC, Petropolis, RJ
- UFF, Rio de Janeiro, RJ
- UFRJ, Rio de Janeiro, RJ

Inria contact: Esther Pacitti

Summary: The overall objective of Dinizia is to develop new data science solutions that will eventually contribute to findings in environmental and related sciences. These solutions will be in terms of methods and real systems. Our technical objective within data science is to help managing complex dataflows by organizing massive and heterogeneous data, in connection with models and making related artifacts (datasets, time series, models, metadata, dataflow components, etc.) easy to search, debug, and parallelize. A technical goal of this project is to make dataflows work as seamlessly with data as queries do in business processing. The work program includes three major research topics: detecting events in large time series, model life-cycle management, and scalable execution of heterogeneous dataflows. To validate our solutions, we capitalize on our previous experience in

developing major systems for scientific applications: PI@ntNet and OpenAlea from Inria; Savime and Harbinger from Brazil. With our main application partners (Cirad and INRAE in France, Fiocruz and Centro de Operações Rio in Brazil), we will validate our results using real datasets and models. The main applications are in agronomy, biodiversity informatics and meteorology.

10.1.2 Participation in other International Programs

IVADO-Inria Program: IROKO & University of Montreal have been selected as one of the 8 projects funded within the 2025 edition of the **IVADO-Inria Program**. Alexis Joly visited the IRBV lab in Montreal two weeks in October and Etienne Laliberté visited IROKO in Montpellier two weeks in November. These exchanges have helped consolidate and structure the scientific collaboration already underway around the interface between artificial intelligence and plant ecology, particularly with regard to the challenges of rapid monitoring of plant biodiversity using drones. It has strengthened exchanges between teams at the University of Montreal (Department of Biological Sciences, IRBV, Mila) and the research teams leading the PI@ntNet platform (Inria IROKO, UMR AMAP).

10.2 International research visitors

10.2.1 Visits of international scientists

Inria International Chair

Participants: Reza Akbarinia, Alexis Joly, Patrick Valduriez.

Fabio Porto, Laboratório Nacional de Computação Científica (LNCC, Brasil), holds an Inria International Chair for a cumulative duration of 12 months, spread over the period from January 2024 to December 2028.

Other international visits to the team

Dennis Shasha

Status Researcher

Institution of origin: University of New-York

Country: USA

Dates: April 7 - June 7

Context of the visit: DeepPEP contract

Mobility program/type of mobility: research stay, lecture

Tiffany Ding

Status PhD Student

Institution of origin: University of Berkeley (California)

Country: USA

Dates: March 1 - June 30

Context of the visit: Chaire ANR CAMELOT

Mobility program/type of mobility: research stay

10.3 European initiatives

10.3.1 Horizon Europe

B3 [B3 project on cordis.europa.eu](https://cordis.europa.eu)

Title: Biodiversity Building Blocks for policy

Duration: From March 1, 2023 to August 31, 2026

Partners:

- INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET AUTOMATIQUE (INRIA), France
- UNIVERSITATEA OVIDIUS DIN CONSTANTA (OVIDIUS UNIVERSITY OF CONSTANTA), Romania
- MARTIN-LUTHER-UNIVERSITAT HALLE-WITTENBERG (MLU), Germany
- Global Biodiversity Information Facility (GBIF), Denmark
- EIGEN VERMOGEN VAN HET INSTITUUT VOOR NATUUR- EN BOSONDERZOEK (EV INBO), Belgium
- LA TROBE UNIVERSITY (LTU), Australia
- JUSTUS-LIEBIG-UNIVERSITAET GIESSEN (JLU), Germany
- UNIVERSIDADE DE AVEIRO (UAveiro), Portugal
- SOUTH AFRICAN NATIONAL BIODIVERSITY INSTITUTE (SANBI), South Africa
- AGENTSCHAP PLANTENTUIN MEISE (AGENCE JARDIN BOTANIQUE DE MEISE), Belgium
- ALMA MATER STUDIORUM - UNIVERSITA DI BOLOGNA (UNIBO), Italy
- PENSOFT PUBLISHERS (PENSOFT), Bulgaria
- STELLENBOSCH UNIVERSITY (SU UNIVERSITY OF STELLENBOSCH), South Africa

Inria contact: Alexis Joly

Summary: The world is changing rapidly; climate change, land use change, pollution and natural resource exploitation are creating a global crisis for biodiversity whose magnitude and dynamics are hard to quantify. Decision makers at all levels need up-to-date information from which to evaluate policy options. For this reason rapid, reliable, repeatable monitoring of biodiversity data is needed at all scales from local to global. Only by leveraging large volumes of data, advanced modeling techniques and powerful computing tools can we hope to synthesize these data within timescales that are relevant to policy.

Data on biodiversity come from a diverse range of sources, citizen scientists, museums, herbaria and researchers are all major contributors, but increasingly new technologies are being deployed, such as automatic sensors, camera traps, eDNA and satellite tracking. Integrating these data is a major challenge, but is necessary if we are to create dependable information on biodiversity change. B3 will use the concept of data cubes to simplify and standardize access to biodiversity data using the Essential Biodiversity Variables framework. These cubes will be used, in conjunction with other environmental data and scenarios, as the basis for models and indicators of past, current and future biodiversity.

The overarching goal of the project is to provide easy access to tools in a cloud computing environment, in real-time and on-demand, with state-of-the-art prediction models of biodiversity, that will output models and indicators of biodiversity status and change. The project envisages a future where primary biodiversity data are seamlessly integrated into monitoring and forecasting such that policy and management can proactively respond to problems while at the same time reduce the costs of monitoring and management, and the negative impacts of biodiversity change.

GUARDEN [GUARDEN project on cordis.europa.eu](https://cordis.europa.eu/project/GUARDEN)**Title:** safeGUARDing biodivErsity aNd critical ecosystem services across sectors and scales**Duration:** From November 1, 2022 to October 31, 2025**Partners:**

- INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET AUTOMATIQUE (INRIA), France
- PARC NATIONAL DE PORT-CROS (CONSERVATOIRE BOTANIQUE NATIONAL MEDITERRANEEN DE PORQUEROLLES), France
- STICHTING NATURALIS BIODIVERSITY CENTER (NATURALIS), Netherlands
- YPOURGEIO GEORGIAS, AGROTIKIS ANAPTYXIS KAI PERIVALLONTOS (MINISTRY OF AGRICULTURE, RURAL DEVELOPMENT AND ENVIRONMENT OF CYPRUS), Cyprus
- DREVEN SRL, Belgium
- PLYMOUTH MARINE LABORATORY LIMITED (PML), United Kingdom
- UNIVERSITY OF ANTANANARIVO, Madagascar
- CHAROKOPEIO PANEPISTIMIO (HAROKOPIO UNIVERSITY OF ATHENS (HUA)), Greece
- INSTITUT METROPOLI (BARCELONA INSTITUTE OF REGIONAL AND METROPOLITAN STUDIES), Spain
- AGENCIA ESTATAL CONSEJO SUPERIOR DE INVESTIGACIONES CIENTIFICAS (CSIC), Spain
- DRAXIS ENVIRONMENTAL SA (DRAXIS), Greece
- EBOS TECHNOLOGIES LIMITED (eBOS), Cyprus
- CENTRE DE COOPERATION INTERNATIONALE EN RECHERCHE AGRONOMIQUE POUR LE DEVELOPPEMENT - C.I.R.A.D. EPIC (CIRAD), France
- AGENTSCHAP PLANTENTUIN MEISE (AGENCE JARDIN BOTANIQUE DE MEISE), Belgium
- ENVECO ANONYMI ETAIRIA PROSTASIAS KAI DIAHIRISIS PERIVALLONTOS A.E. (ENVECO S.A. ENVIRONMENTAL PROTECTION AND MANAGEMENT), Greece
- AREA METROPOLITANA DE BARCELONA (AMB), Spain
- FREDERICK UNIVERSITY FU (FREDERICK UNIVERSITY FU), Cyprus
- EREVNITIKO PANEPISTIMIAKO INSTITOUTO SYSTIMATON EPIKOINONION KAI YPOLOGISTON (RESEARCH UNIVERSITY INSTITUTE OF COMMUNICATION AND COMPUTER SYSTEMS), Greece

Inria contact: Alexis Joly

Summary: GUARDEN's main mission is to safeguard biodiversity and its contributions to people by bringing them at the forefront of policy and decision-making. This will be achieved through the development of user-oriented Decision Support Applications (DSAs), and leveraging on Multi-Stakeholder Partnerships (MSPs). They will take into account policy and management objectives and priorities across sectors and scales, build consensus to tackle data gaps, analytical uncertainties or conflicting objectives, and assess options to implement adaptive transformative change. To do so, GUARDEN will make use of a suite of methods and tools using Deep Learning, Earth Observation, and hybrid modeling to augment the amount of standardized and geo-localized biodiversity data, build-up a new generation of predictive models of biodiversity and ecosystem status indicators under multiple pressures (human and climate), and propose a set of complementary ecological indicators likely to be incorporated into local management and policy. The GUARDEN approach will be applied at sectoral case studies involving end users and stakeholders through Multi-Stakeholder Partnerships, and addressing critical cross-sectoral challenges (at the nexus of biodiversity and deployment of energy/transport infrastructure,

agriculture, and coastal urban development). Thus, the GUARDEN DSAs shall help stakeholders engaged in the challenge to improve their holistic understanding of ecosystem functioning, biodiversity loss and its drivers and explore the potential ecological and societal impacts of alternative decisions. Upon the acquisition of this new knowledge and evidence, the DSAs will help end-users not only navigate but also (re-)shape the policy landscape to make informed all-encompassing decisions through cross-sectoral integration.

MAMBO [MAMBO project on cordis.europa.eu](https://cordis.europa.eu/project/MAMBO)

Title: Modern Approaches to the Monitoring of Biodiversity

Duration: From September 1, 2022 to August 31, 2026

Partners:

- INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET AUTOMATIQUE (INRIA), France
- AARHUS UNIVERSITET (AU), Denmark
- STICHTING NATURALIS BIODIVERSITY CENTER (NATURALIS), Netherlands
- THE UNIVERSITY OF READING, United Kingdom
- HELMHOLTZ-ZENTRUM FÜR UMWELTFORSCHUNG GMBH - UFZ, Germany
- ECOSTACK INNOVATIONS LIMITED, Malta
- UK CENTRE FOR ECOLOGY AND HYDROLOGY, United Kingdom
- CENTRE DE COOPERATION INTERNATIONALE EN RECHERCHE AGRONOMIQUE POUR LE DEVELOPPEMENT - C.I.R.A.D. EPIC (CIRAD), France
- PENSOFT PUBLISHERS (PENSOFT), Bulgaria
- UNIVERSITEIT VAN AMSTERDAM (UvA), Netherlands

Inria contact: Alexis Joly

Summary: EU policies, such as the EU biodiversity strategy 2030 and the Birds and Habitats Directives, demand unbiased, integrated and regularly updated biodiversity and ecosystem service data. However, efforts to monitor wildlife and other species groups are spatially and temporally fragmented, taxonomically biased, and lack integration in Europe. To bridge this gap, the MAMBO project will develop, test and implement enabling tools for monitoring conservation status and ecological requirements of species and habitats for which knowledge gaps still exist. MAMBO brings together the technical expertise of computer science, remote sensing, social science expertise on human-technology interactions, environmental economy, and citizen science, with the biological expertise on species, ecology, and conservation biology. MAMBO is built around stakeholder engagement and knowledge exchange (WP1) and the integration of new technology with existing research infrastructures (WP2). MAMBO will develop, test, and demonstrate new tools for monitoring species (WP3) and habitats (WP4) in a co-design process to create novel standards for species and habitat monitoring across the EU and beyond. MAMBO will work with stakeholders to identify user and policy needs for biodiversity monitoring and investigate the requirements for setting up a virtual lab to automate workflow deployment and efficient computing of the vast data streams (from on-the-ground sensors, and remote sensing) required to improve monitoring activities across Europe (WP4). Together with stakeholders, MAMBO will assess these new tools at demonstration sites distributed across Europe (WP5) to identify bottlenecks, analyze the cost-effectiveness of different tools, integrate data streams and upscale results (WP6). This will feed into the co-design of future, improved and more cost-effective monitoring schemes for species and habitats using novel technologies (WP7), and thus lead to a better management of protected sites and species.

JAMRAI 2

Participants: Reza Akbarinia, Benoit Lange, Florent Masseglia.

Title: Joint Action Antimicrobial Resistance and Healthcare Associated Infections 2

Duration: 2024 to 2027

Partners:

- Institut National de la Santé et de la Recherche Médicale (INSERM), France
- Agence Nationale de Sécurité du Médicament et des Produits de Santé (ANSM), France
- Agence Nationale de Sécurité Sanitaire (Anses), France
- Centre hospitalier universitaire de Nantes (CHUN), France,
- Service public fédéral Santé publique, Sécurité de la Chaîne alimentaire, Belgium
- GESUNDHEIT ÖSTERREICH GMBH, Austria
- Ministry of Health of the Republic of Cyprus, Cyprus
- STATNI ZDRAVOTNI USTAV, Czech Republic
- STATENS SERUM INSTITUT, Denmark
- DANMARKS TEKNISKE UNIVERSITET, Denmark
- And more than 100 more institutions from different European countries.

Inria contact: Reza Akbarinia

Summary: EU-JAMRAI2, in which more than 120 institutions from different European countries participate, is a European collaborative initiative aimed at analyzing and understanding antimicrobial resistance and the infectious diseases associated with it. One of the objectives of the project is to gain a deeper understanding of the mechanisms underlying antimicrobial resistance and its transmission across populations. In this project, we plan to analyze data from humans, animals, and the environment sectors, to support public health policymakers in making informed decisions. Because each country has its own health management system, the initial focus of the project is to evaluate and identify key features that can be applied across European countries. We plan to harmonize analyses among participating nations. To facilitate data analytics, a selection of standardized metrics from diverse domains is essential. The project also aims to consolidate data from multiple countries into a single platform, enabling researchers from different fields to perform integrated analyses.

10.4 National initiatives

PARAD (PARSADA), (2025-2030), 7.7 MEuros.

Participants: Benjamin Bourel, Alexis Joly, Thomas Paillot.

The cross-disciplinary PARAD project aims to anticipate, innovate and support the agroecological transition in weed management by overcoming the obstacles created by the reduction in herbicides and the withdrawal of molecules through (i) a better understanding of the biological characteristics (traits) of weeds that are identified as being responsible for the failure of practices or the conditions that allow species to circumvent practices (WP1), (ii) quantifying/optimizing existing agroecological levers (WP2), (iii) promoting technical and technological innovation in order to detect, identify and manage weeds using alternative methods to herbicides (WP3), (iv) quantitative analyses through simulations and field trials of weed management effectiveness from a multi-criteria assessment (crop yield, GHGs (greenhouse gases) emissions,

impact on biodiversity, etc.) (WP4), (v) support for the collective design of systemic solutions, through case studies involving farmers and other local stakeholders (WP5), (vi) a renewed interest in the recognition and biology of weeds in order to take the right action through initial and ongoing training (WP6).

Led by INRAE, PARAD brings together 19 funded partners and 146 permanent staff from these organizations. For practical reasons, Iroko did not wish to participate directly in this project as a partner. However, Iroko is fully involved in WP3. Benjamin Bourel is PI@ntNet advisor of WP3. Iroko is involved in defining standardized protocols for data acquisition via smartphone and equivalent devices, on a 1m² plot scale. It is also involved in defining annotation formats and metadata. The aim is to ensure the effective integration, storage and reuse of this data within INRIA's PI@ntNet ecosystem. This will enable the project to take full advantage of existing infrastructure, tools and communities. To this end, the PI@ntNet API and batch import tools are being developed to explicitly support plot-related data. This is a mutually beneficial relationship for Iroko and PARAD. PARAD benefits from PI@ntNet's experience and infrastructure for plant identification. For its part, PI@ntNet benefits from the project's numerous partners, infrastructure and resources to acquire large quantities of plant images labeled by professionals and to test these APIs and the new PI@ntNet features (notably the one for multi-species identification).

Past2ECO (PEPR Agroécologie et Numérique), (2026-2031), 3 MEuros.

Participants: Benjamin Bourel, Alexis Joly.

Agroecology relies on the development of new genetic diversity (even lost ones) and practices (including varietal mixture) ensuring ecosystemic services (beyond yield stability) with limited to zero inputs (fertilizers and pesticides) and facing climatic variability and extreme events in the context of ongoing climate change. Past2ECO proposes to investigate both genetic diversity and agricultural practices relevant to agroecology by combining between- and within-crop past (exploiting herbarium specimens) and contemporaneous (leveraging seed banks) genetic diversity of wheat and sorghum, to provide practical and knowledge-driven solutions for climate-resilient agriculture and support the agroecological transition.

Past2ECO brings together complementary expertise in botany, genomics, biology, agronomy, AI and computer vision from eight institutes in integrating historical genetic knowledge, cutting-edge genomics, and AI-based phenotyping tools to guide agroecological crop transitions. Past2ECO aims to decipher historic diversity (WP1), unveil adaptive genomic footprints (WP2) evaluated in the field (WP3), all using AI-based image analysis technologies (WP4). The WP4 of the project is led by Iroko (Benjamin Bourel).

Triticeae and sorghum herbarium collections available are estimated to contain 18 101 specimens from 114 countries spanning 321 years from 1700-2021. After curation and classification, using AI-based morphometry, the analysis of ancient DNA, compared to that of worldwide accessions hosted in seed banks, will allow us to document the genetic diversity and the evolutionary trajectory of the use of varietal mixtures from past to present-day, and assess changes in their soil-root metagenome associations. Exploiting diversity, from the past to the present using innovative genomic offset statistics, we will be able to predict the optimal genotypes and varietal mixtures for current and future climate. We will validate how such predicted adaptive diversity manifests phenotypically in the field, how they can be used to guide the development of agroecological practices such as varietal mixtures in a climatic gradient, and what their benefits and evolutionary dynamics are under realistic on-farm diversity management practices.

Past2ECO builds a strong bridge between computer science and agroecology in unveiling specimen classification from AI-based geometric morphometrics, the development of new machine learning methods using neural networks on hyperspectral images to discriminate between varieties, and digital leaf phenotypes from herbarium specimens as an indicator of adaptation to global climate change.

Overall, Past2ECO will contribute to PEPR 'Agroecology and ICT' by delivering a cutting-edge proof-of-concept project aiming to decipher and exploit past (so-called lost or underused) adaptive diversity of current and future major crop species, wheat and sorghum, to design varieties for agroecological transitions in the context of climate change.

PI@ntAgroEco (PEPR Agroécologie et Numérique), (2023-2027), 1.6 MEuros.

Participants: Antoine Affouard, Christophe Botella, Hervé Goëau, Hugo Gresse, Alexis Joly, Thomas Paillot.

Agroecology necessarily involves crop diversification, but also the early detection of diseases, deficiencies and stresses (hydric, etc.), as well as better management of biodiversity. The main stumbling block is that this paradigm shift in agricultural practices requires expert skills in botany, plant pathology and ecology that are not generally available to those working in the field, such as farmers or agri-food technicians. Digital technologies, and artificial intelligence in particular, can play a crucial role in removing this barrier to access to knowledge.

The aim of the PI@ntAgroEco project will be to design, experiment with and develop new high-impact agro-ecology services within the PI@ntNet platform. This includes : AI and plant science research; agile development of new components within the platform; organizing participatory science programs and animating the PI@ntNet user community. The project is led by Iroko (Alexis Joly).

FishPredict (ANR), (2022-2025), 500 KEuros.

Participants: Benjamin Bourel, Alexis Joly, Maximilien Servajean, Julien Thomazo.

FishPredict ANR project funded in the context of the **IA-Biodiv** challenge. The project aims at predicting the biodiversity of reef fishes using AI technologies. Alexis Joly is co-leading of the whole project jointly with David Mouillot, marine ecologist at the **MARBEC** lab.

DeepPEP (ANR), (2025-2027), 25 KEuros.

Participants: Reza Akbarinia, Dennis Shasha, Patrick Valduriez.

The DeepPEP project, between CNRS, INRAE and Inria Iroko, aims to enhance the fundamental understanding of nutrient homeostasis in plants and develop new biostimulants using signaling peptides. The main objective is to acquire fundamental knowledge on the control of nutrient homeostasis in plants and to develop new biotechnological resources in the form of signaling peptides as biostimulants. The project seeks to create new AI algorithms for designing peptides that interact with any protein and develop potential biostimulants to enhance nitrogen (N) and phosphorus (P) efficiency in agriculture. In this project, Iroko provides its expertise in time series query processing

PPR Antibiorésistance: structuring tool "PROMISE" (2021-2024), 240 KEuros.

Participants: Reza Akbarinia, Florent Masegla.

The objective of the PROMISE (PROfessional coMmunity network on antimicrobial reSistanceE) project is to build a large data warehouse for managing and analyzing antimicrobial resistance (AMR) data. It gathers 21 existing professional networks and 42 academic partners from three sectors, human, animal, and environment. The project is based on the following transdisciplinary and cross-sectoral pillars: i) fostering synergies to improve the One Health surveillance of antibiotic consumption and AMR, ii) data sharing for improving the knowledge of professionals, iii) improving clinical research by analyzing the shared data.

PNR "Beerisk" (2022-2025). 200 KEuros.

Participants: Reza Akbarinia, Florent Masseglia.

The objective of this project is to analyze honeybee daily mortality rates, represented as time series, in order to detect anomalies and study the lethal effects of bees exposure to pesticides.

Plan national Ecoantibio "INTERSECTION" (2024-2028), 175 Keuros

Participants: Reza Akbarinia, Florent Masseglia.

The objective of the INTERSECTION project is to produce intersectoral and territorial indicators for monitoring resistance and use of antibiotics in France, and to facilitate the use and analysis of these indicators, in a One health approach.

PEPR agroécologie et numérique "RootSystemTracker" (2024-2027), 144 Keuros

Participants: Reza Akbarinia, Christophe Pradal, Lo'Ai Gandeel.

Roots play a crucial role in nutrient and water uptake, atmospheric carbon fixation, and soil interactions, significantly influencing resource use efficiency and crop resilience to environmental stresses. The objective of the RootSystemTracker project is to develop efficient methods for the spatio-temporal phenotyping of plant root architectures using heterogeneous data. This involves automatically capturing their topology and geometry over time, despite challenges such as root occlusions and variability in observation conditions.

Inria Challenge OMICFINDER (2023-2027), 1 Engineer - 24 months

Participants: Reza Akbarinia, Rebecca Pontes Salles, Florent Masseglia.

While genomic sequencing is enabling crucial advances in medicine, ecology, and agriculture, the exponential growth of public databases (48 petabytes by 2023) remains largely untapped due to the lack of efficient querying methods. OMICFINDER proposes an innovative global search engine that makes it possible to query nucleotidic sequences against the vast amount of publicly available genomic data. Combining novel algorithms, semantic web technologies, and distributed indexing with a focus on environmental sustainability, it aims to unlock this treasure trove of information – bringing the equivalent of a search engine to genomics at last. The project is led by Pierre Peterlongo (GenScale team, Inria Rennes).

10.4.1 Others

Participants: Alexis Joly, Jean-Christophe Lombardo, Hervé Goëau, Hugo Gresse, Mathias Chouet, Antoine Affouard, David Margery.

PI@ntNet consortium: In 2025, CNRS has joined the PI@ntNet consortium as a new member. This contract, initially signed by four founding research organisms (Inria, CIRAD, IRD, INRAE) aims at sustaining the PI@ntNet platform in the long term. It has been initiated in November 2019 in the context of the InriaSOFT national program of Inria. Each partner subscribes a yearly subscription (10-20K euros per year) to cover engineering costs for maintenance and technological developments. Depending on the membership status, each partner has one vote in the steering committee and/or the technical committee of the platform. He can also use the platform in his own projects and benefit from a certain number of service days within the platform. The consortium is not fixed and is intended to be extended to other members in the coming years.

10.5 Regional initiatives

Regional project "DACLIM" (2023-2026), 70 Keuros

Participants: Reza Akbarinia, Florent Masegla, Guillaume Coulaud.

The objective of this project is to develop scalable techniques based on massive data distribution to enable the efficient detection of anomalies in large climate databases. The detection of anomalies in climate data can provide climatologists with insights into the behavior of various climatological variables, understanding of extreme events such as heatwaves and cold snaps, as well as the prediction of these types of events.

10.6 Public policy support

CESE consultation on the impact of AI on the environment The CESE (Conseil Economique, Social et Environnemental) is one of the 3 assemblies of the French constitution, made up of elected representatives of civil society (unions, associations, companies, students, etc.). Its role is to provide advice on economic, social and environmental policies to guide public decision-making (governmental in particular). Alexis Joly took part in the consultation entitled "Impacts of artificial intelligence: risks and opportunities for the environment". He was consulted and interviewed on several occasions and was one of the 3 experts invited to the final plenary session that voted on the recommendations.

OECD report on the advancement of the productivity of science with citizen science and artificial intelligence Alexis Joly is a co-author of the chapter "Advancing the productivity of science with citizen science and artificial intelligence" in the OECD report *Artificial Intelligence in Science: Challenges, Opportunities and the Future of Research* (PDF). He participated in all preparatory meetings and contributed approximately 15% of the chapter, based on Iroko's expertise in citizen science, large-scale ecological data, and AI-driven biodiversity monitoring.

11 Dissemination

11.1 Promoting scientific activities

11.1.1 Scientific events: organization

General chair, scientific chair

- Alexis Joly: Main organizer of DeepSDM 2025, first conference on Deep Species Distribution Models (100 attendees).
- Joseph Salmon: Main organizer of MLMTP (Machine Learning seminar)

11.1.2 Scientific events: selection

Member of the conference program committees

- Reza Akbarinia: ICDM 2025, ECML-PKDD 2025, IEEE BigData 2025.
- Florent Masegla : ICDM 2025, ECML-PKDD 2025, DS 2025, PAKDD 2025, SAC 2025.

Reviewer

- Alexis Joly: CLEF 2025, ACM MM 2025

11.1.3 Journal

Editor, Associate editor

- Reza Akbarinia: associate editor of IEEE Transactions on Knowledge and Data Engineering (TKDE).
- Joseph Salmon: Associate Editor of IEEE Transaction on Image Processing
- Joseph Salmon: Action Editor of the Journal of Machine Learning Research
- Joseph Salmon: Associate Editor of the Electronic Journal of Statistics

Member of the editorial boards

- Reza Akbarinia: Transactions on Large Scale Data and Knowledge Centered Systems (TLDKS).
- Patrick Valduriez: Distributed and Parallel Databases.

Reviewer - reviewing activities

- Florent Masegla : Data & Knowledge Engineering.

11.1.4 Invited talks

- Patrick Valduriez gave an invited talk on AI and Scientific Research on:
 - July 2, 2025 at Cirad, Montpellier;
 - October 16, 2025 at Inria, Montpellier;
 - July 17, 2025 at the Dinizia workshop, CEFET-RJ, Rio de Janeiro, Brazil;
 - November 6, 2025 at [Workshop of the Artificial Intelligence Institute](#), LNCC, Petropolis, Brazil.
- Joseph Salmon
 - invited talk for IA Connect (launch of IA Montpellier Méditerranée)
 - invited talk at OSCII 2025
- Alexis Joly
 - [MILA Tea talk](#) - November 17, 2025 - Montreal
 - [IRBV](#) seminar - October 23, 2025 - Montreal
 - [Congrès annuel du Collège des Sociétés savantes académiques de France](#) - February 4, 2025 - Montpellier
 - EU JRC workshop on trees and habitat mapping - November 18-19, 2025 - Ispra

11.1.5 Leadership within the scientific community

- Esther Pacitti: Member of the Steering Committee of the BDA conference.
- Reza Akbarinia: Member of the Steering Committee of the BDA conference.
- Alexis Joly
 - Scientific and Technical director of Pl@ntNet platform
 - Coordinator of the LifeCLEF international virtual lab
- Joseph Salmon head of the Statistics and Data Science specialty, doctoral school EDI2S

11.1.6 Scientific expertise

- Christophe Pradal: member of the INRAE evaluation committee CSS (Scientific Specialist Commission) in Plant Integrated Biology
- Reza Akbarinia: member of the evaluation committee (section 27), University of Montpellier.
- Alexis Joly:
 - GENCI expert committee (**AI thematic**)
 - Scientific Advisory Board of the chaire "Angèle St-Pierre / Hugo Larochelle" related to AI applied to environment
 - Expert for SNSF (Swiss National Science Foundation) - project scientific evaluation
- Joseph Salmon :
 - elected member of the "Commission de Section 26", Univ. Montpellier.
 - head of the jury for hiring assistant professor (Univ. Montpellier, Faculté des Science)
 - Member of the Steering Committee for MathPhDInFrance
- Patrick Valduriez: consultant on big data for the **Software Heritage project**

11.1.7 Research administration

- Florent Masseglia: deputy scientific director of Inria for the domain "Perception, Cognition and Interaction", 50% of his time until September 2025.
- Reza Akbarinia: Scientific referent for research data at Inria branch of Montpellier; Member of Inria national commission for research data.
- Esther Pacitti: manager of Polytech' Montpellier's International Relations for the computer science department (100 students).
- Patrick Valduriez: scientific manager for the Latin America zone at Inria's Direction of Foreign Relationships (DRI) and scientific director of the **Inria-Brasil** strategic partnership.
- Christophe Pradal: Team leader with C. Granier of the **PhenoMEn** team of the AGAP Institute.
- Alexis Joly: co-manager of a **Collaborative Doctoral Partnership** between the **EU Joint Research Centre of Ispra** and the University of Montpellier

11.2 Teaching - Supervision - Juries - Educational and pedagogical outreach

11.2.1 Teaching

Esther Pacitti:

- IG3: Database design, physical organization, 54h, level, L3, 50 students.
- IG4: Distributed Databases and NoSQL, 80h , level M1, 50 students.
- Large Scale Information Management (Iot, Recommendation Systems, Graph Databases), 27h, level M2, 20 students.
- Supervision of industrial projects
- Supervision of master internships.
- Supervision of computer science discovery projects.

Joseph Salmon:

- HAX603X: Stochastic Modeling, 20h, level L3, 50 students.
- Supervision of master internships.
- Supervision of data science discovery projects.

11.2.2 Supervision

PhD & HDR:

- PhD (defended): Cesar Leblanc, Predicting biodiversity future trajectories through deep learning. Advisors: Alexis Joly, Maximilien Servajean, Pierre Bonnet.
- PhD (defended [58]): Matteo Contini, Multi-scale monitoring of coastal marine biodiversity. Advisors: Sylvain Bonhommeau, Alexis Joly.
- PhD in progress: Kawtar Zaher, Novel class retrieval through interactive learning. Advisors: Olivier Buisson, Alexis Joly.
- PhD in progress: Guillaume Coulaud, Anomaly Detection in Big Climate Data. Advisors: Reza Akbarinia, Audrey Brouillet, Florent Maseglia.
- PhD in progress: Loai Gandeel, Automatic methods for spatio-temporal reconstruction of root architecture. Advisors: Reza Akbarinia, Romain Fernandez, Christophe Pradal.
- PhD in progress: Raphaël Benerradi, species trends estimation from citizen science data. Advisors: Christophe Botella, Alexis Joly, Maximilien Servajean.
- PhD in progress: Théo Larcher, multi-scale species prediction. Advisors: Alexis Joly, Joseph Salmon, Pierre Bonnet, Marijn Van der Velde.
- PhD in progress: Sébastien Gigot-Leandri, decision-oriented site occupancy models. Advisors: Alexis Joly, Maximilien Servajean, David Mouillot.
- PhD in progress: Alex Maleknia, Influence functions and their applications to machine learning. Advisors: Joseph Salmon, E. Chzhen.

11.2.3 Juries

Members of the team participated in the following PhD or HDR committees:

- Reza Akbarinia:
 - Sara Jarrad, Sorbonne University (PhD reviewer)
 - Omar Ghannou, Aix-Marseille University (PhD reviewer)
- Joseph Salmon:
 - Benjamin Charlier (directeur du jury d’HDR, Univ. Montpellier)
 - Grégoire Pacreau (rapporteur de la thèse, École Polytechnique)
- Alexis Joly:
 - Cesar Leblanc PhD defense, Univ. of Montpellier, (PhD director)
 - Matteo Contini PhD defense, Univ. of Montpellier (as PhD director)

11.2.4 Educational and pedagogical outreach

- Patrick Valduriez, Chiche (2 actions): Lycée Mermoz, Lycée Clémenceau, Montpellier.
- Florent Maseglia, Chiche (4 actions): lycée Philippe de Girard, Avignon.

11.3 Popularization

11.3.1 Specific official responsibilities in science outreach structures

- Alexis Joly:
 - Member of the steering committee of PI@ntNet citizen science platform
 - Scientific Advisory Board of project [Le Féral](#)

11.3.2 Productions (articles, videos, podcasts, serious games, ...)

- Joseph Salmon: [Scientific blogging](#)
- PI@ntNet team:
 - [PI@ntNet website](#), in particular the [news section](#)
 - [PI@ntNet documentation](#) (new in 2025)
 - Videos: [Participation in Xprize](#), [Identify a plant with PI@ntNet](#), [PlantNet for dummies](#)
 - Articles: [Telabotanica news](#), [CIRAD blog](#), [Frandroid app of the week](#), [Gabon Actu](#).

11.3.3 Participation in Live events

- Joseph Salmon (2 actions): Mois des mathématiques appliquées et industrielles. (Lycée Joffre, Montpellier), IAUM (Univ. Montpellier, audience high schoolers)

11.3.4 Others science outreach relevant activities

- Joseph Salmon: Fête des sciences (carasciences), Montpellier

12 Scientific production

12.1 Major publications

- [1] C. Botella, A. Joly, P. Bonnet, F. Munoz and P. Monestiez. ‘Jointly estimating spatial sampling effort and habitat suitability for multiple species from opportunistic presence-only data’. In: *Methods in Ecology and Evolution* 12.5 (1st Feb. 2021), pp. 933–945. DOI: [10.1111/2041-210X.13565](https://doi.org/10.1111/2041-210X.13565). URL: <https://hal.umontpellier.fr/hal-03150701>.
- [2] B. Bourel, A. Joly, M. Servajean, S. Bettinger, J. A. Sanabria-Fernández and D. Mouillot. ‘From Presence-Only to Abundance Species Distribution Models Using Transfer Learning’. In: *Ecology Letters* 28.7 (24th July 2025), e70177. DOI: [10.1111/ele.70177](https://doi.org/10.1111/ele.70177). URL: <https://inria.hal.science/hal-05380189>.
- [3] B. Deneu, M. Servajean, P. Bonnet, C. Botella, F. Munoz and A. Joly. ‘Convolutional neural networks improve species distribution modelling by capturing the spatial structure of the environment’. In: *PLoS Computational Biology* 17.4 (19th Apr. 2021), e1008856. DOI: [10.1371/journal.pcbi.1008856](https://doi.org/10.1371/journal.pcbi.1008856). URL: <https://hal.inrae.fr/hal-03220977>.
- [4] C. Garcin, M. Servajean, A. Joly and J. Salmon. ‘Stochastic smoothing of the top-K calibrated hinge loss for deep imbalanced classification’. In: *Proceedings of ICML 2022*. ICML 2022 - 39th International Conference on Machine Learning. Vol. 162. Baltimore, United States: PMLR, 2022, pp. 7208–7222. URL: <https://inria.hal.science/hal-03828747>.
- [5] G. Heidsieck, D. de Oliveira, E. Pacitti, C. Pradal, F. Tardieu and P. Valduriez. ‘Cache-aware scheduling of scientific workflows in a multisite cloud’. In: *Future Generation Computer Systems* 122 (2021), pp. 172–186. DOI: [10.1016/j.future.2021.03.012](https://doi.org/10.1016/j.future.2021.03.012). URL: <https://hal.science/hal-03189130>.

- [6] C. Leblanc, P. Bonnet, M. Servajean, W. Thuiller, M. Chytrý, S. Aćić, O. Argagnon, I. Biurrun, G. Bonari, H. Bruelheide, J. A. Campos, A. Čarni, R. Čušterevska, M. de Sanctis, J. Dengler, T. Dziuba, E. Garbolino, U. Jandt, F. Jansen, J. Lenoir, J. E. Moeslund, A. Pérez-Haase, R. Pielech, J. Sibik, Z. Stančić, D. Uogintas, T. Wohlgemuth and A. Joly. ‘Learning the syntax of plant assemblages’. In: *Nature Plants* 11 (13th Oct. 2025), pp. 2026–2040. DOI: [10.1038/s41477-025-02105-7](https://doi.org/10.1038/s41477-025-02105-7). URL: <https://hal.inrae.fr/hal-05313746>.
- [7] T. Lefort, A. Affouard, B. Charlier, J.-C. Lombardo, M. Chouet, H. Goëau, J. Salmon, P. Bonnet and A. Joly. ‘Cooperative learning of PI@ntNet’s Artificial Intelligence algorithm: how does it work and how can we improve it?’ In: *Methods in Ecology and Evolution* (4th Feb. 2025). DOI: [10.1111/2041-210X.14486](https://doi.org/10.1111/2041-210X.14486). URL: <https://hal.science/hal-04603038>. In press.
- [8] M. Metz, M. Lesnoff, F. Abdelghafour, R. Akbarinia, F. Masegla and J.-M. Roger. ‘A “big-data” algorithm for KNN-PLS’. In: *Chemometrics and Intelligent Laboratory Systems* 203 (15th Aug. 2020), p. 104076. DOI: [10.1016/j.chemolab.2020.104076](https://doi.org/10.1016/j.chemolab.2020.104076). URL: <https://hal.inrae.fr/hal-02899789>.
- [9] T. Mondal, R. Akbarinia and F. Masegla. ‘kNN matrix profile for knowledge discovery from time series’. In: *Data Mining and Knowledge Discovery* 37.3 (May 2023), pp. 1055–1089. DOI: [10.1007/s10618-022-00883-8](https://doi.org/10.1007/s10618-022-00883-8). URL: <https://hal-lirmm.ccsd.cnrs.fr/lirmm-04225369>.
- [10] D. Oliveira, J. Liu and E. Pacitti. *Data-Intensive Workflow Management: For Clouds and Data-Intensive and Scalable Computing Environments*. Vol. 14. Synthesis Lectures on Data Management 4. Morgan&Claypool Publishers, 13th May 2019, pp. 1–179. DOI: [10.2200/S00915ED1V01Y201904DTM060](https://doi.org/10.2200/S00915ED1V01Y201904DTM060). URL: <https://hal-lirmm.ccsd.cnrs.fr/lirmm-02128444>.
- [11] T. Özsu and P. Valduriez. *Principles of Distributed Database Systems - Fourth Edition*. Springer, 2020, pp. 1–674. DOI: [10.1007/978-3-030-26253-2](https://doi.org/10.1007/978-3-030-26253-2). URL: <https://hal-lirmm.ccsd.cnrs.fr/lirmm-02265930>.
- [12] D.-E. E. Yagoubi, R. Akbarinia, F. Masegla and T. Palpanas. ‘Massively Distributed Time Series Indexing and Querying’. In: *IEEE Transactions on Knowledge and Data Engineering* 32.1 (2020), pp. 108–120. DOI: [10.1109/TKDE.2018.2880215](https://doi.org/10.1109/TKDE.2018.2880215). URL: <https://hal-lirmm.ccsd.cnrs.fr/lirmm-02197618>.
- [13] C. Zhang, R. Akbarinia and F. Toumani. ‘Efficient Incremental Computation of Aggregations over Sliding Windows’. In: *KDD 2021 - 27th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. Singapore (Virtual), Singapore, 2021, pp. 2136–2144. DOI: [10.1145/3447548.3467360](https://doi.org/10.1145/3447548.3467360). URL: <https://hal-lirmm.ccsd.cnrs.fr/lirmm-03359490>.

12.2 Publications of the year

International journals

- [14] B. Bourel, A. Joly, M. Servajean, S. Bettinger, J. A. Sanabria-Fernández and D. Mouillot. ‘From Presence-Only to Abundance Species Distribution Models Using Transfer Learning’. In: *Ecology Letters* 28.7 (24th July 2025), e70177. DOI: [10.1111/ele.70177](https://doi.org/10.1111/ele.70177). URL: <https://inria.hal.science/hal-05380189> (cit. on p. 19).
- [15] A. Brouillet, G. Coulaud, D. Shasha, R. Akbarinia and F. Masegla. ‘ClimBurst: A Novel Method to Detect Climatological Anomalies Over Time and Space’. In: *Geophysical Research Letters* 52.19 (3rd Oct. 2025), e2025GL117095. DOI: [10.1029/2025gl117095](https://doi.org/10.1029/2025gl117095). URL: <https://inria.hal.science/hal-05379952> (cit. on p. 16).
- [16] A. Y. H. Chai, S. H. Lee, F. S. Tay, H. Goëau, P. Bonnet and A. Joly. ‘PlantAIM: A new baseline model integrating global attention and local features for enhanced plant disease identification’. In: *Smart Agricultural Technology* 10 (Mar. 2025), p. 100813. DOI: [10.1016/j.atech.2025.100813](https://doi.org/10.1016/j.atech.2025.100813). URL: <https://hal.inrae.fr/hal-05017656> (cit. on p. 18).
- [17] M. Contini, V. Illien, J. Barde, S. Poulain, S. Bernard, A. Joly and S. Bonhommeau. ‘From underwater to drone: A novel multi-scale knowledge distillation approach for coral reef monitoring’. In: *Ecological Informatics* 89 (Nov. 2025), 103149 (16p.) DOI: [10.1016/j.ecoinf.2025.103149](https://doi.org/10.1016/j.ecoinf.2025.103149). URL: <https://hal.science/hal-05105687> (cit. on p. 19).

- [18] M. Contini, V. Illien, M. Julien, M. Ravitchandirane, V. Russias, A. Lazennec, T. Chevrier, C. L. Rintz, L. Carpentier, P. Gogendeau, C. Leblanc, S. Bernard, A. Boyer, J. Talpaert Daudon, S. Poulain, J. Barde, A. Joly, S. Bonhommeau and T. Chevrier. ‘Seatizen Atlas: a collaborative dataset of underwater and aerial marine imagery’. In: *Scientific Data* 12.1 (2025), p. 67. doi: [10.1038/s41597-024-04267-z](https://doi.org/10.1038/s41597-024-04267-z). URL: <https://hal.umontpellier.fr/hal-04918784> (cit. on pp. 12, 14, 19).
- [19] S. Demotes-Mainard, H. Autret, C. Pradal, J. Le Gall, V. Guérin, N. Leduc, D. Combes, C. Renaud, M. Chelle and J. Bertheloot. ‘Simulating light quantity and quality over plant organs using a ray-tracing method to investigate plant responses in growth chambers’. In: *Biosystems Engineering* 258 (Sept. 2025), p. 104256. doi: [10.1016/j.biosystemseng.2025.104256](https://doi.org/10.1016/j.biosystemseng.2025.104256). URL: <https://hal.science/hal-05212812>.
- [20] V. Espitalier, J.-C. Lombardo, H. Goëau, C. Botella, T. T. Høye, M. Dyrmann, P. Bonnet and A. Joly. ‘Adapting a global plant identification model to detect invasive alien plant species in high-resolution road side images’. In: *Ecological Informatics* 89 (2nd May 2025), p. 103129. doi: [10.1016/j.ecoinf.2025.103129](https://doi.org/10.1016/j.ecoinf.2025.103129). URL: <https://hal.inrae.fr/hal-05046527> (cit. on p. 18).
- [21] S. Kumschick, L. Journiac, O. Boulesnane-Genguant, C. Botella, R. Pouteau and M. Rouget. ‘Mapping potential environmental impacts of alien species in the face of climate change’. In: *Biological Invasions* 27.1 (1st Jan. 2025), p. 43. doi: [10.1007/s10530-024-03490-4](https://doi.org/10.1007/s10530-024-03490-4). URL: <https://hal.inrae.fr/hal-04866133>.
- [22] L. Latchoumane, M. Ecartot, R. Bendoula, J.-M. J. -. Roger, S. Mas Garcia, H. Villesseche, F. Tavernier, M. Ryckewaert, N. J. B. Gorretta, P. Roumet and E. Ballini. ‘Early detection of *Zymoseptoria tritici* infection on wheat leaves using hyperspectral imaging data’. In: *Data in Brief* 59 (Apr. 2025), p. 111404. doi: [10.1016/j.dib.2025.111404](https://doi.org/10.1016/j.dib.2025.111404). URL: <https://hal.science/hal-04959064>.
- [23] C. Leblanc, P. Bonnet, M. Servajean, W. Thuiller, M. Chytrý, S. Ačić, O. Argagnon, I. Biurrun, G. Bonari, H. Bruelheide, J. A. Campos, A. Čarni, R. Čušterevska, M. de Sanctis, J. Dengler, T. Dziuba, E. Garbolino, U. Jandt, F. Jansen, J. Lenoir, J. E. Moeslund, A. Pérez-Haase, R. Pielech, J. Sibik, Z. Stančić, D. Uogintas, T. Wohlgemuth and A. Joly. ‘Learning the syntax of plant assemblages’. In: *Nature Plants* 11 (13th Oct. 2025), pp. 2026–2040. doi: [10.1038/s41477-025-02105-7](https://doi.org/10.1038/s41477-025-02105-7). URL: <https://hal.inrae.fr/hal-05313746> (cit. on pp. 12, 18).
- [24] T. Lefort, A. Affouard, B. Charlier, J.-C. Lombardo, M. Chouet, H. Goëau, J. Salmon, P. Bonnet and A. Joly. ‘Cooperative learning of PL@ntNet’s Artificial Intelligence algorithm: how does it work and how can we improve it?’ In: *Methods in Ecology and Evolution* (4th Feb. 2025). doi: [10.1111/2041-210X.14486](https://doi.org/10.1111/2041-210X.14486). URL: <https://hal.science/hal-04603038>. In press.
- [25] S. Levionnois, N. Gaudio, R. Mahmoud, C. Pradal and C. Robert. ‘Impact of wheat-legume mix intercrops on wheat epidemics by modelling’. In: *Field Crops Research* 336 (1st Feb. 2026), p. 110212. doi: [10.1016/j.fcr.2025.110212](https://doi.org/10.1016/j.fcr.2025.110212). URL: <https://hal.science/hal-05328274> (cit. on p. 15).
- [26] J. Liu, B. Ma, Q. Yu, Y. Zhou, J. Zhou, R. Jin, D. Dou, H. Dai, H. Wang and P. Valduriez. ‘Efficient Federated Learning with Heterogeneous Data and Adaptive Dropout’. In: *ACM Transactions on Knowledge Discovery from Data (TKDD)* 19.8 (Sept. 2025), pp. 1–31. doi: [10.1145/3749376](https://doi.org/10.1145/3749376). URL: <https://hal-lirmm.ccsd.cnrs.fr/lirmm-05161239> (cit. on p. 15).
- [27] D. Marcos, R. van de Vlasakker, I. N. Athanasiadis, P. Bonnet, H. Goëau, A. Joly, W. D. Kissling, C. Leblanc, A. S. van Proosdij and K. P. Panousis. ‘Fully automatic extraction of morphological traits from the web: Utopia or reality?’ In: *Applications in Plant Sciences* 13.3 (June 2025). doi: [10.1002/aps3.70005](https://doi.org/10.1002/aps3.70005). URL: <https://hal.inrae.fr/hal-05101346> (cit. on p. 18).
- [28] G. Martellucci, H. Goëau, P. Bonnet, F. Vinatier and A. Joly. ‘PlantCLEF 2025: Advancing AI-based Multi-Species Plant Identification in Vegetation Quadrats for Supporting Environmental Law and Biodiversity Monitoring’. In: *Biodiversity Information Science and Standards* 9 (15th Dec. 2025), e181733. doi: [10.3897/biss.9.181733](https://doi.org/10.3897/biss.9.181733). URL: <https://hal.inrae.fr/hal-05442222>.

- [29] B. M. Neto-bradley, P. Bonnet, H. Goëau, A. Joly, J. Cavender-bares and D. A. Coomes. ‘Using reflectance spectra and Pl@ntNet to identify herbarium specimens: a case study with *Lithocarpus*’. In: *New Phytologist* (4th June 2025). DOI: [10.1111/nph.70258](https://doi.org/10.1111/nph.70258). URL: <https://hal.inrae.fr/hal-05155975> (cit. on p. 20).
- [30] E. Paixão, H. Balbi, E. Pacitti, F. Porto, J. Santos and E. Ogasawara. ‘FFT-Based Anomaly Detectors: Cutoff Frequency Adjustment and SMA-Based Approach’. In: *Journal of Information and Data Management* (2025). URL: <https://hal-lirmm.ccsd.cnrs.fr/lirmm-05332270>. In press.
- [31] L. Rauch, R. Heinrich, I. Moummad, A. Joly, B. Sick and C. Scholz. ‘Can Masked Autoencoders Also Listen to Birds?’ In: *Transactions on Machine Learning Research Journal* (1st Aug. 2025). URL: <https://inria.hal.science/hal-05457126>.
- [32] F. Reyes, B. Pitchers, C. Pradal and P.-É. Lauri. ‘Young apple tree development under agroforestry radiative conditions: a multi-scale morphological and architectural dataset’. In: *AoB Plants* 17 (Aug. 2025), plaf029. DOI: [10.1093/aobpla/plaf029](https://doi.org/10.1093/aobpla/plaf029). URL: <https://hal.inrae.fr/hal-05129574>.
- [33] R. Salles, B. Lange, R. Akbarinia, F. Masseglia, E. Ogasawara and E. Pacitti. ‘Scalable and accurate online multivariate anomaly detection’. In: *Information Systems* 131 (June 2025), p. 102524. DOI: [10.1016/j.is.2025.102524](https://doi.org/10.1016/j.is.2025.102524). URL: <https://hal-lirmm.ccsd.cnrs.fr/lirmm-05332273> (cit. on p. 16).
- [34] R. A. P. Silva, W. Ferreira, E. Pacitti, Y. Y. Frota and D. de Oliveira. ‘A Greedy Constructive Heuristic for Executing Cloud-based Workflows with Data Confidentiality Restrictions’. In: *SN Computer Science* 7 (Jan. 2026), pp. 1–52/92. DOI: [10.1007/s42979-025-04582-1](https://doi.org/10.1007/s42979-025-04582-1). URL: <https://hal-lirmm.ccsd.cnrs.fr/lirmm-05380013>.
- [35] J. Xu, K. Low, C. Botella, B. Deneu, D. Shasha and A. Joly. ‘Cascading predictions from common to uncommon species improves species distribution models for plants’. In: *Ecological Informatics* 93 (Feb. 2026), p. 103424. DOI: [10.1016/j.ecoinf.2025.103424](https://doi.org/10.1016/j.ecoinf.2025.103424). URL: <https://inria.hal.science/hal-05457107> (cit. on p. 19).

International peer-reviewed conferences

- [36] A. Aniraj, C. F. Dantas, D. Ienco and D. Marcos. ‘PDiscoFormer: Relaxing Part Discovery Constraints with Vision Transformers’. In: *Proceedings of the European Conference on Computer Vision (ECCV)*. ECCV 2024 - 18th European Conference on Computer Vision. Vol. 15143. Lecture Notes in Computer Science. Milano, Italy: Springer Nature Switzerland, 27th Nov. 2025, pp. 256–272. DOI: [10.1007/978-3-031-73013-9_15](https://doi.org/10.1007/978-3-031-73013-9_15). URL: <https://hal.science/hal-04659631>.
- [37] J. S. Cañas, S. Kahl, T. Denton, M. P. Toro-Gómez, S. Rodriguez-Buritica, J. L. Benavides-Lopez, J. S. Ulloa, P. Caycedo-Rosales, H. Klinck, H. Goëau, R. Planqué, W.-P. Vellinga and A. Joly. ‘Overview of BirdCLEF+ 2025: Multi-Taxonomic Sound Identification in the Middle Magdalena, Colombia’. In: *CEUR workshop*. CLEF 2025 - Working Notes of the Conference and Labs of the Evaluation Forum. Vol. CEUR-4038. CEUR Workshop Proceedings 232. Madrid, Spain: CEUR, 2025, pp. 2909–2919. URL: <https://hal.inrae.fr/hal-05319248>.
- [38] A. Y. H. Chai, K. L. Z. Jee, S. H. Lee, F. S. Tay, J. Vandeputte, H. Goeau, P. Bonnet and A. Joly. ‘Deep-Plant-Disease Dataset Is All You Need for Plant Disease Identification’. In: *ACM digital library*. ACM MM 2025 - 33. ACM International Conference on Multimedia. MM ’25: Proceedings of the 33rd ACM International Conference on Multimedia. Dublin, Ireland: ACM, 27th Oct. 2025, pp. 12578–12584. DOI: [10.1145/3746027.3758192](https://doi.org/10.1145/3746027.3758192). URL: <https://inria.hal.science/hal-05340010> (cit. on p. 14).
- [39] M. Contini, V. Illien, S. Poulain, S. Bernard, J. Barde, S. Bonhommeau and A. Joly. ‘The point is the mask: scaling coral reef segmentation with weak supervision’. In: *ICCV 2025 - Joint Workshop on Marine Vision*. Honolulu, Hawaii, United States, 19th Oct. 2025. URL: <https://inria.hal.science/hal-05457113> (cit. on p. 20).

- [40] G. Coulaud, B. Lange, D. Shasha, A. Brouillet, R. Akbarinia and F. Masseglia. ‘ClimBurst: A Dynamic Visualization Tool to Display Climatological Anomalies over Time and Space’. In: *Proceedings of the 34th ACM International Conference on Information and Knowledge Management*. CIKM 2025 - 34th ACM International Conference on Information and Knowledge Management. Seoul, South Korea: ACM, 10th Nov. 2025, pp. 6629–6633. DOI: [10.1145/3746252.3761466](https://doi.org/10.1145/3746252.3761466). URL: <https://inria.hal.science/hal-05379898> (cit. on p. 17).
- [41] J.-B. Fermanian, M. Hebiri and J. Salmon. ‘Class conditional conformal prediction for multiple inputs by p-value aggregation’. In: *NeurIPS 2025, The Thirty-Ninth Annual Conference on Neural Information Processing Systems*. San Diego (CA), United States, 1st Dec. 2025. URL: <https://hal.science/hal-05151303> (cit. on p. 19).
- [42] J.-B. Fermanian, P. Humbert and G. Blanchard. ‘Transductive Conformal Inference for Full Ranking’. In: *NeurIPS 2025, The Thirty-Ninth Annual Conference on Neural Information Processing Systems*. San Diego (CA), United States, 1st Dec. 2025. URL: <https://hal.science/hal-04895043> (cit. on p. 19).
- [43] C. Garcin, M. Servajean, A. Joly and J. Salmon. ‘A two-head loss function for deep Average-K classification’. In: *WACV 2025 - IEEE/CVF Winter Conference on Applications of Computer Vision*. Tucson, United States: IEEE, 2025, pp. 7358–7367. DOI: [10.1109/WACV61041.2025.00715](https://doi.org/10.1109/WACV61041.2025.00715). URL: <https://inria.hal.science/hal-04204318>.
- [44] L. Giusti Tavares, J. Lima, M. Melo, C. Chen, J. M. Garibaldi, G. dos Santos Scatena, A. H. Reali Costa, E. S. Gomi, R. Salles, E. Pacitti, I. Santos, I. G. Siqueira, F. Porto, D. Carvalho, R. Coutinho and E. Ogasawara. ‘Fuzzy-Based Ensemble Method for Robust Concept Drift Detection in Multivariate Time Series’. In: *IEEE Xplore. IJCNN 2025 - International Joint Conference on Neural Networks*. Rome, Italy: IEEE, 2025, pp. 1–12. DOI: [10.1109/IJCNN64981.2025.11228919](https://doi.org/10.1109/IJCNN64981.2025.11228919). URL: <https://hal-lirmm.ccsd.cnrs.fr/lirmm-05332316> (cit. on p. 16).
- [45] A. Joly, L. Picek, S. Kahl, H. Goëau, L. Adam, C. Botella, M. Servajean, D. Marcos, C. Leblanc, T. Larcher, J. Matas, K. Janoušková, V. Čermák, K. Papafitsoros, R. Planqué, W.-P. Vellinga, H. Klinck, T. Denton, P. Bonnet and H. Müller. ‘LifeCLEF 2025 Teaser: Challenges on Species Presence Prediction and Identification, and Individual Animal Identification’. In: *Lecture notes in computer science*. 47. European Conference on Information Retrieval - ECIR 2025. Advances in Information Retrieval. ECIR 2025 : part V LNCS-15576. Lucca, Italy: Springer Nature Switzerland, 3rd Apr. 2025, pp. 373–381. DOI: [10.1007/978-3-031-88720-8_57](https://doi.org/10.1007/978-3-031-88720-8_57). URL: <https://hal.inrae.fr/hal-05135815> (cit. on p. 20).
- [46] C. Leblanc, L. Picek, B. Deneu, P. Bonnet, M. Servajean, R. Palard and A. Joly. ‘Mapping Biodiversity at Very-High Resolution in Europe’. In: *CVPR 2025 - IEEE/CVF Computer Vision and Pattern Recognition Conference*. Nashville, TN, United States: IEEE, 15th June 2025, pp. 2340–2349. DOI: [10.1109/CVPRW67362.2025.00221](https://doi.org/10.1109/CVPRW67362.2025.00221). URL: <https://inria.hal.science/hal-05457128> (cit. on p. 19).
- [47] J. Z. Liaw, A. Y. H. Chai, S. H. Lee, P. Bonnet and A. Joly. ‘Can Language Improve Visual Features For Distinguishing Unseen Plant Diseases?’ In: *ICPR 2024 - 27th International Conference on Pattern Recognition*. Vol. 15330. Lecture Notes in Computer Science. Kolkata, India: Springer Nature Switzerland, 1st Jan. 2025, pp. 296–311. DOI: [10.1007/978-3-031-78113-1_20](https://doi.org/10.1007/978-3-031-78113-1_20). URL: <https://inria.hal.science/hal-04851648>.
- [48] J. Lima, H. Castro, L. Oliveira, E. Paixão, L. Baroni, R. Salles, R. Vargas and E. Ogasawara. ‘UniTED: A Unified Time Series Event Detection Repository’. In: *Brazilian e-Science Workshop - SBCOPENLIB. BreSci 2025 - XIX Brazilian e-Science Workshop*. Anais do XIX Brazilian e-Science Workshop 2025. Brasil, Brazil, 29th Sept. 2025, pp. 1–8. DOI: [10.5753/bresci.2025.247972](https://doi.org/10.5753/bresci.2025.247972). URL: <https://hal-lirmm.ccsd.cnrs.fr/lirmm-05332248>.
- [49] G. Martellucci, H. Goëau, P. Bonnet, F. Vinatier and A. Joly. ‘Overview of PlantCLEF 2025: Multi-Species Plant Identification in Vegetation Quadrat Images ★ Notebook for the LifeCLEF Lab at CLEF 2025’. In: *CLEF 2025 - Working Notes of the Conference and Labs of the Evaluation Forum*. Vol. 4038. CEUR - Workshop Proceedings. Madrid, Spain: CEUR, 2025, pp. 2942–2954. URL: <https://hal.inrae.fr/hal-05319328> (cit. on p. 20).

- [50] R. Parracho, F. Alexandrino, M. de Souza Figueiredo, L. Pereira da Silva, B. Dutra de Macedo, A. Lamblet Vaz, D. Louback, V. C. Desouzart, R. Salles, F. Porto, D. Carvalho and E. Ogasawara. ‘Leveraging Large Language Models for Time Series Prediction on Low-Frequency Data’. In: SBBDD 2025 - 40^o Simpósio Brasileiro de Banco de Dados. Fortaleza, Brazil: Sociedade Brasileira de Computação - SBC, 2025, pp. 196–208. DOI: [10.5753/sbbd.2025.247062](https://doi.org/10.5753/sbbd.2025.247062). URL: <https://hal-lirmm.ccsd.cnrs.fr/lirmm-05332267>.
- [51] L. Picek, C. Leblanc, T. Larcher, M. Servajean, P. Bonnet and A. Joly. ‘Overview of GeoLifeCLEF 2025: Plant Species Presence Prediction with Environmental and High-resolution Remote Sensing Data’. In: CLEF 2025 - Working Notes of the Conference and Labs of the Evaluation Forum. Vol. 4038. CEUR Workshop Proceedings. Madrid, Spain: CEUR, 2025, pp. 2932–2941. URL: <https://hal.inrae.fr/hal-05319448> (cit. on p. 20).
- [52] E. Pinto Sobrinho, J. Souza, J. Lima, L. Giusti, E. Bezerra, R. Coutinho, L. Baroni, E. Pacitti, F. Porto, K. Belloze and E. Ogasawara. ‘Fine-Tuning Detection Criteria for Enhancing Anomaly Detection in Time Series’. In: SBBDD 2025 - 40^o Simpósio Brasileiro de Banco de Dados. Fortaleza, Brazil: Sociedade Brasileira de Computação - SBC, 2025, pp. 209–221. DOI: [10.5753/sbbd.2025.247063](https://doi.org/10.5753/sbbd.2025.247063). URL: <https://hal-lirmm.ccsd.cnrs.fr/lirmm-05332275> (cit. on p. 16).
- [53] R. Salles, B. Lange, R. Akbarinia, F. Massegliia and E. Pacitti. ‘Energy Efficient Time Series Anomaly Detection’. In: ICECET 2025 - International Conference on Electrical, Computer and Energy Technologies. Paris, France, 3rd July 2025. URL: <https://hal-lirmm.ccsd.cnrs.fr/lirmm-05495463> (cit. on p. 17).

Conferences without proceedings

- [54] M. Gawinowski, L. Perthame, F. Rees, C. Richard-Molard, C. Pradal, P. Barbillon and A. Jullien. ‘Méta-modélisation des interactions plante-plante en 3D : application à l’association colza-féverole’. In: Journées scientifiques 2025 du PEPR Agroécologie et Numérique. Dijon, France, 28th Jan. 2025. URL: <https://hal.science/hal-04936942>.
- [55] G. Moraes, F. Porto, F. Ulliana, J.-F. Baget, M. Leclère, P. Bisquert, B. Gonçalves and P. Valduriez. ‘Gypscie-KG: Building a Logic-Based Approach for Knowledge Graph Data Integration View in ML Systems’. In: LAGO 2025 – LLMs, Análise de Grafos e Ontologias (SBBDD). Fortaleza, Brazil, 29th Sept. 2025. URL: <https://hal-lirmm.ccsd.cnrs.fr/lirmm-05229486> (cit. on p. 14).

Scientific books

- [56] E. S. Ogasawara, R. Salles, F. Porto and E. Pacitti. *Event Detection in Time Series*. Synthesis Lectures on Data Management (SLDM). Springer, Feb. 2025, pp. 1–178. URL: <https://hal-lirmm.ccsd.cnrs.fr/lirmm-04776385> (cit. on p. 15).

Scientific book chapters

- [57] L. Picek, S. Kahl, H. Goëau, L. Adam, T. Larcher, C. Leblanc, M. Servajean, K. Janoušková, J. Matas, V. Čermák, K. Papafitsoros, R. Planqué, W.-P. Vellinga, H. Klinck, T. Denton, J. Sebastián Cañas, G. Martellucci, F. Vinatier, P. Bonnet and A. Joly. ‘Overview of LifeCLEF 2025: Challenges on Species Presence Prediction and Identification, and Individual Animal Identification’. In: *Experimental IR Meets Multilinguality, Multimodality, and Interaction: 16th International Conference of the CLEF Association, CLEF 2025*. Vol. 16089. Lecture Notes in Computer Science. Springer Nature Switzerland, 1st Jan. 2026, pp. 338–362. DOI: [10.1007/978-3-032-04354-2_19](https://doi.org/10.1007/978-3-032-04354-2_19). URL: <https://hal.inrae.fr/hal-05408591> (cit. on p. 20).

Doctoral dissertations and habilitation theses

- [58] M. Contini. ‘Multi-scale mapping of changes in tropical reefs’. Université de Montpellier, 21st Nov. 2025. URL: <https://theses.hal.science/tel-05448315> (cit. on p. 33).

Reports & preprints

- [59] E. Coindre, R. Boulord, L. Chir, V. Freitas, M. Ryckewaert, T. Laisné, V. Bouckenooghe, M. Lis, L. Cabrera-Bosquet, A. Doligez, T. Simonneau, B. Pallas, A. Coupel-Ledru and V. Segura. *Robustness of high-throughput prediction of leaf ecophysiological traits using near infrared spectroscopy and poro-fluorometry*. 7th Feb. 2025. DOI: [10.1101/2025.02.07.636194](https://doi.org/10.1101/2025.02.07.636194). URL: <https://hal.science/hal-05323445>.
- [60] G. Coulaud, R. Akbarinia, A. Brouillet and F. Masseglia. *Leveraging Data Seasonality and Matrix Profile for Anomaly Detection: Application to Climate Time Series*. RR-9572. Inria, 22nd Jan. 2025. URL: <https://inria.hal.science/hal-04906596> (cit. on p. 17).
- [61] T. Ding, J.-B. Fermanian and J. Salmon. *Conformal Prediction for Long-Tailed Classification*. 9th July 2025. URL: <https://hal.science/hal-05157207> (cit. on p. 19).
- [62] K. Faizi, P. Mehta, A. Maida, T. Humphreys, E. Berrigan, L. Mc Kee-Reid, R. Mc Corkell, A. Tagade, J. Rumbelow, J. Showalter, L. Brent, C. Coroenne, A. Rigaud, A. Chandrasekhar, S. Navlakha, A. Martin, C. Pradal, S. Lee, W. Busch and M. P. Platre. *Growth Cost and Transport Efficiency Tradeoffs Define Root System Optimization Across Varying Developmental Stages and Environments in Arabidopsis*. 26th July 2025. DOI: [10.1101/2025.07.25.666579](https://doi.org/10.1101/2025.07.25.666579). URL: <https://hal.inrae.fr/hal-05214434>.
- [63] H. Huitu, T.-E. Skog, C. Pradal, A. Calatayud, T. Skaslien, B. Linnestad, A. Ronkainen, C. Fournier, M. Labadie, D. Skirvin, M. Pastell, D. Melchior, J. T. Langvatn and B. Nordskog. *Developing a generic DSS model metadata catalogue and APIs for crop protection*. 6th Aug. 2025. DOI: [10.31220/agriRxiv.2025.00350](https://doi.org/10.31220/agriRxiv.2025.00350). URL: <https://hal.inrae.fr/hal-05218411> (cit. on p. 15).
- [64] M. Ryckewaert, D. Marcos, C. Botella, M. Servajean, P. Bonnet and A. Joly. *Applying the maximum entropy principle to neural networks enhances multi-species distribution models*. 7th Jan. 2026. URL: <https://inria.hal.science/hal-05446820>.

Other scientific publications

- [65] R. Dubois, L. Bousset, M. Leclerc, N. Parisey and A. Joly. ‘Weakly supervised segmentation of leaf symptoms in field conditions’. In: Workshop Franco-Britannique organisé par le réseau « Modélisation et statistique pour la santé des animaux et des plantes ». Paris, France, 21st Oct. 2025. URL: <https://hal.science/hal-05478330>.
- [66] T. Gérard, R. Barillot, C. Pradal, M. Gauthier, C. Richard-Molard, B. Andrieu, A. Jullien and F. Rees. ‘Modelling C and N root-soil exchanges with 3D root and shoot growth based on plant ecophysiology: the Wheat-BRIDGES approach’. In: Rhizosphere 6 - Rooting for Earth. Edinburgh, United Kingdom, 2025. URL: <https://hal.inrae.fr/hal-05218470>.
- [67] I. Moummad. *Audiocarnet - Deep Representation Learning from Unlabeled Bioacoustic Data*. 2nd Nov. 2025. URL: <https://hal.science/hal-05342505>.

Scientific popularization

- [68] I. Manolescu and P. Valduriez. ‘De nouvelles architectures pour les Big Data’. In: *Le calcul à découvert*. CNRS Editions, 23rd Jan. 2025. URL: <https://hal-lirmm.ccsd.cnrs.fr/lirmm-05244631>.

Software

- [69] [SW] F. Rees, C. Pradal, M. Gauthier and T. Gérard, *RhizoDep: a Functional-Structural Root Model to simulate rhizodeposition* version 1.0.0, Apr. 2025. LIC: CeCILL-C. HAL: ([hal-05216496](https://hal.inrae.fr/hal-05216496)), URL: <https://hal.inrae.fr/hal-05216496>, vcs: <https://github.com/openalea/rhizodep>, SWHID: ([swh:1:dir:f8ac2404d19debfe5e5878348ef7f565ab31e129](https://sw.hal.archives-ouvertes.fr/hal-05216496);origin=<https://hal.l.archives-ouvertes.fr/hal-05216496>;visit=[swh:1:snp:b330c3c32015434521b0d4f74ec4355bde78072](https://sw.hal.archives-ouvertes.fr/hal-05216496);anchor=[swh:1:rel:958b443498fc793ae9fc8f14fb226007eea98911](https://sw.hal.archives-ouvertes.fr/hal-05216496);path=/).

12.3 Cited publications

- [70] M. Abdar, F. Pourpanah, S. Hussain, D. Rezazadegan, L. Liu, M. Ghavamzadeh, P. Fieguth, X. Cao, A. Khosravi, U. R. Acharya et al. ‘A review of uncertainty quantification in deep learning: Techniques, applications and challenges’. In: *Information Fusion* 76 (2021), pp. 243–297 (cit. on p. 10).
- [71] E. Afgan, J. Goecks, D. Baker, N. Coraor, A. Nekrutenko and J. Taylor. ‘Galaxy: A Gateway to Tools in e-Science’. In: *Guide to e-Science, Next Generation Scientific Research and Discovery*. Computer Communications and Networks. Springer, 2011, pp. 145–177 (cit. on p. 9).
- [72] R. Akbarinia, C. Botella, A. Joly, F. Masegla, M. Mattoso, E. Ogasawara, D. de Oliveira, E. Pacitti, F. Porto, C. Pradal, D. Shasha and P. Valduriez. ‘Life Science Workflow Services (LifeSWS): motivations and architecture’. In: *Transactions on Large-Scale Data- and Knowledge-Centered Systems* (). URL: <https://hal-lirmm.ccsd.cnrs.fr/lirmm-04173545> (cit. on p. 9).
- [73] T. Amano, J. D. Lamming and W. J. Sutherland. ‘Spatial gaps in global biodiversity information and the role of citizen science’. In: *Bioscience* 66.5 (2016), pp. 393–400 (cit. on p. 9).
- [74] M. Besson, J. Alison, K. Bjerge, T. E. Gorochoowski, T. T. Høyve, T. Jucker, H. M. Mann and C. F. Clements. ‘Towards the fully automated monitoring of ecological communities’. In: *Ecology Letters* 25.12 (2022), pp. 2753–2775 (cit. on p. 10).
- [75] C. Botella, P. Bonnet, C. Hui, A. Joly and D. M. Richardson. ‘Dynamic species distribution modeling reveals the pivotal role of human-mediated long-distance dispersal in plant invasion’. In: *Biology* 11.9 (2022), p. 1293 (cit. on pp. 9, 10).
- [76] C. Botella, A. Joly, P. Bonnet, P. Monestiez and F. Munoz. ‘A deep learning approach to species distribution modelling’. In: *Multimedia Tools and Applications for Environmental & Biodiversity Informatics* (2018), pp. 169–199 (cit. on p. 9).
- [77] C. Botella, A. Joly, P. Bonnet, F. Munoz and P. Monestiez. ‘Jointly estimating spatial sampling effort and habitat suitability for multiple species from opportunistic presence-only data’. In: *Methods in Ecology and Evolution* 12.5 (2021), pp. 933–945 (cit. on p. 9).
- [78] C. Botella, A. Joly, P. Monestiez, P. Bonnet and F. Munoz. ‘Bias in presence-only niche models related to sampling effort and species niches: Lessons for background point selection’. In: *PLoS One* 15.5 (2020), e0232078 (cit. on p. 9).
- [79] M. Chandler, L. See, K. Copas, A. M. Bonde, B. C. López, F. Danielsen, J. K. Legind, S. Masinde, A. J. Miller-Rushing, G. Newman et al. ‘Contribution of citizen science towards international biodiversity monitoring’. In: *Biological conservation* 213 (2017), pp. 280–294 (cit. on p. 9).
- [80] S. E. Fick and R. J. Hijmans. ‘WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas’. In: *International journal of climatology* 37.12 (2017), pp. 4302–4315 (cit. on p. 10).
- [81] M. Fontana, G. Zeni and S. Vantini. ‘Conformal prediction: a unified review of theory and new challenges’. In: *Bernoulli* 29.1 (2023), pp. 1–23 (cit. on p. 10).
- [82] C. Garcin, M. Servajean, A. Joly and J. Salmon. ‘A two-head loss function for deep Average-K classification’. In: *arXiv preprint arXiv:2303.18118* (2023) (cit. on p. 10).
- [83] S. Goff et al. ‘The iPlant Collaborative: Cyberinfrastructure for Plant Biology’. In: *Frontiers in Plant Science* 2 (2011) (cit. on p. 9).
- [84] T. Hey, S. Tansley, K. Tolle and J. Gray. *The Fourth Paradigm: Data-Intensive Scientific Discovery*. Microsoft Research, Oct. 2009 (cit. on p. 8).
- [85] W. Jetz, M. A. McGeoch, R. Guralnick, S. Ferrier, J. Beck, M. J. Costello, M. Fernandez, G. N. Geller, P. Keil, C. Merow et al. ‘Essential biodiversity variables for mapping and monitoring species populations’. In: *Nature ecology & evolution* 3.4 (2019), pp. 539–551 (cit. on p. 9).
- [86] H.-C. Kim and Z. Ghahramani. ‘Bayesian classifier combination’. In: *Artificial Intelligence and Statistics*. PMLR. 2012, pp. 619–627 (cit. on p. 9).

- [87] B. Lange, R. Akbarinia and F. Masegla. ‘A One-Health Platform for Antimicrobial Resistance Data Analytics’. In: *CIKM '24: Proceedings of the 33rd ACM International Conference on Information and Knowledge Management*. CIKM '24: Proceedings of the 33rd ACM International Conference on Information and Knowledge Management. Boise, United States, Oct. 2024, pp. 5230–5233. DOI: [10.1145/3627673.3679237](https://doi.org/10.1145/3627673.3679237). URL: <https://hal-lirmm.ccsd.cnrs.fr/lirmm-04774438> (cit. on p. 13).
- [88] T. Lefort, B. Charlier, A. Joly and J. Salmon. *Identify ambiguous tasks combining crowdsourced labels by weighting Areas Under the Margin*. Tech. rep. 2022, arXiv:2209.15380 (cit. on p. 9).
- [89] Y. Liu, S. Wang, J. Chen, B. Chen, X. Wang, D. Hao and L. Sun. ‘Rice Yield Prediction and Model Interpretation Based on Satellite and Climatic Indicators Using a Transformer Method’. In: *Remote Sensing* 14.19 (2022), p. 5045 (cit. on p. 10).
- [90] T. Lorieul. ‘Uncertainty in predictions of Deep Learning models for fine-grained classification’. Theses. Université Montpellier, Dec. 2020. URL: <https://theses.hal.science/tel-03040683> (cit. on p. 10).
- [91] T. Lorieul. ‘Uncertainty in predictions of deep learning models for fine-grained classification’. PhD thesis. Université Montpellier, 2020 (cit. on p. 10).
- [92] T. Mondal, R. Akbarinia and F. Masegla. ‘kNN Matrix Profile for Knowledge Discovery from Time Series’. In: *Data Mining and Knowledge Discovery (DMKD) (2023)* (cit. on p. 10).
- [93] M. Ohlmann, F. Munoz, F. Massol and W. Thuiller. ‘Assessing mutualistic metacommunity capacity by integrating spatial and interaction networks’. In: *arXiv preprint arXiv:2206.11029* (2022) (cit. on p. 10).
- [94] G. Pleiss, T. Zhang, E. R. Elenberg and K. Q. Weinberger. ‘Identifying mislabeled data using the area under the margin ranking’. In: *NeurIPS*. 2020 (cit. on p. 9).
- [95] C. Pradal, C. Fournier, P. Valduriez and S. C. Boulakia. ‘OpenAlea: scientific workflows combining data analysis and simulation’. In: *International Conference on Scientific and Statistical Database Management (SSDBM)*. 2015, 11:1–11:6 (cit. on p. 9).
- [96] G. Shafer and V. Vovk. ‘A Tutorial on Conformal Prediction.’ In: *Journal of Machine Learning Research* 9.3 (2008) (cit. on p. 10).
- [97] A. J. van Strien, T. Termaat, D. Groenendijk, V. Mensing and M. Kéry. ‘Site-occupancy models may offer new opportunities for dragonfly monitoring based on daily species lists’. In: *Basic and Applied Ecology* 11.6 (2010), pp. 495–503 (cit. on p. 9).
- [98] M. Sundararajan and A. Najmi. ‘The many Shapley values for model explanation’. In: *International conference on machine learning*. PMLR. 2020, pp. 9269–9278 (cit. on p. 10).
- [99] C. Violle, P. Choler, B. Borgy, E. Garnier, B. Amiaud, G. Debarros, S. Diquelou, S. Gachet, C. Jolivet, J. Kattge et al. ‘Vegetation ecology meets ecosystem science: Permanent grasslands as a functional biogeography case study’. In: *Science of the Total Environment* 534 (2015), pp. 43–51 (cit. on p. 10).
- [100] D.-E. E. Yagoubi, R. Akbarinia, F. Masegla and T. Palpanas. ‘Massively Distributed Time Series Indexing and Querying’. In: *IEEE Transactions on Knowledge and Data Engineering* 32.1 (2020), pp. 108–120. DOI: [10.1109/TKDE.2018.2880215](https://doi.org/10.1109/TKDE.2018.2880215). URL: <https://hal-lirmm.ccsd.cnrs.fr/lirmm-02197618> (cit. on pp. 9, 10).
- [101] C. M. Yeh, Y. Zhu, L. Ulanova, N. Begum, Y. Ding, H. A. Dau, D. F. Silva, A. Mueen and E. J. Keogh. ‘Matrix Profile I: All Pairs Similarity Joins for Time Series: A Unifying View That Includes Motifs, Discords and Shapelets’. In: *IEEE 16th International Conference on Data Mining, ICDM 2016, December 12-15, 2016, Barcelona, Spain*. Ed. by F. Bonchi, J. Domingo-Ferrer, R. Baeza-Yates, Z. Zhou and X. Wu. IEEE Computer Society, 2016, pp. 1317–1322. DOI: [10.1109/ICDM.2016.0179](https://doi.org/10.1109/ICDM.2016.0179). URL: <https://doi.org/10.1109/ICDM.2016.0179> (cit. on p. 9).