

SBDI Days 2024

Towards Data-driven Ecology



SWEDISH MUSEUM OF NATURAL HISTORY
JANUARY 24-25 2024



Wednesday

Schedule

09.30	Registration & coffee	
10.00	Introduction to the SBDI Days 2024	Margret Steinthorsdottir
10.30	<i>Perils and promises in big data research - examples from macroecology</i> Keynote: Alexandre Antonelli	Chair: Veronika Johansson
11.00	Monitoring biodiversity with non-random samples. Keynote: Rob James Boyd	
11.30	Evaluating forecasts from alternative citizen science data and fitting methods versus an individual-based model. Tord Snäll	
11.45	An insect species multimeter for troubleshooting landscape ecology. Mikkel Brydegaard	
12.00	LUNCH	
13.00	<i>Machine-learning and artificial intelligence applications in conservation and management of ecosystems</i> Biodiversity data requirements of AI tools for biodiversity-positive decision making. Keynote: Tobias Andermann	Chair: Anna Lena Axelsson
13.30	Biodiversity indicators from remote sensing laser data. Keynote: Eva Lindberg & Langning Huo	
14.00	Unoccupied Aerial Vehicles (UAVs) and Deep Learning for Efficient Ecological Tracking of Ant Mounds. Jérémy Monsimet	
14.15	Sensors and AI for automated seabird monitoring. Jonas Hentati-Sundberg	
14.30	COFFEE	
15.00	Posters and Workshops	1. Research using camera traps – Bellatrix 2. Modelling the distribution of alien species in Sweden – Andromeda
16.00	Poster Session	3. FAIR principles in life science research practice – Vintergatan 4. Big data – big sample size? Assessing risk of bias – Lilla hörsalen
19.00	DINNER	

Thursday

Schedule

09.00	<i>Perspectives on digital twins for ecosystems and ecological modelling</i> Ecosystem Digital Twin pipelines with Big Data and AI Keynote: Arne Jørgen Berre	Chair: Matthias Obst
09.30	Digital twins of biodiversity; concepts, workflows and data. Sam Wenaas Perrin	
09.45	Do we need biodiversity in soil biogeochemical models? Stefano Manzoni	
10.00	Biodiversity models based on multi-agent reinforcement learning. Claes Strannegård	
10.15	Lichens from space – Pioneering lichen remote sensing applied in wildlife conservation and management on a landscape scale. Rasmus Erlandsson	
10.30	COFFEE	
11.00	<i>Scaling up DNA-based ecology research</i> Exploring ecological interactions with DNA-based network analysis. Keynote: Monika Quinones Winder	Chair: Anna Rosling
11.30	AI-assisted plankton monitoring with in-flow microscopy & eDNA. Karin Garefelt	
11.45	SBDI's genetic toolkit. Anders Andersson	
12.00	Airborne eDNA captures three decades of ecosystem biodiversity. Per Stenberg	
12.15	aMeta: a computational method for data-driven ancient metagenomic analysis. Nikolay Oskolkov	
12.30	Wrap up and conclusion	Margret Steinthorsdottir
12.45	LUNCH	

Keynotes

Invited speakers

Machine-learning and artificial intelligence applications in conservation and management of ecosystems

Tobias Andermann

Department of Organismal Biology, Systematic Biology, Uppsala University
SciLifeLab and Wallenberg Data Driven Life Science Fellow

Tobias is a biodiversity researcher dedicated to providing data and computational tools for combating the global biodiversity crisis. His group, the Biodiversity Data Lab, is working on the intersection of molecular biology, spatial ecology, and machine learning, with the mission to provide a more comprehensive view on the distribution of biodiversity, including hidden diversity of inconspicuous and even undescribed species through the use of environmental DNA.

More information about Tobias: <https://www.biodiversity.se>



Talk: **Biodiversity data requirements of AI tools for biodiversity-positive decision making** – There is an ever-increasing need for developing standardized ways of quantifying the biodiversity value of a site. This need arises from international and local policies (e.g. the COP15 UN Biodiversity agreement) as well as from the independent investments of companies world-wide in developing biodiversity-positive profiles, motivated by their public perception. This momentum provides an unprecedented opportunity to efficiently protect biodiversity on a large scale to turn around the alarmingly negative biodiversity trends of the recent decades and centuries. While individual biodiversity assessments for specific sites can be done through targeted inventories, either through taxonomic experts or increasingly through the use of environmental DNA, we also need ways of scaling up such biodiversity evaluations to larger scales, covering large areas. AI models provide a unique tool to leverage complex biodiversity and remote sensing data to make predictions of biodiversity on large spatial scales, while at the same time providing high spatial resolution.

In this talk I will present some of these models from my ongoing research and discuss their utility, but also their current shortcomings and limitations. One of the main bottlenecks that we are currently facing is the quality of the biodiversity data that is currently available, which suffer from huge biases in their spatial and taxonomic coverage. Environmental DNA, collected on large spatial scales and made available through standardized platforms such as SBDI are a promising way forward to satisfy these data needs and overcome some of the biggest shortcomings.

Perils and promises in big data research - examples from macroecology

Alexandre Antonelli

Royal Botanic Gardens Kew

Alexandre is a biodiversity scientist studying the distribution, evolution, threats and sustainable uses of species and developing methods to speed up scientific discovery and innovation. Building on extensive fieldwork in the American tropics and integrating his findings with comprehensive datasets, his research has demonstrated how the evolution of landscapes and biodiversity are intrinsically linked, revealed the origins and processes underlying current patterns and synthesised knowledge across many groups of species and regions. Antonelli is the Director of Science at the Royal Botanic Gardens, Kew and Professor in Biodiversity and Systematics at the University of Gothenburg, among other appointments and distinctions. More information about Alexandre: <https://www.kew.org/science/our-science/people/alexandre-antonelli>

Talk: **Illuminating the world's biodiversity darkspots** – Delivering on the Kunming-Montreal Global Biodiversity Framework requires filling the most critical knowledge gaps. Here I will present a framework for prioritising regions around the world with highest chances of yielding new species to science as well as most new geographic records. I will discuss how new technological developments, from remote sensing to citizen science, environmental DNA and semi-automated inventories can (and should) be integrated with expert surveys and analysed under robust methods in order to properly identify the most important areas for biological conservation and restoration needed to halt biodiversity loss.



Perspectives on digital twins for ecosystems and ecological modelling

Arne Jørgen Berre

SINTEF Digital, Smart Data group in Sustainable Communication Technologies department and NorwAI.ai

Arne is task lead of Digital Twins of the Ocean interoperability in the DTO-BioFlow project on bringing Biodiversity into the European Digital Twins of the Ocean. He is scientific coordinator in the Iliad Digital Twins of the Ocean project and further responsible for digital twins and EOSC interoperability in the AqualNFRA EOSC project and in the CLIMAREST Mission Ocean project. He is co-lead in the European Big Data Value Association (BDVA.eu) task force on AI and Big Data standards and benchmarking, and involved in the ADRA organisation on AI, Data and Robotics. He is the lead of the Norwegian ISO SC42 AI group and involved in ISO SC41 IoT and Digital Twin, The Digital Twin Consortium, ISO/TC211 and OGC. He is co-chair of WG5 Digital Twins of the Ocean Framework and architecture (Turtle) in DITTO (Digital Twins of the Ocean) – a Global Program of the UN Decade of Ocean Science for Sustainable Development (2021-2030).

Talk: **Ecosystem Digital Twin pipelines with Big Data and AI** – Presents a standards-based framework for Ecosystem and Biodiversity digital twin pipelines, with examples from European projects like DTO-BioFlow, BioDT, Destination Earth/DestinE and Iliad/EDITO Digital Twins of the Ocean. A Digital Twin is a digital representation of a target enti-



ty with data connections that enable convergence between the physical and digital states at an appropriate rate of synchronization. Digital Twin pipelines are presented with a foundation in big data from new sensors and observations through multiple data types including observations, citizen science, video/images, remote sensing, bioacoustics, environmental DNA and others. Data management is supported through biodiversity data spaces and digital twin data lakes with the use of appropriate data representation standards and APIs and Digital Twin standards. Hybrid and Intelligent/Cognitive Digital Twins are combining data-driven models and simulation models with AI/Machine learning and natural models for predictions and recommendations supported by Cloud, GPU and HPC processing. Graphical digital twins offer various types of 2D/3D/4D visualisation and interaction modes.

Perils and promises in big data research - examples from macroecology.

Rob James Boyd

UK Centre for Ecology & Hydrology

Rob is a quantitative ecologist and methodologist, whose main interests are in statistics and biodiversity monitoring. He works mostly on sampling theory and developing methods to assess, mitigate and communicate the risk of bias that is typical of biodiversity data.

More information about Rob: <https://www.ceh.ac.uk/fr/staff/robin-boyd>

Talk: **Monitoring biodiversity with non-random samples.** Rob will start by explaining why it is challenging to monitor biodiversity without random samples. He will then introduce some methods for assessing the extent of the problem and mitigating it to the extent that it is possible..



Machine-learning and artificial intelligence applications in conservation and management of ecosystems

Eva Lindberg & Langning Huo

Department of Forest Resource Management, Division of Forest Remote Sensing, Swedish University of Agricultural Sciences

Eva is developing new automatic methods to derive information from remotely sensed data, primarily airborne laser scanning data, relevant for vegetation and habitat analysis and mapping. Her work includes the technical aspects for data analysis, but also research about the related natural processes, for example the influence of canopy structure on the habitats of forest birds, insects, and lichens. More information about Eva:

<https://www.slu.se/en/ew-cv/eva-lindberg/> and www.mistradigitalforest.se/en/news/promising-results-for-a-new-method-of-assessing-biodiversity

Langning is a researcher in forest remote sensing, specializing in developing remote sensing solutions for monitoring forest health and disturbance, from individual-tree scale to landscape scale. She has experience with analyzing indicators from remote sensing data on forest biodiversity and conservation values. Her work also includes developing remote sensing tools to early identify forest disturbances such as bark beetle infestations and drought damages.

More information about Langning:

<https://intern.slu.se/cv-originalen/langning-huo>



Talk: **Biodiversity indicators from remote sensing laser data.** Remote sensing can provide information to map vegetation states and changes for large areas. Recent developments in remote sensing has resulted in data with increasingly higher resolution, both spectrally, spatially, and temporally. This talk will give an introduction to mapping of biodiversity indicators from remote sensing and present how dead wood and trees with high conservation value can be mapped from high-resolution airborne laser scanning data.

Scaling up DNA-based ecology research

Monika Quinones Winder

Department of Ecology, Environment and Plant Sciences, Stockholm University



Monika is a professor in marine ecology at Stockholm University and a leader in aquatic food web interactions. By combining DNA metabarcoding, sustained monitoring datasets and network modelling, her group developed a novel approach to study the importance of diversity within community interactions. Current projects extend food web interactions from protists to fish building one of the most comprehensive and resolved marine food webs. She also explores the hidden role of zooplankton symbiotic interactions and impacts of shifting phenologies on trophic coupling. Monika earned a PhD from ETH, Switzerland and has worked at a number of international universities, including the University of Washington, University of Davis, USA and Geomar, Germany, before moving to Stockholm University.

Talk: **Exploring ecological interactions with DNA-based network analysis**

Food webs, composed of a vast diversity of organisms and their interactions are the engine of healthy, resilient and biodiverse coastal ecosystems. Yet, knowledge about organization of food webs is limited because of many methodological challenges to study ecological interactions in nature. Here I will present a new approach that combines DNA metabarcoding and energetic food web modelling to quantify carbon fluxes and ecosystem functioning in real and dynamic natural landscapes. This is supported by automated hydroacoustic data collection to investigate spatial distribution of plankton and fish in the water column in real time. Overall, these findings illustrate that an improved understanding of species interactions are key for endorsing essential measures to promote a resilient and biodiverse Baltic Sea, able to support socioeconomic activities.

Talks

Abstracts

Talks are either 30 min or 15 min. Please leave 3-5 min for questions. The audience consists largely of academic scientists. Please keep in mind that the audience may not be experts in your field of research.

SBDI's genetic toolkit

Anders Andersson

(Maria Prager^{1,2}, Daniel Lundin³, Jeanette Tångrot⁴, Anna Rosling⁵, Anders F. Andersson^{6*})

1. Stockholm University, Sweden; 2. Karolinska Institute, Sweden; 3. Linnaeus University, Sweden; 4. Umeå University / NBIS; 5. Uppsala University; 6. KTH Royal Institute of Technology / SciLifeLab, Sweden, *Presenting author

The Swedish Biodiversity Data Infrastructure (SBDI) develops tools for publishing, harvesting and analysing biodiversity data. Since molecular methods for species observation, such as metabarcoding of environmental DNA, are becoming increasingly important for documenting taxonomic observations across all kingdoms of life and biomes, SBDI has developed a set of services and tools for genetic-based biodiversity data. The core components of these services are 1) A flexible user-friendly bioinformatics pipeline (nf-core/ampliseq) for reproducible processing of metabarcoding data; 2) The Swedish Amplicon Sequence Variant (ASV) portal: a web interface for metabarcoding-based species observations in the SBDI Biotlas; 3) Curated reference sequence databases for improved taxonomic annotations. In this presentation we will demonstrate how the services provided by SBDI can be used to process, publish, search, download and analyse metabarcoding-based biodiversity data.

An insect species multimeter for troubleshooting landscape ecology, OR: Measuring insect species richness in-situ with photonics.

Mikkel Brydegaard

Mikkel Brydegaard^{1,2,3}, Meng Li¹, Klas Rydhmer⁴, Benoit Kouakou⁵, Adolpe Gbogbo⁵, Jeremie Zoueu⁵, Carsten Kirkeby⁴, Anna Runemark², Henrik Smith²

1 Dept. Physics, Lund University, Sölvegatan 14c, 22363 Lund, Sweden; 2 Dept. Biology, Lund University, Sölvegatan 35, 22362 Lund, Sweden; 3 Norsk Elektro Op2kk, Østensjøveien 34, 0667 Oslo, Norway; 4 Copenhagen University, Copenhagen, Denmark; 5 Institut National Polytechnique Félix Houphouët-Boigny, Yamoussoukro, Ivory Coast

Monitoring insect diversity and its decline can be a costly and time-consuming business, both in terms of trapping campaigns and post-capture identification. Automated approaches by distributed sensors^[1] could imply better coverage and resolution in time^[2] and space^[3], instant classification and potentially speed up the evaluation of insect decline mitigations. We present progress and current state of automated alternatives based on lasers and photonics^[4]. We can currently remotely count hundreds of thousands of insects daily in the field^[5], and retrieve their oscillatory properties individually and differentiate hundreds of taxonomic groups^[6]. How and why does it work? How well does it work? What are the advantages, challenges and limitations? What molecular, nano- and micro-structural features^[7] can be measured on free flying insects in nature^[8], and to what extent can these observational features be used to distinguish species richness and taxonomic group^[9]?

[1] K. Rydhmer, E. Bick, L. S2ll, A. Strand, R. Luciano, S. Helmreich, B. D. Beck, C. Grønne, L. Malmros, K. Poulsen, F. Elbæk, M. Brydegaard, J. Lemmich, and T. Nikolajsen, "Automating insect monitoring using unsupervised near-infrared sensors," *Scientific Reports*, vol. 12, pp. 1-11, 2022. [2] E. Malmqvist, S. Jansson, S. Zhu, W. Li, K. Svanberg, S. Svanberg, J. Rydell, Z. Song, J. Bood, M. Brydegaard, and S. Åkesson, "The batbird-bug baBLE: daily flight activity of insects and their predators over a rice field revealed by high resolution Scheimpflug Lidar," *Royal Society Open Science*, vol. 5, 2018. [3] S. Jansson, E. Malmqvist, Y. Mlacha, R. Ignell, F. Okumu, G. Killeen, C. Kirkeby, and M. Brydegaard, "Real-time dispersal of malaria vectors in rural Africa monitored with lidar," *PLoS One*, 2020. [4] M. Brydegaard and S. Svanberg, "Photonic monitoring of atmospheric and aquatic fauna," *Laser & Photonics Reviews*, 2018. [5] M. Brydegaard, S. Jansson, E. Malmqvist, Y. Mlacha, A. Gebru, F. Okumu, G. Killeen, and C. Kirkeby, "Lidar reveals Activity Anomaly of Malaria Vectors during Pan-African Eclipse," *Science Advances*, 2020. [6] B. K. Kouakou, S. Jansson, M. Brydegaard, and J. T. Zoueu, "Entomological Scheimpflug lidar for es2ma2ng unique insect classes insitu field test from Ivory Coast," *OSA Commun*, vol. 3, pp. 2362-2371, 2020/09/15 2020. [7] M. Li, C. Seinsche, S. Jansson, J. Hernandez, J. Rota, E. Warrant, and M. Brydegaard, "Potential for identification of wild night-flying moths by remote infrared microscopy," *Royal Society Interface*, 2022. [8] L. Müller, M. Li, H. MåneUord, J. Salvador, N. Reistad, J. Hernandez, C. Kirkeby, A. Runemark, and M. Brydegaard, "Remote Nanoscopy with Infrared Elase2c Hyperspectral Lidar," *Advanced Science*, 2023. [9] M. Li, A. Runemark, J. Hernandez, J. Rota, R. Bygebjerg, and M. Brydegaard, "Discrimination of hover fly species and sexes by wing interference signals," *Advanced Science*, 2023.

From chirps to algorithms: using machine learning to interpret eco-acoustic data in Norway.

Benjamin Cretois (cannot participate)

Norwegian Institute for Nature Research

Advancements in machine learning (ML) and artificial intelligence (AI) have begun to play a pivotal role in ecological conservation, providing sophisticated tools for the analysis of environmental acoustic data. Using state-of-the-art machine learning tools, we have analyzed the Sound of Norway dataset, a large eco-acoustic dataset which is the result of two years of continuous recording at 30 sites across Norway. We applied AI-driven techniques to three distinct but interconnected challenges in the acoustic characterization of ecosystems.

Initially, we addressed the ethical issue of anonymizing human vocalizations captured within ecological recordings. Employing state-of-the-art AI algorithms, we have effectively anonymized human auditory data, maintaining the privacy imperative while preserving the ecological validity of ambient soundscapes. Our method underscores the potential for AI to support ethical research practices in eco-acoustic data collection.

Subsequently, we incorporated BirdNET – a deep learning based model for bird classification - to evaluate bird species diversity across Norway's varied ecosystems. BirdNET results were validated by an ornithologist proving that species recognition and classification models can enable comprehensive biodiversity assessments over extensive geographical areas and temporal spans.

Finally, we analyze the ecological impact of snowmobile noise on avian acoustic communication. Developing a deep learning-based model for detecting snowmobiles, we were able to discern alterations in bird vocalization patterns and infer behavioral changes within bird populations subjected to mechanized noise. This analysis provides critical insights into the resilience and vulnerability of avian species to anthropogenic acoustic disruptions.

Our integrated AI approach, applied to the Sound of Norway's extensive dataset, showcase the potential of AI to significantly augment the capability for effective ecosystem conservation and management strategies.

Lichens from space – Pioneering lichen remote sensing applied in wildlife conservation and management on a landscape scale.

Rasmus Erlandsson

Department of Ecology, Environment and Plant Sciences, Stockholm University

Remote sensing is an indispensable tool in landscape ecology. While satellite-based vegetation indices have allowed for efficient global monitoring of green vegetation since the 1970's, this is not the case for lichens and other non-green vegetation. I will present a new, high resolution, satellite based remote sensing model for assessment of lichen biomass, based on a deep neural network trained with ground data collected for more than 20 years. I will show examples of how the model has been implemented in the conservation program for the endangered European wild reindeer in southern Norway, and how it can be used for future management. Because despite their importance, the lack of efficient tools has left many ecological questions related to lichens unanswered. Although often overlooked in vegetation studies, ground dwelling lichens form a prominent vegetation type in many high latitude ecosystems. Changes in biomass of mat-forming pale lichens have the potential to affect not only vegetation, fauna and human activities related to reindeer husbandry, but also climate due to their high albedo. Yet, in contrast to green vegetation, the complex spectral signal of lichens constitutes a considerable methodological challenge regarding development of remote sensing techniques. With the recent development of deep neural networks, however, we now have the tools to address this complexity. I will show how the back catalogue of the Landsat program (>38 years) allows quantification of recent historic dynamics in lichen vegetation across large scales and gives us the possibility to expand our knowledge in the ecological and climatic causes and consequences of changes in lichen biomass.

AI-assisted plankton monitoring with in-flow microscopy & eDNA.

Karin Garefelt

KTH - Royal Institute of Technology, SciLifeLab, Stockholm

Single-celled eukaryotic plankton (protists) form the productive base of marine ecosystems and are key drivers of global biogeochemical cycles of carbon and nutrients. Monitoring of eukaryotic plankton has traditionally been conducted by manual microscopic detection. Recently, alternative approaches have emerged such as high-throughput imaging and DNA metabarcoding. While promising, these methods have their challenges, not least in how to translate between these disparate datasets.

In the AMIME project we utilize state-of-the-art image analysis and deep learning approaches to maximize the information gained from these types of data and to translate between them. The project leverages on existing imaging data from the new Imaging FlowCytobot (IFCB) instrument mounted on the research vessel Svea as well as a parallel IFCB and DNA metabarcoding dataset for 500 water samples spanning the Baltic Sea, Kattegat and Skagerrak. The methodology developed in the project aims to advance plankton research and ecology in general and plankton monitoring in Sweden in particular. It will bridge the gap between imaging and DNA-based diversity data and increase the information output from both approaches.

In the early results of the project, we observe strong correlation between counts from DNA metabarcoding and Imaging FlowCytobot for some taxa. The presentation will introduce the project, and also present early results.

Sensors and AI for automated seabird monitoring.

Jonas Hentati-Sundberg

Department of Aquatic Resources, Swedish University of Agricultural Sciences

Ecological research and monitoring need to be able to rapidly convey information that can form the basis of scientifically sound management. Automated sensor systems, especially if combined with artificial intelligence, can contribute to such rapid high-resolution data retrieval. We have developed a number of automated methods for monitoring of seabirds, a group of animals often monitored for their high conservation value and for being sentinels for marine ecosystem changes. Using video surveillance combined with automated image processing, we are able to continuously monitor breeding seabirds in a semi-natural environment – an artificial cliff in the middle of a large seabird colony on Stora Karlsö, Sweden. We use a deep-learning based object detection algorithm (YOLO) for identifying adult birds, chicks and eggs, and outputs time, location, size and confidence level of all detections, frame-by-frame, in the supplied video material. We demonstrate how object detection can be used to accurately monitor breeding phenology and chick growth. Our automated monitoring approach can also identify and quantify rare events that are easily missed in traditional monitoring, such as disturbances from predators. Further, combining automated video analysis with continuous measurements from a temperature logger allows us to study impacts of heat waves on nest attendance in high detail.

Our automated system produces comparable, and in several cases significantly more detailed, data than those generated from observational field studies. By running in real time on the camera streams, it has the potential to supply researchers and managers with high-resolution up-to-date information on seabird population status. Ongoing further development of the approach includes automated identification and quantification of behaviors, prey deliveries, automated weighing of birds and tracking and identification of individuals from video data through combining track algorithms with computer vision based automated bird ring reading.

Do we need biodiversity in soil biogeochemical models?

Stefano Manzoni

(Stefano Manzoni, Malin Forsberg, Caroline Greiser, Xiankun Li, Pierre Quévieux, Erik Schwarz)

Department of Physical Geography and Bolin Centre for Climate Research, Stockholm University

Process-based mathematical models have emerged as valuable tools to quantify carbon and nutrient cycling in soils. These models are applied to predict soil carbon storage potential (linked to climate mitigation), nitrogen and phosphorus availability (linked to plant nutrition and water quality), emission of greenhouse gases (relevant for climate change projections), and to study the implications of land management on these processes. Biodiversity is typically neglected in these models, as they were developed for prediction of bulk soil properties rather than for assessment of changes in biodiversity and its consequences on biogeochemical cycling. Recognizing that climatic changes affect soil biota and thus soil functioning, new models now capture microbial processes, but we are still far from describing microbial and faunal diversity in soil models coupled to ecosystem and climate models. In this contribution, we

review the state-of-the-art in soil biogeochemical modelling, with a focus on how biodiversity is being integrated (or not) in models. We offer examples on microbial diversity (dynamic microbial communities during decomposition and in response to drying and rewetting), faunal diversity (role of trophic interactions for carbon cycling), and diversity of microbe-plant interactions (mycorrhizal symbiosis). We discuss how recent advances can pave the way to a new generation of models where biodiversity in soil is described in the model equations, but also highlight the challenges and risks of undesired outcomes of this process.

Unoccupied Aerial Vehicles (UAVs) and Deep Learning for Efficient Ecological Tracking of Ant Mounds.

Jérémy Monsimet

(Jérémy Monsimet^{1*}, Sofie Sjögersten², Johan Olofsson¹, Matthias Siewert¹)

¹ Department of Ecology and Environmental Science, Umeå University, Sweden; ² School of Biosciences, University of Nottingham, Loughborough, UK

Unoccupied Aerial Vehicles (UAVs) and deep learning represent new technologies that can provide novel insights in ecological research. High resolution UAV data have alleviated the mismatch between the scale of ecological processes and the scale of remotely sensed data. Ant mounds play key roles in ecosystem functioning, yet their distribution and effects on entire landscapes remain poorly understood in part because their mounds are too small for satellite remote sensing. This research maps the distribution and impact of ant mounds in a 20 ha treeline ecotone. We evaluate the detectability from UAV imagery using a deep learning model for object detection and different combinations of RGB, thermal and multispectral sensor data. We were able to detect ant mounds in all imagery using manual detection and deep learning. However, the highest precision rates were achieved by deep learning using only RGB data which has the highest spatial resolution (1.9 cm) at comparable UAV flight height. This was followed by a model using data-fusion of multispectral and thermal data at 10 cm resolution. The RGB model can surpass manual detection performance with high precision rates (89%) and recall rates (81%). While multispectral data was outperformed for detection, in part due to lower spatial resolution, it allows for novel insights in the ecology of ants and their spatial impact on vegetation productivity using the NDVI. Scaling up, this suggests that ant mounds impact vegetation productivity for 6% of our study area and up to 12.5% of the *Betula nana* vegetation communities with high abundance of ant mounds. Our work demonstrates that ants are an important component of the tree line ecotone with substantial impact on vegetation productivity. Further, we show the powerful combination of UAV multisensor data and deep learning for efficient ecological tracking and monitoring of particular taxa and their spatial impact

aMeta: a computational method for data-driven ancient metagenomic analysis.

Nikolay Oskolkov

Biology Department, Science for Life Laboratory, National Bioinformatics Infrastructure Sweden, Lund University, Lund, Sweden

Analysis of microbial data from archaeological samples is a growing field with great potential for understanding ancient environments, lifestyles, and diseases. However, high error rates have been a challenge in ancient metagenomics, and the availability of computational frameworks that meet the demands of the field is limited. Here, we propose aMeta [1], an accurate metagenomic profiling workflow for ancient DNA designed to minimize the amount of false discoveries and computer memory requirements. Using simulated data, we benchmark aMeta against a current state-of-the-art workflow and demonstrate its superiority in microbial detection and authentication, as well as substantially lower usage of computer memory.

[1] Zoé Pochon, Nora Bergfeldt, Emrah Kırdök, Mário Vicente, Thijessen Naidoo, Tom van der Valk, N. Ezgi Altınışık, Maja Krzewińska, Love Dalen, Anders Götherström, Claudio Mirabello, Per Unneberg and Nikolay Oskolkov, aMeta: an accurate and memory-efficient ancient Metagenomic profiling workflow, *Genome Biology* 2023, 24 (242), <https://doi.org/10.1186/s13059-023-03083-9>

Digital twins of biodiversity; concepts, workflows and data.

Sam Wenaas Perrin

Gjørvoll Center for Biodiversity Analysis, Norwegian University of Science and Technology

Policy makers increasingly are forced to make trade-offs between societal and ecological transformations in order to meet sustainability targets. A comprehensive and up-to-date overview of biodiversity values at informative scales is crucial in enabling data-driven decisions when encountering these trade-offs. Here, we examine pitfalls and methodological gaps to be amended when using GBIF mediated data to create a national-scale digital twin of Norwegian biodiversity as a decision-making tool.

Tools used in decision making should be open, transparent and reproducible. It is therefore imperative that both modelling and data processing frameworks are documented and available for both inspection and subsequent re-use. Whereas there is a rapidly increasing availability of open full-coverage environmental data from remote sensing, full-coverage mapping of relevant biodiversity will not happen in the foreseen future. Also, biodiversity is simply too complex and stochastic to be predicted by functional models based upon first principles. Here we instead rely on statistical modelling to estimate the probability of occurrence of species at sites not covered by observations.

Such models must take care of the inherent taxonomic, spatial and temporal biases of the underlying sampling process of biodiversity data. In order to do this, data with explicit or implicit information on the sampling process is needed. For practical purposes, this means data with quantitative information or presence / absence information, often labelled “sampling-event data” in biodiversity informatics. While this data is available through GBIF, often the examination of metadata and further processing of data can be used to extract more detailed sampling event information. As such, our workflow does not treat GBIF mediated data as a single data source, but rather a collection of datasets each with its own peculiarities and biases. This allows for the application of complex integrated modelling, allowing for more confidence in subsequent decision making.

Evaluating forecasts from alternative citizen science data and fitting methods versus an individual-based model.

Tord Snäll

(Ute Bradter, Arpat Ozgul, Michael Griesser, Kate Layton-Matthews, Jeannette Eggers, Alexander Singer, Brett K. Sandercock, Paul J. Haverkamp, Tord Snäll)

SLU Swedish Species Information Centre, Swedish University of Agricultural Sciences

The aim of the study presented was to evaluate the utility of opportunistic big data from citizen science programmes for forecasting species distributions against forecasts with a model of individual-based population dynamics. We evaluated whether alternative methods for building habitat suitability models (HSMs) based on opportunistic data from citizen science programmes produced forecasts that were consistent with forecasts from two benchmark models: (1) a HSM based on data from systematic monitoring and (2) an individual-based model for spatially explicit population dynamics based on extensive empirical demographic and movement data. We forecasted population numbers and habitat suitability for three realistic, future forest landscapes for a forest bird, the Siberian jay (*Perisoreus infaustus*). We ranked simulated forest landscapes with respect to their benefits to Siberian jays for each modelling method and compared the agreement of the rankings among methods. Forecasts based on our two benchmark models were consistent with each other and with expectations based on the species’ ecology. Forecasts from logistic regression models based on opportunistic data were consistent with the benchmark models if species detections were combined with high-quality inferred absences derived via retrospective interviews with experienced “super-reporters.” In contrast, forecasts with three other widely used statistical methods were inconsistent with the benchmark models, sometimes with misleading rankings of future scenarios. Our critical evaluation of alternative HSMs against a spatially explicit IBM demonstrates that information on species absences critically improves forecasts of species distributions using opportunistic data from citizen science programmes. Moreover, high-quality information on species absences can be retrospectively inferred from surveys of the consistency of reporting of individual species and the identification skills of participating reporters. We recommend that citizen science projects incorporate procedures to evaluate reporting behaviour. Inferred absences may be especially useful for improving forecasts for species and regions poorly covered by systematic monitoring schemes.

Airborne eDNA captures three decades of ecosystem biodiversity.

Per Stenberg

Department of Ecology and Environmental Science (EMG), Umeå University

Conserving biodiversity is a global imperative, yet our capacity to quantify and understand species occurrences has been limited. To help address this challenge, we develop a novel monitoring approach based on deep sequencing of airborne eDNA. When applied to a 34-year archive of weekly filters from an aerosol sampling station in northern Sweden, our methods enabled robust detection of over 2,700 genera across all domains of life and estimates of eDNA catchment areas. We show that reconstructed trends of birds correlate well with external observational data and reconstructed time series of all taxa revealed regional biodiversity declines consistent with contemporary, large-scale transformations of forest composition and structure. This study demonstrates that airborne eDNA can reliably monitor biodiversity and underscore the immense latent potential in the thousands of aerosol monitoring stations deployed worldwide. For more see: <https://www.biorxiv.org/content/10.1101/2023.12.06.569882v1>

Biodiversity models based on multi-agent reinforcement learning.

Claes Strannegård

Gothenburg University

The interplay between animal behavior and ecological systems, encompassing aspects like reproduction, migration, predation, and grazing, significantly impacts population dynamics and biodiversity. Traditional methodologies in ecosystem modeling, both agent-based and population-based, predominantly utilize hand-coded rules to simulate animal behavior. These rules, particularly the if-then-else constructs, form the foundation of the expert systems approach to AI. This approach, however, saw a decline in popularity in the 1980s, following insights that capturing complex interactions in diverse domains by using hand-coded rules is frequently infeasible. More recently, there has been a pronounced shift towards machine learning-based modeling techniques. Despite this evolution in other fields, the expert system approach seems to be lingering on in ecosystem modeling, despite the risk of using fundamentally flawed models.

In my presentation, I propose an approach to ecosystem modeling that leverages machine learning, more precisely multi-agent reinforcement learning, to simulate animal behavior. This approach is encapsulated in a computational framework comprising two components: firstly, an environmental model represented by a grid derived from geographical data, such as a topographic or hydrographic map, where the grid is divided into cells, each having its own set of properties; secondly, a collection of agents that represent individual organisms or groups of organisms. Each agent has a set of properties and a decision function, which is responsible for action selection, based on the agent’s perception of the closest area surrounding it, including other agents and cells. Crucially, this decision function undergoes training through deep reinforcement learning, using a generic reward signal that encourages survival.

I will demonstrate the application of this framework in modeling one terrestrial and one aquatic ecosystem. By altering the populations and land cover classes, one can investigate the local impact of various economic activities — such as fishing, hunting, agriculture, forestry, and urbanization — on biodiversity.

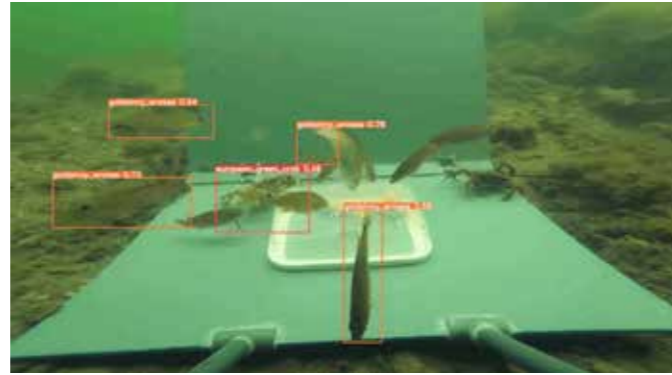
Workshops

Content

1. Research using camera traps.

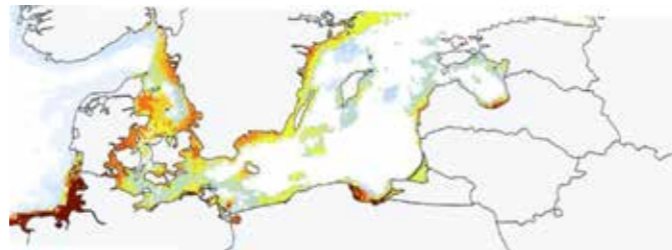
Automatic camera traps linked to AI-based image recognition are rapidly emerging power tools that provide repeatable and standardized sampling methods capable of expanding the extent and resolution of biodiversity monitoring. During the last few years, such traps in Sweden have begun providing data for moths, invasive alien plants and marine hard-bottom habitats. In this workshop we will showcase workflows, discuss the future of such data-driven, image-based approaches in ecology, and explore opportunities for future collaboration.

The workshop will 60 min and moderated by Lars B. Pettersson, Lund university.



2. Modelling the distribution of alien species in Sweden

The number of alien species in terrestrial, marine and freshwater ecosystems is accumulating quickly and requires a systematic response and coordination between public authorities, the research community, and the private sector. Today new monitoring methods for alien species (e.g. eDNA, citizen science, etc) can be coupled with predictive analytical approaches such as Species distribution models to provide biological forecasts for distribution of alien species and invasive hotspots. Currently there are several initiatives which model invasive alien species in Sweden using new and conventional observation methods. The goal of this workshop is to bring these groups together and discuss links to the newly established Digital Twin projects BioDT, DTO-bioflow, and Illiad.



The workshop will be moderated by Matthias Obst, University of Gothenburg.

3. FAIR principles in life science research practice.

NBIS in collaboration with the SciLifeLab Data Centre.

This workshop will showcase how to leverage the FAIR principles to guide effective data management practices for research groups and communities. It is specifically designed to cater to SBDI community members at all career stages and is organized by SciLifeLab Data Centre and NBIS – National Bioinformatics Infrastructure Sweden.

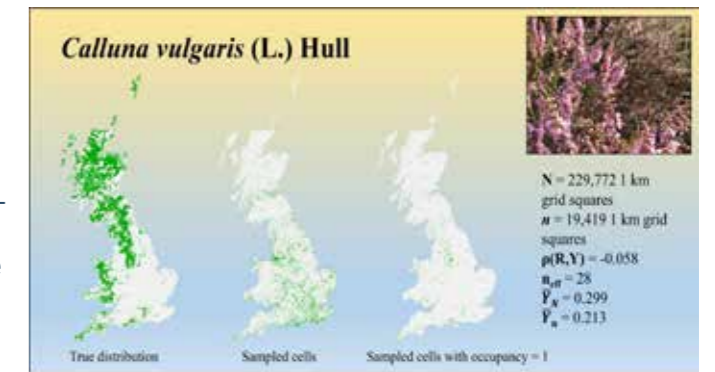
Instructors: Angela Fuentes Pardo, Arnold Kochari, Wolmar Nyberg Åkerström



4. Big data – big sample size? Assessing risk of bias

The effects of selection bias on effective sample size. We will develop on the talk by Rob Boyd (Symposium 1: Perils and promises in big data research) and investigate deeper into why it is challenging to monitor biodiversity without random samples. We will assess risk of bias and investigate consequences for sample size and interpretation of results. See also We need to talk about nonprobability samples, ROBITT: A tool for assessing the risk-of-bias in studies of temporal trends in ecology, Descriptive inference using large, unrepresentative nonprobability samples: An introduction for ecologists, occAssess: An R package for assessing potential biases in species occurrence data.

Workshop held by Oliver Pescott and Rob James Boyd, UK Centre for Ecology & Hydrology.



The effects of selection bias on effective sample size.

Posters

Abstracts

Please use the format 70x100 cm for your poster, which fits best with the boards we have.

Tracking signal crayfish in Lake Vättern.

Patrik Bohman

Department of Aquatic Resources, Swedish University of Agricultural Sciences

In this pilot we investigate how telemetry technology can be used to better study the movements of crayfish in a deep lake. The study is divided into several parts. Firstly, we develop a strategy for attaching transmitters to the crayfish's shell. Secondly, we perform a mortality study on crayfish with and without transmitters. Thirdly, field work is carried out to study how accurate telemetry can be used for crayfish in the lake. We studied several technical details regarding 1) depth, angle and proximity to receivers, 2) if the signal is disturbed when transmitters are hidden in "nesting holes", 3) characteristics of the transmitters (frequency and size etc). Results will be presented in a report later in 2024 which aims to propose a strategy for tracking crayfish in deep lakes.

Comparing ecosystems with and without mycorrhizae.

Malin Forsberg

Malin Forsberg¹, Birgit Wild², Stefano Manzoni¹

¹ Department of Physical Geography and Bolin Centre for Climate Research, Stockholm University, Stockholm, Sweden

² Department of Environmental Science, Stockholm University, Stockholm, Sweden

Symbiotic associations between mycorrhizal fungi, and plants are fundamental for both plant nutrition and carbon (C) exchange between above- and belowground compartments. Plants invest photosynthetically acquired C in symbiotic relations in exchange for nutrients that mycorrhizae mine from organic matter, and this plant-derived C can eventually be stabilized in the soil. This relation is favourable when the mycorrhizal network can access compartments in the soil that the plants cannot either because nutrients are physically protected or stored in chemically recalcitrant organic matter. This nutrient mining is made possible by the decomposition capacity of mycorrhizae and the mycorrhizal network's ability to extend into and explore the soil environment.

Capturing these interactions between the symbiotic partners in process-based models is vital for both quantifying the C and nutrient cycles and understanding how mycorrhizae support ecosystem functions. By including mycorrhizae in ecosystem models, one can study the importance and effects of mycorrhizae in ecosystems. Here we use an ecosystem model to compare systems with and without mycorrhizae. We found higher respiration rates when including mycorrhizae. However, due to the decomposition ability of the mycorrhizae we found a soil C vs plant productivity trade-off in the form of less C in the soil but higher plant biomass when including mycorrhizae.

Autonomous Species Identification Based on Bioacoustic Data to Facilitate Environmental Monitoring.

Dag Glebe

IVL, Swedish Environmental Research Institute

In December 2022 IVL Swedish Environmental Research Institute compiled a review report for the Swedish EPA and the Swedish Agency for Marine and Water Management of how digitalization, such as Artificial Intelligence (AI), can facilitate work with conservation and management of ecosystems. Passive acoustic monitoring (PAM) in bioacoustics can be realized with cheap and simple devices as sound boxes or with collars or implanted devices on individual

animal individuals. Acoustically active animals align themselves with each other and other sounds, which is why soundscapes are considered to provide information about the health of the biosystem. Also, camera traps and other devices for acquisition of image data are frequently used to monitor wildlife. Such devices combined with automatic AI-based analysis systems opens up new possibilities, which were reviewed in the report. The study focus was on new Machine-Learning technologies that have been applied in various projects within the fields of ecosystems and biodiversity, their application, and their success.

The most explored area within bio-acoustic AI has been bird classification, but great progress is made in most fields today. The most common AI methods for resource-demanding listening and image review are employing Deep Learning and, particularly, convolutional neural network, (CNN) but new variants are constantly being developed. The study reviewed various CNN applications and approaches, such as using different realisations of spectrograms (image representations of sound) as input to CNN-models.

The study scope included: 1) Methods not requiring manual classification of training data, 2) Annotation scarcity of training data, 3) Usage of open data and open design, 4) Active learning methods, i.e. methods where experts actively participate in the learning process, quickly produce powerful results, 5) Automatic annotation based on animal behaviour. This provides a model representing the animals' own perception, but must be used with caution.

New climate data at organism-relevant scales.

Caroline Greiser

(Caroline Greiser (Stockholm University, Sweden) – presenting author; David Klinges (University of Florida, USA); Maroof Hamid (University of Kashmir, India); Ilya Maclean (University of Exeter, UK); Ivan Nijs (University of Antwerp, Belgium); Jonathan Lenoir (CNRS, Université de Picardie Jules Verne, France); Juha Aalto (Finnish Meteorological Institute; University of Helsinki, Finland); Julia Kemppinen (University of Oulu, Finland); Koenraad Van Meerbeek (KULeuven, Belgium); Liesbeth van den Brink (University of Concepcion, Chile); Martin Kopecký (Institute of Botany of the Czech Academy of Sciences, Czech Republic); Mick Ashcroft (University of Wollongong, Australia); Miska Luoto (University of Helsinki, Finland); Pieter De Frenne (Ghent University, Belgium); Rachel Penczykowski (Washington University in St. Louis, USA); Rémy Beugnon (German Center for Integrative Biodiversity Research - iDiv, Germany); Eva Gril (Université de Picardie Jules Verne, France); Stijn Van de Vondel (University of Antwerp, Belgium); Stef Haesen (KULeuven, Belgium); Jonas Lembrechts (University of Antwerp, Belgium); Matej Man (Institute of Botany of the Czech Academy of Sciences, Czech Republic))

We present SoilTemp, a geospatial database compiling soil and near-surface air temperature data from all terrestrial habitat types over the world. Temperature-organism relationships lie at the heart of understanding, predicting and mitigating climate change impacts on local, regional and global biodiversity. Mechanistic understanding of temperature-driven processes and patterns like growth, performance and distribution of species requires climate data at organism-relevant scales. Standard weather stations record temperatures and other atmospheric parameters approximately at 2 meters above a flat, open surface, and most gridded climate data only represent long-term average thermal conditions at coarse spatial resolutions. Consequently, factors influencing the proximal climate experienced by organisms within their habitats and at fine spatial resolutions, such as topography, vegetation, snow, or soil characteristics, are often overlooked. These factors, which modify exposure to solar radiation, moisture, and wind, have a profound influence on climate from an organism's perspective. Resultantly, microclimate data are essential and unavoidable for realistic forecasts of the response of biodiversity and the functioning of ecosystems in the face of anthropogenic climate change. High-resolution local microclimate data (e.g., air and soil temperature, humidity, light intensity) from small consumer-grade sensors are becoming increasingly available and are now being compiled in the global SoilTemp database (<https://www.soiltempproject.com>). Currently housing 70,000 temperature logger time series, often complemented with local vegetation data, from 80 countries across all major biomes, the database continues to expand. We present examples of database products, including global soil temperature maps and European forest microclimate maps, as well as highlight key studies on how microclimate is improving our mechanistic understanding of organism and ecosystem responses to climate variation and change. SoilTemp is a large collaborative venture by the scientific community for the scientific community serving as a shining example of how efforts can be pooled to address the climate and ecological crisis.

Measuring biodiversity from molecules to space; Advantages and pitfalls using large-scale DNA based analyses.

Ann Micaela Hellström
MIX Research Sweden AB

TBA

DNA-metabarcoding of marine phytoplankton.

Bengt Karlson

(Bengt Karlson*, Agneta Andersson, Anders F. Andersson, Sonia Brugel, Mikael Hedblom, Krzysztof Jurdzinski, Meike A.C. Latz, Markus Lindh and Anders Torstensson)
SMHI / Swedish Meteorological and Hydrological Institute, Research & Development, Oceanography

Phytoplankton constitute the base of the marine food web and is used worldwide to assess the quality status of marine and freshwater systems. There are long time series of data from monitoring programs, where phytoplankton are analyzed using microscopy. DNA-based methods are quickly emerging. The aim of this project was to develop a practical and robust method to analyze phytoplankton that can be implemented in Swedish marine monitoring programs. The project was performed by conducting sampling during the regular national marine monitoring cruises and for a period of one year (2019-2020). We collected >250 seawater samples for DNA metabarcoding and for analyses of phytoplankton using microscopy. Nineteen stations were sampled, spanning from the Bothnian Bay in the northern Baltic Sea to the Skagerrak. Best practices were developed for field sampling, DNA extraction, sequencing, in bioinformatics and taxonomic annotation. We also produced a system for data handling as part of the SBDI and at the National Oceanographic Data Centre at SMHI. The target genes were 18S (V4 region) and 16S rRNA. Metabarcoding revealed a previously unknown diversity of plankton, especially among the pico- and nanoplankton, small organisms that are difficult to discriminate based on morphology using light microscopy. Environmental variables were investigated as structuring factors of the plankton communities and salinity was identified as an important factor. One part of the project involved comparing results from DNA metabarcoding and microscopy. In general DNA metabarcoding results showed higher biodiversity measures than microscopy, but some taxa were overlooked using metabarcoding. We found that the reference databases have shortcomings and need to be further developed. The project was jointly funded by the Swedish Agency for Marine and Water Management and the Swedish Environmental Protection Agency.

Determining the state of habitat types in Swedish freshwater habitats with eDNA methods – lessons from a pilot project.

Karl Lundén

Department of Forest Mycology and Plant Pathology, Swedish University of Agricultural Sciences

The EU's biodiversity strategy for 2030 requires member nations to establish comprehensive monitoring of their natural habitats to aid their protection and restoration. Evaluating the status of freshwater habitats in a boreal country like Sweden, where almost one tenth of the surface is covered by aquatic habitats including more than 100,000 lakes is extremely challenging. Therefore, the aim of our project is to design and test methods that allow for efficient assessment of the state and biodiversity of Swedish freshwater habitats using scalable sampling and sample processing techniques.

Known and unknown habitat types of Swedish waters were sampled in a north-south gradient in a pilot project summer 2023. The sampling campaign covered the freshwater biogeographical regions of Sweden.

Lessons of methods of data handling, logistics as well as remaining hurdles to establish eDNA based environmental assessment will be discussed.

SciLifeLab Serve - share applications and machine learning models.

Arnold Kochari
Data Centre, SciLifeLab, Sweden

This poster will present SciLifeLab Serve (<https://serve.scilifelab.se>), a national platform for sharing data science and machine learning/AI applications that is available to Swedish researchers in evolution and biodiversity. The platform is suitable for hosting data science applications such as Shiny and Dash apps. SciLifeLab Serve also allows to serve trained machine learning models to allow predictions based on custom input. Such models can be made available

with a graphical user interface (for example, using Gradio and Streamlit frameworks) or through an API endpoint to allow inferences by the research community and general public. SciLifeLab Serve is free for researchers affiliated to a Swedish research institute and their international collaborators, funded by the Data-Driven Life Science Programme by Knut and Alice Wallenberg Foundation.

Soil microbial carbon use efficiency after rewetting – Results from a global database.

Xiankun Li

(Xiankun Li a,b,*; Ainara Leizeagac,d, Johannes Rouskd, Gustaf Hugelius a,b, Stefano Manzoni a,b)
Ställe

Soil drying and rewetting (DRW) events are becoming more intense due to climate change, which affects the soil carbon balance. A useful proxy of microbial contribution to soil carbon is the microbial carbon-use efficiency (CUE=microbial growth over carbon use), but we do not know CUE recovery after rewetting and how fungal to bacterial dominance affects it. If CUE recovers rapidly, carbon mobilized during DRW can be stabilized. Here, we compiled data from 13 studies containing 123 time-series measuring the growth of the microbial community (G), the sum of the bacterial and fungal growth, and respiration (R) after rewetting. First, we characterized microbial carbon use efficiency (CUE=G/(G+R)) by a single flexible function, then we obtained three metrics (time to maximum CUE, maximum CUE, and cumulative CUE) to characterize CUE dynamics after rewetting. Next, we correlated these three metrics with fungal to bacterial dominance, environmental conditions, and experimental drying and rewetting characteristics. These data collectively showed that CUE dynamics after rewetting were less predictable by the selected environmental variables as well as fungal to bacterial dominance. Instead, CUE dynamics could be explained by microbial CUE before drying and rewetting. These results indicate that CUE responses to DRW may be more dependent on the inherent characteristics of the microbial community or previous states rather than solely on the environmental conditions or drought patterns experienced by the microbes.

A reproducible workflow for the field of integrated species distribution models.

Philip Stanley Mostert

(Philip Mostert; Wouter Koch; Anne Bruls; Ellen Claire Martin; Ragnhild Bjørkås; Robert O'Hara)
Norwegian University of Science and Technology - NTNU

Integration of data is needed to address many of the current threats to biodiversity. There has been an exponential increase in quantity and type of biodiversity data in recent years, including presence-absence, counts, and presence-only citizen science data. Species Distribution Models (SDMs) are frequently used in ecology to predict current and future ranges of species, and are a common tool used when making conservation prioritisation decisions. Current SDM practice typically underutilizes the large amount of publicly available biodiversity data and does not often follow a set of standard best practices. Integrating data types with open-source tools and reproducible workflows saves time, increases collaboration opportunities, and increases the power of data inference in SDMs. Here, we address the discipline-wide call for open science and standards in SDMs by (1) proposing methods and (2) producing a reproducible workflow to integrate different available data types to increase the power of SDMs. Integration of datasets is done through a state-space point process formulation: which combines a process model, describing the actual distribution of the species, with data-specific observation models, describing the data collection process. We provide an R package and guidance on how to accommodate users' diverse needs and ecological questions with different data types available on the Global Biodiversity Information Facility (GBIF), the largest biodiversity data aggregator in the world. Finally, we provide a case study of the application of our proposed reproducible workflow by creating SDMs for vascular plants in Norway, integrating presence-only and presence-absence species occurrence data, climate, and habitat data.

Depth Learning - Using Machine Learning to Estimate Vertical Zonation in a Submarine Canyon in Norther Skagerrak.

Christian Nilsson
Gothenburg University

Vertical zonation in marine environments occurs when sessile species are primarily found at specific depth intervals, dividing the habitat into distinct horizontal layers. The community assemblage and size of these different zones is determined by environmental conditions as well as interactions between organisms. This phenomenon has primarily been studied in intertidal waters, but also occurs at greater depths, beneath the halocline. Here, low spatiotemporal variation in environmental conditions suggest increased ecosystem sensitivity to recent climate change. However,

subtidal environments and their zonation patterns have been understudied historically due to technical limitations. Therefore, in this study we apply machine learning based object detection software to ROV footage to estimate and quantify vertical zonation and community structure along the sides of the Koster submarine canyon. We also plan to investigate possible changes in zonation structure as well as any possible correlations to changing environmental conditions over the past 30 years.

New methods for monitoring invasive alien species.

Matthias Obst

Department of Marine Sciences; Gothenburg University

In this study we tested a range of methods for use in early warning, monitoring, and prediction of non-indigenous species. The goal of these tests was to develop a blueprint for regular genetic surveillance programs in the Sweden, improve data flows from genetic monitoring programs to environmental agencies, and increase consistency in monitoring protocols across regions. Our results suggest that genetic monitoring offers mature and cost-effective methods for NIS surveillance and management, especially when combined with citizen alerts and modelling approaches.

SUBSIM - a national platform for SUBSea IMage analysis

Matthias Obst

Department of Marine Sciences; Gothenburg University

SUBSIM is an open-source platform for subsea image analysis to support research and monitoring in marine ecosystems. The platform provides essential functions for machine learning, data management, digital collaboration, and high-performance computing. These functions enable scientists and the general public to study and understand changes in marine environments. SUBSIM's image analysis services are gradually integrated with national and global e-infrastructures..

Fungal communities in Swedish forest soils.

Carina Josefsson Ortiz

Department of Soil and Environment - Biogeochemistry of Forest Soils, SLU Swediosh University of Agricultural Sciences

The Swedish Forest Soil Inventory monitors the soil across all of Sweden. Since 2015 an extra sample is collected from humus samples for DNA sequencing. We are thus able to gain knowledge of which species of fungi occurs in our forest soils. We use ITS2 markers, read with PacBio SMRT sequencing which are matched to the UNITE database for genus and species identification. A DNA occurrence is defined as a species when it constitutes at least 1% of the total amount of markers in a sample. The results from the fungus inventory are summarized at the webpage svamparisverige.se, where one can explore the data both by taxonomy and environmental data.

Swedish Forest Soil Inventory.

Carina Josefsson Ortiz

Department of Soil and Environment - Biogeochemistry of Forest Soils, SLU Swediosh University of Agricultural Sciences

The Swedish Forest Soil Inventory is a national soil monitoring programme that covers all of Sweden except arable land, urban areas and the high mountains. The inventory plots are organized in square clusters with a more dense sample towards the south of Sweden. Soil samples are collected from up to five soil layers. At the laboratory, the samples are dried and prepared before a number of chemical analyses are made, including soil biodiversity of fungi which is determined by DNA methods.

SciLifeLab Data Centre services and support.

Angela Fuentes Pardo

SciLifeLab Data Centre

Life science research is increasingly becoming not only technology-driven, but also data-driven. SciLifeLab Data Centre is a central unit within SciLifeLab with responsibility for IT- and data management since 2016. Our aim is to strive to make data FAIR, handled according to Open Science standards and reusable. This poster presents the variety

of services we offer to the Swedish research community throughout the life cycle of data, from project planning, data production, data analysis, data sharing, to publishing and reuse of data, where researchers are dependent on advanced data analysis and e-infrastructures.

Biodiversity shifts: Data-driven insights from modern ecology, archaeology, and quaternary sciences.

Francesca Pilotto

(Francesca Pilotto¹, Ershad Gholamrezaie², Rebecka Weegar², Philip Buckland²)

1. Norwegian Institute for Nature Research – NINA (Oslo, Norway).2. Umeå University (Umeå, Sweden).

To understand the implications of past changes in climate, landscape and human activity on contemporary biodiversity patterns, data from modern and palaeoecological studies must be connected. The Strategic Environmental Archaeology Database (SEAD) provides access to big data from archaeology and Quaternary science and is an enormous potential resource for investigating past changes in biodiversity. By linking SEAD to SBDI, past species distributions can be analysed for their implications for landscape and climate change. Recent macroecological research using SEAD/SBDI illustrates trends in Late Holocene anthropogenic landscape change in north-western Europe. Over the past few thousand years, humans have impacted insect biodiversity as much as climate change did after the last Ice Age. This demonstrates that data from archaeology, and the consequences of human activity, are essential for fulfilling the promise of using data driven ecology for guiding future conservation practices in response to climate change.

Machine-learning and artificial intelligence applications in conservation and management of ecosystems

Florent Rumiano

SLU Grimsö Wildlife Research Station, Department of Ecology, Swediosh University of Agricultural Sciences

A machine learning-based image analysis method is currently developed to identify and quantify, at very-high spatial resolution, areas, yield losses and spatial distribution of wildlife damages (i.e., large ungulates, particularly wild boar) on different type of crops (i.e., wheat, oat and grassland) in Sweden. To calibrate and validate the developed method, photogrammetry (i.e., the derivation of 2D orthomosaics and 3D point-clouds) and machine learning (e.g., Random Forest, Support Vector Machine) using optical images (i.e., RGB and infrared) from Unmanned Aerial Vehicle (UAV) are considered. These data will be matched against measurements collected in-situ (i.e., crop damages exact locations and different types, using centimetre-precise global positioning systems) in three different areas (i.e., Boo in Orebrö, Mörkö Island in Stockholms and Christinehof in Skåne). This research is interdisciplinary, being both intensive in quantitative modelling and requiring an ability to understand ecology, technology and agriculture to trigger discussions and decisions among different stakeholders.

Next-generation freshwater biomonitoring: Applications of DNA metabarcoding for community-based ecological assessment.

Joeselle Serrana

Joeselle M. Serrana^{1,a} and Kozo Watanabe^{1, *}

1 Center for Marine Environmental Studies (CMES), Ehime University, Matsuyama, Ehime, Japan. a Present address: Department of Environmental Science (ACES), and the Stockholm University Center for Circular and Sustainable Systems (SUCCeSS), Stockholm University, Stockholm, Sweden.

Freshwater ecosystems face increasing pressures due to habitat degradation, biodiversity loss, increasing rates of extinction, and emerging challenges from anthropogenic stress, e.g., dam impoundment. This warrants reliable, verifiable, and efficient biomonitoring and ecological assessment schemes for the management of freshwater resources and restoration of damaged ecosystems. DNA metabarcoding is a powerful tool at the service of taxonomists and ecologists. However, the development and validation of DNA-based monitoring methods are not trivial and involve many steps, from sample collection to laboratory protocols and bioinformatics. The main goal of this thesis was to explore the application of DNA metabarcoding for freshwater biomonitoring, with a specific focus on river ecosystems impacted by dam fragmentation that have undergone or are currently undergoing restoration programs by sediment management and augmentation. We present here the use of both benthic macroinvertebrates and microbial communities as focal groups for bioindicator assessment of freshwater river ecosystems.

Biodiversity Domain Knowledge for Digital Twins.

Laura Slaughter

(Laura Slaughter & Arne Berre)
SINTEF, Norway

Digital twins can be used to answer a variety of question types: descriptive, predictive, and prescriptive (what-if). The types of models used in them can be informed by domain knowledge that must be synthesized, curated, and continuously made available as updated, accurate knowledge. Biodiversity domain knowledge for environment/climate digital twins is available from numerous sources and comes in a variety of formats, both structured and non-structured. These sources can include local (country) databases for lookup of scientific names, red lists, alien species assessments, sites like GloBI (biotic interaction data), as well as the body of scientific literature, containing mostly text in the form of pdfs. Various projects and organisations provide resources that must be leveraged; we need to consider GBIF data standards and a range of available ontologies including those produced by ILIAD, DTO-Bioflow, and AD4GD. This poster is a map of the landscape surrounding biodiversity domain knowledge for twins, highlighting the advantages of using semantic technologies.

Including herbarium specimens in molecular ecology.

Yannick Woudstra

(Yannick Woudstra^{1,2,3}, Niels C.A.M. Wagemaker⁴, Slavica Ivanovic¹, Tanja Slotte³, Barbara Gravendeel^{2,4}, Koen J.F. Verhoeven¹)

1 Netherlands Institute of Ecology (NIOO-KNAW), Wageningen, The Netherlands. 2 Naturalis Biodiversity Center, Leiden, The Netherlands. 3 Stockholm University, Department of Ecology, Environment & Plant Sciences (DEEP), Stockholm, Sweden. 4 Radboud University Nijmegen, Institute for Biological & Environmental Sciences (RIBES), Nijmegen, The Netherlands.

Natural history collections are treasure troves of information, providing a window into the biodiversity of the past. The genetic information encapsulated in historical specimens, albeit in fragmented form, can provide valuable data on genetic diversity trends from past to present populations. With the advent of target capture sequencing, accessing this information has become cost-effective. To find out which target loci are best suited for genotyping below the species level, we examined the limits of intraspecific variation with a versatile target capture tool. We contrasted the genetic variation captured by conserved nuclear exons (comparable to universal target capture tools) with random nuclear loci (comparable to GBS or RAD-Seq) and variable nuclear introns. We tested this in clonal lineages of dandelions (*Taraxacum officinale* s.l.), in which individual members only differ by a few accumulated mutations. A control was performed with the identification of herbarium specimens belonging to the same clonal lineages. We found that conserved nuclear exons best characterise differences between clonal lineages, comparable to distinct genotypes in a population. Random nuclear loci and variable nuclear introns are better at capturing variation within clonal lineages, comparable to allelic diversity within an individual. Universal target capture tools may therefore be suitable for population-level molecular ecology, enabling the inclusion of herbarium material on a large scale.

BAGS-SHINY: a web-based interactive tool for exploring the Baltic Sea microbial gene set.

Luis Fernando Delgado Zambrano

(Luis F. Delgado^{*1}, Marco Vicari¹, Simon Kebede Merid², Gilbert Osen³, Samah Abousharieha⁴, Matthias Labrenz⁵, Lasse Riemann⁶, Jarone Pinhassi⁷, Anders F. Andersson¹)

1 Department of Gene Technology, Science for Life Laboratory, School of Engineering Sciences, in Chemistry, Biotechnology and Health, KTH Royal Institute of Technology, Stockholm, Sweden. *Presenting author. 2 Department of Clinical Sciences and Education, Södersjukhuset, Karolinska Institutet, Stockholm, Sweden. 3 Department of Infectious Diseases, University of Gothenburg, Gothenburg, Sweden. 4 Department of Materials Science and Engineering, Biomedical Engineering, Uppsala University, Sweden. 5 Department of Biological Oceanography, Leibniz Institute for Baltic Sea Research Warnemünde (IOW), Rostock, Germany. 6 Marine Biological Section, Department of Biology, University of Copenhagen, Copenhagen, Denmark. 7 Centre for Ecology and Evolution in Microbial Model Systems - EEMiS, Linnaeus University, Kalmar, Sweden.

High-throughput sequencing has led to the establishment of the field of metagenomics, allowing the direct analysis of genetic material contained within an environmental sample. This approach offers a detailed characterisation of complex microbial communities without needing cultivation. It can address questions like which microorganisms are

present, what they can do, and how they interact. Biome-specific gene catalogues have been recovered for many environments using shotgun metagenomics, followed by assembly and gene calling on the assembled contigs. Notable examples include the integrated reference catalogue of the human microbiome and the Tara Oceans gene catalogue. These catalogues facilitate novel gene function and variant discovery. They also serve as genomic frameworks for mapping sequencing reads and mass-spectrometry spectra, allowing swift and accurate taxonomic and functional profiling. However, the assembly process, crucial for gene recovery, poses challenges. Current methods involve assembling individual samples or combining reads from multiple samples (co-assembly), each with distinct advantages and limitations. We recently proposed a novel mix-assembly strategy (Delgado et al. Microbiome 2022), combining individual and co-assembly approaches, and used this approach to assemble a vast, non-redundant gene set from 124 Baltic Sea metagenome samples. The Baltic Sea Gene Set (BAGS v1.1) comprises 66 million functionally and taxonomically annotated genes. To enable interactive exploration of this gene catalogue we have developed an RShiny application, BAGS-SHINY, that allows users to perform searches by sequence similarity (BLAST), taxonomy and/or function. The gene catalogue and application will serve as valuable tools for exploring microbial gene functions in brackish ecosystem.

We are EOOSC – Building the web of FAIR data and services for science, research and innovation in Europe

Wolmar Nyberg Åkerström

NBIS - National Bioinformatics Infrastructure Sweden, Uppsala University

The European Open Science Cloud (EOOSC) is a European-wide effort towards a large-scale, cross-disciplinary and federated network of infrastructures. Part of the vision is to realise a web of seamlessly interconnected services that take advantage of FAIR data and other research outputs to support science, research and innovation. This poster will present EOOSC and a selection of opportunities to engage with European projects and the EOOSC Association to strengthen the SBDI community's voice and impact.

Practical information

Venue

The conference will take place at the Swedish Museum of Natural History, Frescativägen 40, Stockholm (<https://maps.app.goo.gl/w8ibTD1T4uqKLJDk7>). The public museum is currently being renovated and is closed for the public. The museum is largely inaccessible and the main entrance cannot be used. Instead you will use the restaurant entrance on the courtyard. See images below for directions or use the provided google map link above. The conference is mainly held in 'Stora Hörsalen'.



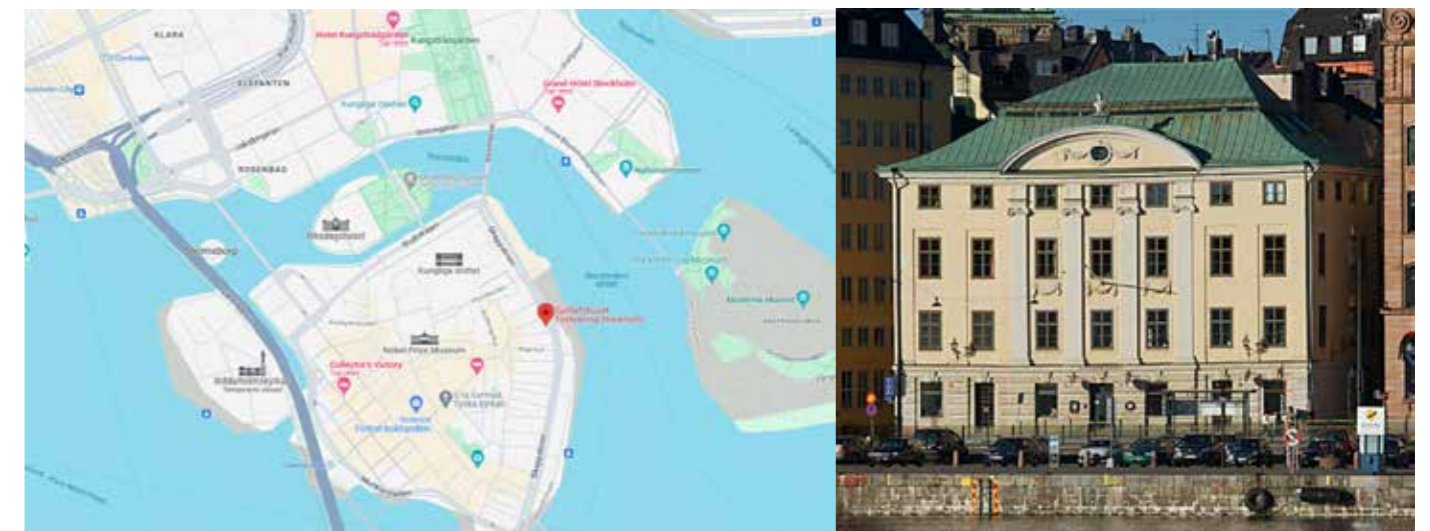
Restaurant Fossilen

Lunch and 'fika' during morning and afternoon sessions will be served in the museum restaurant, Fossilen, which is just next to the conference room. During lunch, two rooms will be reserved in the restaurant for conference participants. The tables will be set so you can go directly to kitchen counter where food is distributed. We will serve two vegetarian options for lunch. The kitchen is informed about registered allergies so indicate your preferences to them. Please respect the table reservations as there will be external restaurant visitors as well.

Conference dinner

The conference dinner will be held at Sjöfartshuset (the Seafarer house), Skeppsbron 10, in Gamla stan (old town), for those who have registered to participate. <https://maps.app.goo.gl/XHgEdeZSXHKgd6QT9>. You can walk there from the Metro stations: Gamla stan or Kungsträdgården.

If you are unable to join the dinner but have registered, please let us know, as there is a waitlist to join.



Getting around

Stockholm has a well working and extensive public transport system. The easiest way to go to the conference venue is to go by the metro to the station called Universitetet. It is the red line towards Mörby Centrum. You can pay directly at the metro barriers with any credit/debit card. There is also an application for buying tickets called [Travis](#).

From the airport it is easiest to go by Arlanda Express. Follow yellow markings on the floor and make sure you buy a ticket before getting on the train. If you prefer to go by taxi, make sure to choose a fixed fee and a taxi with logos printed on the cars. Examples are Taxi Stockholm, Taxi Kurir or Taxi 020.



SBDI Days 2024: Towards Data-driven Ecology January 24-25

hosted by:



SWEDISH MUSEUM OF
NATURAL HISTORY

SBDI is made possible by contributions from its partners and financiers:

