



## Zürich-Basel Plant Science Center

-----  
Contact:

Dr. Barbara Templ

barbara.templ@usys.ethz.ch  
-----

This summer school  
receives funding from  
the swissuniversities  
innovation program  
under the P-8 "Stärkung  
von Digital Skills in der  
Lehre" grant.

swissuniversities  
-----

# PSC Summer School 2022

Application of Machine Learning in Plant  
Sciences

12-16 September 2022

Einsiedeln, Switzerland

# Introduction

## Welcome to Application of Machine Learning in Plant Sciences

10<sup>th</sup> Summer school of the Zurich-Basel Plant Science Center

Technological developments have advanced measurements on plants, leading to routine production of large and complex datasets. This has led to increased efforts to extract meaning from these measurements and to integrate various datasets. At the same time, machine learning has rapidly evolved and is now widely applied in science in general and in plant sciences.

Machine Learning (ML) is a type of Artificial Intelligence (AI) that allows software applications to become more accurate at predicting outcomes without being explicitly programmed to do so. Nowadays, ML algorithms are successfully employed for classification, regression, clustering, or dimensionality reduction tasks of large sets of especially high-dimensional input data. One subset of ML algorithms, deep learning has revolutionized image recognition.

The Summer school aims to facilitate the understanding of the concept of machine learning, to get insights in what it can do and what it can be used for demonstrated by best practice examples.

Participants will also learn, how similarly and differently they are used in different disciplines, for instance in soil science, ecology, biodiversity, agriculture, plant breeding and plant pathology.

The program of the study week will include sessions on the fundamentals of ML, applications of deep learning, ML in plant breeding, ML in ecology and soil sciences and ML in agriculture.

Theoretical inputs to understand the concepts and methodologies, hands-on programming session, exemplary insights in good practices and working on a hackathon educational challenge will help the participants to understand and apply machine learning in various areas of plant sciences.

Invited national and international speakers will make presentations on the topic of their research, give insight into their research field and conduct interactive programming sessions. The founder of AICrowd will act as mentor in the hackathon challenge.

# Organization of Student Work

## Learning objectives

By the end of the summer school, participants will:

- Understand the fundamentals of machine learning and its application in plant science research.
- Understand the basics of deep learning and several applications of it various fields including plant phenotyping, plant identification, detecting plant disease or deforestation from images
- Know how machine learning is applied in plant breeding, macro-ecology and soil science
- Hear about agricultural robotics and how drones are applied in smart farming
- Be able to explore and run machine learning code with R and/or python
- Can implement various machine learning methods in their research
- Know what a Hackathon challenge is and practice their skills with a real-life problem

---

*Number of  
participants: 20  
ECTS: 2*

---

---

*Hackathon  
participants will  
work in teams  
during the  
Summer school  
on a challenge. A  
drone will be  
awarded to the  
best solution.*

---

### Before Summer school

- Application for the Summer school should include description of motivation, background (including experience in programming (n. of years, type of language)) and description of own research project.
- Preparatory reading: Students will have to read the assigned literature before the Summer school.
- Preparatory computer set-up: Students must install R and python before the summer school (software available for free).
- All Summer school participants are expected to present a poster of their research at the beginning of the Summer school.

### During Summer school

- Sessions are composed as lectures, related hands-on programming sessions and case-study examples.
- Teamwork will be done on a Hackathon challenge, individual working time on this is expected to be about 10 h.
- Hackathon presentation on day 5 is 5-10 min per team.

### Group enrollment

- Enrollment to the case studies and group work at OLAT learning platform. Link will be provided after registration.



Day 1 (Monday, 12.09.2022)

## The fundamentals of machine learning

Time	Title	Speaker
9:15-9:30	Welcome	Organizer
9:30-12:30	<b>Hands-on programming session:</b> The basics of python programming	Dr. Carol Alexandru
12:30-13:30	<i>Lunch</i>	
13:30-15:00	The fundamentals of machine learning	Prof. Manuel Günther
15:30-17:00	<b>Hands-on programming session:</b> The fundamentals of machine learning using python	Prof. Manuel Günther
17:00-17:30	Introduction to the Hackathon work	Sharada P. Mohanty
18:00-19:00	<i>Dinner</i>	
19:30-	Poster session & Welcome Aperitif	

Day 2 (Tuesday, 13.09.2022)

## Applications of deep learning in plant sciences

Time	Title	Speaker
8:00-9:00	Analysis of vegetation parameters at global scale with deep learning	Prof. Jan Dirk Wegner
9:00-10:30	The fundamentals of deep learning	Prof. Fernando Perez Cruz
11:00-12:30	<b>Hands-on programming session:</b> Applications of deep learning using python	Manuel Knott
12:30-13:30	<i>Lunch</i>	
13:30-14:00	Deep learning for plant identification	Dr. Michael Rzanny
14:10-14:40	-	
14:45-15:15	Detecting deforestation from satellite images using deep learning	Dr. Thales Sehn Körting
17:00-17:30	Introduction to the Hackathon challenge	Sharada P. Mohanty
18:00-19:00	<i>Dinner</i>	
19:00-	Hackathon work	

Day 3 (Wednesday, 14.09.2022)

## Applications of machine learning in plant breeding

Time	Title	Speaker
8:00-08:30	Deep learning for plant genomics and crop improvement	Dr. Hai Wang
08:30-09:00	Identification of novel short protein coding genes in prokaryotes by proteogenomics – implications for biocontrol	Dr. Christian Ahrens
09:00-09:30	Machine learning in plant -omics data	Dr. Aalt-Jan van Dijk
09:30-10:30	<b>Hands-on programming session:</b> Predicting plant gene interactive networks using python	Dr. Aalt-Jan van Dijk
11:00-11:30	Image-based plant phenotyping using deep learning	Dr. Gert Kootstra
11:30-12:30	<b>Hands-on programming session:</b> Image-based plant phenotyping	Dr. Gert Kootstra
12:30-13:30	<i>Lunch</i>	
13:30-14:30	Interpretable machine learning: Examples and practices	Prof. Shinhan Shiu

15:15-16:30	<i>Excursion: Guided tour at the Milch Manufaktur Einsiedeln</i>	
18:00-19:30	<i>Barbecue</i>	
19:30-	Evening free / Hackathon work	

Day 4 (Thursday, 15.09.2022)

## Applications of machine learning in ecology and soil sciences

Time	Title	Speaker
8:00-09:00	Data mining and machine learning in macro-ecological research	Prof. Niklaus E. Zimmermann
09:00-09:45	Using machine learning to predict ecosystem-atmosphere fluxes	Prof. Benjamin Stocker
09:45-12:30	<b>Hands-on programming session:</b> Applying machine learning on eddy covariance data using R	Prof. Benjamin Stocker
12:30-13:30	<i>Lunch</i>	
13:30-14:00	Using machine learning for spatial mapping demonstrated with soil maps	Dr. Madlene Nussbaum
14:00-15:00	<b>Hands-on programming session:</b> Spatial mapping using R	Dr. Madlene Nussbaum
15:30-16:00	Observing and predicting patterns of biodiversity using remote sensing data in a machine learning framework	Dr. Andrea Paz
16:00-16:30	BioDetect: deep learning for biodiversity detection and classification	Dr. Luca Pegoraro
16:30-17:00	Improving biodiversity protection through AI	Dr. Daniele Silvestro
17:00-18:00	Hackathon work	
18:00-19:00	<i>Dinner</i>	
19:30-	Hackathon work	

Day 5 (Friday, 16.09.2022)

## Applications of machine learning in agriculture

Time	Title	Speaker
8:30-09:30	Machine learning and modelling in the context of crop phenotyping	Dr. Andreas Hund
09:30-10:00	Identification of stress, based on machine learning applied to plant electro-physiology.	Dr. Elena Najdenovska
10:00-10:30	Polyploid plants in natura studied by machine learning	Prof. Kentaro Shimizu
11:00-11:30	Can we use machine learning for agricultural land suitability assessment?	Dr. Anders Bjorn
11:30-12:30	Hackathon work	
12:30-13:30	<i>Lunch</i>	
13:30-14:15	Robotic methods for precision agriculture and environmental monitoring	Prof. Stefano Mintchev
14:15-14:30	TraitSpotting: drone-based phenotyping	Dr. Andreas Hund
14:40-14:50	Hackathon team presentations (Team1 and Team 2, 5-5 min)	
14:50-15:00	Hackathon team presentations (Team3 and Team 4, 5-5 min)	
15:00-15:30	Hackathon winner announcement & award ceremony	
15:30-	Wrap up	





# Abstracts

Dr. Christian Ahrens  
Dr. Carol Alexandru  
Dr. Anders Bjørn Møller  
Prof. Fernando Perez Cruz  
Prof. Manuel Günther  
Dr. Andreas Hund  
Dr. Gert Kootstra  
Dr. Thales Sehn Körting  
Manuel Knott  
Prof. Stefano Mintchev  
Sharada P. Mohanty  
Dr. Elena Najdenovska  
Dr. Madlene Nussbaum

Dr. Andrea Paz  
  
Dr. Luca Pegoraro  
Dr. Michael Rzanny  
Prof. Kentaro Shimizu  
Prof. Shinhan Shiu  
Dr. Daniele Silvestro  
Dr. Benjamin Stocker  
  
Dr. Aalt-Jan van Dijk  
Dr. Hai Wang  
Prof. Jan Dirk Wegner  
Prof. Niklaus Zimmermann



# Identification of novel short protein coding genes in prokaryotes by proteogenomics – implications for biocontrol

Dr. Christian H. Ahrens, Agroscope & SIB, Switzerland

Using the latest NGS technologies from Pacific Biosciences and Oxford Nanopore Technologies (ONT), we *de novo* assemble complete genomes of functionally relevant microorganisms isolated from complex microbiomes such as those present on plant surfaces or in the soil. We thereby aim to eventually bring microbiome research into applied practice, e.g., for plant growth promotion or biocontrol of important crop pests.

*Keywords: Biocontrol,  
plant growth promotion,  
de novo genome  
assembly,  
proteogenomics, novel  
sProteins, essential  
genes, public web server*

In stark contrast to the advances in NGS technologies, the accurate prediction of all protein coding genes is still an unsolved issue. In particular, small proteins, which carry out numerous important biological functions (Storz et al. 2014), are often missed as gene prediction tools use arbitrary length cut-offs of 50-100 amino acids and standard proteomics approaches require adaptation to facilitate small protein identification (Ahrens et al. 2022). We have developed a proteogenomics framework for the identification of small proteins in prokaryotic genomes

using tandem mass spectrometry data (Omasits et al. 2017), aiming to enable experimental groups to study these small proteins, some of which have been shown to have a function in biocontrol (Van der Voort et al. 2015). In this talk, I will highlight some aspects of this framework that can be generically applied for novel small protein identification in prokaryotes.

## Literature

Ahrens, C.H. et al. (2022) A practical guide to small protein discovery and characterization using mass spectrometry. *Journal of Bacteriology*, 204 (1): e0035321. <https://doi.org/10.1128/JB.00353-21>

Storz, G. et al. (2014) Small proteins can no longer be ignored. *Annual Review of Biochemistry*, 83 (1), 753-777. <https://doi.org/10.1146/annurev-biochem-070611-102400>

Omasits, U. et al. (2017) An integrative strategy to identify the entire protein coding potential of prokaryotic genomes by proteogenomics. *Genome Research*, 27 (12), 2083-2095. <https://doi.org/10.1101/gr.218255.116>. (\*)

Van der Voort, M. et al. (2015) Genome mining and metabolic profiling of the rhizosphere bacterium *Pseudomonas* sp. SH-C52 for antimicrobial compounds. *Frontiers in Microbiology*, 6: 693. <https://doi.org/10.3389/fmicb.2015.00693>

\* *useful to read before the presentation*

**Christian H. Ahrens** is senior scientist at the Microbial Genomics and Bioinformatics group of Agroscope and group leader at the Swiss Institute of Bioinformatics (SIB, Bioinformatics & Proteogenomics group).

Since obtaining his PhD in the US (1995), he has focused on bioinformatic data analysis, mining and integration of genomics, transcriptomics, proteomics, metabolomics and TnSeq data. With his team, he has developed solutions to seamlessly integrate these data types, to achieve extensive proteome coverage (including a complete membrane proteome), to analyse subcellular localization and to identify novel short ORFs with proteogenomics (e.g., the public iPTgxDB web server). He has exploited the advantages of functional genomics data, as collaborator or lead scientist, in host-symbiont, host-pathogen interaction, biofilm formation and antibiotic resistance, and microbiome studies.

# The basics of python programming

Dr. Carol Alexandru, University of Zurich, Switzerland

During three 50-minute sessions, you will learn the basics of the Python programming language. First, we'll discuss variable types and scoping, conditionals and loops, functions as 1st-class citizens, and how to organize source code. In the second session, we look at data structures, classes, and inheritance. You'll also learn how to read and write files, particularly CSV and JSON. In the third session, you will hear about comprehensions and generators, how to parse program arguments and testing. We'll finish with a broad tour through the available literature, particularly the Python Standard Library (<https://docs.python.org/3/library/index.html>).

To prepare for these sessions, it is recommended that you install Python if it is not already installed (<https://www.python.org/downloads/>), so you can follow along with the examples presented. You may also install a Python IDE (PyCharm is popular these days).

**Carol V. Alexandru** is a senior research associate working with Prof. Dr. Harald Gall in the Software Evolution and Architecture Lab at the University of Zurich. His research focuses on software analysis, quality assessment, visualization, and command line interfaces. He has published in venues such as the International Conference on Software Engineering (ICSE), Foundations of Software Engineering (FSE) and the Empirical Software Engineering Journal (EMSE). He manages and teaches "Informatics 1", the introductory course to programming at the Department of Informatics, UZH.

# Can we use machine learning for agricultural land suitability assessment?

Dr. Anders Bjorn Moller, Aarhus University, Denmark

Farmers generally need to know if their land is suitable for the crops that they plan to grow. An increasing number of studies have proposed machine learning as a means to provide this information. These studies often base their models on land use observations, assuming that farmers mainly grow crops in the most suitable locations.

This work aimed to test this crucial assumption. Firstly, we mapped land suitability for 41 specialty crops in Denmark using Maxent models based on farmers' registrations combined with environmental and socioeconomic covariates. Secondly, we mapped land suitability for the same crops based on the ecological crops requirements (ECOCROP) database and compared the results. The results from the two methods frequently conflicted with each other, highlighting a fundamental discrepancy.

Most importantly, farmers make land use decisions based on growing conditions as well as the socioeconomic context. Models based on land use observations therefore tend to reflect the socioeconomic reality, which sets them apart from models such as ECOCROP, which focus on growing conditions. It is therefore necessary to distinguish these different phenomena with the terms *ecological* and *socioeconomic* suitability, respectively. Otherwise, the resulting ambiguity would incur a serious risk for misunderstandings and poor land use decisions.

## Literature

Estes, L.D. et al. (2013) Comparing mechanistic and empirical model projections of crop suitability and productivity: implications for ecological forecasting. *Global Ecology and Biogeography*, 22(8): 1007-1018. <https://dx.doi.org/10.1111/geb.12034>

Møller, A.B. et al. (2021) Can we use machine learning for agricultural land suitability assessment? *Agronomy*, 11(4). <https://dx.doi.org/10.3390/agronomy11040703>

Overmars, K.P. et al. (2007) Comparison of a deductive and an inductive approach to specify land suitability in a spatially explicit land use model. *Land Use Policy*, 24(3): 584-599. <https://dx.doi.org/10.1016/j.landusepol.2005.09.008>

**Anders Bjorn Moller** holds a strong interest in the natural world and the ways in which it interacts with humanity across space and time. With a background combining archaeology, physical geography and soil science, his research focuses on the use of machine learning to map soil properties, to provide crucial soil information to landowners and decision makers. In this context, he generally works to bring forth, test new methods and applications. Moreover, he uses the same tools to discover and discuss relationships between the soil and the general environment, including human actions and their consequences for society.

# The fundamentals of deep learning

Prof. Fernando Perez Cruz, ETH Zurich & SDSC, Switzerland

In this tutorial, we will cover recent developments in deep learning and how they have changed machine learning research, in which human expertise is only encoded at a meta-level, architecture design, not at a feature design level. This has opened to solving long-standing problems in computer vision and NLP that it is trickling down to many other science domains. We will travel in time from fully connected neural networks to transformer architecture. We will also touch upon the promise and pitfalls of generative deep neural networks.

## Literature

Goodfellow, I., Bengio, Y. and Courville, A. (2016) Deep Learning  
<https://www.deeplearningbook.org> . MIT Press.

**Fernando Perez Cruz** received a PhD. in Electrical Engineering from the Technical University of Madrid. He has been a member of the technical staff at Bell Labs and a Machine Learning Research Scientist at Amazon. Fernando has been a visiting professor at Princeton University under a Marie Curie Fellowship and an associate professor at University Carlos III in Madrid. He held positions at the Gatsby Unit (London), Max Planck Institute for Biological Cybernetics (Tuebingen), and BioWulf Technologies (New York). Since 2022, Fernando is the Deputy Executive Director of the SDSC. Furthermore, he is Professor at Dept. of Computer Science, ETH Zürich.

# The fundamentals of machine learning

Prof. Manuel Günther, University of Zurich, Switzerland

In the theoretical session, we will discuss which kinds of concepts of machine learning exist and why they are usually better than classical methods. Brief introductions to some basic and more advanced machine learning techniques will be provided to solve the problems of linear regression, dimensionality reduction, clustering, binary classification, categorical classification and sequence modeling. Also, we will showcase some examples of data preprocessing and feature extraction, which enables to use these methods on different kinds of data. Finally, we will discuss some evaluation metrics to assess how well our machine learning model has learned the desired task.

*Keywords: unsupervised  
learning, supervised  
learning, feature  
extraction, evaluation  
metrics*

In the hands-on session we will make use of our acquired knowledge and implement a few data processing, feature extraction and machine learning methods. This includes examples for linear regression, clustering and categorical classification, using python and the scikit-learn python module. We will provide Jupyter notebooks and run our code in Google Colab including some instructions that we will fill together. For preparation, students should make sure to have a google account and register at Google Colaboratory (<https://colab.research.google.com>) with a free (student) account.

## Literature

Andriy Burkov (2019). The Hundred-Page Machine Learning Book. ISBN 199957950X (\*)

\*you can download from OLAT

**Manuel Günther** is assistant professor for Artificial Intelligence and Machine Learning at the Department of Informatics of the University of Zurich. He has studied Computer Science with a major in Machine Learning and minor in Math at the Technical University of Ilmenau, Germany. His main interests are on machine learning for image processing, which he uses in topics like automatic face recognition and open-set classification. His current focus is on the improvement of deep learning systems in order to achieve these tasks.

# Machine learning and modelling in the context of plant phenotyping

Dr. Andreas Hund, ETH Zurich, Switzerland

Crop phenotyping aims to quantify traits like photosynthesis, development, architecture, biomass or quality of crops using a broad variety of sensors and analysis procedures. The classical analysis workflow requires as sequence of machine learning techniques starting with feature extraction from images or hyperspectral data, followed by the modelling of the dynamic development of these features throughout critical phases of crop development. The major aim is to train models enabling breeders or decision makers to select the best varieties given the targeted production environment. I will explain the different steps of the modelling workflow giving the example of our wheat experiments in the field phenotyping platform of ETH Zurich ([kp.ethz.ch/FIP](https://kp.ethz.ch/FIP)).

## Literature

Roth, L. et al. (2021) Phenomics Data Processing: A Plot-Level Model for Repeated Measurements to Extract the Timing of Key Stages and Quantities at Defined Time Points. *Field Crops Research*, 274, 108314. <https://doi.org/10.1016/j.fcr.2021.108314>.

Zenkl, R. et al. (2022) Outdoor Plant Segmentation with Deep Learning for High-Throughput Field Phenotyping on a Diverse Wheat Dataset. *Frontiers in Plant Science*, 12: 774068.

<https://doi.org/10.3389/fpls.2021.774068>

David, E. et al. (2021) Global Wheat Head Detection 2021: An Improved Dataset for Benchmarking Wheat Head Detection Methods. *Plant Phenomics*, 2021, 1–9. <https://doi.org/10.34133/2021/9846158>.

Pérez-Valencia, D. M. et al. (2022) A Two-Stage Approach for the Spatio-Temporal Analysis of High-Throughput Phenotyping Data. *Scientific Reports*, 12 (1), 3177. <https://doi.org/10.1038/s41598-022-06935-9>.

**Andreas Hund** is senior scientist in the group of Crop Science at ETH Zurich. His research focus is the adaptation of crop varieties to a changing climate. By evaluating crop development under field conditions, he aims to i) achieve a deeper understanding about the genetic control of developmental processes during critical phases of crop development and ii) to develop genetic crop models to predict the performance of new breeding lines in a certain target environment and climate scenario. To achieve this, the group monitors wheat throughout its development using the Field Phenotyping Platform (FIP) and the [Phenofly](#) infrastructure.



# Image-based plant phenotyping using deep learning: moving into the 3rd dimension

Prof. Gert Kootstra, Wageningen University and Research, The Netherlands

In the theoretical session, I will present some of the ongoing research at Wageningen University and Research on the topic of image-based plant phenotyping. We will start by looking at deep-learning approaches for two-dimensional (2D) image processing and then move into the field of three-dimensional (3D) point-cloud processing to get more information about the 3D structure of plants. The main challenges will be discussed as well as approaches about how to tackle these challenges.

*Keywords: image-based  
plant phenotyping, deep  
neural network, 3D*

In the hands-on session, you will learn how a deep neural network can be used to learn segment a 3D point-cloud of a plant into leaf and stem/branch points. You will get a python notebook with information and exercises, which you can run Google Colab. No prior software installation is needed. You only need a Google account.

## Literature

- Boogaard, F. P. et al. (2020) Robust node detection and tracking in fruit-vegetable crops using deep learning and multi-view imaging. *Biosystems Engineering*, 192, 117–132. <https://doi.org/10.1016/j.biosystemseng.2020.01.023>
- Jia, N. et al. (2020) Segmentation of body parts of cows in rgb-depth images 2 based on template matching. *Computers and Electronics in Agriculture*, 180: 105897. <https://doi.org/10.1016/j.compag.2020.105897>
- Boogaard, F. P. et al. (2022). Improved point-cloud segmentation for plant phenotyping through class-dependent sampling of training data to battle class imbalance. *Frontiers in Plant Science*, 13: 838190. <https://doi.org/10.3389/fpls.2022.838190>

**Gert Kootstra** is assistant professor at the Farm Technology group of Wageningen University and Research (WUR). His expertise is in the area of computer vision and robotics with applications in agriculture and food production. Gert studied Artificial Intelligence at the University of Groningen, where he also did his PhD on human and robotic vision. He then moved to the Royal Institute of Technology (KTH) in Stockholm, focusing on computer vision for robotic grasping. Since 2012, Gert works at WUR where he established a group of 15 PhD students and 5 postdocs working on machine vision related to various application domains in agriculture. He coordinates a Dutch research program, FlexCRAFT, developing novel technologies to advance agri-food robotics.

# Detecting deforestation from satellite images using deep learning

Dr. Thales Sehn Körting, National Institute for Space Research, Brazil

The aim of this talk is to describe some of the Remote Sensing initiatives of INPE related to deforestation mapping. Before describing deep learning approaches, some basics of optical

*Keywords:* remote sensing will be covered, including time series and data cubes, in order to provide the baseline for actual techniques. Some initiatives currently applied at INPE, related to deep learning, will be described, including the challenges related to the different biomes present in Brazil.

## Literature

Maretto, R. V. et al. (2021) Spatio-Temporal Deep Learning Approach to Map Deforestation in Amazon Rainforest. *IEEE Geoscience and Remote Sensing Letters*, 18 (5). <https://doi.org/10.1109/LGRS.2020.2986407>

Neves, A. K. et al. (2021) Hierarchical mapping of Brazilian Savanna (Cerrado) physiognomies based on deep learning. *Journal of Applied Remote Sensing*, 15 (4): 044504. <https://doi.org/10.1117/1.JRS.15.044504>

De Souza, A. A. et al. (2021) On a Data-Driven Approach for Detecting Disturbance in the Brazilian Savannas Using Time Series of Vegetation Indices. *Remote Sensing*, 13 (24): 4959. <https://doi.org/10.3390/rs13244959>

**Thales Sehn Körting** is Ph.D. in Remote Sensing, with a M.S. degree in Applied Computing (both titles obtained at the Brazil's National Institute for Space Research - INPE), and a B.S. degree in Computer Engineering (at FURG, Brazil). The research areas include remote sensing image segmentation, multi-temporal analysis, image classification, data mining algorithms and artificial intelligence. Thales is currently the president of the Remote Sensing Postgraduate Course Council at INPE, and also Executive Secretariat of The International Charter Space and Major Disasters.

# Application of deep learning using python

Manuel Knott, EMPA, Switzerland

This hands-on programming session will be about Vision Transformers pre-trained with the DINO method. We will use these models to convert images to lower-dimensional representations that enable the usage of shallow machine learning models for downstream tasks such as classification or clustering. This procedure can be superior to other approaches when the number of available training samples is small or when no GPUs for training Deep Neural Networks are available. We will demonstrate the method by classifying fruit quality based on public image data sets. Participants are welcome to experiment with their own data sets during the session. We will work with Python/Jupyter, PyTorch, and scikit-learn.

## Literature

Caron, M. et al. (2021) Emerging properties in self-supervised vision transformers. *arXiv*: 2104.14294. <https://doi.org/10.48550/arXiv.2104.14294>

Dosovitskiy, A. et al. (2022) An image is worth 16x16 words: transformers for image recognition at scale. *arXiv*: 2010.11929v2. <https://doi.org/10.48550/arXiv.2010.11929>

Knott, M. et al. (2022) Facilitated machine learning for image-based fruit quality assessment. *arXiv*: 227.04523. <https://doi.org/10.48550/ARXIV.2207.04523>

**Manuel Knott** received a MSc in Data Science from the University of Amsterdam. He is currently doing his PhD at ETH Zurich, working in a joint position with the Swiss Data Science Center and Empa. His research is about the application of machine learning to reduce food waste in the postharvest supply chain of fresh fruits and vegetables.

# Robotic methods for precision agriculture and environmental monitoring

Prof. Stefano Mintchev, ETH Zurich

Today agriculture faces the unprecedented challenge of feeding a growing population under very tight constraints, such as lack of space and water, new pests, loss of biodiversity, and climate change. These antagonistic pressures require innovative solutions to improve agriculture efficiency and meet the challenge of producing more food with fewer resources.

*Keywords: robotics,  
smart farming,  
environmental  
monitoring*

Robotics has emerged as a promising methodology to address these challenges. Indeed, robots are achieving the high levels of autonomy needed to operate in complex natural environments, characterized by high variability in both crops and environmental conditions. For this reason, robots are becoming a key technology in the transition to smart farming and in supporting the

agricultural needs of the 21st century. In the first part of the presentation, I will explain what a robot is and how it perceives and move in the environment. I will present examples of robots for precision agriculture and discuss limitations and future challenges. In the second part of the presentation, I will highlight the preliminary results of two projects that aim to remotely detect biodiversity and plant volatiles with aerial robots.

## Literature

Bechar, A. and C. Vigneault (2016). Agricultural robots for field operations: Concepts and components. *Biosystems Engineering*, 149, 94–111. <https://doi.org/10.1016/j.biosystemseng.2016.06.014>

Asseng, S. and Asche, F. (2019) Future farms without farmers. *Science Robotics*, 4 (27): eaaw1875. <https://doi.org/10.1126/scirobotics.aaw1875>

**Stefano Mintchev** is assistant professor of Environmental Robotics at ETH Zurich. He received his Ph.D. degree in biorobotics in 2014 at the BioRobotics Institute, Scuola Superiore Sant'Anna, Italy. During his postdoctoral activity at the Laboratory of Intelligent Systems at EPFL, he worked on new design principles, soft materials and manufacturing solutions for aerial robots. In 2018, he co-founded the company Foldaway Haptics, where he acted as CTO until April 2020. He then joined ETH Zurich with an SNSF Eccellenza Professorial Fellowships. He is currently studying robotics applied to today's environmental challenges with the aim of developing concrete robotic solutions to some of the United Nations Sustainable Development Goals.

# Deep learning for image-based plant disease detection

Sharada P. Mohanty, AI Crowd

Abstract and literature to follow.

## Literature

**Sharada Mohanty** is the CEO and Founder of Alcrowd, a platform for crowdsourcing Artificial Intelligence for real world problems. His research focuses on using Artificial Intelligence for [diagnosing plant diseases](#), [teaching simulated skeletons how to walk](#), [scheduling trains in simulated railway networks](#), and on [AI agents which can perform complex tasks in Minecraft](#).

He is extremely passionate about benchmarks and building communities. He has led the design and execution of many large-scale machine learning competitions and benchmarks, such as [NeurIPS 2017: Learning to Run Challenge](#), [NeurIPS 2018: AI for Prosthetics Challenge](#), [NeurIPS 2018: Adversarial Vision Challenge](#), [NeurIPS 2019: MineRL Competition](#), [NeurIPS 2019: Disentanglement Challenge](#), [NeurIPS 2020: Flatland Competition](#), [NeurIPS 2020: Procggen Competition](#), [NeurIPS 2021 NetHack Challenge](#), to name a few.

During his Ph.D. at EPFL, he worked on numerous problems at the intersection of AI and health, with a strong interest in reinforcement learning. In his previous roles, he has worked at the Theoretical Physics department at CERN on crowdsourcing compute for PYTHIA powered Monte-Carlo simulations; he has had a brief stint at UNOSAT building GeoTag-X, a platform for crowdsourcing analysis of media coming out of disasters to assist in disaster relief efforts. In his current role, he focuses on building better engineering tools for AI researchers and making research in AI accessible to a larger community of engineers.

# Identification of stress, based on machine learning applied to plant electro-physiology

Dr. Elena Najdenovska, HES-SO, Switzerland

The work that will be presented involves the application of signal processing and machine learning algorithms on plant electrophysiological data to evaluate the health status of plants by identifying their physiological responses to environmental changes. This work is part of three research projects (funded by Innosuisse and FOAG) conducted with Agroscope, HEIA-FR, and Vivent SA. The main objective is to move towards automated and improved crop monitoring to increase yields while using fewer resources, thus contributing to sustainable agriculture.

## Literature

Najdenovska, E. et al. (2021) Identifying general stress in commercial tomatoes based on machine learning applied to plant electrophysiology. *Applied Sciences*, 11: 5640. <https://doi.org/10.3390/app11125640>

Najdenovska, E. et al. (2021) Classification of plant electrophysiology signals for detection of spider mites infestation in tomatoes. *Applied Sciences*, 11: 1414, <https://doi.org/10.3390/app11041414>

**Elena Najdenovska** is an electrical engineer with a Master's degree in Information Technologies obtained in 2012 from EPFL, followed by a Ph.D. in Life Sciences in 2018 from UNIL. She is currently a postdoctoral researcher at the Health Engineering and Economics (HEE) group at HEIG-VD. Her research mainly focuses on signal and image processing and data analysis related to the life science and medical domain.



# Using machine learning for spatial mapping demonstrated with soil maps

Dr. Madlene Nussbaum, Bern University of Applied Sciences, Switzerland

Machine learning refers to a set of tools to establish models of linear or non-linear relations or other previously unknown relationships in complex data. In soil science it is widely used to create soil maps. Machine learning algorithms based on decision trees are among the most popular techniques. In this block you will get an overview how to create continuous maps

*Keywords: spatial prediction, soil map*

from point observations by using spatial predictions. We will cover potential explanatory covariates, type of responses and how to include spatial location in non-spatial machine learning models.

## Literature

Nussbaum, M. (2022) Machine learning and processing of large data. In: Reference Module in Earth Systems and Environmental Sciences, *Elsevier*, <https://doi.org/10.1016/B978-0-12-822974-3.00065-3>.

Gareth, J. et al. (2017) An introduction to statistical learning: With applications in R. In: Springer Texts in Statistics, 8th edn. New York, Heidelberg, Dordrecht, London: Springer.

**Madlene Nussbaum** is a research associate at Bern University of Applied Sciences with focus on methodological developments for digital soil mapping. Areas of expertise: predictive statistics, spatial prediction with geostatistics and machine learning, large geodata processing, sampling design optimization, communication of prediction uncertainties to stakeholders.

# Observing and predicting patterns of biodiversity using remote sensing data in a machine learning framework: A study case from the Brazilian Atlantic rainforest.

Dr. Andrea Paz, ETH Zurich, Switzerland

Understanding how species are distributed is essential for accurate protection, management and monitoring of biodiversity. For this, it is key to understand the multiple dimensions of biodiversity and their environmental correlates. However, there is little consensus on which environmental variables are best at predicting multiple dimensions of diversity. Using a machine learning framework, I show that diversity (taxonomic and phylogenetic) commonly correlates with environments, despite ecological differences, across nine clades along a single rainforest domain.

*Keywords: species richness, phylogenetic diversity, phylogenetic endemism, remote sensing*

Specifically, five clades of plants (Bromelioideae, Miconieae, Bertolonia, Cambessedesieae, and the Fridericia and allies) and four clades of animals (butterflies in the tribe Ithomiini, frogs in the genera Boana and Proceratophrys, and birds in the subfamily Thraupinae). In my talk, I show that while it is

difficult to establish in situ biodiversity monitoring programs at broad geographical scales, remote sensing advances allow for near-real time Earth observations that can help with this goal. Using data from biological collections and information from repeated field inventories, I show the potential of this machine learning framework to accurately predict trends of biodiversity change for both taxonomic and phylogenetic diversity.

## Literature

Paz, A. et al. (2021) Environmental correlates of taxonomic and phylogenetic diversity in the Atlantic Forest. *Journal of Biogeography*, 48, 1377–1391. <https://doi.org/10.1111/jbi.14083>

Paz, A. et al. (2020) Predicting patterns of plant diversity and endemism in the tropics using remote sensing data: A study case from the Brazilian Atlantic rainforest. In J. Cavender-Bares, J. Gamon, & P. Townsend (Eds.), *Remote Sensing of Plant Biodiversity* (pp. 255–266). Springer, Cham.

Paz, A. et al. (in press). A framework for near-real time monitoring of diversity patterns based on indirect remote sensing, with an application in the Brazilian Atlantic rainforest. In press. Peer J.

**Andrea Paz** is a postdoctoral researcher at the Crowther Lab at ETH Zurich. She is interested in mapping diversity patterns and understanding the relation between environments and the distribution of biodiversity on earth. She is also interested in how remote sensing data can help us improve biodiversity mapping and monitoring. She holds a BSc and MSc in Biology from Universidad de los Andes in Colombia. She studied the environmental and ecological determinants of population genetic divergence in amphibians of Panama.

# BioDetect: Deep learning for biodiversity detection and classification

Dr. Luca Pegoraro, WSL, Switzerland

Biodiversity is declining worldwide at an alarming rate, but our ability to gather critical data from ecosystems has not kept pace. In fact, collecting information on biological interactions, for instance pollination, or on morphological traits, is still largely done by hand.

Technological advances in computer vision and deep learning have tremendous potential to automate ecological data collection, however tools developed for real-world biological scenarios are lacking.

*Keywords detection  
algorithm, computer  
vision, biodiversity*

Here we present two software modules to detect flower-visiting animals and to automate counting and measuring of aquatic macroinvertebrates. We developed a detection algorithm that extracts frames with visitors from field monitoring footage, using a hummingbird-plant interactions database from Ecuador. We are extending this module using more diverse training data, with the goal to

provide a robust general solution for detection of visitors to focal items (i.e., flowers, nests, etc.) in visually cluttered environments.

In another module, we use computer vision to segment images of aquatic macroinvertebrate samples. We leverage simple thresholding algorithms to count individual organisms and apply morphological transformations to compute individual-level phenotypical features (e.g., length, width, area).

Further modules can be developed that predict the identity of organisms (to the level of functional group or taxon) and that can be fine-tuned by the user with their own labelled data. Our aim is to provide the ecology and evolution community with cutting-edge tools to gather more and better data, aiding in monitoring and tackling biodiversity loss.

## Literature

Høye, T. T. et al. (2021) Deep learning and computer vision will transform entomology. *Proceedings of the National Academy of Sciences of the United States of America*, 118 (2), 1–10. <https://doi.org/10.1073/pnas.200254511>

Lürig, M. D. et al. (2021) Computer Vision, Machine Learning, and the Promise of Phenomics in Ecology and Evolutionary Biology. *Frontiers in Ecology and Evolution*, 9(April). <https://doi.org/10.3389/fevo.2021.642774>

**Luca Pegoraro** received a PhD in evolutionary ecology from Queen Mary University of London and the Royal Botanic Gardens, Kew (UK), where he studied the evolution of one of the largest plant families (Asteraceae) in the Alps. That's also when he became interested in technological approaches for ecology & evolution. He is currently a postdoc at the Federal Research Institute WSL, working on implementing hardware (e.g., cameras) and software (e.g., computer vision & deep learning) solutions for studying biological interactions, especially pollination (both hummingbirds in the Neotropics and insects in alpine environments in Switzerland).

# Deep learning for plant identification

Dr. Michael Rzanny, Max Planck Institute for Biogeochemistry, Germany

The mobile App “Flora Incognita” leverages machine learning algorithms to identify wild plant species in the field. Since its release in 2018, more than 20 mio valid observations have been collected by citizen science contributors all over the world (but mostly in Central Europe). In this presentation I will give an overview on the Flora Incognita research project, where we developed a mobile application for automated plant species identification and observation recording (Mäder et al. 2021). The application was downloaded more than 4 million times so far and is capable of automatically identifying almost every central European plant species. During the vegetation season, we collect tens of thousands of geolocated plant observations each day. Our approach combines novel advances in machine learning-enabled identification with a plant species field guide. The image classifier itself is trained on currently more than 1 million plant images on a cluster of GP-GPUs over a period of several months. Accompanied location data allows for additionally utilizing presence–absence maps, occurrence records acquired by our users and other sources of information such as soil type and land cover databases. The identification process itself is realized by a cascade of deep networks that combines the information provided by the image and the associated location data.

Flora Incognita differs from other plant identification systems by taking a multi-modal and interactive approach that not only analyses a single image depicting an unknown plant but queries the user for images of one or more complementary plant organs to deliver a precise identification (Rzanny et al. 2019). Some taxa, for example Poaceae, are exceptionally difficult to distinguish, especially if flowers (which are usually the most discriminative) are not available. In order to improve the identification accuracy of the

*Keywords Flora  
Incognita, identification,  
image classifier*

Flora Incognita app we evaluated several perspectives and identified the most important ones. This allows for accuracies above 90 % even in the absence of flowers (Rzanny et al. 2022). High numbers of accurate plant identifications in combination with their metadata provide an important source of data for, e.g., automated biodiversity monitoring. With this, the research group provides data to better understand biodiversity and its change due to natural and anthropogenic drivers.

## Literature

Rzanny, M. et al. (2022) Image-based automated recognition of 31 Poaceae species: The most relevant perspectives. *Frontiers in Plant Science*, 12: 804140. <https://doi.org/10.3389/fpls.2021.804140>

Mäder, P. et al. (2021) The Flora Incognita app – interactive plant species identification. *Methods in Ecology and Evolution*, 12, 1335–1342. <https://doi.org/10.1111/2041-210X.13611>

Rzanny, M. et al. (2019) Flowers, leaves or both? How to obtain suitable images for automated plant identification. *Plant Methods*, 15: 77. <https://doi.org/10.1186/s13007-019-0462-4>

**Rzanny Michael:** In 2014, the Flora Incognita project was launched, and he joined as a botanist in the interdisciplinary team after finishing his dissertation on the complexity of plant-based ecological networks. His contribution focuses on improving the identification accuracy of the application, for example by running experiments and determining the most successful combination of image perspectives, leading to an improvement of the overall identification accuracy for challenging plant groups.

# Machine learning of transcriptome data in natura: drought as a trigger for flowering in aseasonal tropics

Prof. Kentaro Shimizu, University of Zurich, Switzerland

Plant phenotypes in the regulated chamber conditions are often distinct from those in the naturally fluctuating field environments, or *in natura*. My group has focused on natural polyploid *Arabidopsis* and crop polyploid wheat to understand the advantages and disadvantages of polyploidy. It has long been hypothesized that polyploid plants can be a generalist in changing environments. A major barrier to test such hypotheses is the technology to observe plant phenotypes continuously *in natura*. We recently developed hardware and software for phenotyping cart. To detect morphological and physiological traits, machine learning was employed. The possibility of machine learning to test biological questions will be discussed.

## Literature

Akagi, T. et al. (2022) Polyploidy before and after domestication of crop species. *Current Opinion in Plant Biology*, 69:102255. <https://doi.org/10.1016/j.pbi.2022.102255>.

Kuroki, K. et al. (2022) Development of a high-throughput field phenotyping rover optimized for size-limited breeding fields as open-source hardware. *Breeding Science*, 72 (1): 66-74. <https://doi.org/10.5167/uzh-216875>

**Kentaro Shimizu** is the director of the PhD Program in Evolutionary Biology, a Co-Director of University Research Priority Program of Evolution in Action, deputy director of the Department of Evolutionary Biology and Environmental Studies at the University of Zurich. His PhD study was about molecular genetics of *Arabidopsis* sexual reproduction at Kyoto University, Japan. He employed statistical genetics with molecular genetics as a postdoc at the North Carolina State University, USA. His current focus is evolutionary genomics of natural and crop polyploid species.

# Interpretable machine learning: Examples and practices

Prof. Shin-Han Shiu, Michigan State University, USA

Machine learning is used in almost every daily application in our lives and in scientific research. In plant sciences, there is an ever-increasing number of publications where machine learning models are created to make predictions of, for example, molecular activities, physiology, phenotypes, inter-specific interactions, and ecosystem functions. Beyond making predictions, these models can be interpreted to reveal how and why the models work. Through this understanding, we gain insights on the potential mechanisms so data-informed hypotheses can be generated and tested. In addition, such interpretations may also reveal issues with the data, particularly data quality and biases, which allow us to further improve our models and a better understanding of the biological phenomena of interests.

*Keywords: trends in genetics, interpretability*

In this session, we will discuss examples of machine learning model interpretation methods, their theoretical bases, and their applications in addressing plant science questions. In addition, there will be a hands-on session for using a few interpretation approaches that are specific or agnostic to the machine learning approaches. The goals are for the participants to not only appreciate the importance of model interpretation, but also pave the way for their own research applications using model interpretation methods.

## Literature

Azodi, C B. et al. (2020) Opening the Black Box: Interpretable Machine Learning for Geneticists. *Trends in Genetics*, 36, 442-455. <https://doi.org/10.1016/j.tig.2020.03.005>

Cho, Y-R. and Kang, M. (2020) Interpretable machine learning in bioinformatics. *Methods*, 179: 1-2. <https://doi.org/10.1016/j.ymeth.2020.05.024>

Molnar, C. (2022) Chapter 2 Introduction in “Interpretable machine learning: a guide for making black box models explainable”. [GitHub: <https://christophm.github.io/interpretable-ml-book/> ].

**Shin-Han Shiu** is a professor for plant biology in two academic departments: Plant Biology and Computational Mathematics, Science, and Engineering. His core research questions center on how genomes evolve and contribute to novelty, how to identify functional genomic regions, and how to predict molecular activity, physiological functions, phenotypes, and fitness traits in plants through integrating heterogeneous data with diverse approaches.



# Improving biodiversity protection through AI

Dr. Daniele Silvestro, University of Fribourg, Switzerland

Over a million species face extinction, urging the need for conservation policies that maximize the protection of biodiversity to sustain its manifold contributions to people. Here I present a suite of new methods aimed to help guiding conservation efforts using artificial intelligence. Specifically, we develop machine learning methods to predict the effect of current and forecasted extinction risks on biodiversity and compare future trends with historical extinction trajectories. I present a deep learning

*Keywords: conservation  
planning, extinction risk,  
reinforcement learning*

approach to evaluate the extinction risk across thousands of species, complementing the Red List compiled by the International Union for Conservation of Nature (IUCN). Finally, I will introduce a novel framework for spatial conservation prioritization based on reinforcement learning that consistently outperforms available state-of-the-art software using simulated and empirical data. This model, CAPTAIN (Conservation Area Prioritization Through Artificial INtelligence), quantifies the trade-off between the costs and benefits of area and biodiversity protection, allowing the exploration of multiple biodiversity metrics. Under a limited budget, the model protects significantly more species from extinction than areas selected randomly or naively and meets conservation targets more reliably than alternative software. Artificial intelligence holds great promise for improving the conservation and sustainable use of biological and ecosystem values in a rapidly changing and resource-limited world.

## Literature

Silvestro, D. et al. (2022) Improving biodiversity protection through artificial intelligence. *Nature Sustainability*, 5: 415-424. <https://doi.org/10.1038/s41893-022-00851-6>

Zizka, A. et al. (2021) IUCNN – Deep learning approaches to approximate species' extinction risk. *Diversity and Distributions*, 28 (2): 227-241. <https://doi.org/10.1111/ddi.13450>

**Daniele Silvestro** is an SNF assistant professor in computational biology at the University Fribourg (Switzerland) and associate scientist at the university of Gothenburg (Sweden). His research focuses on the development of new models and software with applications in evolutionary biology, geology, sociology, and conservation biology. His most recent programs include new tools to perform phylogenetic analyses and artificial intelligence approaches to estimate patterns of species richness and guide biodiversity conservation policies. Daniele authored more than 90 scientific publications across specialized and general high-impact journals, including *Science*, *Nature Sustainability*, and *Proceedings of the National Academy of Sciences USA*. He is an associate editor for *Methods in Ecology and Evolution* and *Systematic Biology*. Beside academic research he is an avid fossil hunter and mountain hiker.

# Using machine learning to predict ecosystem-atmosphere fluxes

Prof. Benjamin Stocker, University of Bern, Switzerland

In this workshop, we use ecosystem flux data and parallel measurements of meteorological variables to model ecosystem gross primary production (the ecosystem-level CO<sub>2</sub> uptake by photosynthesis). These data and prediction task is used to introduce fundamental methods of machine learning (data splitting, model training, random forest algorithm) and their implementations in R. After this course, you will ...

- Understand how overfitting models can happen and how it can be avoided.
- Implement a typical workflow using a machine learning model for a supervised regression problem.
- Evaluate the power of the model.
- Visualise results.

## Literature

To be followed.

**Prof. Stocker** received his PhD from the University of Bern in 2013. He then worked at Imperial College London, at ETH Zurich, also at the Centre for Ecology and Forestry, CREAM, in Barcelona, Spain. After a Visiting Researcher at Stanford University in the USA, he took up an Assistant Professorship at ETH Zurich in 2019, endowed with an SNF Eccellenza Fellowship. Since Aug. 1, 2022, he holds the tenure track assistant professorship in Geocomputation and Earth Observation at the Institute of Geography, University of Bern. Prof. Stocker's research focuses on process understanding and prediction of impacts of climate change and extreme meteorological events on terrestrial ecosystems, biogeochemical cycles, and on interactions and feedbacks between the terrestrial biosphere, climate, and land use. To this end, his research group develops process-based and data-driven models and uses them in combination with a variety of data, including satellite-based Earth observations, continuous ecosystem monitoring, field data collections, and atmospheric composition measurements.

# Machine learning on plant -omics data

Dr. Aalt-Jan van Dijk, Wageningen University & Research Center, The Netherlands

My talk will cover how machine learning enables progress in plant science and plant breeding, focusing on applications related to -omics data. Machine learning offers a suite of methods enabling us to find meaningful patterns in relevant plant data. This includes genome data, genetic variation, concentrations and interactions of molecules in the cell such as RNA, proteins, and metabolites, etc. More recently, high-throughput, automated measurements of macroscopic phenotypes have become available, i.e., quantification of the development, morphology, growth, or yield of plant tissues, organs, or whole plants. A challenge in both fundamental and applied research (e.g., breeding applications) is to explain or predict phenotypes from the underlying genotypes under different environmental conditions. Genotypic variation leads to differences in the biochemical makeup of cells, which in turn

*Keywords: phenotype,  
omics data, supervised  
machine learning*

together with the environment influence organ formation, plant growth, and eventually traits relevant in agriculture, such as yield and tolerance to stresses and pests. Unraveling the effects of genotypic variation and environment on phenotypes yields fundamental insights into the regulation of important processes in plant development and physiology and the ability to predict yield and quality traits from genotypes in specific environments, which is essential in modern molecular plant breeding. Analyzing phenotypes measured at these different levels or linking these phenotypes to genotypes increasingly calls for processing and integration of large, noisy, and heterogeneous data sets. Machine learning plays an increasingly important role in these efforts. In various scientific and engineering domains, machine learning has driven a spur of recent innovations, and as I will illustrate in this talk, it is set to do the same in plant research. In the practical session following the talk, we will look at various ways in which -omics data can be analyzed with machine learning. One dataset we will use involves plant sesquiterpene synthases, for which a sequence- and/or structure-based model can predict their function. A second dataset involves prediction of crossovers, based on genome-sequence. These two examples involve supervised machine learning; if time allows, you can also look at unsupervised machine learning, in particular using tSNE to analyze expression data.

## Literature

A. D. J. van Dijk et al. (2021) Machine learning in plant science and plant breeding. *iScience*, 24(1): 101890. <https://doi.org/10.1016/j.isci.2020.101890>

Demirci, S. et al. (2016) DNA sequence and shape are predictive for meiotic crossovers throughout the plant kingdom. *The Plant Journal*, 95 (4), 686-699. <https://doi.org/10.1111/tpj.13979>

**Aalt-Jan van Dijk** performed his PhD in computational biology at Utrecht University (2002-2006). He then moved to Wageningen University, where he currently is associate professor in the Bioinformatics group. His research involves modelling and analysis of plant -omics data. This includes the development of algorithms, in particular based on machine learning, and application to various plant datasets. Recent examples include analysis of plant protein interactions, prediction of meiotic crossovers in plants, and the use of sequence and structure based features to predict functions for plant enzymes.

# Deep learning for plant genomics and crop improvement.

Prof. Hai Wang, China Agricultural University, China

Our era has witnessed tremendous advances in plant genomics, characterized by an explosion of high-throughput techniques to identify multi-dimensional genome-wide molecular phenotypes at low costs. In recent years, deep learning has been found extremely effective in modeling the flow of information from biological sequences to molecular phenotypes, and interpretation of these deep learning models provides quantitative functional annotations for

*Keywords: plant  
genomics, crop  
improvement, deep  
learning*

biological sequences at a high resolution. Deep learning models predicting molecular phenotypes will revolutionize crop genetic improvement by 1) identifying functional genomic variants that are useful for crop breeding, 2) designing artificial genomic variants for genome editing, 3) designing novel genomic elements or proteins with desirable functions for crop synthetic

biology. Taken together, we propose a central role of deep learning in future plant genomics research and crop genetic improvement.

## Literature

Wang, H. et al. (2020) Deep learning for plant genomics and crop improvement. *Current Opinion in Plant Biology*, 54, 34-41. <https://doi.org/10.1016/j.pbi.2019.12.010>

Washburn, J. D. et al. (2019) Evolutionarily informed deep learning methods for predicting relative transcript abundance from DNA sequence. *PNAS*, 116(12): 5542-5549. <https://doi.org/10.1073/pnas.1814551116>

**Hai Wang** received his PhD degree from Fudan University at 2012. During his PhD training, he worked as an exchange student at Donald Danforth Plant Science Center in St. Louis, MO (2008-2011) to dissect novel modulators of Arabidopsis innate immune response and redox stress response. After graduation, he worked in the field of maize functional genomics in Chinese Academy of Agricultural Sciences as an assistant professor (2012-2015) and later as an associate professor (2016-2019). Although trained as an experimental molecular biologist, he is an enthusiast for mathematics and computer sciences, and spend most of his spare time teaching himself programming, bioinformatics and machine learning. From 2017 to 2019, he collaborated with Professor Edward Buckler in Cornell University to apply deep learning in plant genomics. Since 2019 he works as a full professor at the National Maize Improvement Center, China Agricultural University.

# Analysis of vegetation parameters at global scale with deep learning

Prof. Dr. Jan Dirk Wegner, University of Zurich, Switzerland

Worldwide analyzes and estimates of vegetation parameters such as biomass or vegetation height are essential for modeling climate change and biodiversity. Traditional allometric approaches usually must be adapted for specific ecosystems and regions. It is therefore very

*Keywords: biomass,  
vegetation height, deep  
convolutional neural  
network*

difficult to carry out homogeneous, global modeling with high spatial and temporal resolution and, at the same time, good accuracy. Data-driven approaches, especially modern deep learning methods, promise great potential here. In this talk, new research results on the large-scale determination of vegetation parameters will be presented.

## Literature

- Lang, N. et al. (2022) A high-resolution canopy height model of the Earth. arXiv:2204.08322  
Lang, N. et al. (2022) Global canopy height regression and uncertainty estimation from GEDI LIDAR waveforms with deep ensembles. Remote Sensing of Environment, 268: 112760. <https://doi.org/10.1016/j.rse.2021.112760>  
Lang, N. et al. (2019) Country-wide high-resolution vegetation height mapping with Sentinel-2. Remote Sensing of Environment, 233: 111347. <https://doi.org/10.1016/j.rse.2019.111347>

**Jan Dirk Wegner** holds the "Data Science for Sciences" chair at the Institute for Computational Science, University of Zurich, as an associate professor and is head of the EcoVision Lab at ETH Zurich. Jan was PostDoc (2012-2016) and senior scientist (2017-2020) in the Photogrammetry and Remote Sensing group at ETH Zurich after completing his PhD (with distinction) at Leibniz Universität Hannover in 2011. His main research interests are at the frontier of machine learning, computer vision, and remote sensing to solve scientific questions in the environmental sciences and geosciences. Jan was granted multiple awards, among others an ETH Postdoctoral fellowship and the science award of the German Geodetic Commission. He was selected for the WEF Young Scientist Class 2020 as one of the 25 best researchers world-wide under the age of 40 committed to integrating scientific knowledge into society for the public good. Jan is vice-president of ISPRS Technical Commission II, chair of ISPRS II/WG 6 "Large-scale machine learning for geospatial data analysis", director of the PhD graduate school "Data Science" at University of Zurich, and his professorship is part of the Digital Society Initiative at University of Zurich. Together with colleagues, Jan is chairing the CVPR EarthVision workshops.

# Data mining and machine learning in macro-ecological research

Prof. Niklaus Zimmermann, WSL, Switzerland

Data mining and machine learning have become essential elements of macroecological research. Challenges are manifold as we basically aim to collect data across large spatial scales, and to understand the drivers behind the observed patterns in species, functional and phylogenetic richness and turnover with the help of large numbers of predictors. Classical regression is one way to analyze such relationships, machine learning techniques are powerful

*Keywords:*  
*Biodiversity, citizen*  
*science, data*  
*collection, image*  
*classification, niche*  
*theory, plausibility*  
*assessment,*  
*recommender*  
*systems, species*  
*distributions*

alternatives. Yet, although we often use enormous amounts of data, matrices remain sparse because of the high dimensionality along which we analyze these patterns. This poses challenges to machine learning and requires us to combine modern methods with theoretical considerations. In my presentation, I summarize the main challenges and solutions to analyze such biodiversity patterns. But I also explore opportunities to enhance the trustful collection of new biodiversity data, both for professional and citizen science campaigns.

## Literature

Brun P. et al. (2020) Model complexity affects species distribution projections under climate change. *Journal of Biogeography*, 47: 130-142. <https://doi.org/10.1111/jbi.13734>

In 2006, **Niklaus Zimmermann** was elected as head of the research unit "Landscape Dynamics", which grew to four research groups and included ca. 100 scientists, postdocs, Ph.D. students, technicians and visiting scientists. He has resigned from this research unit leader position in 2013 and became a senior scientist at WSL. In 2014, he was awarded adjunct professor for macroecology at the Institute of Terrestrial Ecosystems at ETH. Since 2007, Niklaus Zimmermann is a member of the directorate of WSL. Since 2012, he is a member of the science commission of InfoFlora, the Swiss National Floristic Data Center. Since 2014, he is a member of the C2SM Steering Board, the ETH center for climate systems modelling, in which he represents WSL.

Prof. Zimmermann's main research topics include (1) plant and ecosystems ecology, (2) macroecology, (3) biodiversity, and (4) evolutionary and functional ecology. He has considerable expertise in the conceptual and scientific advancement of predictive distribution modeling, in modelling climate change impacts on species, ecosystems and biodiversity, and in functional and evolutionary analyses related to biodiversity and community ecology. Niklaus E. Zimmermann has published >180 ISI publications. He has a Web of Science (google scholar) h-index of 55 (65), has amassed >17'000 (>30'000) citations, and he is listed as a "highly cited researcher" (highlycited.com) annually since 2014. Two of his publications are cited >3'600 and 17 of his papers are listed as "highly cited" by the WoS/Clarivate Analytics statistic. 38 of his papers are cited >100 times.



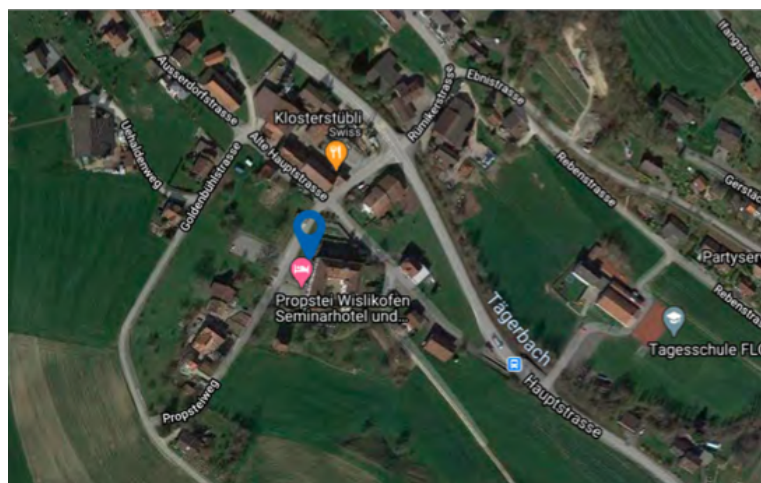
## General information

### Accommodation

We are staying (12-16.09.2022) at the Pilgerhaus Allegro, Einsiedeln. Shared rooms (3 to 4 persons) with shared bathrooms. Rooms will be allocated by the organizers. Full board from Monday lunch to Friday with breakfast, lunch and coffee breaks.

#### Contact information

Hotel Allegro  
Lincolnweg 23  
8840 Einsiedeln  
Tel: +41554188888



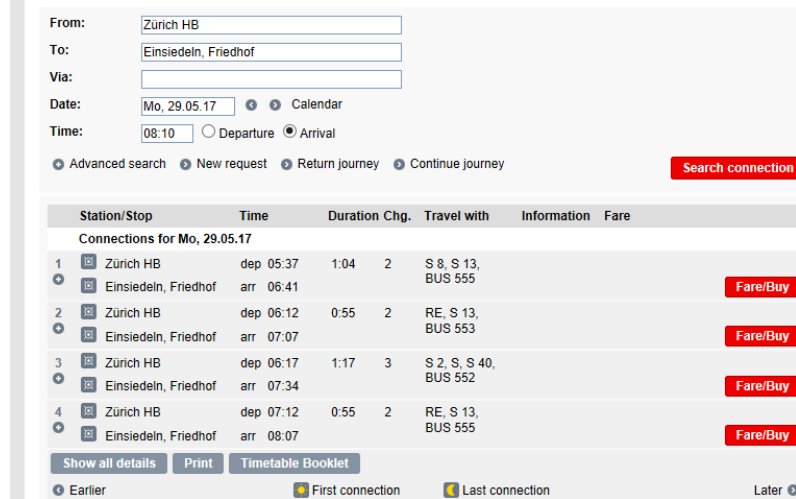
## How to get to the venue

There is a bus stop close (approx. 250 m) to Hotel Allegro in Einsiedeln, this bus stop is called „Friedhof“. Check the SBB online timetable for your detailed connections:

<http://fahrplan.sbb.ch/bin/query.exe/en>

[Buy a single train ticket to Einsiedeln, Friedhof](#)

Example: Travel plan from Zurich (Main station) to Einsiedeln (Bus stop: Friedhof):



The screenshot shows the SBB online timetable search interface. The search criteria are: From: Zürich HB, To: Einsiedeln, Friedhof, Date: Mo, 29.05.17, Time: 08:10. The search results show four connections for Monday, 29.05.17. Each connection is a train from Zürich HB to Einsiedeln, Friedhof, with a duration of 1:04, 0:55, 1:17, and 0:55 respectively. The connections are: 1. Zürich HB dep 05:37, arr 06:41, S 8, S 13, BUS 555. 2. Zürich HB dep 06:12, arr 07:07, RE, S 13, BUS 553. 3. Zürich HB dep 06:17, arr 07:34, S 2, S, S 40, BUS 552. 4. Zürich HB dep 07:12, arr 08:07, RE, S 13, BUS 555. Each connection has a 'Fare/Buy' button. The interface also includes a 'Search connection' button and a 'Show all details' button.

Station/Stop	Time	Duration	Chg.	Travel with	Information	Fare
Connections for Mo, 29.05.17						
1 Zürich HB	dep 05:37	1:04	2	S 8, S 13, BUS 555		
Einsiedeln, Friedhof	arr 06:41					Fare/Buy
2 Zürich HB	dep 06:12	0:55	2	RE, S 13, BUS 553		
Einsiedeln, Friedhof	arr 07:07					Fare/Buy
3 Zürich HB	dep 06:17	1:17	3	S 2, S, S 40, BUS 552		
Einsiedeln, Friedhof	arr 07:34					Fare/Buy
4 Zürich HB	dep 07:12	0:55	2	RE, S 13, BUS 555		
Einsiedeln, Friedhof	arr 08:07					Fare/Buy

In Einsiedeln, take the postal bus in the direction of “Hoch-Ybrig” (or Studen), and get off at the stop “Friedhof”. From here, the Hotel Allegro is just a three-minute walk. Upon arrival at the Hotel, go to the main desk and ask for Barbara Templ.

## Region

The area is geographically interesting and beautiful with several high mountains. Einsiedeln is located up a plateau (ca. 880 m (2,890 ft) above sea level) and situated near the artificial mountain lake Sihlsee. The dam, which retains the lake, produces electricity for the trains and protects the city of Zurich further down the valley from the flood of the Sihl. The village is a popular tourist destination in central Switzerland. The Benedictine Einsiedeln Abbey, located within the village, is considered one of the most important Roman Catholic pilgrimage sites in Europe. Since the Middle Ages the Graces Chapel and a statue of the Black Madonna have been the centerpiece of the pilgrimage. Einsiedeln is also a popular destination for sports year-round.

## Organizer

The Zurich-Basel Plant Science Center (PSC) is a competence center for the plant science research community at ETH Zurich, University of Zurich and at University of Basel. The center promotes research, education and outreach and provides platforms for interactions with peers, policymakers, industry, stakeholders and the public.

Zurich-Basel Plant Science Center

ETH Zurich

Tannenstrasse 1, 8092 Zurich

[www.plantsciences.ch](http://www.plantsciences.ch)