New framework for analysis of aquatic ecosystems

Ankita Ravi Vaswani and Klas Ove Möller[1]
[1] Biological carbon pump group, Institute of Carbon Cycles, Helmholtz-Zentrum Hereon, Geesthacht

Date of submission: 13.05.2022
Required funding: 6 months

Abstract

Advances in high-throughput *in situ* imaging offer unprecedented insights into aquatic ecosystems by observing organisms in their natural habitats. However, unlocking this potential requires new analysis tools that transcend species identification to reveal morphological, behavioral, physiological and life-history traits. We will develop, document and validate an image analysis pipeline for semi-automated functional trait annotation, apply it to zooplankton in a continuously monitored North Sea region, and train a neural network for full automation. We foresee that these tools will enable new avenues of investigation in aquatic research, ecosystem modelling and global biogeochemical flux estimations, revealing previously inaccessible relationships between species biodiversity, zooplankton traits and seasonal variations in environmental conditions.

Introduction

Zooplankton are essential for aquatic food webs and biogeochemical cycles[1,2]. Understanding how changing environmental conditions affect distribution and abundance of zooplankton traits allows us to decode the effects of climate change on biodiversity, ecosystem dynamics and global carbon cycles[3–5]. Body size is a master trait linked to developmental stage, respiration, metabolism, excretion, motility and predator-prey interactions[6]. *In situ* images also provide information about individual organisms' survival, growth, reproduction and resource acquisition from the visual signatures of the underlying traits[4,7] such as lipid reserves[8], egg clutch sizes[9], appendage extension[7], and body posture[7]. Recently, zooplankton traits analyzed during spring ice-melts in the Arctic Ocean have revealed complex ecosystem responses to environmental changes[7]. However, manual annotation procedures cannot scale to analyze the billions of images now produced by imaging technologies across diverse environmental conditions and long durations. Instead, we require new analysis pipelines to efficiently annotate a fraction of these data, to produce training data for fully automated deep-learning based feature extraction[5,10].

We will construct an image analysis pipeline (deliverable **D1**), combining an existing neural network for taxonomic classification[11] with semi-automated segmentation of zooplankton traits. Such a consolidated pipeline does not currently exist for trait extraction and will be an invaluable tool for marine researchers. We will apply D1 to a continuously monitored North Sea region to produce an annotated trait dataset (**D2**), and use this to train a convolutional neural network (CNN, **D3**) in collaboration with Hereon's Model-Driven Machine Learning group (MDML). We will document each step of this process using interactive tutorials (**D4**) and publish open-source software (**D5**). These deliverables will accelerate trait extraction and increase reproducibility.

During my PhD, I designed a semi-automated analysis pipeline to extract morphological features from ~15,000 embryonic brain cells, revealing complex relationships between cell morphology and migration speed[12]. Despite the new application domain, the technical aspects and challenges of semi-automated segmentation align well with my skill set well, allowing me to design an efficient feature extraction pipeline for the annotation of zooplankton traits.

I. *Incubator Project description*

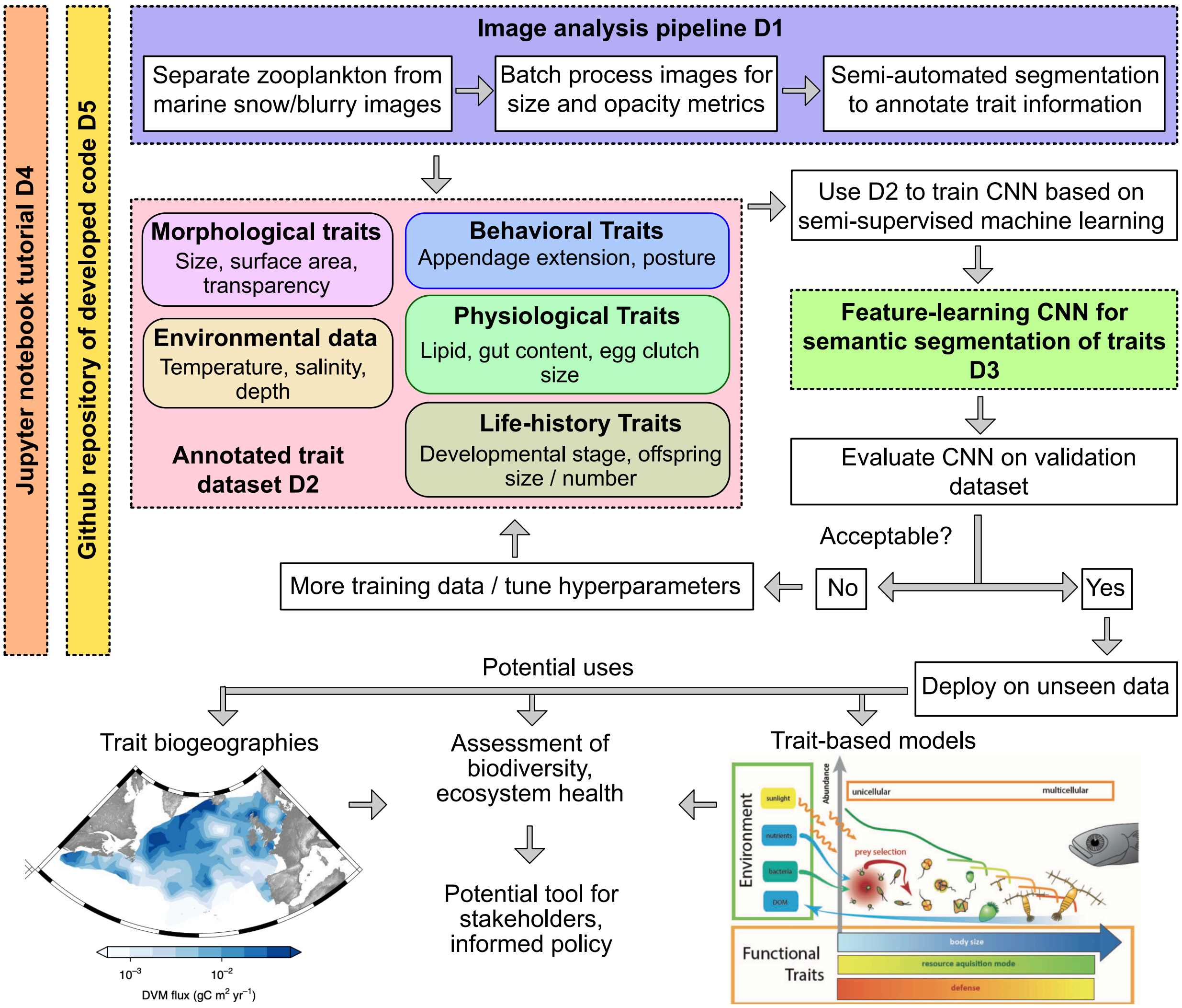

Figure 1: A consolidated data analysis pipeline D1 will be built to classify *in situ* plankton images, extract morphological features and guide users through semi-automated trait segmentation. D1 will be applied to HPT data to generate annotations D2. D2 will be used to train a trait extraction CNN (D3). A Jupyter notebook (D4) will guide users through customizing and implementing D1 and D3 for diverse image data. Wide-scale application of these tools will lead to detailed trait-biogeographies, better ecosystem models and accurate assessment of the impact of climate change on biodiversity and ecosystem health. Image adapted from previous studies[13,16].

The Möller lab has deployed underwater observatories in the North Sea to acquire high resolution plankton time series (HPT) imaged at 6-10 Hz. ~10% of HPT data will be analyzed using our analysis pipeline (D1) to generate an annotated trait dataset (D2).

1. In the first 2 weeks, zooplankton will be separated from marine snow and blurry images and classified into relevant taxon units (copepods, jellyfish, marine snow, etc.) using an existing CNN developed by the Möller and MDML groups[11].
2. Quantitative traits such as body length, surface area, volume and opacity metrics will be automatically extracted from images[13]. Taxonomic classification will be confirmed and semi-automated segmentation will be used to extract complex traits such as lipid body area, lipid fraction[8], egg-clutch size[9], body posture and extension of appendages and antennae[7].
3. D1 will combine all automated and semi-automated methods from steps 1-2 into one program, where imported images will be sequentially classified, processed to extract numerical descriptors and inspected by a user for trait segmentation. All measurements and trait segmentation masks will be stored as metadata linked to the images. This pipeline will run in a browser with a Jupyter-based rich graphical interface (D4). Development of the image analysis pipeline (D1) will require ~2-3 months, culminating in an annotated trait dataset (D2).
4. D2 will be used for semi-supervised training of a new CNN (distinct from the existing CNN of step 1), resulting in software for automatic trait extraction (D3). This step will be carried out over ~2-3 months in collaboration with the MDML group and support from Helmholtz AI consultants (see attached letters of support). If further annotated data are needed, step 3 will be carried out on additional images from the HPT data in parallel to CNN training.
5. Tutorials for the the data analysis pipeline and analyses in steps 1-3 will be documented in Jupyter notebooks (D4). Code will be documented and shared in open-source repositories (Github; D5). This step will be carried out in parallel to previous steps.

This project has a high chance of success given the participants' expertise and experience and immediate availability of required tools. D1-2 will provide value to end users prior to the completion of subsequent steps. Beyond the scope of this incubator project, follow-up work is foreseen investigating seasonal variations in trait abundances in the North Sea using D2 and D3.

II. *Relevance for the NFDI4Earth*

- Automatic taxonomic classification and trait extraction (D1/D3) will be valuable for marine biologists, ecologists and image analysts.
- Global trait biogeographies, generated using tools developed here, could reveal relationships between trait abundances, biodiversity and ecosystem health, helping scientists, stakeholders, and policy makers to guide sustainable ecosystem management[14,15].
- Detailed trait-biogeographies could also be overlayed on data from NFDI4Earth repositories (e.g., HZG marine geoportal coastal maps) to reveal relationships between the biology and geography of aquatic ecosystems.
- Our tools could increase incorporation of functional trait data in detailed, predictive models of aquatic ecosystems and carbon flux[16,17].

III. *Deliverables*

- D1: Data-analysis pipeline (month 3)
- D2: Annotated trait dataset (month 3)
- D3: CNN for automated feature learning and extraction (month 6)

- D4: Tutorials for design and implementation of D1 (month 3), generation of D2 (month 3) and D3 (month 6)
- D5: Open-source code repository (GitHub)

References

1. Frederiksen, M., Edwards, M., Richardson, A. J., Halliday, N. C. & Wanless, S. From plankton to top predators: bottom-up control of a marine food web across four trophic levels. *J Anim Ecology* **75**, 1259–1268 (2006).
2. Boyd, P. W. Multi-faceted particle pumps drive carbon sequestration in the ocean. *Nature* **9** (2016).
3. Violle, C. *et al.* Let the concept of trait be functional! *Oikos* **116**, 882–892 (2007).
4. Ohman, M. D. A sea of tentacles: optically discernible traits resolved from planktonic organisms in situ. *ICES Journal of Marine Science* **76**, 1959–1972 (2019).
5. Orenstein, E. *et al.* Machine learning techniques to characterize functional traits of plankton from image data. *In press* (2022).
6. Hatton, Heneghan, Bar-On & Galbraith. The global ocean size-spectrum from bacteria to whales. *Sci. Adv.* **7**, (2021).
7. Vilgrain, L. *et al.* Trait-based approach using in situ copepod images reveals contrasting ecological patterns across an Arctic ice melt zone. *Limnol Oceanogr.* **66**, 1155–1167 (2021).
8. Schmid, M. S., Maps, F. & Fortier, L. Lipid load triggers migration to diapause in Arctic Calanus copepods—insights from underwater imaging. *Journal of Plankton Research* **40**, 311–325 (2018).
9. Möller, K. O. *et al.* Effects of climate-induced habitat changes on a key zooplankton species. *Journal of Plankton Research* **37**, 530–541 (2015).
10. Irisson, J.-O., Ayata, S.-D., Lindsay, D. J., Karp-Boss, L. & Stemmann, L. Machine Learning for the Study of Plankton and Marine Snow from Images. *Annu. Rev. Mar. Sci.* **14**, 18.1–18.25 (2021).
11. Schanz, T., Möller, K. O., Ruehl, S. & Greenberg, D. Robust Detection of Marine Life with Label-free Image Feature Learning and Probability Calibration. *Helmholtz AI Conference, Dresden* (2022).
12. Vaswani, A. R. *et al.* Correct setup of the substantia nigra requires Reelin-mediated fast, laterally-directed migration of dopaminergic neurons. *eLife* **8**, e41623 (2019).
13. Kiørboe, T., Visser, A. & Andersen, K. H. A trait-based approach to ocean ecology. *ICES Journal of Marine Science* **75**, 1849–1863 (2018).
14. Gorsky, G. *et al.* Digital zooplankton image analysis using the ZooScan integrated system. *Journal of Plankton Research* **32**, 285–303 (2010).
15. Prowe, A. E. F., Visser, A. W., Andersen, K. H., Chiba, S. & Kiørboe, T. Biogeography of zooplankton feeding strategy. *Limnol. Oceanogr.* **64**, 661–678 (2019).
16. Martini, S. *et al.* Functional trait-based approaches as a common framework for aquatic ecologists. *Limnol Oceanogr* **66**, 965–994 (2021).
17. Brun, P. *et al.* Climate change has altered zooplankton-fuelled carbon export in the North Atlantic. *Nat Ecol Evol* **3**, 416–423 (2019).
18. Serra-Pompei, C., Soudijn, F., Visser, A. W., Kiørboe, T. & Andersen, K. H. A general size- and trait-based model of plankton communities. *Progress in Oceanography* **189**, 102473 (2020).

Helmholtz-Zentrum
hereon

Helmholtz-Zentrum hereon | Max-Planck-Straße 1 | 21502 Geesthacht, DE

Geesthacht, 12.05.2022

Letter of support for the NFDI4Earth incubator project "New framework for analysis of aquatic ecosystems"

Institut für
Dynamik der Küstenmeere
Mariner Schnee und Plankton
Institute of
Coastal Ocean Dynamics
Marine Snow and Plankton

www.hereon.de

Abteilungsleitung | Department Head
Dr. Klas Ove Möller
T +49 4152 87-2371
klas.moeller@hereon.de

To whom it may concern,

This letter expresses my strong and full support for the NFDI4Earth Incubator project proposal "New framework for analysis of aquatic ecosystems" put forward by Dr. Ankita Vaswani. The proposal outlines an ambitious and timely approach which will provide new tools to examine biodiversity, ecosystem functioning and ecosystem health in the ocean.

The candidate will work in the "Biological Carbon Pump" group at the Institute of Carbon Cycles, Helmholtz Zentrum Hereon. The Biological Carbon Pump group has a strong background in exploring marine plankton dynamics, biological-physical interactions and the role of the marine environment in regulating global climate. At the Institute of Carbon Cycles, the candidate will find a wealth of complimentary expertise ranging over diverse topics including biology, oceanography, physics and ecology.

All required infrastructure, computing and data storage resources are available at Hereon including cluster access, large computation and storage resources, high bandwidth data transfers, support for code and software exchange via HiFis, and access to additional resources via HiCore.

The candidate is well qualified and experienced in processing and analysing large image datasets derived from optical instruments, which is the key technical skill required for the success of the project. With respect to this highly interdisciplinary project; the candidate will have support and collaboration from the Model-driven Machine Learning Group, at the Helmholtz-Zentrum Hereon, giving her a strong chance to successfully complete the project and achieve her career goals.

Klas Möller

……………………………………………

(Dr. Klas Ove Möller)

Helmholtz-Zentrum
hereon

Helmholtz-Zentrum Hereon | Max-Planck-Straße 1 | 21502 Geesthacht, DE

NFDI4Earth Coordination Office
Incubator Projects

Institut for Coastal Resarch
Modellbasiertes
Maschinelles Lernen
Institute of Coastal Systems
Analysis and Modeling
Model-Driven Machine Learning

www.hereon.de

Abteilungsleitung | Head of Department
Dr. David S. Greenberg
T +49 4152 87-2133
david.greenberg@hereon.de

Letter of Support for Project "New framework for analysis of aquatic ecosystem"
Geesthacht, 12 May 2021

Dear NFDI4Earth Project Committee,

I'm writing to express strong support for the project "New framework for analysis of aquatic ecosystem." Since 2020 I have lead the Model-Driven Machine Learning research group at Helmholtz Centre Hereon, focused on developing and applying artificial intelligence and machine learning (AI/ML) to improve understanding and predictability both physical and biological aspects of coastal systems.

A major challenge in effectively applying AI/ML methods in the earth sciences is the wide gulf between the artificial, sanitized datasets used by the deep learning community and real field data. *In situ* observations of biological systems exhibit complexity, noise, nonstationarity and class imbalance far beyond AI/ML benchmarks, so that promising AI/ML methods often fail in the real world.

By annotating image datasets with labeled morphological features and functional traits, and by developing a pipeline for further annotation, this project would significantly aid and accelerate the development of AI/ML methods for studying marine organisms, with considerable benefits for both marine biology and deep learning.

I am therefore happy to pledge our research group's support in training deep learning models on these datasets. In particular, we can provide existing convolutional networks for detecting and classifying zooplankton image features, developed and trained in an ongoing collaboration with Dr. Klas Ove Möller at Hereon. We will provide access to multi-GPU compute nodes for training and deploying these networks.

Best regards,

David Greenberg

DKRZ
DEUTSCHES KLIMARECHENZENTRUM

DKRZ GmbH · Bundesstr. 45a · D-20146 Hamburg

NFDI4Earth Coordination Office
Incubator Projects

Ust-IdNr.: DE118713406
Steuernummer: 17/413/00377

Deutsche Bank Hamburg
BLZ 200 700 00
Konto-Nr.: 367518800
IBAN: DE51 2007 0000 0367 5188 00
BIC: DEUTDEHHXXX
Telefon: 040 - 460094 -
Telefax: 040 - 460094 - 270

E-Mail: weigel@dkrz.de

Ansprechpartner: Tobias Weigel

Ihr Zeichen	Ihre Nachricht vom	Unser Zeichen	Durchwahl	Datum
			-	13. Mai 2022

Letter of Support for project “New framework for analysis of aquatic ecosystem”

Dear NFDI4Earth project committee,

the Helmholtz AI Cooperation Unit (Helmholtz AI) is set up as a hub for applied artificial intelligence and machine learning across the whole Helmholtz Association. It consists of six local units, representing the six research fields of the Helmholtz Association, each of which hosts research groups and an AI consultant team. The AI consultant teams’ mission is to enable and facilitate the use and implementation of AI/ML methods through short and mid-term collaborations of up to six months. Their expertise can be solicited through a Voucher System accessible at no cost to all researchers; decisions to engage in collaboration are based on capability and capacity.

As consultants focussing on “Earth and Environment” based at DKRZ we are happy to support the project “New framework for analysis of aquatic ecosystem” according to the procedures of the voucher support system. We recognize the feasibility of the proposed approach and see the potential of transfer to other application cases in the marine sciences, but also beyond. We have worked on Computer Vision problems including semantic segmentation in several prior support activities, and also set up efficient procedures for the practical training and tuning workflows, and would be happy to lend such additional expertise to the project.

Yours sincerely,

Tobias Weigel

Digital signiert von Tobias Weigel
DN: cn=Tobias Weigel, c=DE, o=Deutsches Klimarechenzentrum GmbH, ou=Anwendungen, email=weigel@dkrz.de
Datum: 2022.05.13 14:22:15 +02'00'

Dr. Tobias Weigel