Issue 36
March - April 2023

HEIBRIDS NEWS

HELMHOLTZ EINSTEIN INTERNATIONAL BERLIN RESEARCH SCHOOL IN DATA SCIENCE

"People worry that computers will get too smart and take over the world, but the real problem is that they're too stupid and they've already taken over the world"

- Pedro Domingos, Professor Emeritus of Computer Science & Engineering at University of Washington and author of *The Master Algorithm*

Latest Highlights

New Publications and Preprints

- P. Graniero, M. Khenkin, H. Köbler, N. T. Putri Hartono, R. Schlatmann, A. Abate, E. Unger, T. J. Jacobsson and C. Ulbrich (2023). **The challenge of studying perovskite solar cells' stability with machine learning.** *Front. Energy Res.,* Sec. Solar Energy, 11. https://doi.org/10.3389/fenrg.2023.1118654
- N. H. Chan, M.Langer, B. Juhls, T. Rettelbach**,** P. Overduin, K. Huppert and J. Braun (2023). **An Arctic Delta Reduced Complexity Model and its Reproduction of Key Geomorphological Structures.** *Earth Surface Dynamics*, 259–285. https://doi.org/10.5194/esurf-11-259-2023
- W. J. Foster, B. J. Allen, N. H. Kitzmann, J. Münchmeyer, T. Rettelbach, J. D. Witts, , R. J. Whittle, E. Larina, M. E. Clapham and A. M. Dunhill (2023). **How predictable are mass extinction events?** *Royal Society Open Science*, 10 (3), 221507. https://doi.org/10.1098/rsos.221507
- K. Palczynski, T. Kirschbaum, A. Bande, J. Dzubiella (2023). **Hydration Structure of Diamondoids from Reactive Force Fields.** *J. Phys. Chem. C*, 127, 6, 3217–3227. https://doi.org/10.1021/acs.jpcc.2c07777
- T. Kirschbaum, B. von Seggern, J. Dzubiella, A. Bande, F. Noé (2023). **Machine Learning Frontier Orbital Energies of Nanodiamonds.** J. *Chem. Theory Comput.* https://doi.org/10.1021/acs.jctc.2c01275

Recent Presentations at Conferences

- B. Groenke, M. Langer, G. Gallego, and J. Boike: **Exploring physics-informed machine learning for accelerated simulation of permafrost processes.** *EGU General Assembly*, Vienna, Austria, 24–28 April 2023. https://doi.org/10.5194/egusphere-egu23-10135

HEIBRiDS Outreach

HEIBRiDS PhD student, Daniel León Periñán, participated in a panel discussion on **"Spark: Creative processes in the arts and sciences"**, an event that was part of the Breaking Boundaries Series and took place in MDC-BIMSB on **March 13**. The series invites world-class musicians, composers and scientists to share their insights and experiences in lighting the creative fire in their respective fields.

HEIBRiDS Annual Retreat

The **HEIBRiDS Annual Retreat** took place this year on **March 27-28** at Schloss Diedersdorf and spanned through two days of scientific discussions and networking among HEIBRiDS PIs and PhD researchers.

Announcements

HEIBRiDS Spring School

Our next HEIBRiDS Spring School will take place on **June 7-9** at the Conference Room of ECDF. You can read more information about it here. Note that the first two days are mandatory only for the last cohort of HEIBRiDS students, while the last day is for everyone.

From Slack to Mattermost

Our workspace in Slack will soon discontinue due to changes in their license policy. We have created a HEIBRiDS Team under Mattermost and we invite all HEIBRiDS PhD students and associated students to join.

2nd AI-HERO Hackathon on Energy Efficient AI

NHR@KIT, Helmholtz AI, Helmholtz Imaging and the Helmholtz Metadata Collaboration are jointly hosting the 2nd AI-HERO Hackathon on Energy Efficient AI which will take place on **June 19-21** as an in-person event at the Marsilius Arkaden in Heidelberg. The challenge of the Hackathon is developing and optimizing an AI model, while being as energy efficient as possible. The Hackathon is targeting AI-researchers with experience in model development, training and optimization. Participants will join into teams of three to tackle one of two use-cases, from the research field health or from the research field energy, and evaluate their approaches on the HoreKa supercomputer system at KIT. The registration is open until **June 1,** or until all seats are taken. Detailed information on registration and schedule can be found here.

Two specialized courses in earth system / data sciences

AWI received funding from HIDA to invite Dr. Jonathan Lilly as a guest scientist in data sciences. Dr. Lilly is a reknown expert in earth system sciences, working with large, complicated, multivariate datasets and has ample teaching experience.
He will teach on site on the following dates:

June 5 – 16: Ocean/Atmosphere Time Series Analysis at **GEOMAR**
June 19 – 23: Data Analysis Intensive Course at **AWI**

Participation is free of charge, however, participants would need to cover their own travel expenses. The organizers are happy to offer support with finding budget accomodation. Online participation is not possible. You can find attached both announcements as pdf including contact details for registration.

DDSA Visit Grant

Would you like to spend time with new research groups and fields of interest within a Danish data science environment? The new Friend of HIDA, the Danish Data Science Academy (DDSA) would like to share with you their Visit Grant**,** which is open all year for applications. The purpose of the grant is to give students and young researchers the opportunity to form the basis for a future DDSA PhD or postdoc fellowship application in collaboration with a potential supervisor from a Danish university or a Danish research institution.

The Nordic AI Conference

The Nordic AI Conference, a joint conference of all major AI players in the Nordic countries, will take place in Copenhagen on **October 2-3**. The call for proposals is unfortunately only open to young researchers from the Nordic countries. However, general participation at the conference is possible.

hub.berlin 2023

hub.berlin, the business festival for tech & digitalization, will open its doors again this year on **June 28-29**. For tickets and detailed schedule have a look here.

MarDATA HELMHOLTZ School for Marine Data Science

HiDA HELMHOLTZ Information & Data Science Academy

Ocean/Atmosphere Time Series Analysis: Theory and Practice

Jonathan Lilly, Planetary Science Institute

June 05 - 16, 2023 10:00-15:00
Monday through Friday, onsite @GEOMAR
Sponsored by Helmholtz Information & Data Science Academy

Statistics — Fourier Spectra — Wavelet Analysis — Mental Factors

Course content:
This course will introduce students to classical as well as cutting-edge techniques for analysing time series in oceanographic, atmospheric science, and climate applications. Beginning with a solid understanding of the link between time-domain and frequency-domain analyses, we will proceed from simple smoothing, to Fourier spectral estimation, to time-frequency methods such as the continuous wavelet transform. The chosen techniques are those that experience has shown to be the most useful in dealing with time series from the ocean and the atmosphere. Emphasis will be given to hands-on, practical application of methods, as well as to understanding the theory behind various methods. Extensive course notes may be found at http://jmlilly.net/course/index.html. Students can choose to participate in the course using either Python, Matlab, or for the first time, Julia.

Students will bring with them a dataset of their choice that they would like to investigate in detail. A final project will consist of applying the methods taught in the course to this dataset and interpreting the results. The students will receive personalized feedback, tailored to their specific datasets, through one-on-one meeting sessions with the instructor.

Learning modules/structure:
The course will run Monday through Friday from June 05 to June 16 onsite at GEOMAR (Kiel). We will meet for five hours a day for ten days. Class time will be roughly split between lectures and interactive lab sessions. There will be a long half-hour break each day as well as several shorter breaks.

Target group:
The course is open to doctoral researchers in marine and climate sciences or related data sciences who are working with time series data sets. Place availability permitting, early postdoctoral researchers are equally invited.

MarDATA HELMHOLTZ School for Marine Data Science

HiDA HELMHOLTZ Information & Data Science Academy

Ocean/Atmosphere Time Series Analysis: Theory and Practice

Prerequisites:
Students should have a working knowledge of one of the programming languages the course is offered in (Matlab, Python, or Julia). The course endeavours to be self-contained in that needed information will be covered during the course itself, however, prior experience with complex numbers, e.g. u+iv, as well as linear algebra is expected. Those with questions should contact the instructor at jmlilly@psi.edu.

Registration:
Total number of participants is limited to 35 (with 20 slots open for non-MarDATA members). Participation will be allocated on a first come - first serve basis. **The registration is considered binding**!

Please note that any costs for accommodation in and travel to Kiel are to be borne by the participants themselves!

Registration closes on May 22, 23:59

Please register by mail to info@mardata.de

"If I had taken this course at the beginning of my PhD, I could have accomplished everything I did in the first year in four months."

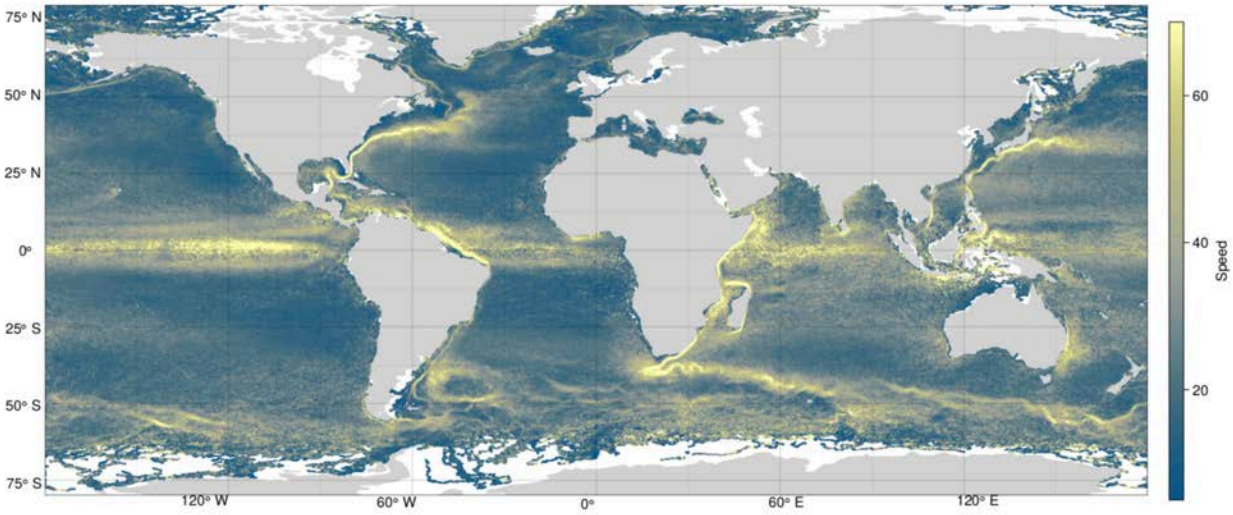
"The course is amazing, I can recommend it to everyone dealing with time series."

"There are a few moments when something changes in your person and you note those moments. This has been one of those transformations. Which is the definition of deep, integrated learning."

Data Analysis Intensive

Dr. Jonathan Lilly, Planetary Science Institute

(Kindly sponsored by Helmholtz Information & Data Science Academy HIDA)



Date & Time:	June 19 – 23, 2023 10 a.m. – 3 p.m. (lunch will be provided)
Location:	AWI Bremerhaven, room E-4025
Language:	English
Target audients:	Master's students, doctoral students, or postdocs
POLMAR credit points:	5
Registration:	sea.courses@awi.de

Learning modules/structure: There will be roughly 2.5 hours of lectures and a two-hour lab session each day.

Course description: This course addresses the mental, technical, and tactical skills necessary to discover the stories hidden within datasets. Each participant will bring a dataset of their own choosing, and will dive into its analysis with the help of the lectures, examples from labs, and regular one-on-one meetings with the instructor.

The lectures emphasize fundamental statistical tools, visualization techniques, and data analysis principles, together with training in the essential mental factors of curiosity, imagination, and objectivity.

The course is highly interactive and customizable to individual needs and experience. In the one-on-one meetings, the instructor will make suggestions for further analysis steps, drawing from a large grab bag of possible methods, depending on the nature of the dataset and the student's background and interests.

Participants may choose to work in Matlab, Python, or — for the first time — Julia, with numerous analysis tools and detailed scripts being provided in each of these three languages. The global surface drifter dataset, the current speed from which is pictured above, will be used as an illustrative example of a large and complex dataset.

The final project consists of a written report, ideally in the form of a Jupyter notebook, documenting the participant's progress on their dataset. On the final day, students will present results from the week.

There are no constraints on the type of dataset. Large, complicated, multivariate datasets, time series data, and model output are all welcomed. Participants will benefit by taking advantage of supporting or ancillary data, rather than relying exclusively on, for example, a single measurement from a single location.

Prerequisites: Students should have prior experience in one of the programming languages the course is offered in (Matlab, Python, or Julia). We will use complex numbers to represent bivariate data, e.g. u+iv, so those not familiar with these would benefit from some reading ahead of time. For the analysis of multivariate datasets we will need to use matrices, so some prior familiarity with linear algebra is expected. Those with questions should contact the instructor at jmlilly@psi.edu. Students will not be turned away for lacking prerequisites.

Course materials: All course notes and labs can be found online at http://www.jmlilly.net/course/index.html, which will be updated prior to the course (in particular for the Julia labs). This course will focus on Part I, though individual students may be pointed to topics from Part II.