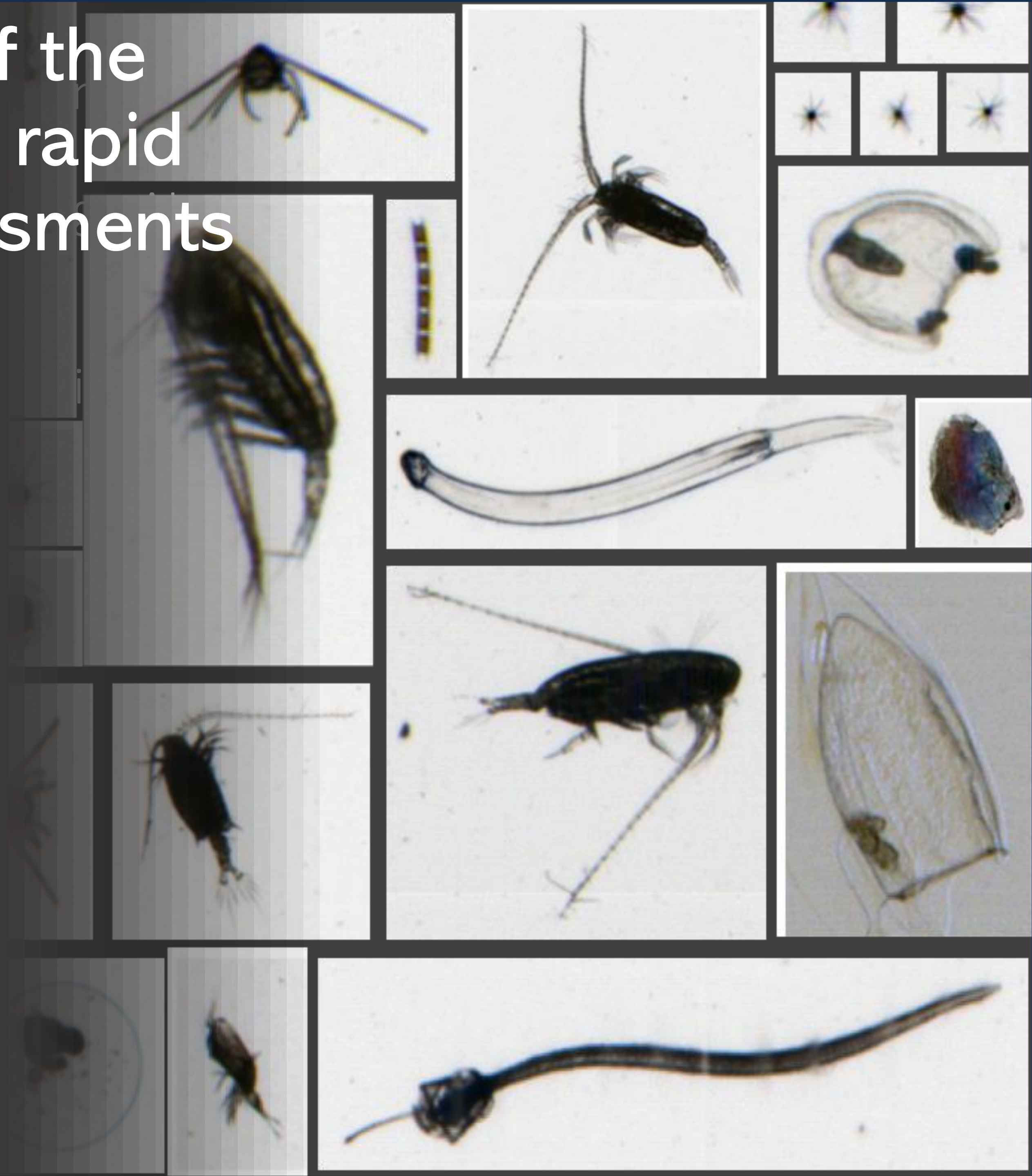
Are plankton nets a thing of the past? How we can use AI for rapid plankton and ecosystem assessments



Sophie Pitois

PICES – 2024

Oct 26 – Nov 1, 2024 Honolulu, USA

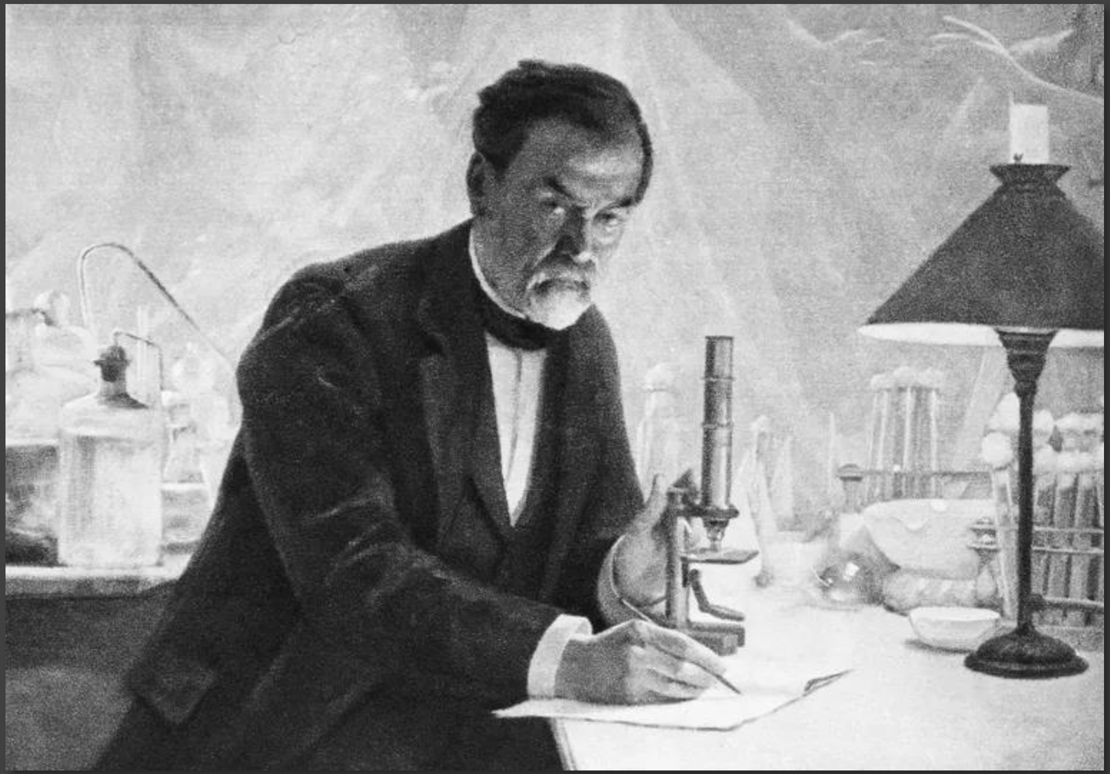
Centre for Environment Fisheries & Aquaculture Science

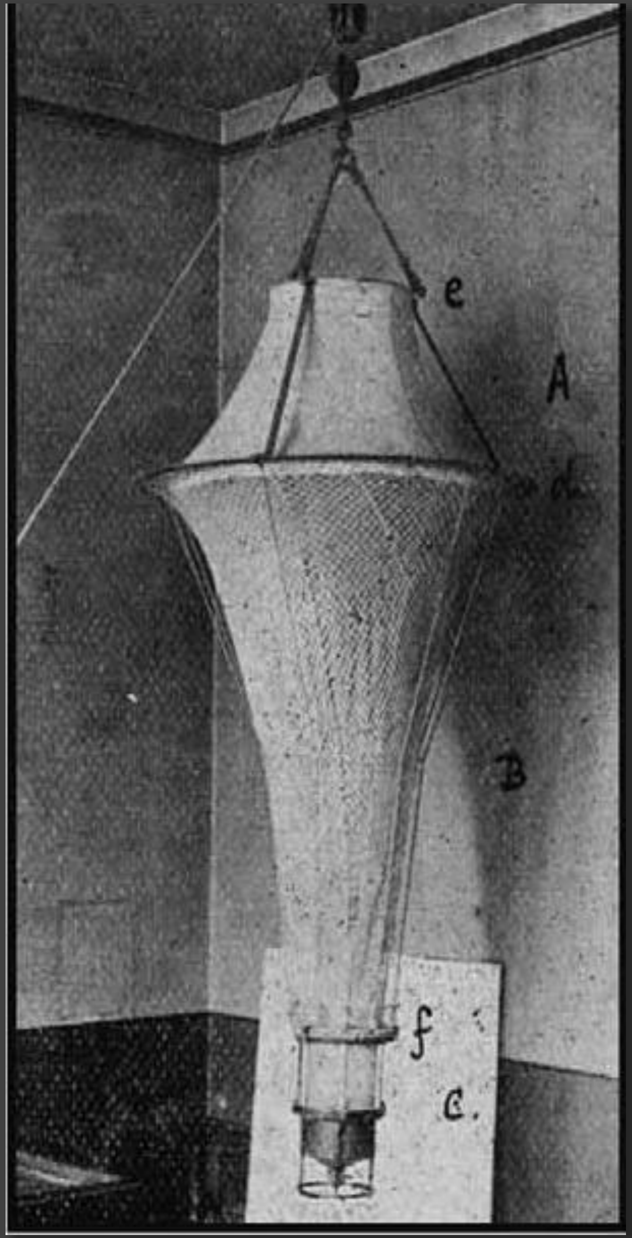
Cefas

The Alan Turing Institute

In the beginning.....

- Ring net first developed in 1880s
- Microscope analysis

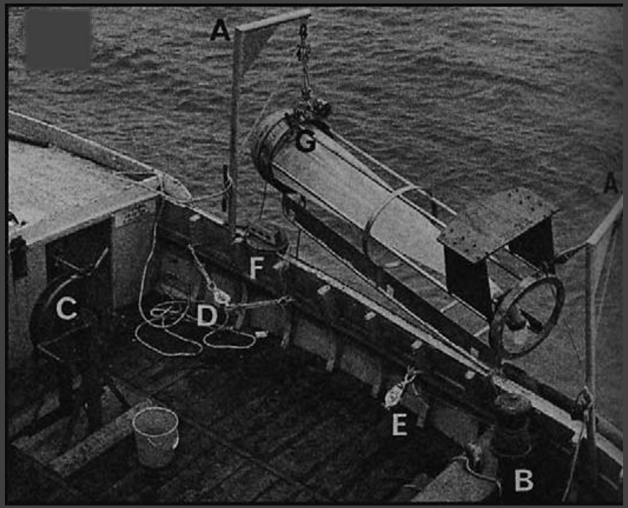




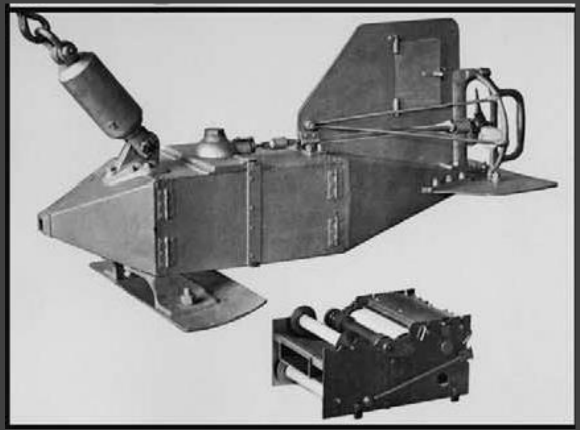
The first Hensen net developed in the 1880s (Wiebe and Benfield, 2003)

For the following 100 years.....

- Nets evolved, to include high speed samplers etc…
- CPR appeared in late 1930s
- Microscopes became more performant



Gulf high-speed sampler



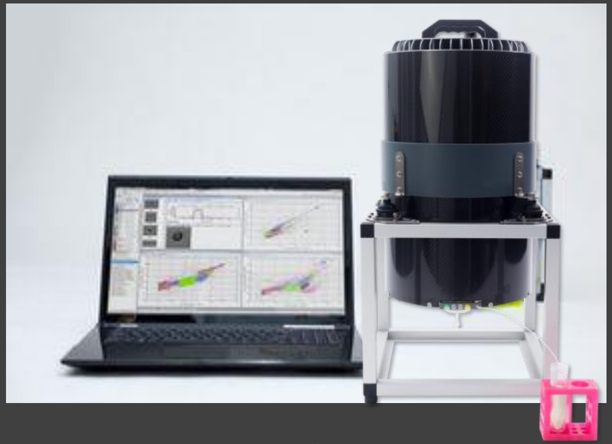
CPR in 1936 (Wiebe and Benfield, 2003)

Deployment of ring net in the 1960s and 2000s

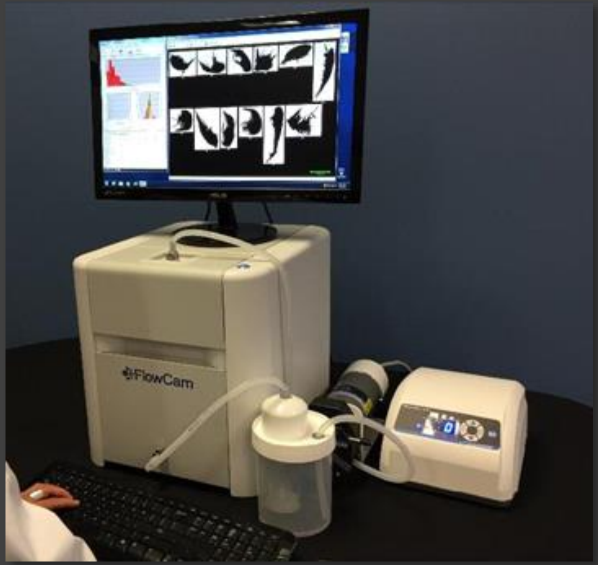
Sir Alistair Hardy looking through microscope (photo from CPR survey)

Scientist looking through microscope (2015)

Today…



CytoSense Flow cytometry

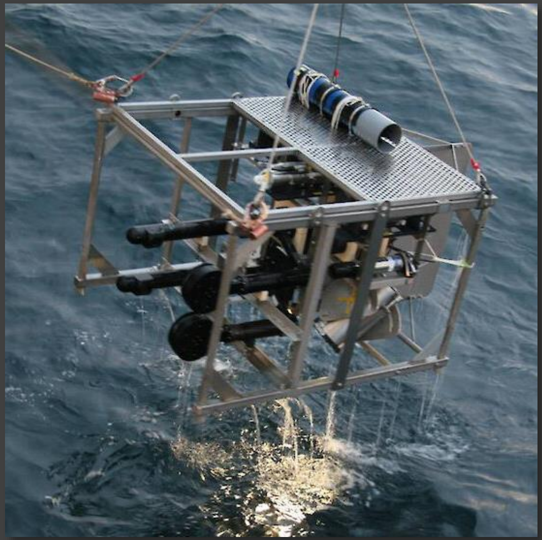


FlowCam (Fluid Imaging Techonlogies)

ZooSCAN (Hydroptic)

Plankton Imager PI-10 (Plankton Analytics)

Underwater Vision Profiler (UVP)



Video Plankton Recorder (VPR)

In Situ Ichthyoplankton Imaging system (ISIIS)

Review: Lombard et al., 2019

The Plankton Imager (PI): an all-in-one tool that combine zooplankton sampling and image analysis

Cefas

The Plankton Imager

The PI is a **high-speed line-scan** camera that **images all particles continuously** in a through-flow sampling system

Image capture

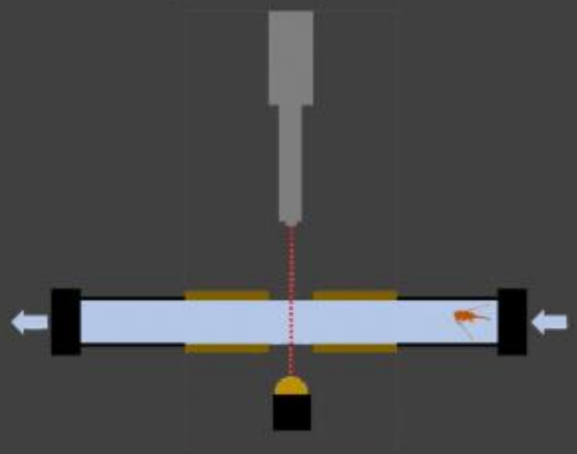
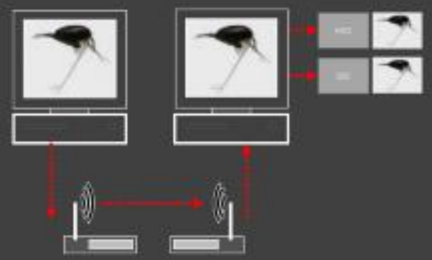


Image storage



Classification

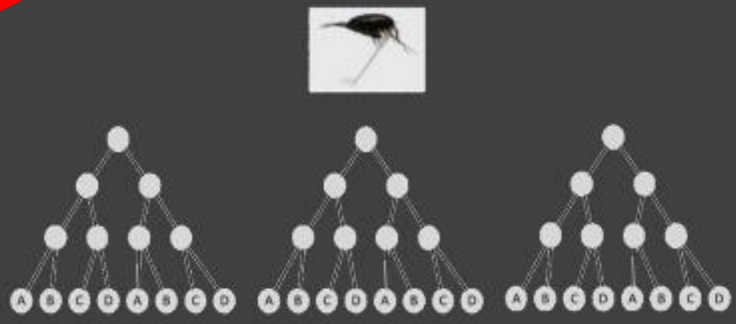


Image capture

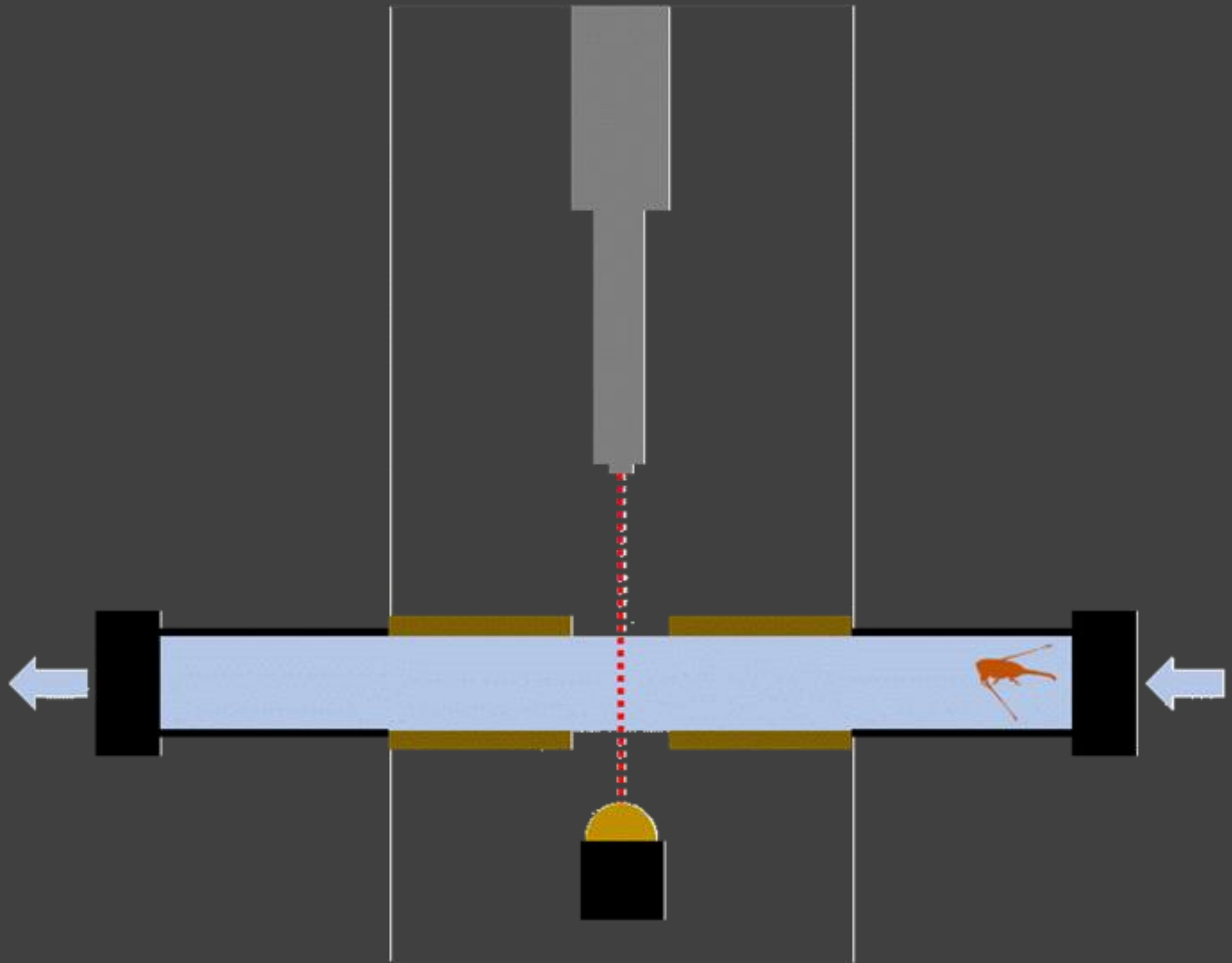
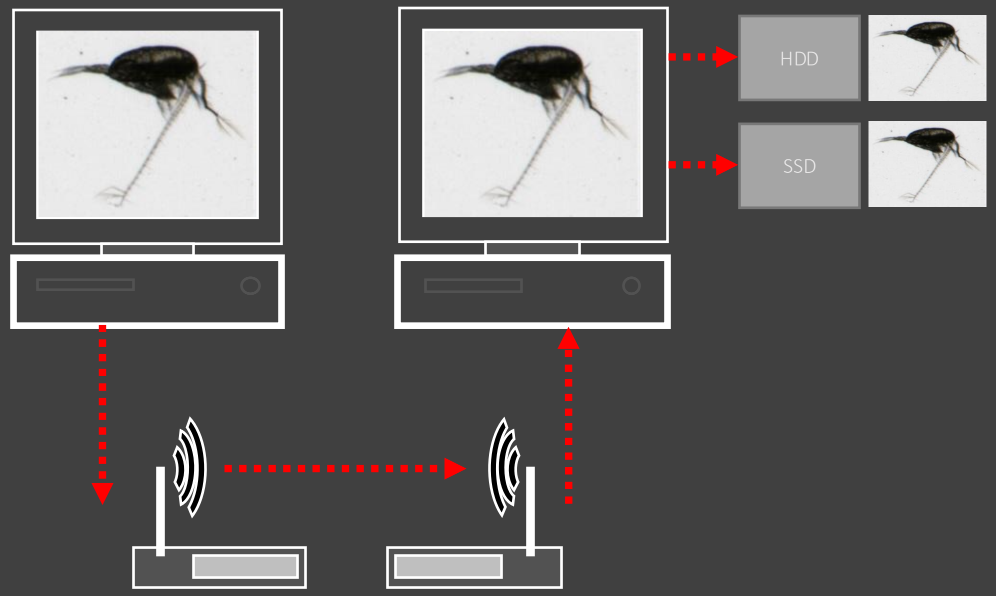
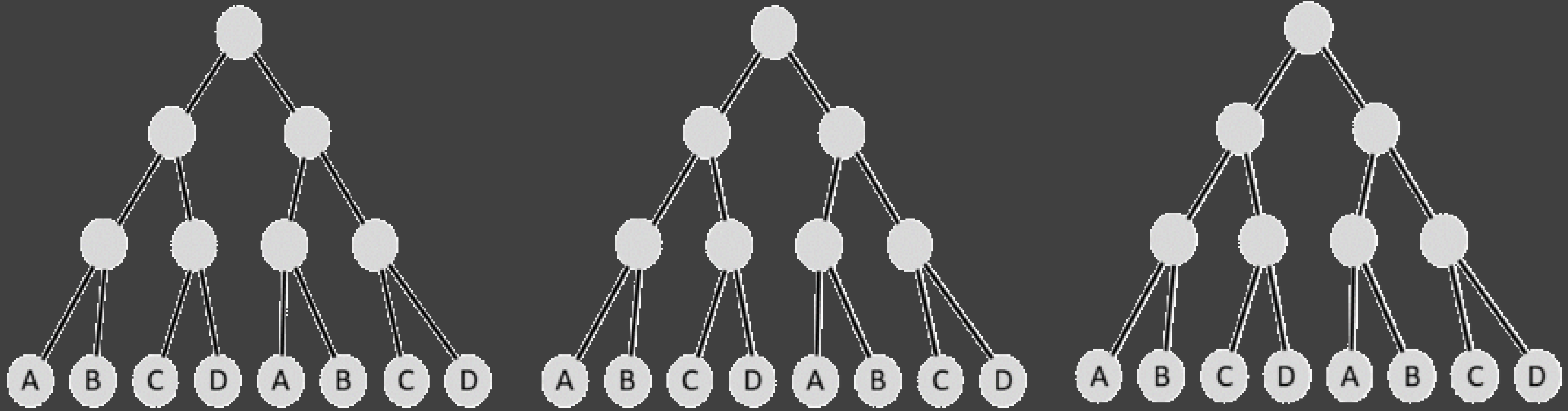
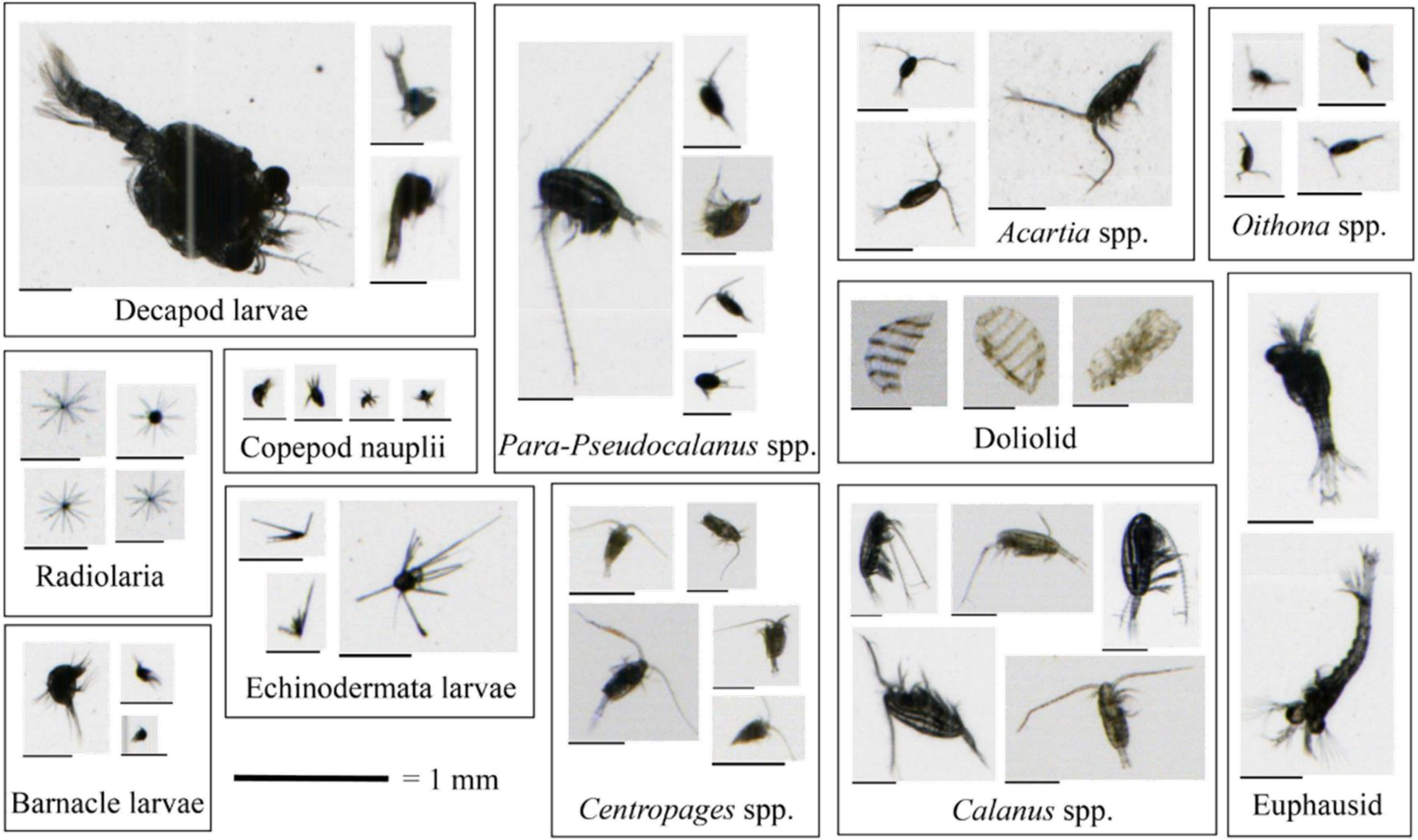


Image storage



Classification





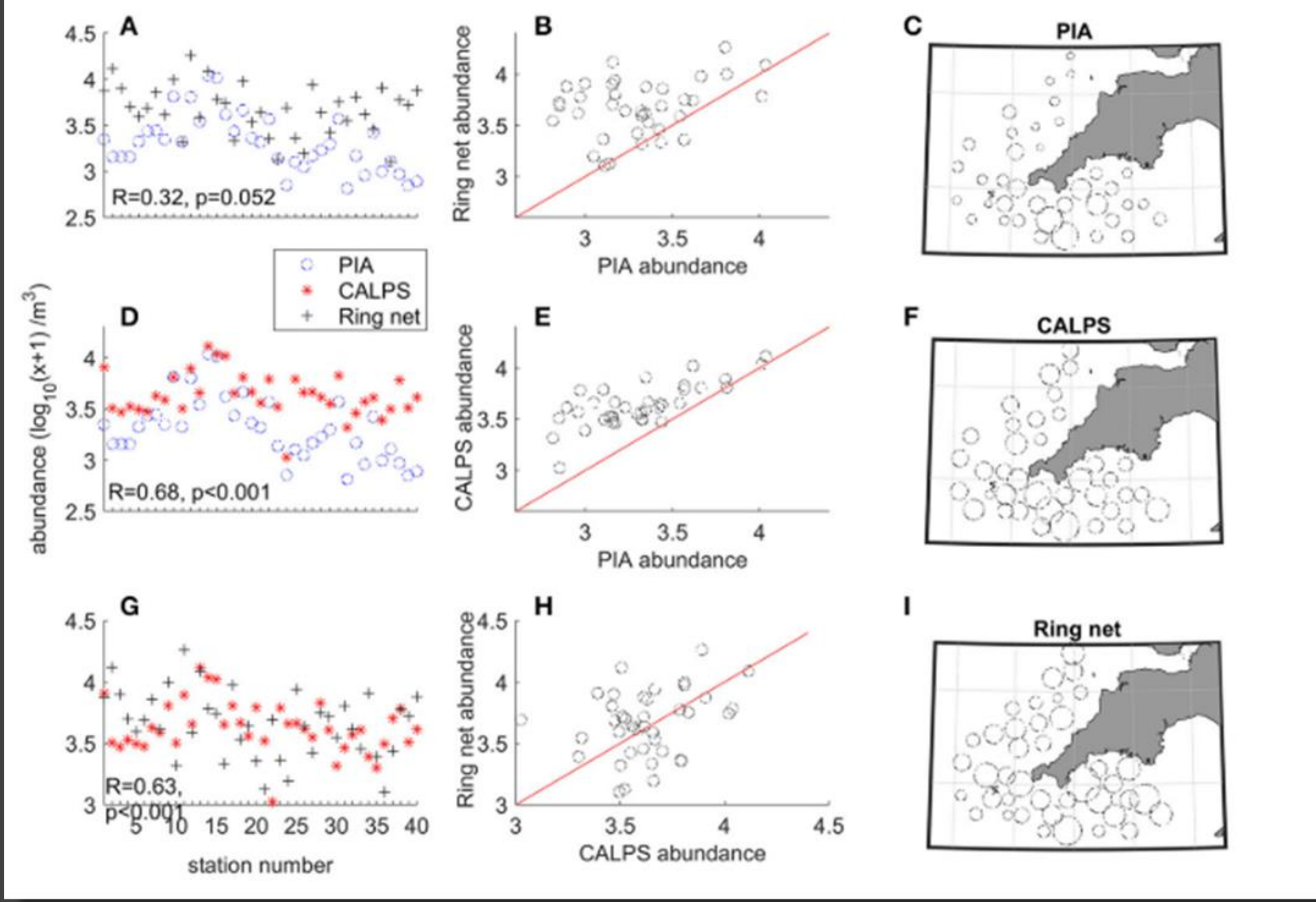
Decapod larvae
Para-Pseudocalanus spp.
Acartia spp.
Oithona spp.
Radiolaria
Copepod nauplii
Doliolid
Euphausid
Echinodermata larvae
Barnacle larvae
= 1 mm
Centropages spp.
Calanus spp.

Method in numbers

- Continuous sampling in all weather, no ‘down-time’
- Camera resolution = 10ɥm
- Current size range for recorded particles is 0.20 mm – 3.5 cm (changeable)
- Tiff metadata includes GPS and ships time
- Survey 2024: 7 days = 3 billion particles captured, 128 million images saved.

Proof of concept

- The PI provide a robust description of zooplankton
- Comparison across 3 methods
- Differences in catchability but similar patterns of distribution



Pitois et al. 2018

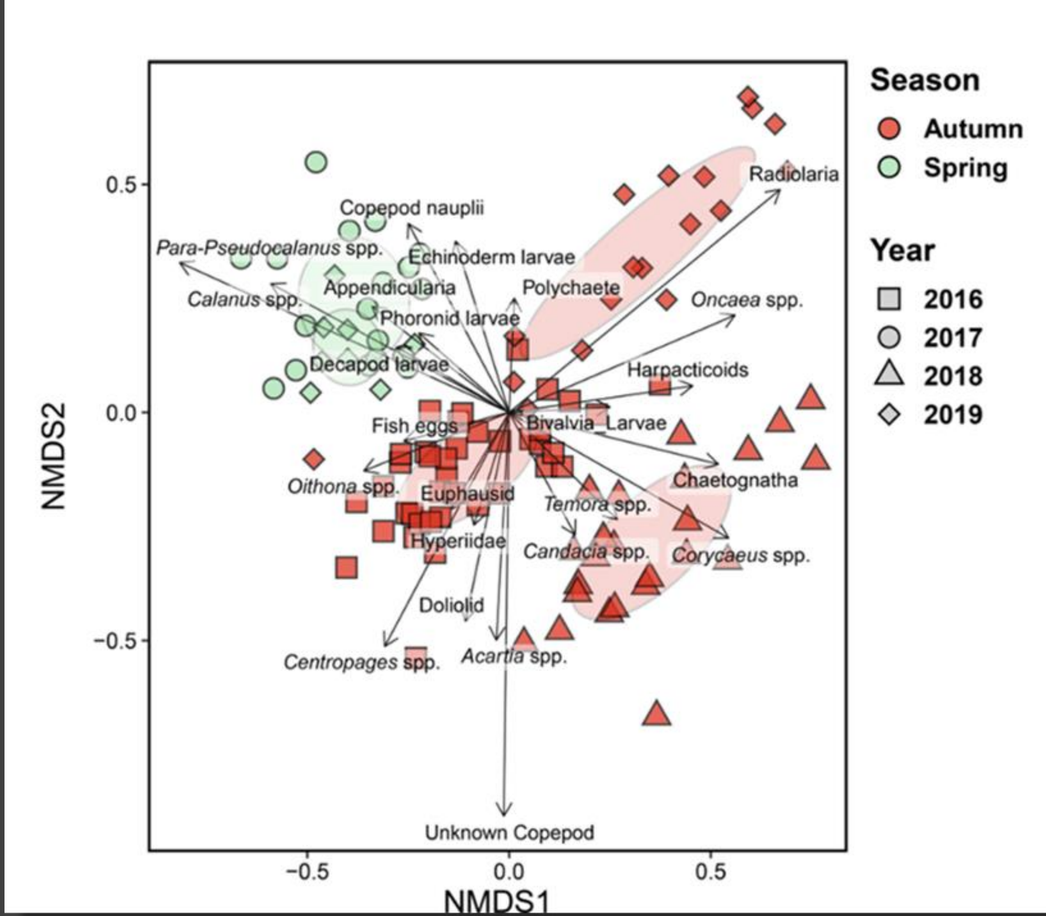
Proof of concept

- the PI can be used as part of zooplankton monitoring

- Aligns well with CPR and PML L4 station:

  - similar variability in abundances between years and seasons

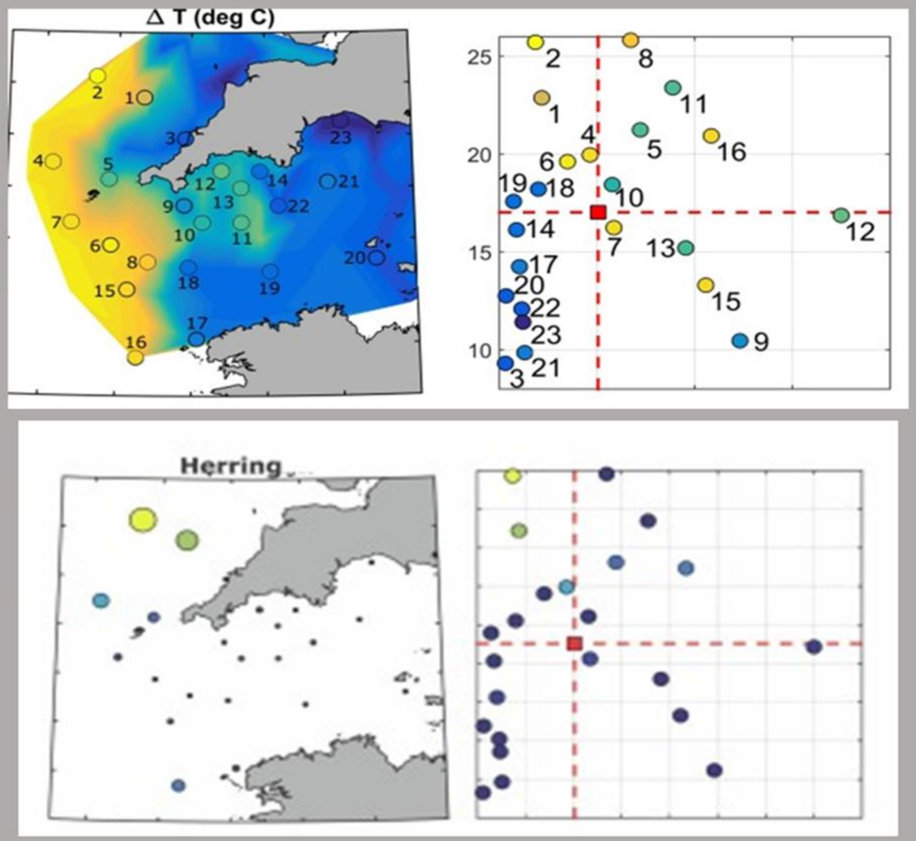
  - Similar variability in community structures between spring and autumn



Scott et al. 2021

Proof of concept

- Application of in-situ size information to ecological indicators
- Colder waters characterised with low abundances of mainly small copepods
- Herring higher biomass located in areas of where copepods are larger but not necessary more abundant than average.



Link 2D indicator with physical parameter, lower and higher trophic levels (pelagic fish)

Pitois et al. 2021

Challenges to overcome

Challenges to overcome

Need robust and quick method to identify all images

Need new methods for data intensive plankton research

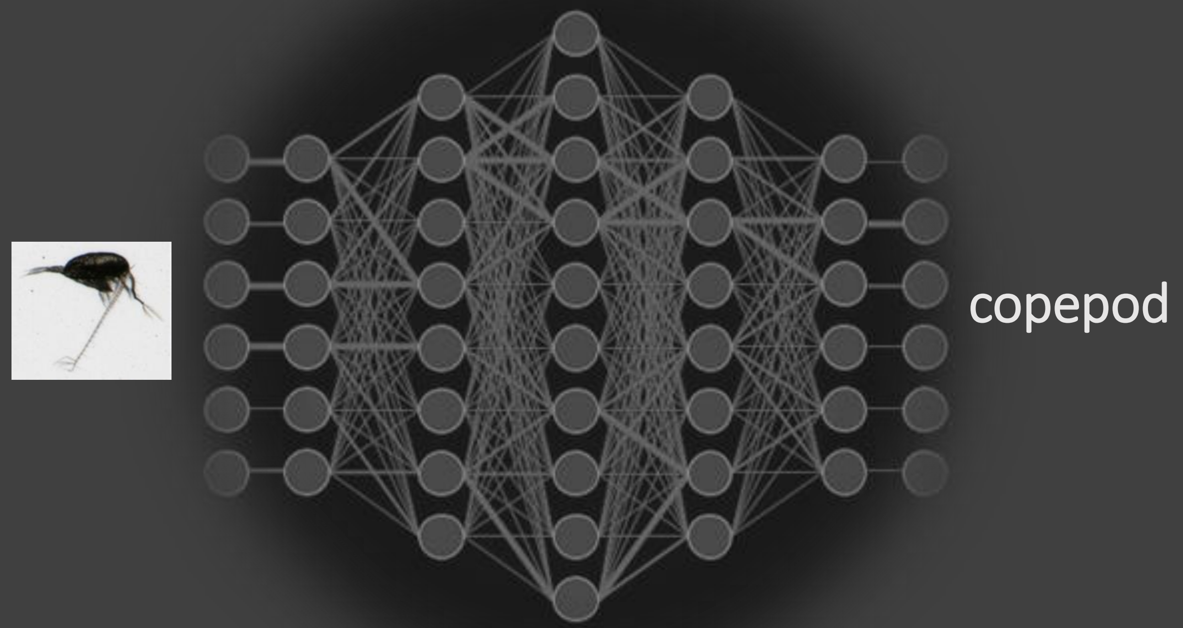
Need robust and quick method to identify all images

Need robust and quick method to identify all images

- Imbalance in the training set
- Building Training set is Resource intensive
- Manual validation required

The Alan Turing Institute

Quick and robust method to identify all images



True label \ Predicted label	Copepod	Detritus	Non-copepod
Copepod	97.10%	2.20%	0.70%
Detritus	0.80%	96.40%	2.80%
Non-copepod	8.00%	2.40%	89.60%

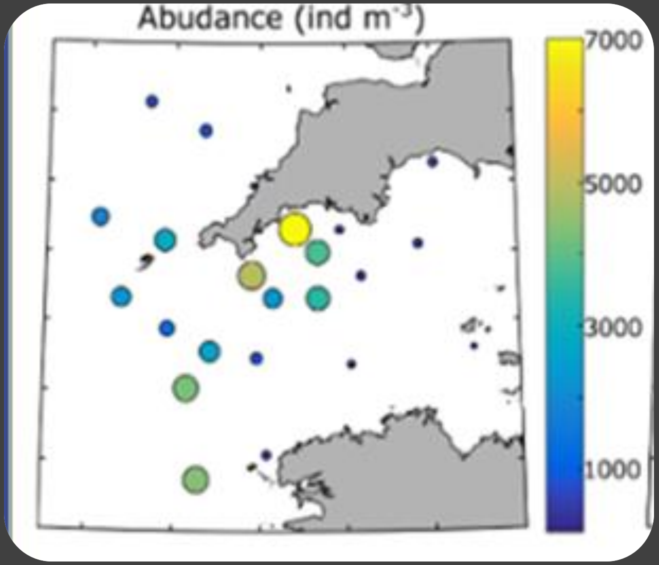
Can identity all images fast and reliably but ....

Low taxonomic resolution

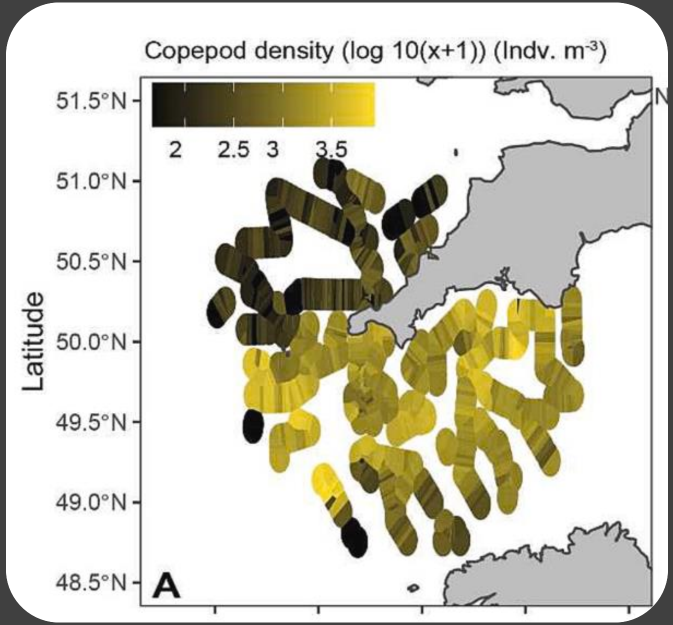
Data Study Group Final Report:
doi.org/10.5281/zenodo.6799166

From low to high spatial resolution

- From 40 to 923 stations
- PI-10 station = 10 min bin of images



40 stations *Pitois et al. 2018*



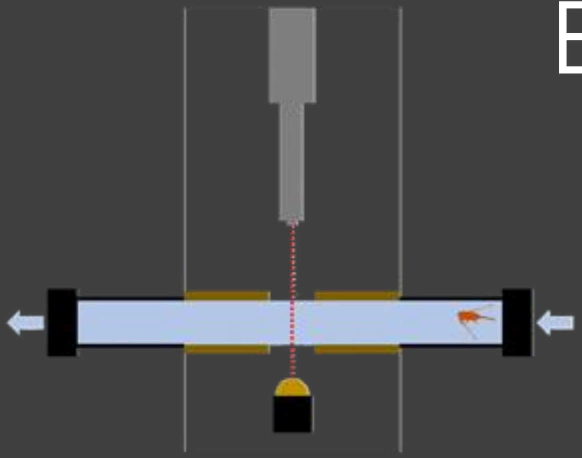
923 stations *Scott et al. 2023*

Need new methods for data intensive plankton research

Need new methods for data intensive plankton research

- Millions of images collected and analysed
- Resource intensive
- Data collection rate still faster than processing rate

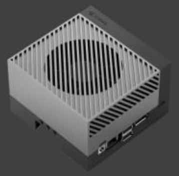
Intensive plankton research



Edge-AI

Cloud based

Real-time data pipeline



RAPID dashboard

Edge AI – e.g., NVIDIA Jetson AGX Orin running a classifier and summarising plankton data in near real time.

Ship's internet, bandwidth budget < 5K per minute

Terrestrial digital **dashboard** showing near-real time plankton abundance and distribution.

Home page - Cefas RV Dashboard (cefastest.co.uk)

Regular data pipeline

External storage captures data directly from instrument and uploaded to Azure

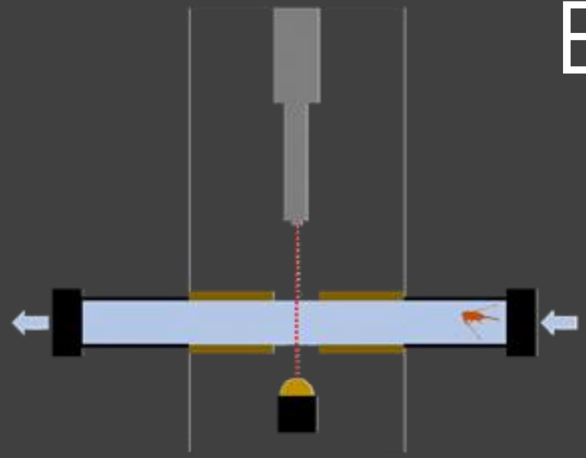
Data are classified in **Azure Machine learning** workspace, the classifier / new classifier can be run multiple times

Classification results returned to local servers for analysis

The Alan Turing Institute

Cefas

Intensive plankton research



Edge-AI

Cloud based

Real-time data pipeline

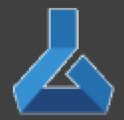
Edge AI – e.g., NVIDIA Jetson AGX Orin running a classifier and summarising plankton data in near real time.

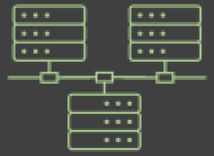
Ship's internet, bandwidth budget < 5K per minute

Terrestrial digital **dashboard** showing near-real time plankton abundance and distribution.

[Home page - Cefas RV Dashboard (cefastest.co.uk)]

Cloud data pipeline





External storage captures data directly from instrument and uploaded to Azure

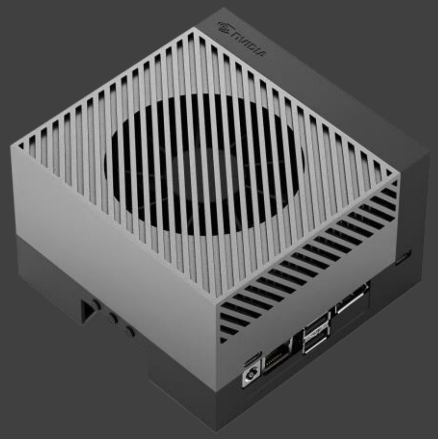
Data are classified in **Azure Machine learning** workspace, the classifier / new classifier can be run multiple times

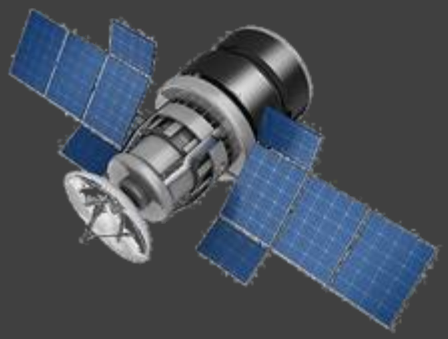
Classification results returned to local servers for analysis

The Alan Turing Institute

Cefas

Real-time data pipeline





RAPID dashboard

Home page - Cefas RV Dashboard (cefastest.co.uk)

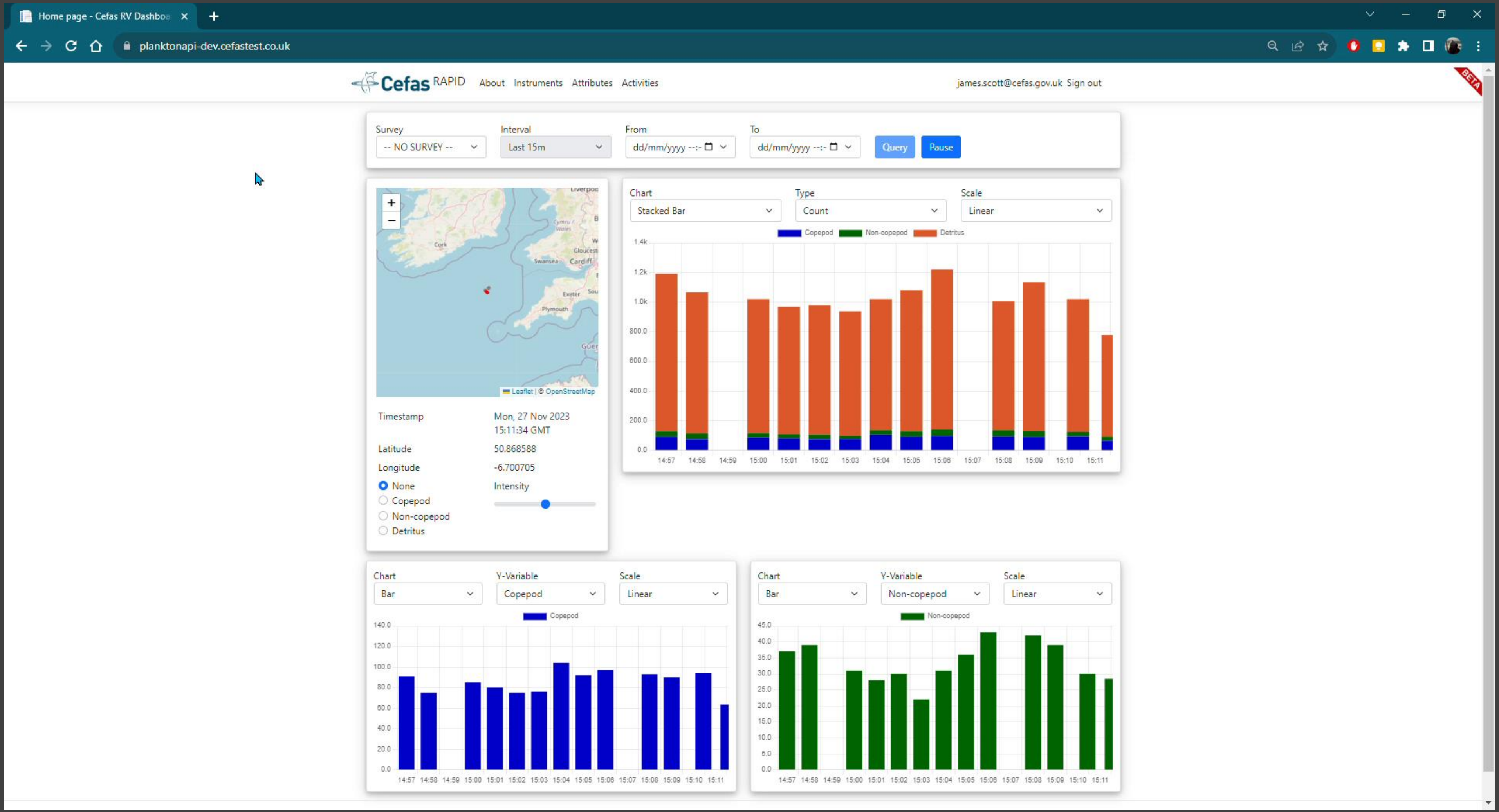
Edge AI – e.g., NVIDIA Jetson AGX Orin running a classifier and summarising plankton data in near real time.

Ship's internet, bandwidth budget < 5K per minute

Terrestrial digital **dashboard** showing near-real time plankton abundance and distribution.

[Home page - Cefas RV Dashboard (cefastest.co.uk)](cefastest.co.uk)

RAPID dashboard



[Home page - Cefas RV Dashboard (cefastest.co.uk)]

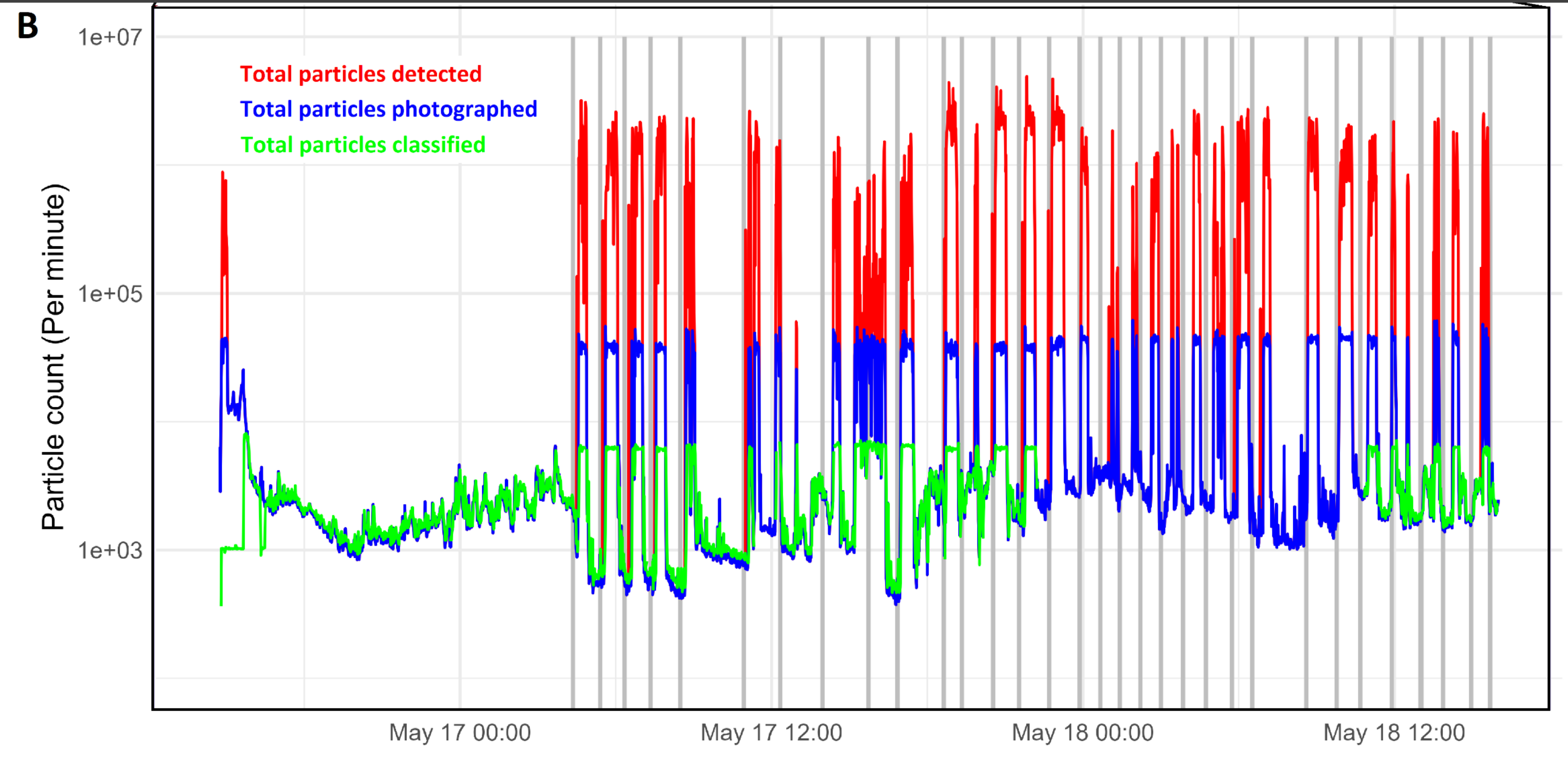
Cloud data pipeline

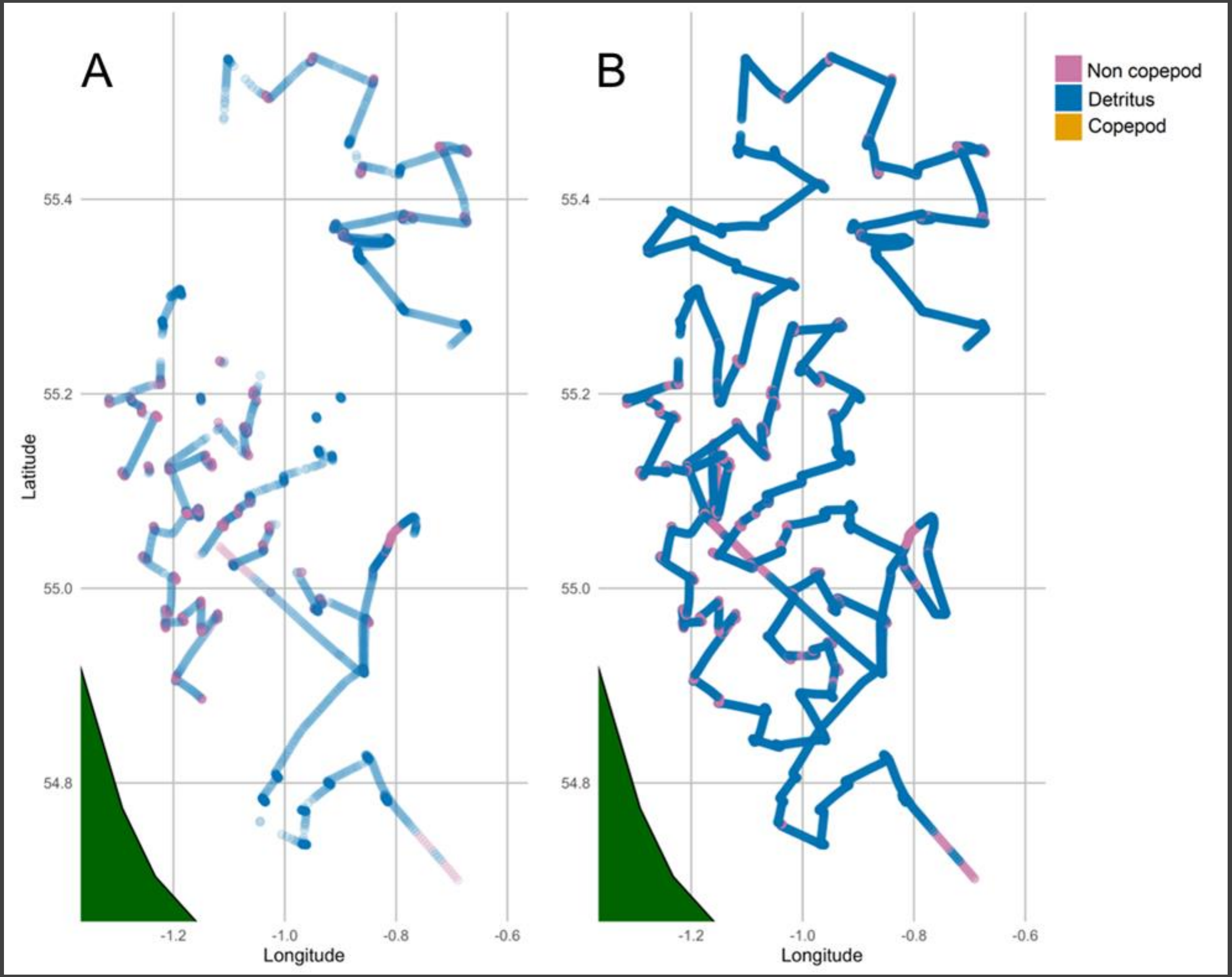
External storage captures data directly from instrument and uploaded to Azure

Data are classified in **Azure Machine learning** workspace, the classifier / new classifier can be run multiple times

Classification results returned to local servers for analysis

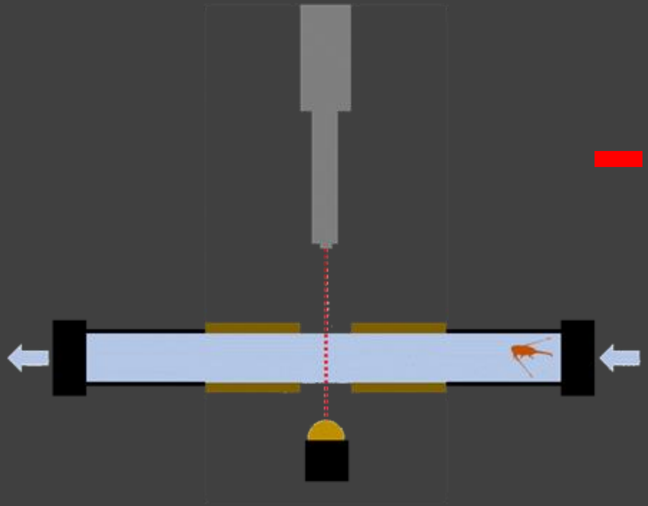
Edge AI vs Post survey computing



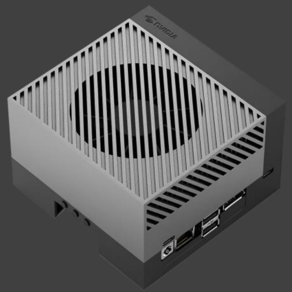


Real-time pipeline **Post-survey processing**

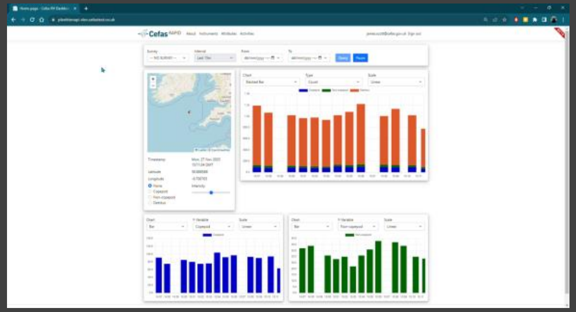
Summary



3 billion particles captured
128 million images
(4.8%)

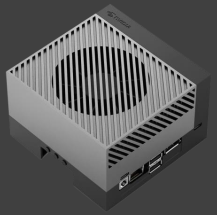


17 million images processed
(13.3%)



15 million datapoints visualised
87%

128 million images processed
(100%)



Edge AI	Cloud compute
Real-time or near real-time	Can take days / weeks to run survey
Low taxonomic resolution (broad groups)	Expanded algorithm: detailed taxonomic resolution
No computing cost	cloud computing cost
Images not saved	Archive for future ML improvements methods
processing limit (8,000 / min)	All data collected

Use of AI for future integrated monitoring

- Development of real time monitoring for zooplankton:
  - Automated 24/7 collection of images
  - Edge-AI for fast processing and data transfer
  - Near real time visualisation of zooplankton abundances, size distribution etc…
  - Adaptive sampling
  - Cloud computing for in depth analysis

- Expand to other components of the pelagic system
  - Physical parameters
  - Phytoplankton
  - Pelagic fish from acoustics

Automated Sampling

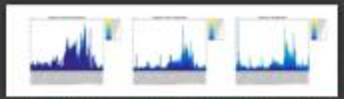
Continuous flow

Flowcytometer
Phytoplankton

Connected to ship water supply

Phytoplankton every 30-60 mins

Particle sizes and pigmentation



FerryBox
Physical parameters

Live data for
- Temperature
- Salinity
- Fluorescence
- Oxygen
- Turbidity

The Plankton Imager

The PI is a **high-speed line-scan** camera that **images all particles continuously** in a through-flow sampling system

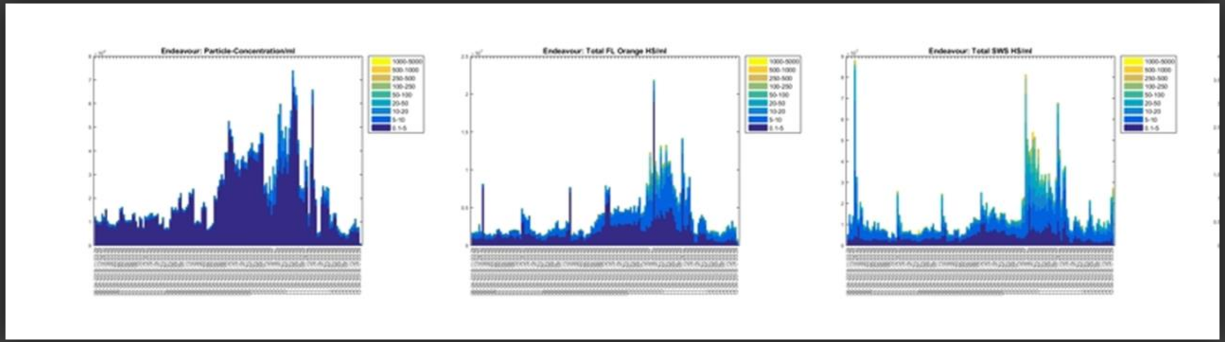
Flowcytometer

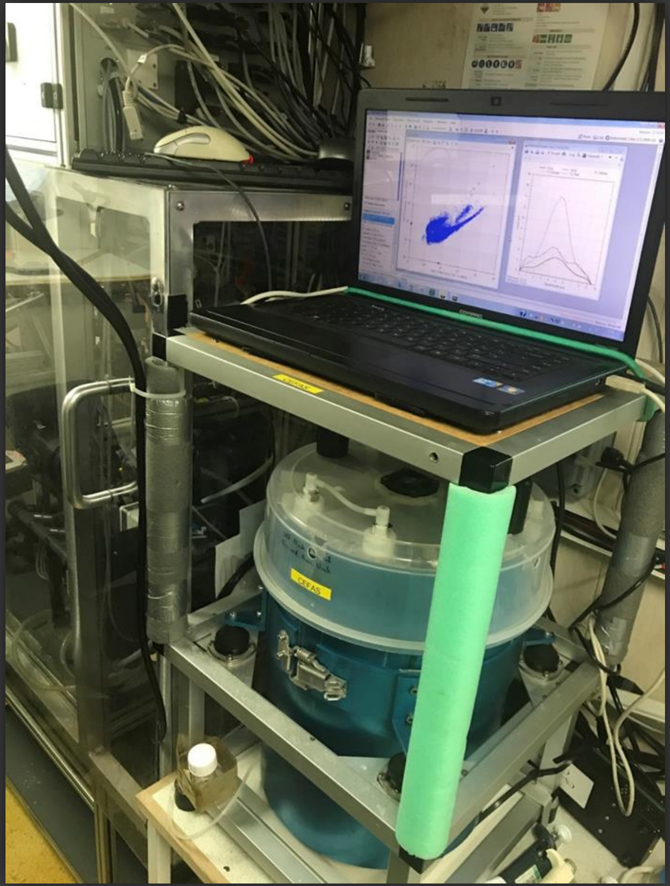
Phytoplankton

Connected to ship water supply

Phytoplankton every 30-60 mins

Particle sizes and pigmentation





Number of particles/ml and total red and orange fluorescence/ml per size fraction

FerryBox
Physical parameters

Live data for
- Temperature
- Salinity
- Fluorescence
- Oxygen
- Turbidity

Automated Sampling

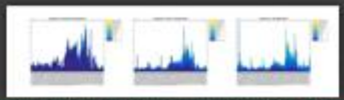
Continuous flow

Flowcytometer
Phytoplankton

Connected to ship water supply

Phytoplankton every 30-60 mins

Particle sizes and pigmentation



FerryBox
Physical parameters

Live data for
- Temperature
- Salinity
- Fluorescence
- Oxygen
- Turbidity

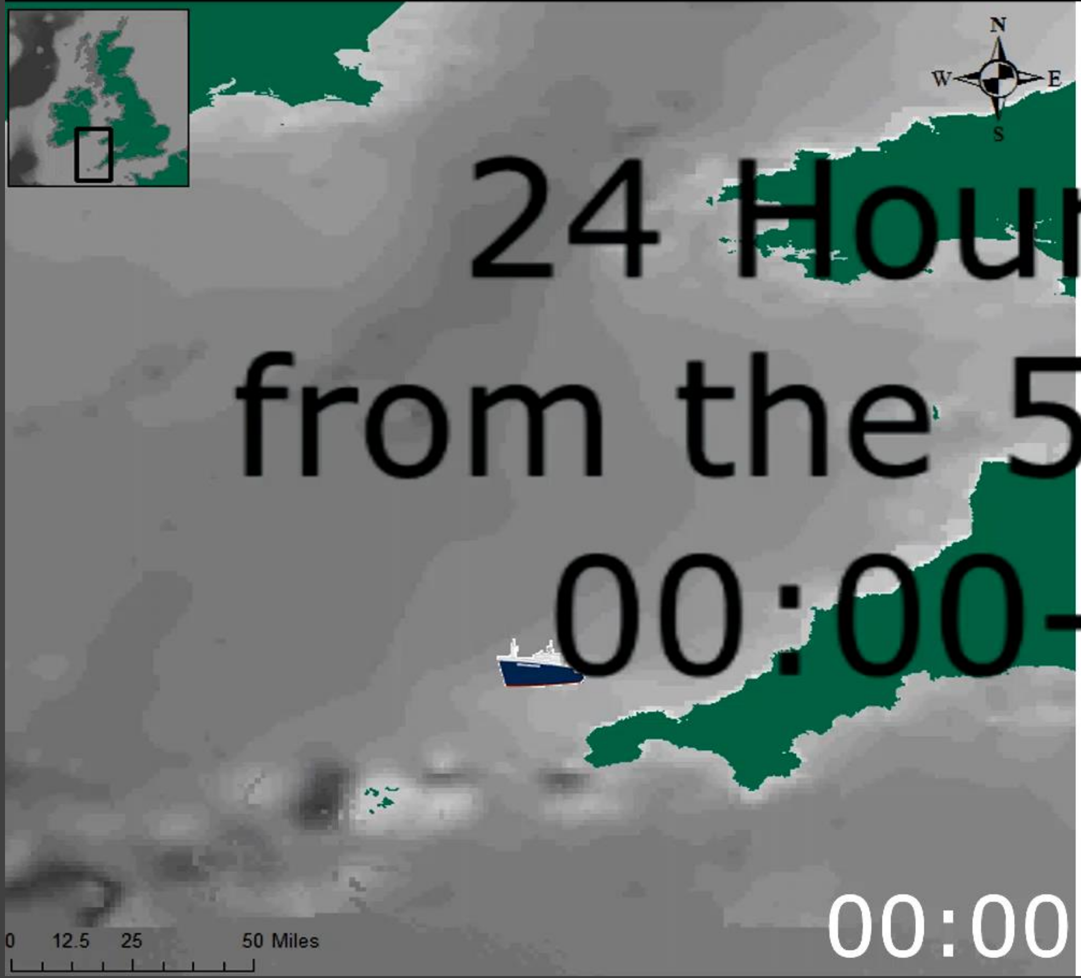
The Plankton Imager

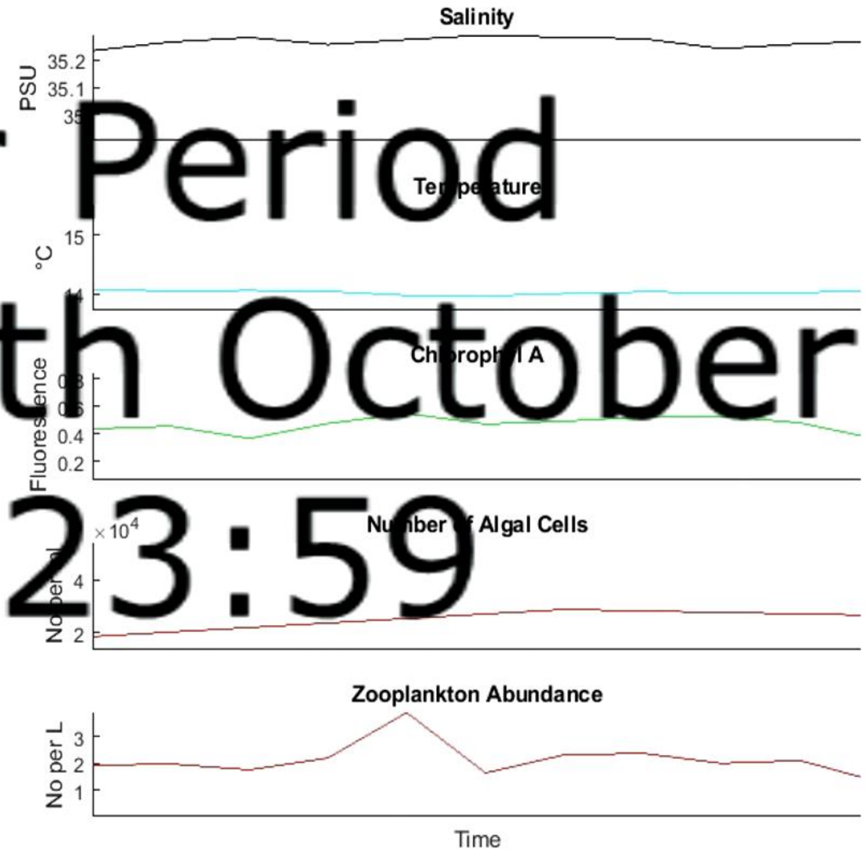
The PI is a high-speed line-scan camera that images all particles continuously in a through-flow sampling system

Data from the 5th October

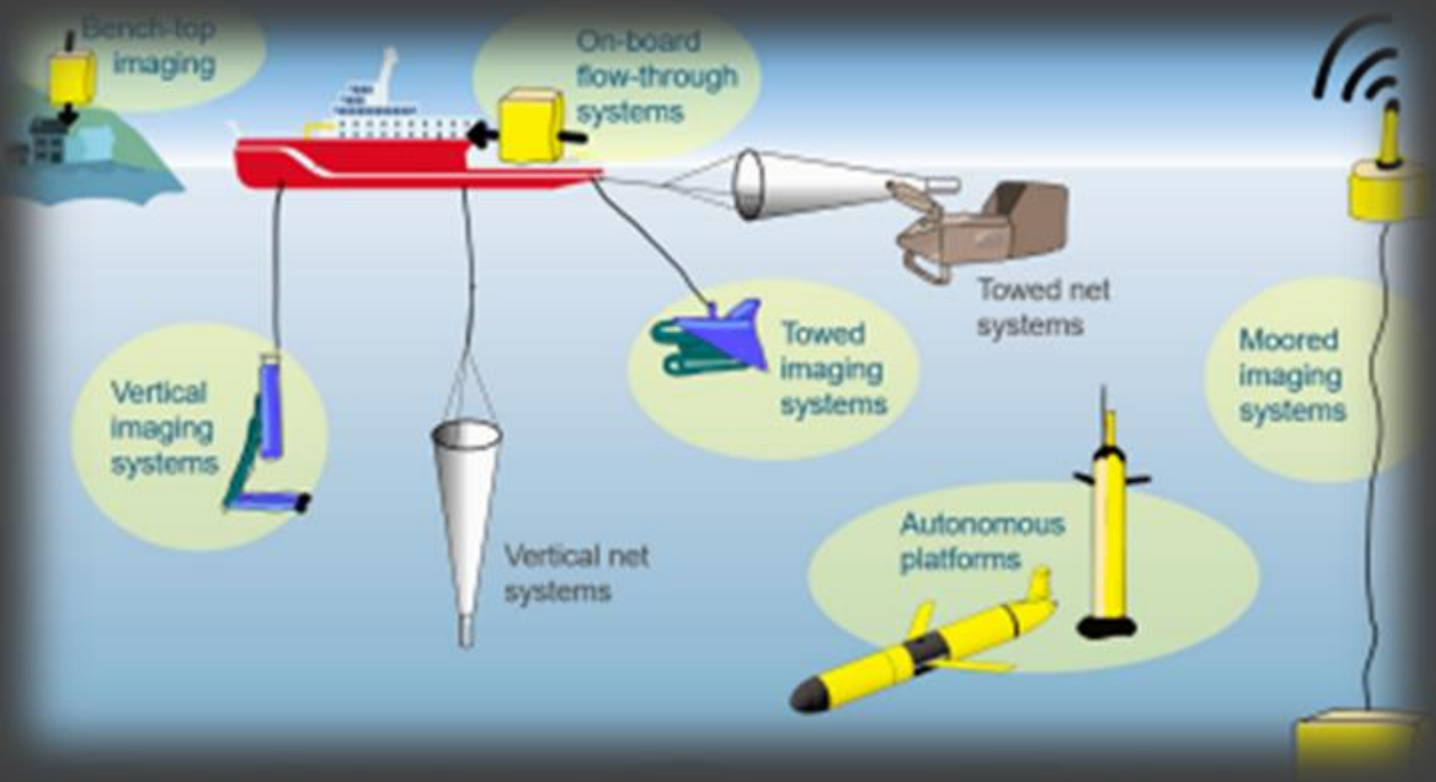
Cefas

24 Hour Period from the 5th October 00:00-23:59





AI to change how we conduct monitoring



Giering et al. Front. Mar. Sci., 2022

Many technologies

Big data analytics for multimodal data

Exponential rise in AI capabilities to provide new possibilities for fully integrated monitoring of the pelagic environment

Are nets a thing of the past?

Thank you !

Joseph Ribeiro (Data scientist, Cefas)
James Scott (Plankton ecologist, Cefas)
Hayden Close (Marine scientist, Cefas)
Eric Payne (Software engineer, Cefas)
James Pettigrew (taxonomist, Cefas)
Nevena Almeida (Taxonomist, Cefas)

Robert Blackwell (Data scientist, Alan Turing Institute)
Noushin Eftekhari (Data scientist, Alan Turing Institute)

Sari Giering (National Oceanographic Centre)
Mojtaba Masoudi (National Oceanographic Centre)

The Alan Turing Institute

Centre for Environment Fisheries & Aquaculture Science

Cefas