Statistics and Data Science

AgriFoRwArdS Summer School 05/07/2022

David Maxwell, Data Science & Statistics Team Leader

Together we are working for
a sustainable blue future

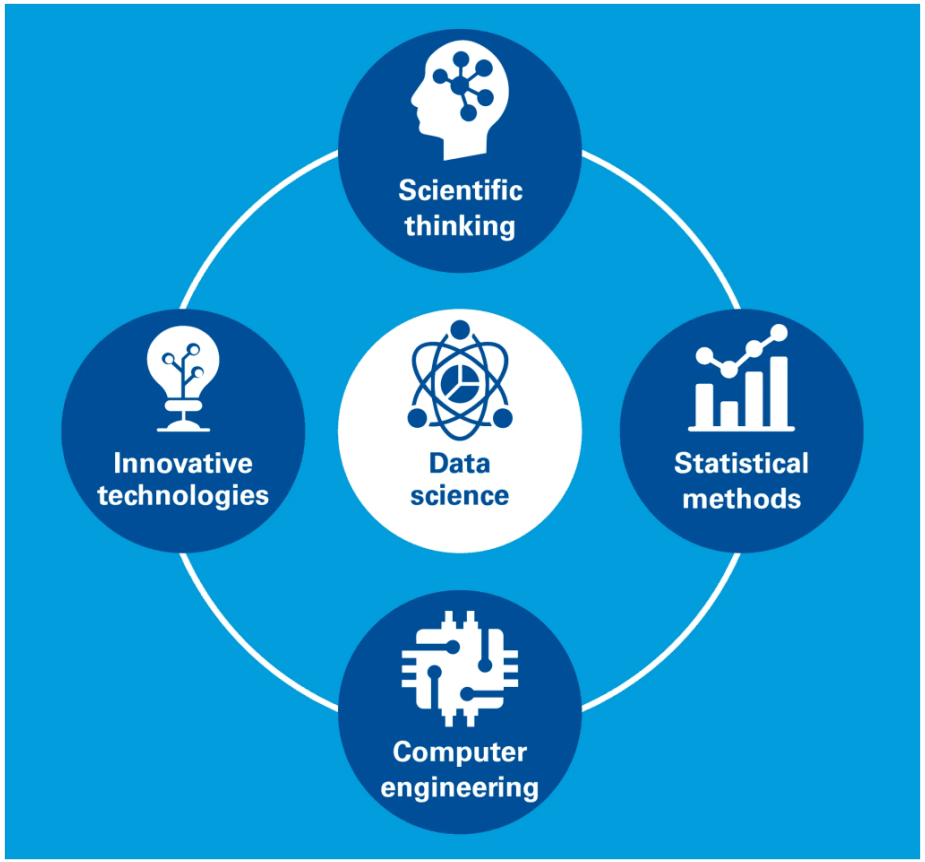
Centre for Environment
Fisheries & Aquaculture
Science

Cefas

Data science and statistics

*Data science combines the **scientific method**, maths and **statistics**, specialized programming, advanced analytics, AI, and even storytelling to uncover and explain insights from data.*

(IBM)



- Links to your skills, current and future work

Cefas Data Science and Statistics Team - Backgrounds

Current and previous team members subject backgrounds include:

Statistics (environmental & medical), Biometry, Computing, Physics, Oceanography, Ecology, Evolutionary Biology







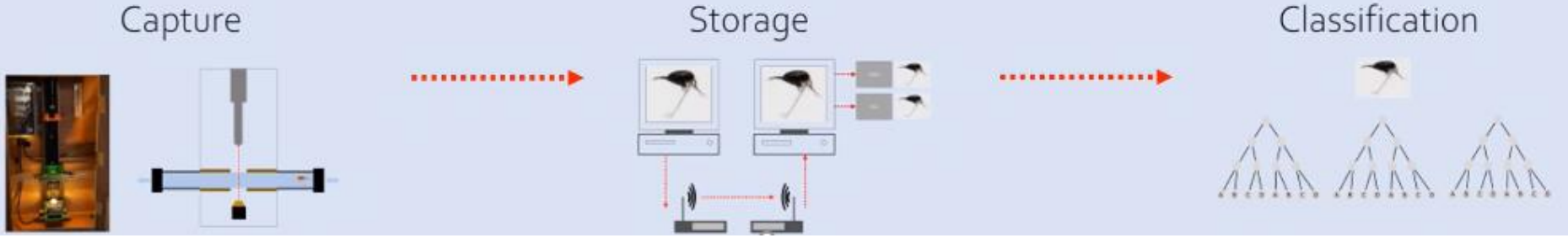


Data science - Digital Twins

Digital Twins of the Ocean

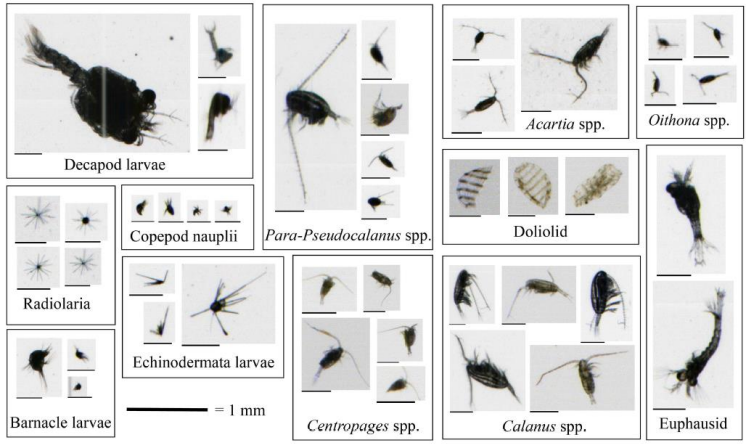
Data science – plankton imager

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



Cefas working with Plankton Analytics and The Alan Turing Institute to automatically id images.

Convolutional neural network.



Statistics

Official statistics,
reporting & visualisation

Applied statistics,
design & analysis

Method development,
Statistical theory

Cefas Statistics - examples

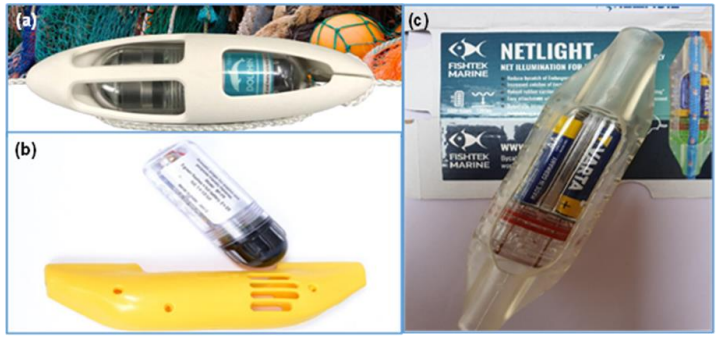
- Automated reports for Seychelles Fishing Authority

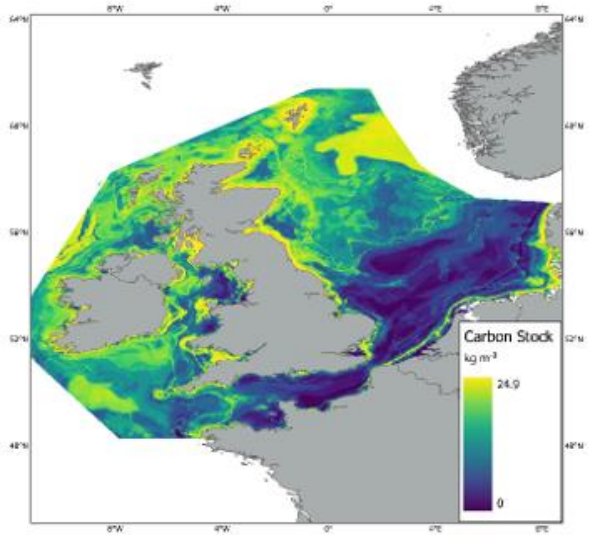
- Streamlining processes for salmon assessment data

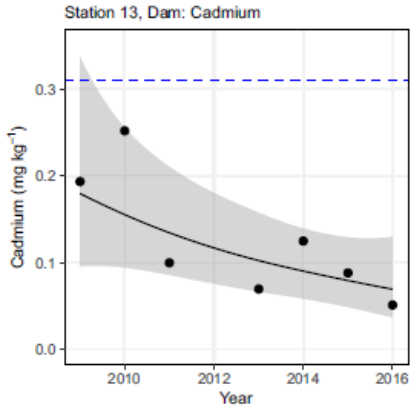
➡ Reproducible analytical pipelines

Cefas statistics - examples

- Design – bycatch mitigation trials
- Carbon maps (boosting, cross-validation with buffer)
- Contaminant trends in coastal and marine sediments of Bahrain (generalised additive models)
  Nicolaus et al. (2022) 10.1007/s10661-021-09722-7



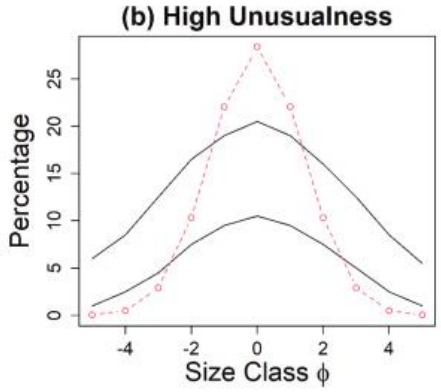


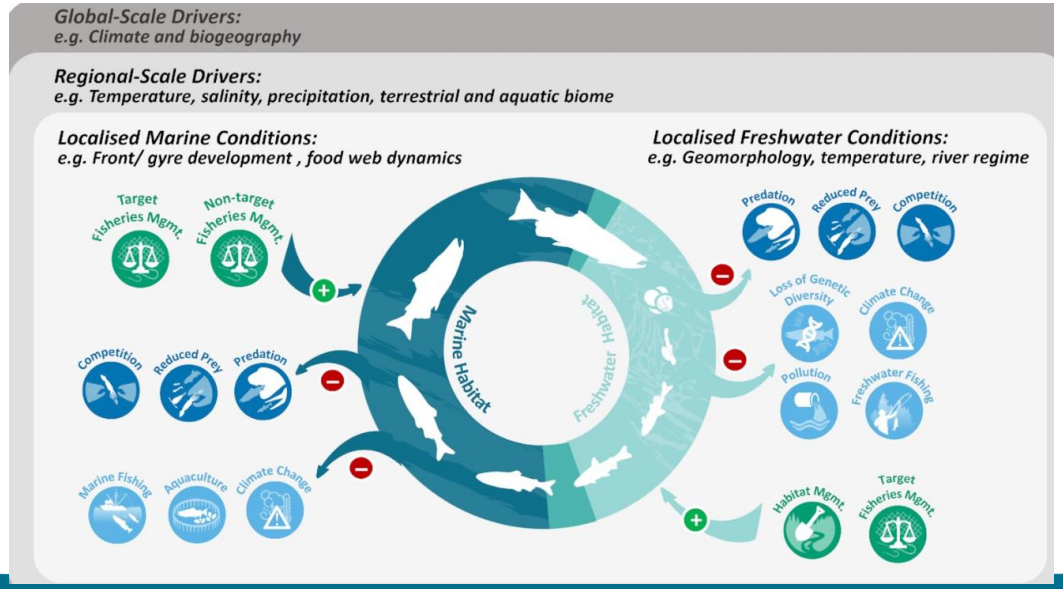


Cefas statistics - examples

- Methods for comparing sediment compositions
  Barry et al. (2021) 10.1016/j.csr.2021.104548

- Life cycle models and ecosystem-based management Bull et al. (2022) 10.1093/icesjms/fsac099





Statistics - Design

Applications from Rothamsted field trials to design of computer experiments

Plenty of guidelines and tools available in different fields e.g.

Data Quality Objectives

PREPARE guidelines https://norecopa.no/prepare

Experimental design assistant https://eda.nc3rs.org.uk/

Clinical trials toolkit

- Objectives, hypothesis, potential drivers
- Defining experiment unit and levels of variation
- Avoiding confounding
- Before-After-Control-Impact
- Blocking and randomisation
- Missing values
- Sample sizes - statistical power and precision
- Simulation – our version: emon R package

Statistics and data science – consultancy and context

Trend modelling example

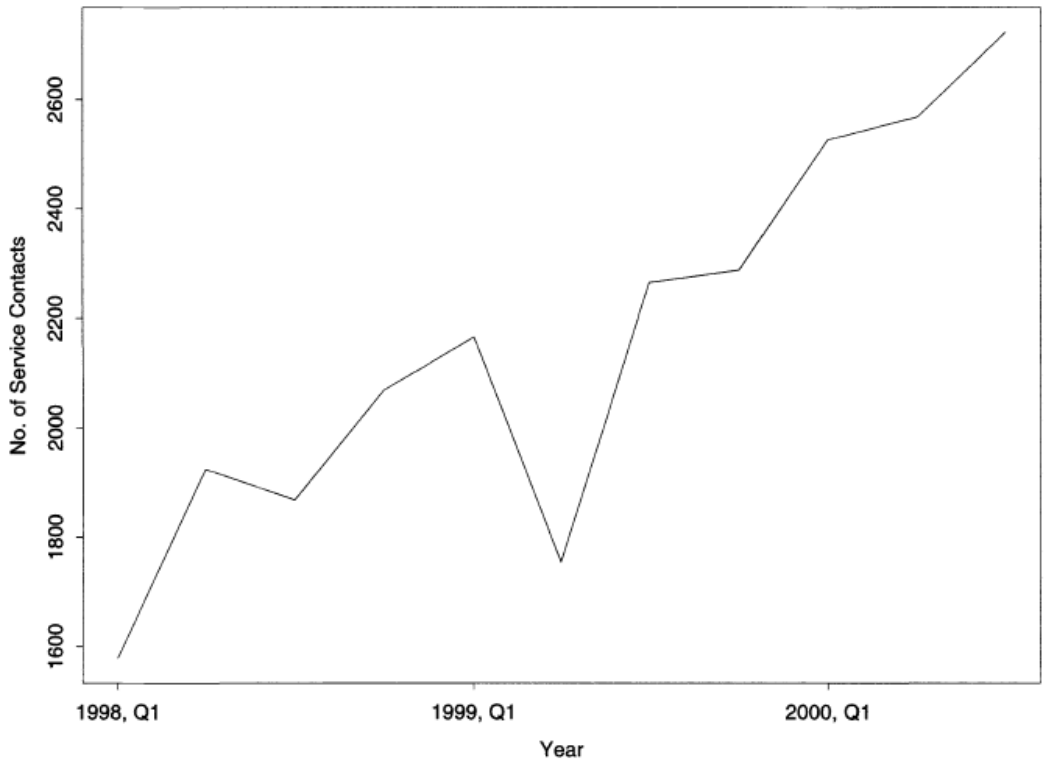
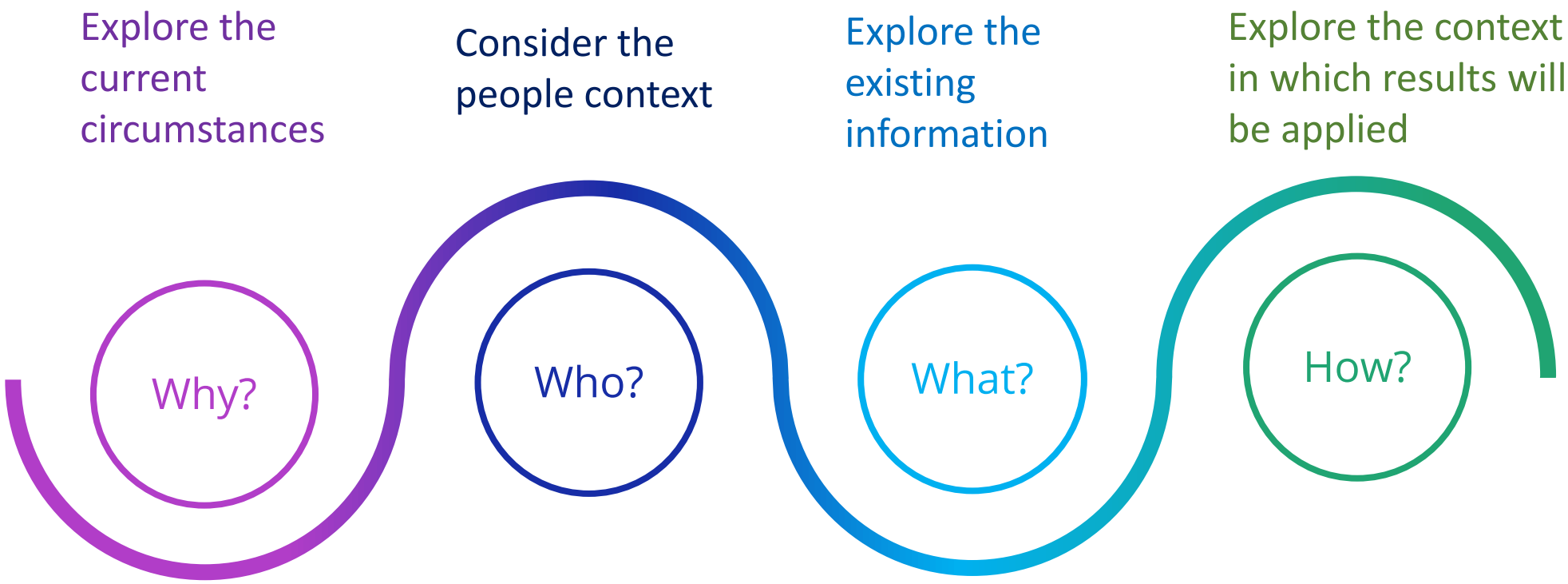


Fig. 1. Number of services required for a particular type of consumer durable in successive quarters from 1998, quarter 1, to 2000, quarter 3

Chatfield (2002) Confessions of a pragmatic statistician.

Statistics and data science – consultancy and context



Technical expertise plus teamworking skills

Thank you for listening

You are welcome to contact us for further information

david.maxwell@cefas.co.uk
robert.blackwell@cefas.co.uk

Follow @CefasGovUK

Visit cefas.co.uk

Subscribe to our newsletters

Together we are working for
a sustainable blue future

Centre for Environment Fisheries & Aquaculture Science

Cefas