Numerical ecology and time series analysis of marine proxy data

Sofia Ribeiro[1], M. Heikkilä[2], K. Weckström[1,2] and A. Pieńkowski[3]

ACME workshop, Copenhagen, Denmark, 14–16 November 2022

The ACME (Arctic Cryosphere Change and Coastal Marine Ecosystems; pastglobal-changes.org/acme) working group held its second workshop from 14–16 November 2022 at the Geological Survey of Denmark and Greenland (GEUS), in Copenhagen, Denmark. The workshop "Numerical ecology and time series analysis of marine proxy data" focused on statistical treatment of sedimentary proxy data and correct data handling as a foundation for accurate and robust paleoenvironmental and paleoclimatic reconstructions.

The overarching aim of ACME is to assess and refine available marine proxies in the Arctic coastal zone. A particular focus is placed on the techniques (field, analytical and numerical) and the quality of data, on the establishment of new community-driven protocols and on the training of early-career researchers (ECRs). The ACME Community Survey, carried out in late 2020, revealed that nearly half (45% of respondents) of the community consider that we do not, yet, have adequate methods for the quantitative determination of widely reconstructed environmental parameters, and over half (65% of respondents) consider that we do not currently have an adequate understanding of proxy behavior (Heikkilä et al. 2022). There seems to be a misalignment between the statistical knowledge of method users and method developers, which often precludes a more critical evaluation of the approaches used by the community.

This ACME workshop aimed to overcome this challenge by inviting an expert in numerical ecology with extensive experience in paleoecological proxy data and paleocommunities. Dr. Gavin Simpson, professor in applied biometrics at Aarhus University, Denmark, has over two decades of experience in quantitative environmental science across timescales, with a specific focus on aquatic environments and complex environmental problems. He has developed several R packages, and worked with large databases (e.g. Neotoma).

During the workshop, Dr. Simpson shared his knowledge, discussed possibilities and common pitfalls related to numerical approaches in paleoecology, and here, specifically, coastal paleoceanography. In addition, hands-on tutorials and computer exercises using representative coastal data sets were provided, to serve as a real-world training, especially directed at ECRs with some prior experience in using R software.

The workshop included 22 participants from eight different countries (Canada, Denmark,

Figure 1: Diagram showing the workflow in data science. Image credit: Gavin Simpson and Steve Juggins.

Finland, Iceland, Italy, Poland, Portugal and UK) of which 16 (73%) were ECRs. The majority of participants were female (68% female/ 32% male). Altogether, the participants covered a broad range of proxy methods, ranging from different groups of microfossils (diatoms, dinoflagellate cysts, foraminifera), biogeochemical markers, and environmental DNA (eDNA). The participants unanimously expressed the wish to become more familiar with data handling, visualization, and numerical methods used to interpret and reconstruct past climate and environmental changes based on proxy data.

The first day of the workshop (14 November) was dedicated to an overview of R studio and data handling in R, including the Tidyverse collection of packages, importing data, data wrangling, and visualization. At the end of the first day, Dr. Nicolas van Nieuwenhove (University of New Brunswick, Canada) gave an invited lecture on the Neotoma database and efforts to migrate the widely used Northern Hemisphere modern dinoflagellate cyst reference dataset there. He provided a historical perspective on the use of dinoflagellate cysts in Northern Hemisphere paleoceanography and an overview of the current practices using quantitative methods.

The second day of the workshop was dedicated to multivariate ordination methods (unconstrained and constrained ordination), generalized additive models (GAMs), and their use in ecological modeling, in particular to identify rates and periods of change in time series. The participants had the opportunity to work with their own datasets and discuss the outcome of the analyses.

The third and last day of the workshop was dedicated to quantitative paleoenvironmental reconstruction methods and the assessment of reconstruction validity and accuracy. The different types of paleoecological transfer functions were covered, including weighted averaging-based regressions and the modern analogues technique, as well as evaluation of model performance and diagnostics.

Key take home messages from this workshop were the importance of carefully considering study and sampling design ahead of numerical analyses (e.g. number of samples, replicates, environmental gradients), and the importance of critically assessing proxy data (e.g. what are species responding to?, what processes affect the formation and preservation of this proxy?) beyond statistical and method knowledge.

All the course materials and tutorials are publicly available at: github.com/gavinsimpson/acme-stats

AFFILIATIONS
[1]Department of Glaciology and Climate, Geological Survey of Denmark and Greenland (GEUS), Copenhagen, Denmark
[2]Environmental Change Research Unit (ECRU) Ecosystems and Environment Research Programme, Faculty of Biological and Environmental Sciences and Helsinki Institute of Sustainability Science (HELSUS), Helsinki, Finland
[3]Institute of Geology, Adam Mickiewicz University, Poznań, Poland

CONTACT
Sofia Ribeiro: sri@geus.dk

REFERENCES
Heikkilä M et al. (2022) Anthropocene 37: 100319