

IOW press release, August 07, 2026

Reading ecosystems like a book: AI method reveals hidden patterns in microbial communities in the river Warnow

Microbial communities are highly sensitive to environmental changes, yet their enormous diversity makes it difficult to identify ecological patterns. A research team from the Leibniz Institute for Baltic Sea Research Warnemünde (IOW) has now shown that an AI method originally developed for text analysis can be successfully applied to the analysis of complex environmental samples. The method identified five seasonally successive microbial sub-communities in the Warnow estuary while preserving ecological and functional information just as well as conventional methods – in some cases they even performed better. The study has now been published in the journal “Environmental Microbiome”.

Microorganisms are found almost everywhere and play a central role in ecosystems: They break down organic matter, influence matter cycles and often react very quickly to changing environmental conditions. In aquatic ecology therefore, the composition of microbial communities can provide valuable insights into the condition of a water body. However, studying these communities generates extremely large data sets. A single water sample can often yield up to hundreds of thousands of different genetic signatures of microbial organisms, the presence of which may change depending on temperature, salinity, nutrients, the season and numerous other factors.

A research team led by Anna Kujat, an IOW PhD student and lead author of the newly published study, and the corresponding author, Theodor Sperlea, investigated whether so-called topic modelling is suitable for analysing such complex data. This type of unsupervised machine learning was originally developed to identify patterns in large collections of text. This involves grouping words that frequently occur together into ‘topics’, which otherwise could not have been identified. When applied to microbial communities – also known as microbiomes – this means that, instead of words occurring together, microorganisms are grouped based on similar occurrence patterns within the analysed samples. “We have applied the basic idea of text analysis to an ecological problem to extract a small number of biologically interpretable patterns from a biodiversity that is almost impossible to grasp,” Kujat explains the approach.

From more than 1,200 genetic signatures to interpretable patterns

As a particularly challenging test case, the researchers selected the Warnow estuary and the adjacent coastal waters of the south-western Baltic Sea. Here, river discharge, wind and the interplay of fresh- and salty Baltic Sea water cause environmental fluctuations, often on a daily basis. At the same time, urban wastewater, industry, tourism and other human activities impact the system. The data set comprised 1,551 water samples taken between April 2022 and April 2023 at 14 fixed stations as part of the [OTC Genomics 2*](#) project, covering the ecological gradient from freshwater habitats to marine Baltic Sea habitats. In addition to sampling the microbial communities, the team recorded, salinity, water temperature, chlorophyll-a and nutrient concentrations, amongst other things, and utilised information on the Warnow's freshwater discharge provided by the State Agencies for Agriculture and the Environment, Mecklenburg-Western Pomerania. After processing, the entire dataset contained 1,236 different bacterial genetic signatures.

The researchers tested two topic modelling variants and compared them with established methods that condense large and complex datasets into a few key patterns. For the quantitative assessment, the team examined how well the condensed data representations still reflected ecological variables such as salinity, chlorophyll-a concentration or sampling location. In addition, the study investigated the extent to which functional similarities between the microorganisms were preserved. The results showed that, overall, topic modelling performed at least as well as conventional methods in terms of analytical output and even achieved better results in some of the analyses.

Five sub-communities follow the change of the seasons

For a more in-depth ecological analysis, the team selected a topic modelling method, which proved particularly effective at identifying both, ecological and functional information. Of a total of eleven identified microbial sub-communities, five were characteristic for the Warnow estuary. They did not occur simultaneously, but followed one another in a distinct seasonal sequence, with peak occurrences in spring, summer, early autumn and during cold periods up to the end of November. Unlike in the freshwater sections of the Warnow and along the coast, the team found no sub-community in the estuary that dominated consistently throughout the year.

The five groups differed both in their taxonomic composition and in the environmental conditions, under which they predominantly occurred. Two summer sub-communities, for example, were primarily associated with temperatures above 20 °C and salinity levels between 15 and 20 PSU. The spring group, by contrast, occurred at salinity levels of 8 to 15 PSU and temperatures between 8 and 17 °C. In autumn, the supply of freshwater increased and salinity declined for a longer period of time. As a result, the sub-communities typical of the estuary were gradually replaced by communities more characteristic of freshwater or coastal waters respectively. "The method has not only shown that the microbial community undergoes change, but also reveals which groups of organisms respond collectively to specific environmental conditions. In a habitat as dynamic as the Warnow estuary in particular, this opens up an additional ecological perspective," emphasises Theodor Sperlea.

New prospects for environmental monitoring

What is new about this study is the application of an approach originally derived from linguistics for the analysis of dynamic natural habitat in an ecological context, which, on top of that, was methodologically evaluated systematically. The researchers thereby demonstrated convincingly that topic modelling is really suitable for condensing complex microbiome data significantly while still preserving important ecological patterns.

"In the future, topic modelling could help to detect environmental changes at an earlier stage than before and thus monitor water bodies more effectively. But for the time being, it is not intended to replace tried-and-tested analytical methods, but rather to complement them," comments Theodor Sperlea. "Our study, however, opens the door to incorporating further methods from computational linguistics into the analysis of highly complex datasets such as those from environmental microbiology – including the algorithms behind the currently much-discussed large language models," the IOW researcher concludes.

Original publication:

Anna S. Kujat, Christiane Hassenrück, Stefan Lüdtkke, Matthias Labrenz, Theodor Sperlea (2026): *Enhancing the understanding of environmental microbiomes through topic modeling: a quantitative and qualitative analysis*. Environ. Microbiome, 21: 90. doi.org/10.1186/s40793-026-00927-2

Scientific contact

Dr. Theodor Sperlea | Tel.: +49 381 5197 3402 | theodor.sperlea@iow.de

* The OTC Genomics 2 project is funded by the Federal Ministry of Research, Technology and Space as part of the Ocean Technology Campus Rostock.

IOW Media contact:

Dr. Kristin Beck | Tel.: +49 (0)381 – 5197 135 | presse@iow.de

The Leibniz Institute for Baltic Sea Research Warnemünde (IOW) investigates natural and anthropogenic changes in coastal and marginal seas using a system-wide and interdisciplinary approach, ranging from fundamental to applied research. The Baltic Sea serves as an ideal case study on the institute's doorstep. An important mission of the IOW is to engage in knowledge-based dialogue with policy-makers, stakeholders, and society, thereby contributing to the sustainable development of coastal seas.
www.iow.de