University of Zagreb

Faculty of Science

Marija Purgar Filjak

VIBRIO SPP. IN MARICULTURE: PREDICTIVE MODELING AND WATER QUALITY ASSESSMENT

DOCTORAL THESIS

Zagreb, 2025

University of Zagreb

Faculty of Science

Marija Purgar Filjak

VIBRIO SPP. IN MARICULTURE: PREDICTIVE MODELING AND WATER QUALITY ASSESSMENT

DOCTORAL THESIS

Supervisors:

Tin Klanjšček, PhD

Antica Čulina, PhD

Zagreb, 2025

Prirodoslovno-matematički fakultet

Marija Purgar Filjak

VIBRIO SPP. U MARIKULTURI: PREDIKTIVNO MODELIRANJE I PROCJENA KAKVOĆE VODE

DOKTORSKI RAD

Mentori:

Dr. sc. Tin Klanjšček

Dr. sc. Antica Čulina

Zagreb, 2025

This doctoral dissertation was carried out at the Laboratory for Informatics and Environmental Modelling, Division for Marine and Environmental Research, Ruđer Bošković Institute, under the supervision of Dr. Tin Klanjšček and Dr. Antica Čulina, as part of the PhD program in Oceanology at the Faculty of Science, University of Zagreb. The research was conducted as part of the project "AqADAPT – Adapting Marine Finfish Aquaculture to a Changing Climate" (IP-2018-01-3150), and DOK-202102-6688, funded by the Croatian Science Foundation to Dr. Tin Klanjšček.

Information about the supervisors

Dr. Tin Klanjšček is a senior scientist and head of the Laboratory for Informatics and Environmental Modelling at the Ruđer Bošković Institute in Zagreb, Croatia. He is a biological oceanographer with a PhD from the joint program between Woods Hole Oceanographic Institution and the Massachusetts Institute of Technology.

Dr. Klanjšček's research focuses on dynamic energy budget (DEB) theory, applying it to optimize mariculture and investigate how pollution and food availability impact individual and population growth. Dr. Klanjšček has authored or co-authored 52 scientific papers, and been a principal investigator of 12 projects.

Dr. Antica Čulina is a senior research associate at the Ruđer Bošković Institute and an honorary fellow at the Netherlands Institute of Ecology. She is a co-founder and executive committee member of SPI-Birds and the Society for Open, Reliable, and Transparent Ecology and Evolutionary Biology (SORTEE). She also serves on the advisory board of the FAIRsFAIR project and Open Knowledge Maps and is a member of the UNESCO Open Science Initiative.

Dr. Čulina's expertise spans evidence synthesis, data and code standards, evolutionary ecology of bonding, and life-history trade-offs. She is one of the pioneers in studying and promoting Open Science practices in ecological and evolutionary research. Much of her current work is dedicated to meta-research, an emerging multidisciplinary field that aims to identify and address challenges within the research ecosystem. Through meta-research and Open Science, she strives to help ecology to solve issues currently present in the research and publishing systems, increase diversity of researchers, and thus improve the scope, reach, and value of research. Dr. Čulina has authored or co-authored 35 scientific papers.

“Well, here at last, dear friends, on the shores of the Sea comes the end of our fellowship in Middle-earth.” — J.R.R. Tolkien

To my supervisors and mentors, Tin and Antica, thank you for guiding me through my PhD journey with both patience and high expectations. I’m deeply grateful for every discussion, thoughtful suggestion, and for your support not just in research, but in life.

To my colleagues from Laboratory for Informatics and Environmental Modelling, and to all my collaborators, thank you for all ideas we exchanged, the late-night debugging sessions, and the unmeasurable support.

To my colleagues at SORTEE, thank you for showing me how community can drive meaningful change in science.

To all the professors who helped shape my academic path, thank you for the knowledge you shared, the questions you asked, and the sparks of curiosity you ignited along the way.

To Emory University, Rollins School of Public Health, and Department of Epidemiology, thank you for welcoming me with such warmth during my Fulbright scholarship. To my Fulbright supervisor and mentor, Josh, thank you for guiding me through my research in the U.S. I’m grateful for the discussions, classes, and encouragement throughout my stay. To my friends, Hannah, Blake, Kelly, David, Luis, Matt, and Erika, thank you for making Atlanta feel like home, for the laughter, the road trips, memories, and the kindness I’ll always cherish.

Katarina, Nikolina and Karla, thank you for always reminding me of what truly matters in life.

To my mother Dragica and brother Ivan, thank you for your quiet strength, endless support, and unconditional love. To my late father, Zdenko, I hope you are proud. Wherever I go, you are with me always.

And finally, to my husband Filip, my greatest joy. You have seen me at my best and at my most uncertain, and met both with unwavering love and support. Thank you for celebrating every small win and reminding me to rest. You are my home, and my heart.

This journey has taken me further than I ever imagined. To all the people who touched my heart, thank you.

BASIC DOCUMENTATION CARD

University of Zagreb
Faculty of Science
Department of Geology

Doctoral Thesis

VIBRIO SPP. IN MARICULTURE: PREDICTIVE MODELING AND WATER QUALITY ASSESSMENT

Marija Purgar Filjak

Thesis completed in: Institut Ruđer Bošković, Bijenička cesta 54, 10000 Zagreb

Abstract:
Bacteria from the *Vibrio* genus have an important role in marine ecosystems, as they are linked to serious diseases in humans and marine organisms. Mathematical models can be used to predict *Vibrio* spp. dynamics, thus informing coastal management, financial planning, and water quality/food safety regulations in mariculture. This thesis investigated three interrelated components essential for improving science-based decision-making in mariculture: (i) the predictive performance of existing *Vibrio* growth models from the literature in various marine environments, (ii) the suitability of *Vibrio* spp. abundance as a supplementary indicator for marine water quality assessment, and (iii) the overall informative value of ecological research which supports effective coastal management and scientific progress. The first study standardized and tested 28 published models against seven datasets, revealing limited predictive accuracy in coastal environments, particularly in areas subject to strong anthropogenic influence such as mariculture. The second study analyzed open-access dataset from the Adriatic Sea and found that *Vibrio* spp. abundance provided distinct and seasonally relevant information not captured by conventional fecal indicators, underscoring its applicability in water quality monitoring near mariculture. The third study conducted a meta-analysis of 33 meta-studies based on 10,464 individual ecological studies. It revealed that 44.7% of ecological studies remain unpublished, 67.4% have design flaws, and 40.7% underreport results, indicating that only 11-18% reach full informative value. Collectively, the findings of this doctoral dissertation underscore the urgent need for more reliable *Vibrio* spp. predictive models, broader integration of supplementary microbial indicators in water quality assessment, and systemic improvements in scientific practices, particularly in study planning, result reporting, and publication of research, with the goal of increasing the informative value of ecological research.

Keywords: *Vibrio* spp., model evaluation, microbial indicators, water quality monitoring, meta-research, open science

Thesis contains: 79 pages, 100 references

Original in: English

Supervisors: Tin Klanjšček, PhD, scientific adviser
Antica Čulina, PhD, senior research associate

Reviewers: Kristina Pikelj, PhD, associate professor
Vlado Cuculić, PhD, senior research associate
Damir Kapetanović, PhD, scientific adviser

TEMELJNA DOKUMENTACIJSKA KARTICA

Sveučilište u Zagrebu
Prirodoslovno-matematički fakultet
Geološki odsjek

Doktorski rad

VIBRIO SPP. U MARIKULTURI: PREDIKTIVNO MODELIRANJE I PROCJENA KAKVOĆE VODE

Marija Purgar Filjak

Rad je izrađen: Institut Ruđer Bošković, Bijenička cesta 54, 10000 Zagreb

Sažetak:
Bakterije iz roda *Vibrio* imaju važnu ulogu u morskim ekosustavima jer uzrokuju oboljenja kod ljudi i morskih organizama. Matematički modeli mogu služiti za predviđanje brojnosti *Vibrio* spp., čime doprinose planiranju upravljanja obalnim područjima, financijskom planiranju te izradi zakonskih regulativa vezanih uz kakvoću vode i sigurnost hrane u marikulturi. Ova disertacija istražila je tri međusobno povezana aspekta važna za unaprjeđenje znanstveno utemeljenog donošenja odluka u marikulturi i upravljanju obalnim područjima: (i) prediktivnu učinkovitost postojećih matematičkih modela za procjenu brojnosti *Vibrio* spp., (ii) potencijal brojnosti bakterija roda *Vibrio* kao dodatnog pokazatelja kakvoće morske vode, i (iii) informativnu vrijednost ekoloških istraživanja koja čine temelj za učinkovito upravljanje obalnim područjima i znanstveni napredak. Prvo istraživanje standardiziralo je i testirano ukupno 28 objavljenih modela na sedam skupova otvorenih podataka, pri čemu je utvrđena ograničena primjenjivost postojećih modela u obalnim uvjetima, osobito u područjima pod snažnim antropogenim utjecajem, poput marikulture. Drugo istraživanje analiziralo je skup otvorenih podataka iz Jadranskog mora te je utvrđeno da brojnost *Vibrio* spp. pruža sezonski relevantne i komplementarne informacije koje nisu zabilježene konvencionalnim fekalnim pokazateljima, što potvrđuje korisnost brojnosti bakterija roda *Vibrio* kao dodatnog pokazatelja kakvoće vode u blizini marikulture. Treće istraživanje provelo je meta-analizu 33 meta-studije koje su obuhvatile ukupno 10.464 pojedinačnih ekoloških studija. Rezultati su pokazali da 44,7 % studija ostaje neobjavljeno, 67,4 % ima nedostatke u dizajnu, a 40,7 % nepotpuno izvještava o rezultatima, što upućuje na to da samo 11-18 % ekoloških istraživanja doseže punu informativnu vrijednost. Zajedno, nalazi ove doktorske disertacije naglašavaju hitnu potrebu za pouzdanijim prediktivnim modelima za *Vibrio* spp., širu integraciju dopunskih mikrobnih pokazatelja u praćenje kakvoće vode, te sustavna poboljšanja u znanstvenoj praksi, osobito u planiranju istraživanja, izvješćivanju rezultata i objavi radova, s ciljem povećanja informativne vrijednosti ekoloških istraživanja.

Ključne riječi: *Vibrio* spp., evaluacija modela, mikrobni pokazatelji, praćenje kakvoće vode, metaistraživanje, otvorena znanost

Rad sadrži: 79 stranica, 100 literaturnih navoda

Jezik izvornika: engleski

Mentori: Dr. sc. Tin Klanjšček, znanstveni savjetnik
Dr. sc. Antica Čulina, viša znanstvena suradnica

Ocjenjivači: Dr. sc. Kristina Pikelj, izvanredna profesorica
Dr. sc. Vlado Cuculić, viši znanstveni suradnik
Dr. sc. Damir Kapetanović, znanstveni savjetnik

THESIS SUMMARY

Vibrio bacteria are one of the most prominent microorganisms in coastal areas due to their role in biogeochemical cycles, health implications, and adverse impact on the mariculture sector (Martinez-Urtaza et al., 2010; Shafiee et al., 2024). *Vibrio* infections lead to significant economic losses in mariculture and negative environmental impact through massive die-offs of fish, shellfish, and other organisms (Manchanayake et al., 2023; Shafiee et al., 2024). Climate change and other anthropogenic pressures hamper risk management by rapidly altering the dynamics of microbial communities in marine environments. As these communities shift, they can threaten the health of natural ecosystems and contribute to the spread of diseases in mariculture, causing significant economic damage and negative environmental impact. This makes it imperative to develop management strategies that safeguard both the aquaculture industry and the marine environments in which production facilities are located.

Mathematical models are important tool for understanding and forecasting *Vibrio* spp. abundance, as they enable the prediction of future trends in various aquatic systems and marine aquaculture products, support the development of management strategies, and advance knowledge of changes in microbial communities within aquatic ecosystems (Baker-Austin et al., 2013; Riedinger et al., 2025). The results of mathematical models can i) improve decision-making and cost estimation in aquaculture, as well as ii) guide future research and the food safety legislation (Alver & Føre, 2023; Huss et al., 2004). However, many existing models are derived from controlled laboratory experiments and may not translate well to *in situ* conditions, which are highly variable and often lack consistent data. Therefore, before using existing models to assess the dynamics of *Vibrio* populations, their applicability should be checked to avoid misleading predictions and the resulting errors in management strategies, and changes in regulatory frameworks.

In addition to mathematical modeling, monitoring *Vibrio* spp. as part of water quality assessments offers another promising approach to support sustainable mariculture. Current water quality assessment practices include indicators characterizing heterotrophic bacteria but are mostly focused on fecal bacteria (Some et al., 2021), specifically *E. coli* and enterococci (Some et al., 2021; Stewart et al., 2008). Other bacteria, such as *Vibrio* spp., have been largely overlooked despite their

potentially negative effect on humans and marine organisms. Despite the large number of research studies on *Vibrio* spp. and the suggestion to use their abundance as a supplementary indicator of water quality as early as 1984 (Robertson, 1984), uniform guidelines, recommendations, and official regulations for their monitoring still do not exist. Generally, the literature on the practical applicability of *Vibrio* spp. as a supplementary water quality indicator remains limited and inconclusive.

Both the development of mathematical models and the analysis of the applicability of new environmental indicators heavily rely on the availability and quality of existing research (Stein et al., 2001; van den Berg et al., 2022). However, "publish or perish" culture (Udesky, 2025), coupled with insufficient training in study design and statistical methods, often leads to poorly designed studies, incomplete reporting, and unpublished studies. These issues reduce the informative value of research, which is defined as the completeness of the available results of the conducted studies. Informative value of research is reduced by suboptimal study design, unpublished studies, and incomplete reporting in publications. Prior to the research presented in this thesis, informative value of research has only been quantified in medicine, where it was estimated at just 15% (Chalmers & Glasziou, 2009). The problem could be relatively large in all fields of science (Begley & Ellis, 2012; Camerer et al., 2018; Open Science Collaboration, 2015), given that the root causes, such as misaligned research incentive structures, limited methodological training, and inadequate support for transparent practices, are systemic in nature.

Therefore, the aim of this doctoral thesis was to: (i) assess the ability of existing functional models from the literature for predicting the *Vibrio* spp. abundance in various marine areas, (ii) explore the potential of *Vibrio* spp. abundance as an additional indicator of water quality for science-based coastal management and mariculture, (iii) quantify the informative value of existing studies in the field of ecology. The thesis is based on three published articles addressing the following hypotheses:

(H1): Existing models of Vibrio bacteria growth can predict the abundance of Vibrio spp. in mariculture.

(H2): Indicators of Vibrio spp. abundance have the potential to be included in the regular set of indicators for assessing water quality in mariculture.

(H3): The informative value of ecological research is comparable to that estimated in medicine, which is about 15%.

The first publication (Purgar et al., 2022a) assessed the performance of 28 functional *Vibrio* spp. growth models obtained via literature search and extracted from 16 eligible peer-reviewed studies identified via a literature search using Web of Science. The available functional models from the literature were standardized using unified nomenclature and limited to those employing both primary (bacterial growth over time) and secondary (environmental drivers such as temperature, salinity, pH) models to enable model simulation. Model performance was evaluated using seven open datasets of *Vibrio* abundance, including two newly collected in the Adriatic Sea and five from existing literature, spanning four habitat types (aquaculture, urban estuary, estuary, and coastal area). Results of model simulations and model performance analysis demonstrate that, while the models were able to predict *Vibrio* spp. abundance to an extent, their predictive accuracy was generally limited. Models often underperformed, especially in coastal environments under significant anthropogenic influence such as marine aquaculture (around fish farms) and urban estuaries under greater anthropogenic pressure, such as mariculture.

The second publication (Purgar et al., 2023) investigated whether *Vibrio* spp. abundance could and/or should be used as a supplementary bacterial indicator for monitoring water quality near mariculture. The study was conducted on an open dataset collected in Mali Ston Bay (Adriatic Sea), allowing for a comparison of bacterial indicators and environmental parameters at a fish farm and a nearby control site across two seasons. Statistical analysis revealed that *Vibrio* spp. abundance, along with heterotrophic bacteria and enterococci, was unexpectedly more abundant during colder months, while *E. coli* and total coliforms followed the typical pattern of higher abundance in warmer months. Here, *Vibrio* spp. abundance indicated a potential microbial risk and lower water quality that conventional fecal indicators did not capture, providing distinct and complementary information compared to traditional indicators, which is particularly relevant for sustainable mariculture. Given their role as opportunistic pathogens responsible for disease outbreaks in marine organisms, incorporating *Vibrio* monitoring could improve early detection and management of health risks. However, the study also highlighted the need for additional research to establish *Vibrio*-specific water quality thresholds and identify key pathogenic species before incorporating the indicator into regulatory frameworks.

The third publication (Purgar et al., 2022b) quantified the informative value of ecological research based on the results of 33 meta-studies. Meta-studies were

defined as studies that synthesized published or unpublished research with the aim to estimate research waste components occurring at any of the main stages of the research life cycle (study planning, result reporting, and publication). Research waste was categorized into core waste, which represents studies that remain unpublished because of low quality or publication bias, and exploitative waste, which represents published studies that are poorly designed or insufficiently reported, diminishing their informative value. The meta-analysis of 43 estimates of research waste components from 33 meta-studies showed that 44.7% (95% CI: 44.2-46.7%) of studies remained unpublished. Among published studies, 67.4% (95% CI: 66.3-68.4%) had significant issues at the study design stage, and 40.7% (95% CI: 38.7-42.8%) of their results were underreported, meaning that critical information such as effect sizes, sample sizes, or uncertainty measures was missing. Overall, these findings suggest that only 11-18% of ecological research reaches its full informative value, akin to the 15% observed in medicine.

This doctoral thesis provides new insights into modeling and monitoring *Vibrio* spp. abundance in marine environments while also quantifying a broader systemic issue in ecological research: low informative value, which limits the ability to build upon existing research. The thesis presents the following original scientific contributions:

1. A comprehensive review and standardization of existing models for predicting *Vibrio* spp. abundance.
2. Assessment of the predictive performance of existing *Vibrio* spp. growth models near mariculture and other coastal areas.
3. Comparative analysis of the effectiveness of *Vibrio* spp. abundance versus conventional bacterial indicators in assessing water quality.
4. Assessment of the potential of *Vibrio* spp. abundance as a supplementary indicator for monitoring coastal water quality.
5. The first quantitative estimate of the informative value of ecological research.
6. Release of datasets and analytical code in open-access repositories to promote transparency and reproducibility of research findings.

Together, the findings of this PhD thesis systematize available growth models and compare their ability to predict *Vibrio* spp. abundance in mariculture, provide evidence of *Vibrio* spp.'s applicability as a supplemental indicator of water quality in environmental monitoring near mariculture, and underscore the urgent need to improve study design, reporting practices, and publication rates in ecology which are crucial for

supporting evidence-based decision-making in marine aquaculture, coastal management, and scientific progress more broadly.

PROŠIRENI SAŽETAK

Bakterije iz roda *Vibrio* jedne su od najistaknutijih mikroorganizama u obalnim područjima zbog svoje uloge u biogeokemijskim ciklusima, utjecaju na ljudsko zdravlje te negativnom učinku na sektor marikulture (Martinez-Urtaza i sur., 2010; Shafiee i sur., 2024). Infekcije uzrokovane *Vibrio* bakterijama dovode do velikih ekonomskih gubitaka u marikulturi i negativnog utjecaja na okoliš, posebice kroz masovno ugibanje riba, školjkaša i drugih organizama (Manchanayake i sur., 2023; Shafiee i sur., 2024). Klimatske promjene i drugi antropogeni pritisci otežavaju upravljanje rizikom jer ubrzano mijenjaju dinamiku mikrobnih zajednica u morskom okolišu. Kako se te zajednice mijenjaju, mogu ugroziti zdravlje prirodnih ekosustava i pridonijeti širenju bolesti u marikulturi, uzrokujući znatnu štetu i okolišu i gospodarstvu. Stoga je ključno razviti strategije upravljanja koje štite i industriju marikulture i morske ekosustave u kojima se uzgoj odvija.

Matematički modeli predstavljaju važan alat za razumijevanje i predviđanje brojnosti bakterija roda *Vibrio*, jer omogućuju projekcije budućih trendova u vodenim sustavima i proizvodima iz marikulture, potporu razvoju strategija upravljanja te unaprjeđenje znanja o promjenama u mikrobnim zajednicama u vodenim ekosustavima(Baker-Austin i sur., 2013; Riedinger i sur., 2025). Rezultati takvih modela mogu pridonijeti donošenju odluka u marikulturi, procjeni troškova, usmjeravanju budućih istraživanja te oblikovanju zakonodavnog okvira za kakvoću vode i sigurnost hrane (Alver i Føre, 2023; Huss i sur., 2004). Međutim, brojni postojeći modeli temelje se na laboratorijskim eksperimentima i često nisu primjenjivi na stvarne okolišne uvjete koji su vrlo promjenjivi i često nedovoljno dokumentirani. Stoga je, prije njihove primjene u praksi, nužno procijeniti primjenjivost dostupnih modela kako bi se izbjegle pogreške u predviđanjima, upravljanju i zakonskoj regulativi.

Osim modeliranja, praćenje brojnosti *Vibrio* spp. u sklopu procjene kakvoće morske vode nudi dodatni pristup održivoj marikulturi. Trenutne prakse procjene kakvoće vode uglavnom se oslanjaju na heterotrofne bakterije, posebno fekalne pokazatelje (Some et al., 2021) kao što su *Escherichia coli* i enterokoki (Some i sur., 2021; Stewart i sur., 2008). Druge bakterije, poput *Vibrio* spp., uglavnom se zanemaruju, iako mogu imati štetan učinak na zdravlje ljudi i morskih organizama. Unatoč velikom broju istraživanja o *Vibrio* spp. te prijedlogu da se njihova brojnost koristi kao dopunski pokazatelj kvalitete vode već 1984. godine (Robertson, 1984.),

jedinstvene smjernice, preporuke i službeni propisi za praćenje *Vibrio* spp. još uvijek ne postoje. Općenito, može se reći da je literatura o praktičnoj primjeni i prikladnosti brojnosti *Vibrio* spp. kao dodatnog pokazatelja kakvoće vode ograničena.

Razvoj matematičkih modela i analiza prikladnosti novih okolišnih pokazatelja uvelike ovise o kvaliteti i dostupnosti postojećih istraživanja (Stein i sur., 2001; van den Berg i sur., 2022). No, *"objavi ili propadni"* (engl. *publish or perish*) kultura (Udesky, 2025), u kombinaciji s nedovoljnom edukacijom u području dizajna istraživanja i statističkih metoda, često dovodi do sub-optimalnog dizajna istraživanja, nepotpunog izvješćivanja te neobjavljenih rezultata. Sve navedeno umanjuje informativnu vrijednost znanstvenih istraživanja, odnosno potpunost i iskoristivost rezultata provedenih studija. Dosad je informativna vrijednost sustavno kvantificirana samo u području medicine, gdje je procijenjena na svega 15% (Chalmers i Glasziou, 2009). S obzirom na to da su temeljni uzroci niske informativne vrijednosti, poput neusklađenih sustava nagrađivanja znanstvenika, ograničenog metodološkog znanja i nedostatne podrške transparentnim praksama, sistemski, isti problem vjerojatno postoji i u drugim znanstvenim područjima.

Stoga je cilj ove doktorske disertacije bio: (i) procijeniti učinkovitost postojećih funkcionalnih modela iz literature za predviđanje brojnosti *Vibrio* spp. u različitim morskim područjima, (ii) istražiti potencijal brojnosti *Vibrio* spp. kao dodatnog pokazatelja kakvoće morske vode za znanstveno utemeljeno upravljanje obalnim područjima i marikulturom, (iii) kvantificirati informativnu vrijednost postojećih ekoloških istraživanja.

Disertacija se temelji na tri objavljena znanstvena rada, koji istražuju sljedeće hipoteze:

(H1) Postojeći modeli rasta bakterija roda Vibrio mogu se koristiti za predviđanje njihove brojnosti u marikulturi.

(H2) Brojnost Vibrio spp. ima značajan potencijal da se uključi među redovne pokazatelje za procjenu kakvoće morske vode.

(H3) Informativna vrijednost ekoloških istraživanja usporediva je s onom procijenjenom u medicini, otprilike 15%.

Prvi rad (Purgar i sur., 2022a) analizirao je 28 funkcionalnih modela dobivenih literaturnim pregledom i ekstrakcijom iz 16 prihvatljivih recenziranih studija. Modeli su standardizirani prema jedinstvenoj nomenklaturi i ograničeni na one koji su sadržavali primarne (rast bakterija kroz vrijeme) i sekundarne (opisuju utjecaj okolišnih faktora

poput temperature, saliniteta, pH) komponente. Učinkovitost modela testirana je na sedam otvorenih skupova podataka, uključujući dva nova iz Jadranskog mora te pet iz literature, obuhvaćajući četiri tipa staništa (marikultura, urbani estuariji, estuariji i obalna područja). Rezultati su pokazali ograničenu prediktivnu točnost, osobito u obalnim područjima pod snažnim antropogenim utjecajem, poput marikulture.

Drugi rad (Purgar i sur., 2023) istraživao je može li se brojnost *Vibrio* spp. koristiti kao dopunski pokazatelj kakvoće vode u blizini marikulture. Za ovo istraživanje korišten je otvoreni skup podataka iz Malostonskog zaljeva, a uspoređivani su bakteriološki pokazatelji i okolišni parametri između ribogojilišta i obližnje kontrolne točke kroz dvije sezone. Statističkom analizom je utvrđeno da su *Vibrio* spp., zajedno s heterotrofnim bakterijama i enterokokima, bili brojniji tijekom hladnijih mjeseci, dok su *E. coli* i ukupni koliformi pratili očekivani sezonski obrazac s višim vrijednostima ljeti. Usporedbom graničnih vrijednosti kakvoće vode, brojnost bakterija roda *Vibrio* ukazivala je na potencijalni mikrobni rizik i nižu kakvoću vode koju konvencionalni fekalni indikatori nisu zabilježili, dajući jasne i komplementarne informacije u usporedbi s tradicionalnim pokazateljima, što je posebno važno za održivu marikulturu. Rezultati ovog istraživanja upućuju na to da brojnost *Vibrio* spp. može pružiti dodatne i specifične informacije koje standardni pokazatelji ne bilježe, čime se povećava učinkovitost praćenja kakvoće vode.

Treći rad (Purgar i sur., 2022b) kvantificirao je informativnu vrijednost ekoloških istraživanja meta-analizom 33 meta-studije koje su obuhvaćale ukupno 10464 pojedinačnih studija. Rezultati su pokazali da 44,7% studija nikada ne bude objavljeno, 67,4% ima nedostatke u dizajnu istraživanja, a 40,7% ne prikazuje rezultate u cijelosti. Ukupno je procijenjeno da tek 11–18% ekoloških istraživanja doseže svoju punu informativnu vrijednost.

Ova doktorska disertacija donosi nove uvide u modeliranje i praćenje brojnosti *Vibrio* spp. u morskom okolišu te ističe širi problem u ekološkim istraživanjima: nisku informativnu vrijednost, koja ograničava mogućnost iskorištavanja postojećih znanstvenih spoznaj. Znanstveni doprinosi uključuju:

1. Pregled i standardizaciju postojećih modela za predviđanje brojnosti bakterija iz roda *Vibrio*.
2. Procjenu učinkovitosti postojećih modela za predviđanje brojnosti *Vibrio* spp. u marikulturi.

3. Usporedbu učinkovitosti standardnih indikatora s indikatorom brojnosti *Vibrio* spp. za procjenu kakvoće morske vode.
4. Procjenu potrebe korištenja brojnosti *Vibrio* spp. kao dodatnog indikatora kakvoće morske vode.
5. Prvu procjenu informativne vrijednosti postojećih studija za područje ekologije.
6. Objavljene setove podataka i analitičkog koda u javno dostupnim repozitorijima s otvorenim pristupom.

Zajedno, rezultati ove doktorske disertacije standardiziraju postojeće modele rasta bakterija roda *Vibrio* i uspoređuju njihovu učinkovitost u predviđanju brojnosti *Vibrio* spp. u marikulturi, pružaju dokaze o primjenjivosti brojnosti *Vibrio* spp. kao dopunskog pokazatelja kakvoće morske vode te naglašavaju hitnu potrebu za unapređenjem dizajna studija, praksi izvještavanja i stope objavljivanja u ekologiji, što je pak ključno za podršku održivoj marikulturi, upravljanje obalnim područjima i napredak znanosti u širem smislu.

TABLE OF CONTENTS

LIST OF PUBLICATIONS

(I) Purgar, M., Kapetanović, D., Geček, S., Marn, N., Haberle, I., Hackenberger, B. K., Gavrilović, A., Pečar Ilić, J., Hackenberger, D. K., Djerdj, T., Ćaleta, B., Klanjscek, T. (2022a). Investigating the Ability of Growth Models to Predict In Situ *Vibrio* spp. Abundances. Microorganisms, 10(9), 1765. https://doi.org/10.3390/microorganisms10091765

(II) Purgar, M., Gavrilović, A., Kapetanović, D., Klanjšček, J., Jug-Dujaković, J., Kolda, A., Žunić, J., Kazazić, S., Vardić Smrzlić, I., Vukić Lušić, D., Pikelj, K., Listeš, E., El-Matbouli, M., Lillehaug, A., Lončarević, S., Knežević, D., Hengl, B., Geček, S., Klanjscek, T. (2023). Assessment of *Vibrio* spp. abundance as a water quality indicator: Insights from Mali Ston Bay in the Adriatic Sea. Estuarine, Coastal and Shelf Science, 295, 108558. https://doi.org/10.1016/j.ecss.2023.108558

(III) Purgar, M., Klanjscek, T., Culina, A. (2022b). Quantifying research waste in ecology. Nature Ecology & Evolution, 6(9), 1390-1397. https://doi.org/10.1038/s41559-022-01820-0

1. INTRODUCTION

Marine aquaculture plays an important role in advancing multiple Sustainable Development Goals (SDGs) including SDG 2 (Zero Hunger), SDG 8 (Decent Work and Economic Growth), SDG 13 (Climate Action), and SDG 14 (Life Below Water), by contributing to food security, supporting livelihoods and economic development, enhancing climate resilience, and promoting the sustainable use of ocean and marine resources (FAO, 2024; Stead, 2019; Troell et al., 2023). In 2022, global aquaculture production reached 130.9 million tonnes (valued at USD 312.8 billion) and accounted for 59% of total global aquatic production (FAO, 2024). Here, aquaculture for the first time surpassed capture fisheries (i.e., the harvesting of wild fish and shellfish from their natural environment) in aquatic animal output, contributing 51% of the global total. Within this expansion, marine and coastal aquaculture, which includes the farming of aquatic species in ocean and nearshore environments (hereafter referred to as mariculture), accounted for 37.4% of all farmed aquatic animals. While this growth highlights the increasing significance of mariculture in global food systems and aids in alleviating pressure on capture fisheries, it also amplifies the necessity for science-based strategies to tackle emerging environmental and biological risks in marine farming systems.

One of the most pressing challenges in marine aquaculture is the increasing risk posed by microbial communities, particularly in the context of climate change and intensified human activities (FAO, 2024). Changing environmental conditions, such as higher sea temperature, acidification, and increased nutrient loads, are reshaping microbial communities by creating favorable conditions for microbial proliferation, often favoring opportunistic and potentially pathogenic species, notably *Vibrio* spp. (Nogales et al., 2011). *Vibrio* bacteria are one of the most prominent microorganisms in coastal areas due to their role in biogeochemical cycles, health implications, and adverse impact on the mariculture sector (Baker-Austin et al., 2018; Martinez-Urtaza et al., 2010). Monitoring the abundance and dynamics of *Vibrio* spp. enables early detection of pathogenic strains in marine environments and aquaculture systems, particularly those responsible for vibriosis (Sanches-Fernandes et al., 2022). *Vibrio*sis is among the most common diseases in marine aquaculture and can result in large-scale mortality events and economic losses reaching billions of dollars globally each year

(Sanches-Fernandes et al., 2022; Shafiee et al., 2024). These risks highlight the need for effective, science-based approaches to monitoring and managing *Vibrio* spp. abundance in coastal environments and mariculture facilities worldwide.

Mathematical models are important tool for understanding and forecasting *Vibrio* spp. abundance, as they enable the prediction of future trends in various aquatic systems and marine aquaculture products, support the development of management strategies, and advance knowledge of changes in microbial communities within aquatic ecosystems (Baker-Austin et al., 2013; Riedinger et al., 2025). The results of mathematical models can i) improve decision-making and cost estimation in aquaculture, as well as ii) guide future research and the food safety legislation (Alver & Føre, 2023; Huss et al., 2004).

Modeling *Vibrio* abundance typically relies on growth curves that show the number of living cells as a function of time (Peleg & Corradini, 2011). In constant (laboratory) conditions, the growth of bacteria is characterized by a sigmoid curve where the dependent variable is the logarithm of the concentration of live cells (Baranyi et al., 1993). The sigmoid curve describes isothermal growth, i.e., growth as a function of time at constant temperature (Peleg & Corradini, 2011). Secondary models describe the functional dependence of growth with respect to external factors such as temperature or pH (Esser et al., 2015; Peleg & Corradini, 2011). The most common secondary models are the Ratkowsky model (Ratkowsky et al., 1983) and the Arrhenius model (Davey, 1989). Some sources also define tertiary models (Esser et al., 2015), described as software packages that build upon primary and/or secondary models and often have a user interface.

In general, the aforementioned models are useful but are not always applicable. The quality of the model, and thus interpretative value, depends on numerous factors such as knowledge about the modeled system, availability of data, and existing structures and mathematical formulations of the models themselves. Existing models of population dynamics of *Vibrio* spp. are mainly based on information from experiments, so their applicability for *in situ* modeling is questionable. Possible limitations arise from the model formulations: while the equations provide satisfactory descriptions of *Vibrio* growth and its dependence on, e.g., temperature or pH, they usually are not subject to mechanistic interpretation (Esser et al., 2015). For data collected in situ, exact conditions of data collection are often unknown, the intervals between data collection may be long and/or irregular, and environmental conditions

vary greatly (Ramsey, 2021). Therefore, before using existing models to assess the dynamics of *Vibrio* populations, their applicability should be checked to avoid errors in forecasts, planning, and changes in regulatory frameworks.

Sustainable mariculture and regulatory frameworks strongly rely on indices such as water quality, which is important for the success of farming itself (Leung et al., 2015; Liu et al., 2023; Webber et al., 2021). Typical water quality assessment includes indicators characterizing heterotrophic bacteria but is mostly focused on fecal bacteria (Some et al., 2021), specifically *Escherichia coli* (*E. coli)* and enterococci (Some et al., 2021; Stewart et al., 2008). Bacterial indicators explain specific anthropogenic pressures and help with risk assessment. Heterotrophic plate counts (HPC) reflect the general load of different bacteria that need organic nutrients for growth in water bodies (Bartram et al., 2013). Historically, coliform bacteria were widely used as indicators of fecal pollution in water. However, they were eventually replaced by more specific indicators, such as *E. coli* and enterococci, due to their higher reliability in detecting fecal contamination (Price et al., 2017). Other bacteria, such as *Vibrio* spp., have been largely overlooked despite their potentially negative effect on humans and marine organisms (Price et al., 2017). Despite the growing number of studies on *Vibrio* spp. (Onohuean et al., 2022; Zakaria et al., 2025), and the suggestion to use *Vibrio* abundance as a supplementary indicator of water quality as early as 1984 (Robertson, 1984), uniform guidelines, recommendations, and official regulations for the monitoring of *Vibrio* spp. still do not exist. Generally, the literature on the practical applicability of *Vibrio* spp. as a supplementary water quality indicator remains limited and inconclusive.

Predictive modeling and development and implementation of new environmental indicators, such as *Vibrio* spp. for water quality, have the potential to support sustainable mariculture by improving monitoring and informing management decisions. However, effectiveness, reliability, and integration of models and new indicators into monitoring frameworks and regulatory policies ultimately depend on the quality and accessibility of ecological studies. While ecological research routinely generates datasets, analytical workflows, and derived results, only a small and potentially biased fraction of this output is made publicly available through publications (Rothstein et al., 2005) and thus available as information for re-use in e.g. evidence synthesis (Nakagawa, Koricheva, et al., 2020) and for other researchers to build on. The prevailing *publish or perish* culture (Udesky, 2025), combined with limited training

in study design and statistical analysis (Touchon & McCoy, 2016), also contributes to making data collection and analysis suboptimal and biased.

Collectively, issues related to unpublished studies, suboptimal study planning (including flaws in study design, data collection, and data analysis), and inadequate result reporting reduce the informative value of the research, defined as the accessibility, usability, and completeness of the available results in the conducted studies. To date, the informative value of research has been estimated only for one field of science, medicine, where it amounts to 15 percent (Chalmers & Glasziou, 2009) leads to an estimated annual loss exceeding US$170 billion (Glasziou & Chalmers, 2016). Research showed that the problem could be relatively large in all fields of science (Begley & Ellis, 2012; Camerer et al., 2018; Open Science Collaboration, 2015), including ecology, given that the root causes, such as misaligned research incentive structures, limited methodological training, and inadequate support for transparent practices, are systemic in nature.

This doctoral thesis is based on three peer-reviewed scientific publications, each addressing a specific research gap relevant to sustainable mariculture. The first publication systematizes 28 functional growth models of *Vibrio* spp. extracted from the literature and examines their applicability for predicting *Vibrio* abundance in various coastal habitats, including marine aquaculture environments. The second publication explores the potential of *Vibrio* spp. abundance to serve as a supplementary bacterial indicator of water quality by analyzing seasonal patterns, environmental conditions, and bacterial abundance at a mariculture site and a nearby control site in the Adriatic Sea. The third publication quantifies the informative value of ecological research by synthesizing evidence from 33 meta-studies, assessing the extent of unpublished, poorly designed, or incompletely reported studies that limit the informative value of research in ecology. The study did not conduct an assessment focused specifically on mariculture or aquatic ecology but instead examined ecological research as a whole for two main reasons: (i) possible limited number of available meta-research studies quantifying components of research waste across research cycle within ecological subfields, and (ii) scientific disciplines, and their subfields, often share similar structural characteristics, such as research norms and incentive systems, making it reasonable to expect comparable informative patterns across subfields.

These three publications integrate mathematical modeling, statistical analysis, and a meta-research approach to strengthen the scientific foundation for modeling and

integrating *Vibrio* spp. abundance as an indicator in mariculture, and provide the first quantitative estimate of the informative value of research in ecology. Estimates of the informative value of research enable further advocacy for changes in various stakeholder incentives to improve research practices, resulting in a greater amount of knowledge for creating more effective strategies for managing marine ecosystems, mitigating the effects of climate change, and enhancing marine aquaculture practices, ultimately leading to cleaner and safer food production. The final section of the thesis presents a general discussion that synthesizes the findings, reflects on their broader implications, and outlines recommendations for future research in marine aquaculture and the broader field of ecology.

1.1. RESEARCH OBJECTIVES AND HYPOTHESES

The doctoral thesis had three main objectives: (i) to assess the ability of existing functional models from the literature for predicting the *Vibrio* spp. abundance in various marine areas, (ii) to explore the potential of *Vibrio* spp. abundance as an additional indicator of water quality for science-based coastal management and mariculture, (iii) to quantify the informative value (research waste) of existing studies in the field of ecology.

Specifically, the thesis investigated three main hypotheses:

(H1): Existing models of *Vibrio* bacteria growth can be used to predict the abundance of *Vibrio* spp. in mariculture.

(H2): *Vibrio* spp. abundance has significant potential to be included in the regular set of indicators for assessing the water quality in mariculture.

(H3): The informative value of ecological research will be similar to that estimated in medicine (about 15%).

2. RESEARCH ARTICLES

2.1. Publication I: Investigating the Ability of Growth Models to Predict In Situ *Vibrio* spp. Abundances

Purgar, M., Kapetanović, D., Geček, S., Marn, N., Haberle, I., Hackenberger, B. K., Gavrilović, A., Pečar Ilić, J., Hackenberger, D. K., Djerdj, T., Ćaleta, B., Klanjscek, T. (2022a). Investigating the Ability of Growth Models to Predict In Situ *Vibrio* spp. Abundances. Microorganisms, 10(9), 1765. https://doi.org/10.3390/microorganisms10091765

Contributions: conceptualization, formal analysis, investigation, visualization, writing - original draft preparation, and writing - review and editing.

microorganisms

MDPI

Article

Investigating the Ability of Growth Models to Predict In Situ *Vibrio* spp. Abundances

Marija Purgar [1], Damir Kapetanović [1], Sunčana Geček [1], Nina Marn [1,2], Ines Haberle [1], Branimir K. Hackenberger [3], Ana Gavrilović [4], Jadranka Pečar Ilić [1], Domagoj K. Hackenberger [3], Tamara Djerdj [3], Bruno Ćaleta [3] and Tin Klanjscek [1,*]

[1] Ruđer Bošković Institute, 10000 Zagreb, Croatia
[2] School of Biological Sciences, The University of Western Australia, Crawley, WA 6009, Australia
[3] Department of Biology, Josip Juraj Strossmayer University of Osijek, 31000 Osijek, Croatia
[4] Faculty of Agriculture, University of Zagreb, 10000 Zagreb, Croatia
* Correspondence: tin@irb.hr

Abstract: *Vibrio* spp. have an important role in biogeochemical cycles; some species are disease agents for aquatic animals and/or humans. Predicting population dynamics of *Vibrio* spp. in natural environments is crucial to predicting how the future conditions will affect the dynamics of these bacteria. The majority of existing *Vibrio* spp. population growth models were developed in controlled environments, and their applicability to natural environments is unknown. We collected all available functional models from the literature, and distilled them into 28 variants using unified nomenclature. Next, we assessed their ability to predict *Vibrio* spp. abundance using two new and five already published longitudinal datasets on *Vibrio* abundance in four different habitat types. Results demonstrate that, while the models were able to predict *Vibrio* spp. abundance to an extent, the predictions were not reliable. Models often underperformed, especially in environments under significant anthropogenic influence such as aquaculture and urban coastal habitats. We discuss implications and limitations of our analysis, and suggest research priorities; in particular, we advocate for measuring and modeling organic matter.

Keywords: mechanistic modeling; primary and secondary growth models overview; comprehensive datasets; bacterial growth

Citation: Purgar, M.; Kapetanović, D.; Geček, S.; Marn, N.; Haberle, I.; Hackenberger, B.K.; Gavrilović, A.; Pečar Ilić, J.; Hackenberger, D.K.; Djerdj, T.; et al. Investigating the Ability of Growth Models to Predict In Situ *Vibrio* spp. Abundances. *Microorganisms* **2022**, *10*, 1765. https://doi.org/10.3390/microorganisms10091765

Academic Editor: Fabien Joux

Received: 24 July 2022
Accepted: 23 August 2022
Published: 31 August 2022

1. Introduction

Vibrio spp. are naturally occurring aquatic bacteria, highly adaptive and freely associated with a variety of biotic and abiotic surfaces including water, sediment, fish, shellfish, algae, and zooplankton. *Vibrio* spp. comprise a minor portion of the total microbial population, and around 1 percent of the total bacterioplankton in coastal waters [1]. Despite their relatively low abundance, *Vibrio* species are one of key constituents of aquatic heterotrophic bacterial groups [2].

Aquatic heterotrophic microorganisms have an important role in the mineralization of organic matter, and the variations in abundance, community structure, and activities of heterotrophic microbial communities affect both the biotic and the abiotic components of aquatic environments. *Vibrio* spp. participate in biogeochemical processes by utilizing a variety of substrates and mineralization of organic matter, thus directly contributing to the recycling of carbon, nitrogen, and organic matter in the aquatic environment [1,3,4]. Alongside their role in the abiotic cycles, some >140 described species from the genus *Vibrio* can have a strong biotic impact, and consequently pose severe health risks and economic losses. A well-known example is *Vibrio cholerae*, which causes cholera—a global threat to public health with about four million cases of infection every year, leading to over 100,000 deaths [5]. The rise of noncholera *Vibrio* species (*V. parahaemolyticus*, *V. alginolyticus*, and *V. vulnificus*) can cause other potentially lethal infections (vibriosis) in humans [2],

and some of the well-known *Vibrio* pathogens (e.g., *V. anguillarum*, *V. harveyi*, *V. vulnificus*, *V. salmonicida*) are harmful to aquatic (marine) organisms; these species induce vibriosis in fish and other marine species, which results in massive economic losses for the aquaculture industry worldwide.

Outbreaks of vibriosis naturally arise mainly with fluctuations in the physicochemical properties of water such as temperature, salinity, dissolved oxygen, and nutrient pulses (e.g., phytoplankton blooms and dust storms). Fluctuations are supported by the fast response of *Vibrio* spp. to favorable environmental conditions [6,7]. The ongoing climate change adds complexity to the environmental patterns, as it induces shifts in the marine environments by increasing temperature, altering nutrient loads, shifting precipitation patterns, and acidifying the ocean [8–10]. This, in turn, affects the *Vibrio* spp. abundance and alters the distribution of infectious diseases such as vibriosis [11]. Climate change can also initiate the lengthening of the seasonal period of maximal *Vibrio* concentrations and broaden the areas permitting the survival of these pathogens [12]. Therefore, in order to develop informed strategies to minimize vibriosis outbreaks and prevent potential health risks and aquaculture economic losses, it is crucial to take both *Vibrio* spp. dynamics and the environment into account.

Mathematical modeling, analysis, and simulations provide useful insights into *Vibrio* spp. abundance in various natural or anthropogenic systems. They help in developing management strategies, and advance the knowledge of changes in the microbial communities in aquatic environments. Accurate predictions can also advance the decision-making processes of aquaculture and estimation of costs, as well as the enactment of legislation in food safety and water research, which are two major areas in applied microbiology [13,14]. The main approach to modeling the *Vibrio* spp. is based on empirical techniques, where models are analyzed from statistical, numerical, and computational points of view, such as goodness of fit or standard errors of the estimated parameters. Mathematical models in food safety research (e.g., [15–18]) are generally categorized into primary, secondary, and tertiary models [19].

Microbial growth curves show the number of living cells as a function of time [17]. Primary models typically describe isothermal growth, i.e., growth as a function of time at a constant temperature. Under constant (laboratory) conditions, bacterial growth is characterized by a sigmoid curve where the dependent variable is the logarithm of the viable cell concentration [20]. The slope of a sigmoid curve at a given time provides the instantaneous specific growth rate, which can be considered as the "cells' per capita rate of division" [21]. The other two essential parameters of primary models are the maximum specific growth rate at the inflexion point of the sigmoid curve, and the length of the lag phase.

Secondary mathematical models describe the functional dependence of microbial growth on external factors such as temperature or pH [17]. Commonly used secondary models are the square-root model (Ratkowsky model [22]) and the Arrhenius-based model [23], which provide satisfactory descriptions of the dependence of growth on temperature, pH, or other factors [24]. Finally, tertiary mathematical models are software packages that combine primary and/or secondary models and often add a user interface [24].

Generally, all described models are useful, but they are not always applicable. The described mathematical models can be used for description and prediction of *Vibrio* spp. abundance only under certain (known and tested) conditions. The limitations stem from model formulation: while the equations provide satisfactory descriptions of *Vibrio* growth and its dependence on, e.g., temperature or pH, they usually are not subject to mechanistic interpretation [24], and therefore should not be used for explanatory purposes or predictions under untested conditions. The issue of the applicability and usability of a model becomes important when we wish to apply a model to understand or predict *Vibrio* spp. dynamics in situ. For data collected in situ, exact conditions of data collection are unknown, intervals between data collection are long and/or irregular, and environmental conditions are rarely constant.

We aimed to (i) identify and systematize primary and secondary *Vibrio* spp. growth models, and (ii) to analyze and validate the models by applying them to different sets of available data. We reviewed growth models of *Vibrio* spp. using examples from research in food safety and water research, and validated them on several sets of field data. Then, we analyzed whether the model(s) can be used to capture in situ *Vibrio* spp. abundances, with an emphasis on differences between the marine habitat types.

2. Materials and Methods

Our methodological approach can be divided into four main steps: (1) a literature search, (2) data preparation, (3) model simulations, and (4) performance analysis. The methodology overview is graphically presented in Figure 1.

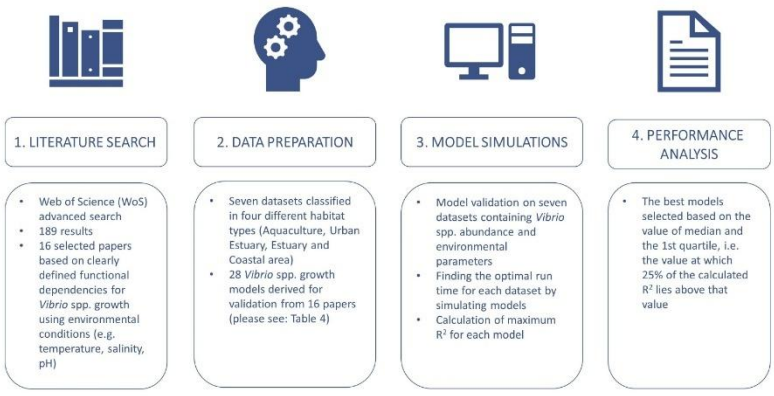

Figure 1. The methodological approach to model analysis. We performed a literature search to find all *Vibrio* spp. growth models and all available datasets suitable for comparison. We found five published datasets, and included two of our own collected through projects AqADAPT and AQUAHEALTH of the Croatian Science Foundation (HRZZ).

2.1. Literature Search and Model Synthesis

The literature search was conducted using the Web of Science (WoS) advanced search in April 2022. The search string was defined based on keywords and Boolean and adjacency operators, and was searched for in Abstracts (Field Tag "AB"). The search string was as follows: AB = (((vibrio*) AND (growth OR abundance) AND (temperature OR salinity OR "pH" OR "COD" OR "organic matter" OR nutrient*) AND (model*))). We obtained 189 results based on our search, which was restricted to the English language. For a detailed description of our literature review, please see Appendix A. We selected 16 papers with *Vibrio* spp. growth models based on clearly defined functional dependencies for using environmental conditions (e.g., temperature, salinity, pH). We did not include models derived purely by regression or similar statistical means because those are not modular, and cannot be differentiated into primary and secondary. From the 16 papers containing growth models, we extracted, systematized, and classified explicit formulations for primary (Table 1) and secondary growth models (Table 2), using parameters listed in (Table 3) and thus arriving at 28 unique *Vibrio* spp. growth models for further analysis (Table 4).

Table 1. Systematized equations for *Vibrio* spp. primary growth models. Of the 12 models listed in this table, one (new logistic model) did not have parameters listed, so only the remaining 11 were used in further analysis. The reference in the column "Model" is the original paper containing the equation. The column "Article" lists all published articles that used the given primary models.

Model	Equation		Article
Modified logistic [25]	$Y(t) = \frac{A}{\left\{1+\exp\left[\frac{4\mu_{\max}}{A}(\lambda - t)+2\right]\right\}}$	(1)	[26–29]
Baranyi [30]	$\begin{cases} Y(t) = Y_0 + \mu_{\max}\mathrm{A(t)} - \ln[1+\frac{\exp[\mu_{\max}\mathrm{A(t)}]-1}{\exp(Y_{\max}-Y_0)}] \\ \mathrm{A(t)} = t + \frac{1}{\mu_{\max}}\ln[\exp(-\mu_{\max}t)+\exp(-\mu_{\max}\lambda)-\exp(-\mu_{\max}t-\mu_{\max}\lambda)] \end{cases}$	(2)	[26,28,29,31–37]
Gompertz [25]	$Y(t) = Y_0 + A\left(e^{\left(-e^{-B(t-D)}\right)}\right)$	(3)	[37–39]
Modified Gompertz [25]	$Y(t) = Y_0 + A\exp\left\{-\exp\left[\frac{\mu_{\max}\cdot e}{A}(\lambda - t)+1\right]\right\}$	(4)	[28,29,31,40–42]
Weibull [43]	$Y(t) = Y_0 - \left(\frac{t}{\delta}\right)^p$	(5)	[28,41]
Three-phase linear [44]	$\begin{cases} Y(t) = Y_0, t \le \lambda \\ Y(t) = Y_0 + \mu_{\max}(t-\lambda), \lambda < t < t_s \\ Y(t) = Y_{\max}, \quad t \ge t_s \end{cases}$	(6)	[29,45]
Huang [46]	$\begin{cases} Y(t) = Y_0 + Y_{\max} - \ln\{\exp(Y_0) + [\exp(Y_{\max}) - \exp(Y_0)]\exp(-\mu_{\max}\mathrm{B(t)})\} \\ \mathrm{B(t)} = t + \frac{1}{4}\ln\frac{1+\exp[-4(t-\lambda)]}{1-\exp(4\lambda)} \end{cases}$	(7)	[29,47]
No-lag phase [48]	$Y(t) = Y_0 + Y_{\max} - \ln\{\exp(Y_0) + [\exp(Y_{\max}) - \exp(Y_0)]\exp(-\mu_{\max}t)\}$	(8)	[47]
Net exponential	$Y(t) = Y_0 \cdot e^{\mu t}$	(9)	[49]
Modified Richards [25]	$Y(t) = A\left\{1+\nu\cdot\exp(1+\nu)\cdot\exp\left[\frac{\mu_{\max}}{A}(1+\nu)\left(1+\frac{1}{\nu}\right)\cdot(\lambda - t)\right]\right\}^{\left(-\frac{1}{\nu}\right)}$	(10)	[29]
Modified Schnute [25]	$Y(t) = \left(\mu_{\max}\frac{(1-b)}{a}\right)\left[\frac{1-b\cdot\exp(a\cdot\lambda+1-b-at)}{1-b}\right]^{\frac{1}{b}}$	(11)	[29]
New logistic [50]	$\frac{\mathrm{d}Y}{\mathrm{d}t} = \mu_{\max}Y\left\{1-\left(\frac{Y}{Y_{\max}}\right)^m\right\}\left\{1-\left(\frac{Y_{min}}{Y}\right)^n\right\}$	(12)	[51]

Table 2. Systematized equations for *Vibrio* spp. secondary growth models. These models modify specific growth rate and lag time in primary models to capture effects of environmental conditions such as temperature, salinity, and pH.

Model	Equation		Article
Square root [52]	$\mu_{max} = [a(T - T_{min})]^2$	(13)	[26,33,35,40,41,45]
Polynomial model	$\lambda \text{ or } \mu_{max} = a + a_1 T + a_2 T^2 + \ldots + a_n T^n$	(14)	[32,34,36,51]
Response surface [37]	$\mu_{max} \text{ or } 1/\lambda = \exp\left(C_0 + C_1 \cdot T + C_2 \cdot a_w + C_3 \cdot T \cdot a_w + C_4 \cdot T^2 + C_5 \cdot a_w^2\right)$	(15)	[37]
Arrhenius-based [23]	$\mu_{max} = \exp\left(C_0 + \frac{C_1}{T} + \frac{C_2}{T^2} + C_3 a_w + C_4 a_w^2\right)$	(16)	[31,37,40,42]
Modified Ratkovsky [22]	$\mu_{max} = b(T - T_{min})^2\{1 - \exp[c(T - T_{max})]\}$	(17)	[31]
Suboptimal Huang square root [53]	$\mu_{max} = \left[a(T - T_{min})^{0.75}\right]^2$	(18)	[47]
Four-parameter square root and water activity [38]	$\mu_{max} = \left(b(T - T_{min})\{1 - \exp[c(T - T_{max})]\}\right)^2 \cdot (a_w - a_{w\,min})\{1 - \exp[d(a_w - a_{w\,max})]\}$	(19)	[38]
Net *Vibrio* growth rate [49]	$\mu_v = [\mu_{max} * \text{fn}(S, S_{opt}, S_{width}) - k_d] * \theta^{T-20}$, with $\text{fn}(S, S_{opt}, S_{width}) = \frac{-(S - S_{opt})^2}{e^2 (S_{width})^2}$, if $S < S_{opt} - 0.5 \cdot S_{width}$, or $S > S_{opt} - 0.5 \cdot S_{width}$	(20)	[49]

Table 3. Parameters used in *Vibrio* spp. primary and secondary models. Last column lists models using the particular parameter.

Parameter	Description	Used in Model
$Y(t)$ Y_0 Y_{max}	Logarithm of real-time initial and maximum bacterial counts	All primary models
μ_{max}	Maximum specific growth rate	All primary models except Weibull and New logistic All secondary models except Net *Vibrio* growth rate
μ_v	Net *Vibrio* growth rate	Net *Vibrio* growth rate
λ	Lag time	All primary models except Gompertz, Weibull and New logistic
t	Time	All models
t_s	Time to reach stationary growth phase	Three-phase linear
A	Maximum increase in microbial cell density	Modified logistic Gompertz Modified Gompertz Modified Richards Modified Schnute
B, D	Maximum relative growth rate and time at which the absolute growth rate is maximum	Gompertz
ν	Shape parameter	Modified Richards
a, b, c, m, n	Fitted coefficients	Modified Schnute and New logistic model
C_0, C_1, C_2 C_3, C_4, C_5	Fitted coefficients	Response surface and Arrhenius-based
T, T_{min}, T_{max}	Temperature, minimum and maximum temperature required for growth of the organism	All secondary models
δ, p	Coefficients in the Weibull model	Weibull
$a_w, a_{w\,min}, a_{w\,max}$	Optimal, the minimum, and maximum water activity	Four-parameter
S, S_{opt}, S_{width}	Salinity, optimal salinity value, and salinity range for optimal growth	Net *Vibrio* growth rate

Table 4. List of models used in the analysis. For each model, *Vibrio* spp. is specified along with the environment where the growth of the organism was observed. The primary model defines growth function and the secondary model describes functional dependencies accounting for environmental conditions (temperature, salinity, and pH). "Temp" stands for temperature, "Sal" for salinity represented in models as the concentration of NaCl (% NaCl), and "Sal (w.a.)" stands for water activity calculated from salinity. In simulations, salinity from datasets was converted to water activity when needed.

Derived Model	*Vibrio* spp.	Environment	Environmental Conditions	Primary Model	Secondary Model
Model 1 [26]	*V. cholerae*	Sea water	Temp	Modified logistic	Square root
Model 2 [26]	*V. cholerae*	Sea water	Temp	Baranyi	Square root
Model 3 [32]	*V. parahaemolyticus*	Soy sauce	Temp	Baranyi	Polynomial
Model 4 [33]	*V. parahaemolyticus*	*C. gigas*	Temp	Baranyi	Square root
Model 5 [35]	*V. parahaemolyticus*	*C. virginica*	Temp	Baranyi	Square root
Model 6 [36]	*V.* cocktail [1]	Table Olives	pH and Sal	Baranyi	Polinomial
Model 7 [34]	*V. cholerae* and *V. vulnificus*	*O. minor*	Temp	Baranyi	Polinomial
Model 8 [37]	*V. harveyi*	TSYEB [2]	Temp and Sal (w.a.)	Baranyi	Response surface
Model 9 [37]	*V. harveyi*	TSYEB [2]	Temp and Sal (w.a.)	Baranyi	Arrhenius-based
Model 10 [31]	*V. parahaemolyticus*	*L. vannamei*	Temp	Baranyi	Modified Ratkowsky
Model 11 [37]	*V. harveyi*	TSYEB [2]	Temp and Sal (w.a.)	Gompertz	Response surface
Model 12 [37]	*V. harveyi*	TSYEB [2]	Temp and Sal (w.a.)	Gompertz	Arrhenius-based
Model 13 [38]	*V. parahaemolyticus*	Model broth system	Temp and Sal (w.a.)	Gompertz	The four-parameter square root
Model 14 [31]	*V. parahaemolyticus*	*L. vannamei*	Temp	Modified Gompertz	Modified Ratkowsky
Model 15 [40]	*V. parahaemolyticus*	Broth	Temp	Modified Gompertz	Square root Arrhenius-based
Model 16 [40]	*V. vulnificus*	Broth	Temp	Modified Gompertz	Square root Arrhenius-based
Model 17 [40]	*V. parahaemolyticus*	Flounder sashimi	Temp	Modified Gompertz	Square root Arrhenius-based
Model 18 [40]	*V. parahaemolyticus*	Salmon sashimi	Temp	Modified Gompertz	Square root Arrhenius-based
Model 19 [40]	*V. vulnificus*	Oyster meat	Temp	Modified Gompertz	Square root Arrhenius-based
Model 20 [42]	*V. parahaemolyticus* [3]	*C. gigas* broth	Temp	Modified Gompertz	Square root Arrhenius-based
Model 21 [42]	*V. parahaemolyticus* [4]	*C. gigas* broth	Temp	Modified Gompertz	Square root Arrhenius-based
Model 22 [42]	*V. parahaemolyticus* [3]	*C. gigas* Oyster slurry	Temp	Modified Gompertz	Square root Arrhenius-based
Model 23 [42]	*V. parahaemolyticus* [4]	*C. gigas* Oyster slurry	Temp	Modified Gompertz	Square root Arrhenius-based
Model 24 [41]	*V. parahaemolyticus*	*Oncorhynchus* spp.	Temp	Modified Gompertz, Weibull	Square root
Model 25 [45]	*V. parahaemolyticus*	*L. vannamei*	Temp	Three-phase linear	Square root
Model 26 [47]	*V. parahaemolyticus*	*L. vannamei*	Temp	Huang primary	Suboptimal Huang square root
Model 27 [47]	*V. parahaemolyticus*	*L. vannamei*	Temp	No-lag	Suboptimal Huang square root
Model 28 [49]	*Vibrio* spp.	NR Estuary	Temp and Sal	Net exponential	Net *Vibrio* growth rate

[1] *V. vulnificus*, *V. furnissii* and *V. fluvialis*, [2] Tryptone Soybean Yeast Extract Broth, [3] pathogenic, [4] nonpathogenic.

2.2. Data Preparation

An additional literature search identified five datasets suitable for model validation; hence, a total of seven datasets were available for analysis (Table 5) once our two previously unpublished datasets were added.

The previously unpublished datasets, AqADAPT [54] and AQUAHEALTH [55], contain observed values for *Vibrio* spp. abundance and environmental parameters from the Adriatic Sea. Sampling was conducted in three floating-cage fish farms in the northern, middle, and southern Adriatic Sea (Croatia), where European sea bass (*Dicentrarchus labrax*) and sea bream (*Sparus aurata*) are cultured. The farms in the northern and central Adriatic are located in the semi-open sea at depths of about 49 m and 22 m, respectively. The fish farm in the southern Adriatic is located in the outer part of the Mali Ston Bay at a depth of 18 m. The Mali Ston Bay is occasionally strongly influenced by the (freshwater) Neretva River. Periodic use of antibiotics is possible on all three fish farms, and this would have affected the *Vibrio* spp. abundance. However, no specific data on antibiotics use are available. We classified these datasets into aquaculture habitat type. Dataset AqADAPT [54] was labeled as AQC1 and dataset AQUAHEALTH [55] was labeled as AQC2.

Table 5. The seven datasets used for model validation. AQC1 and AQC2 were previously unpublished; the other datasets are publicly available and can be accessed through the provided reference. Information for each dataset contains reported values (i.e., the number of entries in a dataset), values used for validation (i.e., the number of observations after the missing values were removed from the dataset), temperature, salinity, and pH range. Seven datasets used for model validation were classified into four habitat types based on the characteristics of the collection sites. Methods used for determining *Vibrio* spp. abundance are listed in Appendix C.1.

Dataset	Reported Values	Values for validation	Temperature Range (°C)	Salinity Range (ppt)	pH	Habitat TYPE	Collection Site
AQC1 [54]	108	99	11.1–27.5	33.5–39.3	8.10–8.61	Aquaculture	Adriatic Sea, Croatia
AQC2 [55]	88	81	7.86–25.23	24.9–38.2	7.56–8.49	Aquaculture	Adriatic Sea, Croatia
URB1 [56]	213	149	22.4–31	7.98–34.74	7.51–8.27	Urban Estuary	Ala Wai Canal in Honolulu, Hawaii
URB2 [57]	243	240	19.2–31.8	1.0–36.0	/	Urban Estuary	Ala Wai Canal in Honolulu, Hawaii
EST1 [58]	249	223	3.1–31.7	0.09–18.56	6.57–9.17	Estuary	Neuse River Estuary, North Carolina (USA)
EST2 [59]	133	127	2.16–25.89	9.32–31.86	6.82–8.41	Estuary	Great Bay Estuary, New Hampshire (USA)
COAST [60]	117	72	8.9–29.4	12.0–40.0	/	Coastal Area	Eastern North Carolina coast (USA)

Bullington et al. [56] and Steward et al. [57] published *Vibrio* spp. abundance and environmental parameters from the Ala Wai Canal in urban Honolulu, Hawaii, on the island of O'ahu. The 3.1 km-long engineered waterway operates as a tidally influenced estuary with freshwater input from a watershed that covers 42.4 km^2 via the Manoa and Palolo streams, which merge to form the Manoa–Palolo Stream prior to entering the canal, and the Makiki Stream, all of which run through urban areas before reaching the canal. Consequently, the streams are contaminated with a variety of anthropogenic substances, and their convergence in the Ala Wai Canal has contributed to its pollution and eutrophication [57]. We classified datasets from this area as the urban estuary habitat type due to the strong anthropogenic influence. The dataset by Bullington et al. [56] was labeled as URB1 and the dataset by Steward et al. [57] was labeled as URB2.

Froelich et al. [58] gathered data from the Neuse River Estuary in Eastern North Carolina (USA). The Neuse River Estuary, located in Eastern NC (USA), is a well-described, lagoonal estuary, with wind-driven mixing characteristics and minimal tidal influence due to the protection offered by the proximal Pamlico Sound. Being broad and shallow (generally less than 3 m in depth), the estuary flow and mixing are dominated by wind and river input. This dataset was classified as an estuary habitat and labeled as EST1.

Urquhart et al. [59] collected data from the Great Bay Estuary, New Hampshire (USA). The Great Bay Estuary (GBE) extends inland from the mouth of the Piscataqua River near Kittery, ME, through Little Bay and eventually into the Great Bay (25 km). The GBE has deep, narrow channels with strong tidal currents, and wide, shallow mudflats. The physical transport regime of the GBE follows the classical estuarine circulation model for drowned river valley estuaries [59]. This dataset was also classified as an estuary habitat and labeled EST2.

Williams et al. [60] presented data from five sites along the Eastern North Carolina coast (USA). Locations were as follows: Harlowe Creek, South River, North River, Hoop Pole Creek, and Jumping Run Creek. These sites were chosen to represent the range of high- and low-salinity environments, some of which experience large salinity fluctuations, while others have very small salinity fluctuations (for more information, please refer to the original manuscript [60]). We classified this dataset as a coastal area, and labeled it COAST.

From the given datasets, we selected variable *Vibrio* abundance and the following environmental parameters: temperature, salinity, and pH. We excluded all of the missing values from datasets and logarithmically transformed *Vibrio* abundance greater than 0 (log10 + 1 in datasets AQC1, AQC2, and COAST, log10 in datasets URB1 and URB2). Datasets EST1

and EST2 already contained logarithmically transformed values. The analytical code can be found in the R script named `data_preparation`. Results of the dataset preparation are summarized in Table 5.

2.3. Model Simulations

Herein, we showcase a model simulation approach aiming to calculate *Vibrio* spp. abundance based on primary and secondary models accounting for environmental parameters (temperature, salinity, and pH). All collected models were developed for controlled environments (i.e., laboratories), where *Vibrio* spp. growth was monitored at regular, short time intervals. Such data lend themselves to time series modeling, where abundance is plotted against time. In contrast, in situ data are typically irregularly collected from variable abiotic microenvironments, at longer time intervals, and with the noise typically inherent to field measurements. Such data could not be modeled as a time series, but had to be modeled as independent abundances—each data point was considered to be a result of bacterial growth that started some time ago.

The model had to be simulated for the time of growth, but determining how long ago the growth started was a challenge which we needed to overcome in order to select the (optimal) model run time for a dataset. Note that having an independent run time for each data point would create an option to fit each data point exactly by choosing the perfect time, thus defeating the whole point of modeling the bacterial dynamics. To minimize the bias introduced by our choice of the model run time, we determined a (run) time that gives the best result for each model-and-dataset combination. Optimal run time is then the simulation duration that produces the best match between the model predictions and the observations.

We used default parameter values (Appendix B, Table A1) listed in their respective references for each of the 28 models (Table 4). First, the values of the specific growth rate and lag time were calculated, which were then used in the primary models: modified logistic, Barayni, Gompertz, modified Gompertz, three-phase linear, Huang, no-lag phase and met exponential models. These primary models adjust the specific growth rate and lag time using one or more of the secondary models, as described in the original literature and summarized in Table 4. The environmental parameters considered in the literature for modeling *Vibrio* spp. growth in dynamic conditions were: temperature, salinity, and pH (Table 4). Secondary models sometimes use water activity or NaCl concentration instead of salinity. As all datasets contained information on salinity, water activity or NaCl concentration were calculated from salinity (Table 5) (for more details, please refer to Appendix B). To determine the optimal run time, we simulated *Vibrio* spp. growth in a selected time range from 1 to 600 hours, and selected the run time that produced the best fit to the data.

2.4. Model Performance

We used the coefficient of determination (R^2) to evaluate the ability of models to describe the observed experimental data:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}, \tag{21}$$

where y_i is the actual value in the dataset, $\hat{y}_i$ is the corresponding predicted value, $\bar{y}$ is the mean value of the dataset, and n is the sample size.

To determine applicability (i.e., model generality) of a specific model to a specific dataset, we compared R^2 values calculated for all models for that specific dataset, and then selected those models for which the dataset-specific R^2 value was higher than the overall median of all R^2 values for all models and all datasets. For example, if the median overall goodness of fit of all models to all datasets was $R^2 = 0.30$, all models with $R^2 > 0.30$ for dataset 1 would be marked as capturing dataset 1. We then tested the robustness of the results by looking at a more stringent requirement, where we marked a certain model as

capturing a dataset only if its R^2 for that dataset was in the top 25% (1st quartile) of the values for all models and all datasets.

Robust ANOVA based on trimmed means [61,62] was used to test the difference in model performance (obtained R^2) between different habitat types (aquaculture, urban estuary, estuary, and coastal area). Robust ANOVA was used to overcome the problems associated with deviations from homoscedasticity/normality and to reduce the influence of outliers observed in the data. Post hoc tests were also performed in the robust WRS2 environment [61], where p-values were adjusted for multiple testing using the Benjamini–Hochberg (BH) method.

The model analysis and simulations were performed in RStudio Integrated Development Environment, Version 4.1.2 [63] using the packages: tidyverse [64], caret [65], Metrics [66], SciViews [67], data.table [68], readxl [69] and xlsx [70]. The exact analytic code can be found at Zenodo [71]. We visualized results using the package ggplot2 [72].

3. Results

3.1. *Vibrio spp.* Growth Models

Vibrio spp. growth models identified by the literature review could be partitioned into 12 primary (Table 1) and 8 secondary (Table 2) models, using the parameters listed in Table 3. In total, we identified 29 models for *Vibrio* spp. growth in a dynamic environment, but we could not find parameter values for Fujikawa et al. [50]. Therefore, we further analyzed only the remaining 28 models (Table 4), using the parameters listed in Table A1.

The model summaries (Table 4) provide an overview of the models used when studying *Vibrio* spp. growth in a dynamic environment. The main findings were as follows:

- Baranyi and modified Gompertz are the most commonly used primary models for describing *Vibrio* spp. growth over time.
- Square root and Arrhenius-based models are the most frequently applied secondary models for *Vibrio* spp. growth in dynamic conditions.
- *V. cholerae*, *V. parahaemolyticus*, *V. harveyi*, and *V. vulnificus* are the species used most often as modeling organisms.
- *Vibrio* growth was monitored in/on various substrates (free water column, within organisms, in broth substrates, etc.), under different temperature, salinity, and pH conditions. This implies that the aquatic environments and organisms (marine and freshwater), as well as food and water health and safety, are the key areas of research and concern.
- Temperature was the prevailing environmental parameter used in secondary models, implying a strong effect of temperature on *Vibrio* spp. abundance. The effect of temperature on the primary model parameters (growth rate and lag time) was most often modeled by the square root or the Arrhenius-based model.

3.2. *Vibrio In Situ Datasets*

We summarize the dataset classification and characteristics of the seven datasets in Table 5, including the number of *Vibrio* spp. abundance data, and range and type of environmental variables used in model validation (temperature, salinity, pH). Datasets URB2 and COAST do not contain information on pH, so simulations of Model 6 (Table 4) were not possible for those datasets.

3.3. Model Performance

The ability of models to describe the data greatly varied between models and datasets (Figure 2). The calculated R^2 values ranged from <0.001 (Model 10 for dataset AQC1) to 0.40 (Model 28, dataset URB2), with the overall median value of all models for all datasets $R^2 = 0.13$. R^2 values also significantly differed between habitat types (Figure 3, Table A2) in all pairwise comparisons (post hoc tests $p < 0.017$; see details in Table A3). Therefore, R^2 values suggest the models performed best for coastal areas, followed by estuaries, urban estuaries, and aquaculture habitats. Comparing R^2 values provides performance

estimates of a particular model on a particular dataset, but does not provide information on generality.

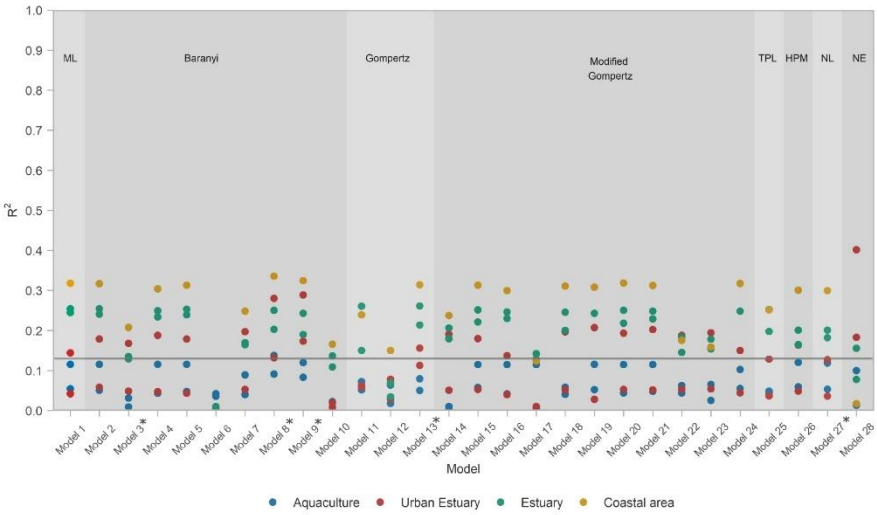

Figure 2. R^2 calculated for all models and for each dataset. Horizontal line depicts the median $R^2 = 0.13$ used to evaluate model performance. Primary models are labeled as follows: ML—modified logistic, Baranyi, Gompertz, modified Gompertz, TPL—three-phase linear, HPM—Huang primary, NL—no-lag and NE—net exponential. Star (*) signifies models that had an evaluation issue with some of the data points in some of the datasets (details in Appendix D).

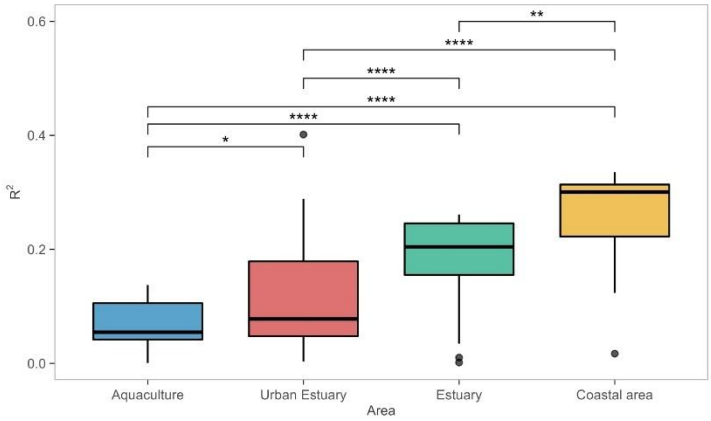

Figure 3. Boxplot of model performance measured as R^2 in different habitat types (as classified in Table 5). Model performance significantly differed between habitat types (robust ANOVA, F(3,75.561) = 49.9, $p < 0.001$, effect size $\xi = 0.77$, confidence interval CI(ξ) = [0.68, 0.84]; Table A2). Significance codes are as follows: $p \leq 0$ '****', $p < 0.01$ '**', and $p < 0.5$ '*'.

Model generality analysis (Figure 4A) shows rankings above the median ($R^2 > 0.13$) where:

Figure 4. Model generality. Stacked bar chart used for graphical representation of model's applicability on the dataset from a specific habitat: colors represent habitats (see the legend), and stars (*) denote models that exhibit the evaluation issue with some of the data points in some of the datasets (details in Appendix D). Values on y axis denote the frequency of occurrence of a particular model whose R^2 value is above the median ($R^2 > 0.13$; Panel **A**), and in the first quartile ($R^2 > 0.22$; Panel **B**). Primary models are labeled as follows: ML—modified logistic, Baranyi, Gompertz, modified Gompertz, TPL—three-phase linear, HPM—Huang primary, NL—no-lag and NE—net exponential. Star (*) signifies models that had an evaluation issue with some of the data points in some of the datasets (details in Appendix D). Note that there is only one coastal area dataset, while other habitats have two datasets each. Hence, scoring a single occurrence of the coastal area habitat represents a 100% success rate, while scoring the same in any other habitat represents a success rate of 50%.

- All models except Model 6 (Baranyi with polinomial pH and salinity secondary models) were able to score above average at least in some habitats, i.e., capture those habitats. Incidentally, Model 6 was the only one *not* applying a temperature correction
- A total of 93% of models captured the coastal area habitat.
- A total of 93% of all models captured estuary habitat, but only 85% of those (i.e., 79% of all models) captured both estuary datasets; 26 models captured EST1, with 22 of them capturing EST2 as well.
- A total of 75% of models captured an urban habitat, but only two (Models 9 and 28) captured both urban habitat datasets; URB1 was captured by all 20 of them; only 3 models managed to capture URB2.
- Only the Baranyi-type model (Model 8, with temperature and salinity secondary models) captured the AQC1 aquaculture habitat.
- The model with the highest R^2 values (Model 28—net exponential, for urban estuary URB2) had low generality, as it captured datasets from only two out of four habitats.
- Of the models that performed well for at least one habitat type, Model 17 had lowest generality as it captured only one EST1 dataset.

Selecting for better capability by scoring only performances within the first quartile ($R^2 > 0.22$, Figure 4B) shows that:

- ≈A total of 72% of the analyzed models had an exceptional ability to capture datasets from the coastal area habitat.
- The ability of models to capture/perform well for estuarine habitats was severely diminished, with only 16/28 (57% of the models) capturing one of the two estuary datasets, and only 10 models (36%) capturing both.
- None of the models were able to capture for the aquaculture habitat datasets, and only three (Models 8, 9, and 28) captured the urban estuary habitat.
- Model 28 seemed even more specialized, as it captured a single urban estuary (URB2)
- Most prolific Baranyi models (Models 8 and 9) remained so by capturing datasets from three habitat types, albeit only a single dataset from each.

4. Discussion

Our work adds to the growing body of knowledge from the past few decades that helped refine a general understanding of the ecology of *Vibrio* spp. We (i) summarized dynamics and unified nomenclature for all published functional models we could find (12 primary and 8 secondary), (ii) added two datasets to the existing five longitudinal datasets on *Vibrio* spp. and their environments, and (iii) used the datasets to asses the ability of existing models to capture *Vibrio* spp. abundances in four different habitat types.

There are no clear winners between the 28 investigated models. Generally, the R^2 values were not particularly large ($\leq$0.40), but the values can be considered acceptable given the difficulty of the task, in particular the multitude of potential factors affecting both the environment and the *Vibrio* populations.

Baranyi-based models (Models 2 to 10 in Table 4) seemed to capture the widest variety of habitats well, with Models 8 and 9 leading the way in diversity. Although both models had a similar R^2 for the AQC1 dataset (0.14 and 0.12, respectively), Model 8 was the only one that crossed the median threshold and therefore captured an aquaculture dataset as an above-average performer. Notably, Models 8 and 9 are the only Baranyi-based models that included both temperature and salinity secondary models. Model 28 (net exponential) also included both factors and performed quite well, giving the highest overall R^2 (0.40 for the URB2 dataset).

Inclusion of temperature and salinity, however, does not guarantee success. All three Gompertz-based models included both factors, but all three underperformed. This would signal Gompertz-based models should be used with caution. Modified Gompertz-based models (Models 14–24), however, seem to lag only slightly behind the best.

While it may be tempting to proclaim Baranyi-based models as the most versatile, they do have significant drawbacks. First, both Models 8 and 9 had issues with simulations:

their fast growth rate sometimes caused very large predictions (see Appendix D). While this has not been a problem for the optimal run time of Model 8, Model 9 did lose up to 3% of data in the evaluation. These issues may not be consequential in the current assessment, but may become so in new datasets.

Second, the secondary models accounting for salinity in Models 8 and 9 yielded unexpected growth rate patterns. For example, at moderate and high salinities, growth rates could be extremely high at the ends of the environmental temperature range (e.g., 3.18 ln(CFU/g)/h for salinity of 20 and temperature of 27 °C, Figure 5, Panel B). Likewise, at 10 °C, the growth rate of Model 8 was moderate and slightly *increased* with salinity; at 30 °C, however, the rate was extremely high for low salinity, and rapidly *decreased* with salinity. This may be plausible as abundance of *Vibrio* spp. increased with water temperatures during periods of reduced salinity [73], but we recommend caution when utilizing Baranyi-based models in highly variable environments.

Overall, predictions for the coastal area and estuary datasets seemed to be more robust than those for urban estuaries and aquaculture (Figure 4, Panel B). We hypothesize this may be due to higher levels of anthropogenic disturbance in aquaculture and urban estuaries, in particular potentially high inputs of organic matter. *Vibrio* spp., as a prototypical copiotroph, dominates in nutrient-rich environments [74]. They exhibit a feast-and-famine lifestyle and swim to colonize sporadic, nutrient-rich patches and particles [75]. Therefore, we think that organic matter has an important role in determining *Vibrio* spp. abundance, especially in these types of habitats.

Our study has some limitations. First, we treated all *Vibrio* spp. the same. While justified in the context of a wide assessment such as ours, *Vibrio* species clearly differ and could potentially exhibit significantly different dynamics, especially since different species favor different environmental conditions. These differences could become important as environmental conditions change, and the *Vibrio* community could shift towards disease-causing species. Unfortunately, current datasets and available models do not allow for a deeper investigation of the issue.

Second, we chose the simulation time that minimized R^2, but relied on a single simulation time for each dataset. In principle, a different simulation time could be chosen for each data point. Such an approach would, however, in effect result in fitting each value by choosing simulation time that gives a desired result, thus defeating the purpose of modeling. There could, nonetheless, be some value in exploring functional dependencies of the simulation time on various environmental factors (temperature in particular), but additional research would be required to suggest a particular form of such a function.

Third, we set out to investigate *published* parameters and models with well-defined functional forms. Perhaps a different set of parameters and/or combinations of models could have described the datasets better. We have made a first step towards such research by systematizing primary and secondary models, parameters, and available—including two previously unpublished—datasets. Alternatively, *statistical* models could possibly be re-fitted to better capture the datasets. Such statistical approaches may be appropriate in some cases, e.g., when aiming to inter- or extrapolate data from a single area.

Clearly, additional research is needed for developing a growth model capable of predicting in situ *Vibrio* spp. abundances in natural environments. We suggest further development of secondary models should be one priority; dynamics of secondary models should be well-defined for the whole range of expected environmental factors, especially temperature and salinity. Modeling could also benefit from research on additional environmental factors such as organic matter, which has been shown to affect *Vibrio* growth [56]. Given the role of *Vibrio* spp. in aquatic environments, it is surprising that less than half datasets include organic matter measurements. We suggest that future modeling development should include organic matter, especially when sampling to ensure measurement of organic matter.

Perhaps a development of a third generation of models based on big data and deep learning could also work in synergy with mechanistic modeling to improve our ability to

predict *Vibrio* spp. dynamics in a changing environment. Improved models could then enhance predictive frameworks, e.g., by replacing the basic ecological niche approach to estimating *Vibrio* presence used in exploration of future risk scenarios by Trinanes and Martinez-Urtaza [76].

In conclusion, none of the investigated models provide a complete solution: Baranyi-based models might be the most versatile, but other models (e.g., net exponential) may provide a better fit for a particular cause. Therefore, the choice of the model should be at least in part guided by the type of the environment and expected ranges of environmental factors; if many data lie outside of the well-described range of a particular secondary model (see Figure 5), perhaps a different model should be tried instead. Our summary and systematization (Tables 1–5) provide currently available primary and secondary models, and data, which can be used as a toolbox for model creation and testing.

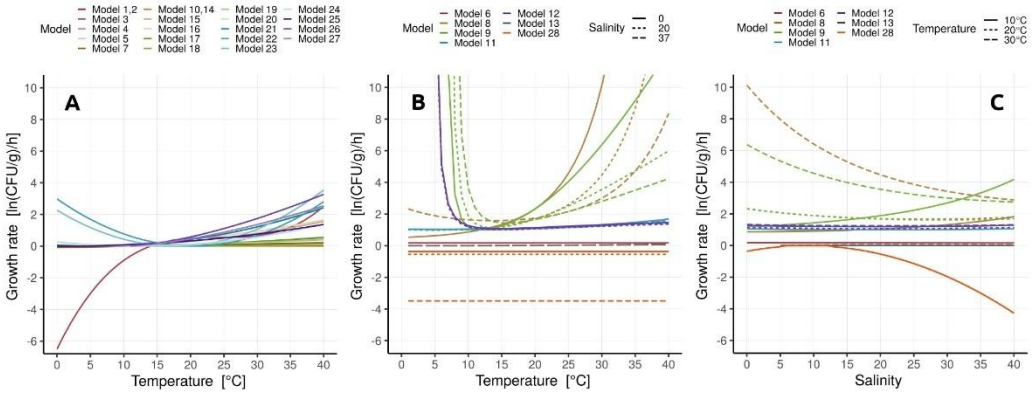

Figure 5. Relationships between growth rate and environmental variables as predicted by secondary models. The models in panel (**A**) include only a temperature correction. Panel (**B**) shows dependence of growth rate on temperature for three salinity levels. Panel (**C**) shows dependence of growth rate on salinity for three temperatures. A pH value of 8.1 was assumed for Model 6. Model 28 had a flat temperature response because it used the default parameter value [49] that minimized temperature correction, $\theta = 1$; increasing θ would increase the temperature dependence.

Author Contributions: Conceptualization, T.K. and M.P.; methodology, T.K. and S.G.; formal analysis, M.P. and S.G.; investigation, M.P. and D.K.; writing—original draft preparation, M.P., S.G., T.K., I.H. and N.M.; data acquisition, A.G., D.K., T.D., D.K.H., and B.K.H.; writing—review and editing, all authors; visualization, M.P. and T.K.; supervision, T.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Croatian Science Foundation (HRZZ) project AqADAPT (no. IP-2018-01-3150 to T.K., and DOK-2021-02-6688 to T.K. for M.P.).

Data Availability Statement: The five datasets from literature search are available at their respective references, and the two new datasets are submitted to PANGAEA open data repository (https://www.pangaea.de/), please search for keywords AqADAPT and/or AQUAHEALTH. The code used to perform analysis and to create plots is deposited at Zenodo, https://doi.org/10.5281/zenodo.7013394 (accessed on 20 August 2022).

Acknowledgments: We thank Kornat-Ittica d.o.o. (Pakoštane, Croatia), Orada Adriatic d.o.o. (Cres, Croatia), and Ciplić d.o.o. (Trpanj, Croatia) for access to sampling sites, and free use of equipment, materials, and labor in sampling.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Literature Search

The literature search was conducted using the Web of Science (WoS) advanced search in April 2022. We accessed all databases in the Web of Science: Web of Science core collection (Arts & Humanities Citation Index, Book Citation Index—Science, Book Citation Index—Social Sciences & Humanities, Conference Proceedings Citation Index—Science, Conference Proceedings Citation Index—Social Science & Humanities, Current Chemical Reactions, Emerging Sources Citation Index, ESCI Backfiles, Chemicus Index, Science Citation Index Expanded, Social Sciences Citation Index), BIOSIS citation index, Medline, Zoological record, Current contents connect, Derwent innovations index, Data citation index, SciELO citation index, BIOSIS previews, CABI–CAB abstracts and global health, Inspec, KCI–Korean journal database, journal citation reports, essential science indicators, EndNote online. The search string was defined based on keywords and Boolean and adjacency operators, and was searched for in Abstracts (Field Tag "AB"). The search string was: AB = (((vibrio*) AND (growth OR abundance) AND (temperature OR salinity OR "pH" OR "COD" OR "organic matter" OR nutrient*) AND (model*))). We obtained 189 results based on our search, which was restricted to the English language. We detected 9 potentially duplicate articles before the screening, which were resolved, and 180 remained. A primary search resulted in 27 articles for the screening. Furthermore, we performed a backward and a forward search based on identified relevant articles. A backward search was performed by searching a list of references at the end of the articles, and a forward search was conducted using Google Scholar. We performed this procedure two times until no new relevant article was identified. Finally, we included six more articles. The screening was conducted in Rayyan collaborative review application [77]. The first screening was performed by a reviewer, MP, and the final screening was carried out by reviewers TK and MP. We selected 16 papers with the *Vibrio* spp. growth models for analysis after the full screening of the papers from the primary search and additional search. Please see the PRISMA diagram for a breakdown of the overall procedure (Figure A1). In the analysis, we included articles with *Vibrio* spp. growth model equations depending on environmental parameters. Data extraction led to 28 *Vibrio* growth models for validation.

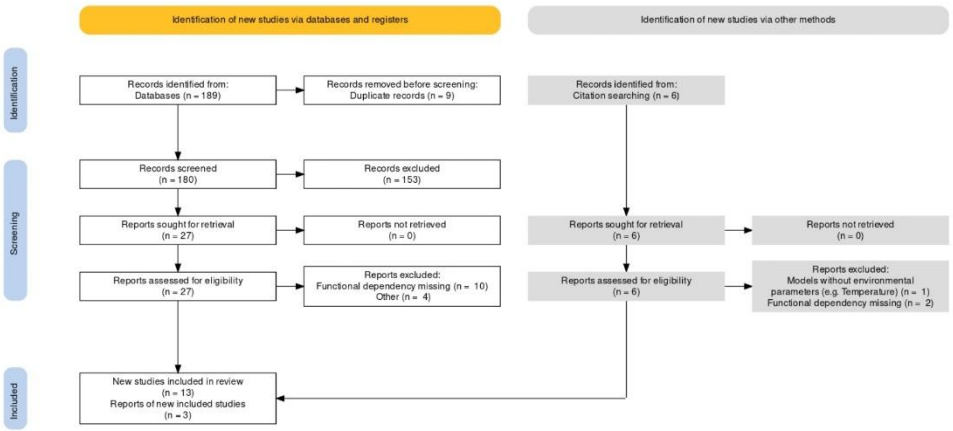

Figure A1. PRISMA flow chart for the literature retrieval and screening.

Appendix B. Data Analysis

Data analysis was based on 28 *Vibrio* spp. growth models extracted from 16 papers that clearly defined functional dependencies for growth rate and lag time using environmental conditions (e.g., temperature, salinity, pH). For example, in Model 1, we used a modified logistic function and square root model describing the temperature; in Model 2, we used the Baranyi function and square root model describing the temperature, etc. For more details, please see Table 2 in the main text. Model 6 used the concentration of sodium chloride calculated from salinity. Model 8, Model 9, Model 11, and Model 12 used water activity determined by the concentration of sodium chloride. The water activity term, used in Model 13, was calculated from the concentration of NaCl (%), i.e., salinity. Values for water activity and the corresponding NaCl concentration (%) can be found in [78]. The function for calculating water activity from the concentration of NaCl (%) is available in R script "water_activity.R". Analytical code for model simulations can be found in the following scripts: "AqADAPT.R", "AQUAHEALTH.R", "Bullington2022.R", "Froelich2019.R", "Steward2022.R", "Urquhart2016.R", and "Williams2017.R".

Table A1. List of parameters used in primary model analysis. Parameter A is the maximum increase in microbial cell density, and Y_0 and Y_{max} represent logarithm of initial and maximum bacterial counts, respectively. Parameter T_{min} is the minimum temperature required for growth of the organism. We used the minimum value from dataset divided by 2 (Est Y_0) to estimate initial bacterial count whenever it was not provided by the authors. In cases where the maximum bacterial count was not provided, we used the estimate of the maximal bacterial count from each dataset (Est Y_{max}). Parameters used in secondary models are available in the provided code.

Derived Model	A	Y_0	Y_{max}	T_{min} (°C)
Model 1 [26]	4	/	/	6.4 °C
Model 2 [26]	/	Est Y_0	Est Y_{max}	6.4
Model 3 [32]	/	Est Y_0	Est Y_{max}	15
Model 4 [33]	/	Est Y_0	Est Y_{max}	8.3
Model 5 [35]	/	Est Y_0	Est Y_{max}	10.0
Model 6 [36]	/	Est Y_0	Est Y_{max}	/
Model 7 [34]	/	Est Y_0	Est Y_{max}	8.0
Model 8 [37]	/	Est Y_0	Est Y_{max}	12.9
Model 9 [37]	/	Est Y_0	Est Y_{max}	12.9
Model 10 [31]	/	Est Y_0	Est Y_{max}	15.0
Model 11 [37]	/	Est Y_0	Est Y_{max}	12.9
Model 12 [37]	/	Est Y_0	Est Y_{max}	12.9
Model 13 [38]	/	Est Y_0	Est Y_{max}	8.0
Model 14 [31]	/	Est Y_0	/	15.0
Model 15 [40]	4	Est Y_0	/	13.0
Model 16 [40]	4	Est Y_0	/	13.0
Model 17 [40]	4	Est Y_0	/	13.0
Model 18 [40]	4	Est Y_0	/	13.0
Model 19 [40]	4	Est Y_0	/	13.0
Model 20 [42]	6	Est Y_0	/	10.0
Model 21 [42]	6	Est Y_0	/	10.0
Model 22 [42]	6	Est Y_0	/	10.0
Model 23 [42]	6	Est Y_0	/	10.0
Model 24 [41]	4	Est Y_0	/	12.1
Model 25 [45]	/	Est Y_0	9.28	12.1
Model 26 [47]	/	Est Y_0	7.64	10.8
Model 27 [47]	/	Est Y_0	7.70	10.5
Model 28 [49]	/	Est Y_0	/	/

Appendix C. Additional Results

Appendix C.1. Methods Used for Determining Vibrio spp. Abundance

In this subsection, we present more details on the techniques used to determine the abundance of *Vibrio* spp. in the studied environments. The results of different techniques (e.g., culture techniques or qPCR) are not always consistent and do not have the same meaning (e.g., possible presence of noncultivable viable bacteria). In datasets AQC1 [54] and AQC2 [55], *Vibrio* spp. abundance from water samples was determined by counting the total number of visible colonies that exhibited relief from the plate surface from the Thiosulphate Citrate Bile Salt Sucrose (TCBS) (Difco™, BD) agar plates. In the dataset URB1 [56], the polymerase chain reaction (qPCR) of the hemolysin A gene (vvhA) was used for determening *V. vulnificus* concentration in water samples. In dataset URB2 [57], the abundance of *V. vulnificus* was also determined by quantitative PCR (qPCR) of the

hemolysin gene (vvhA). In dataset EST1 [58], *Vibrio* spp. concentrations were determined by counting the total number of visible yellow and green colonies that exhibited relief from the plate surface from TCBS, adjusting for dilution, and expressing the results as colony-forming units (CFUs) per 100 mL. In the dataset EST2 [59], oyster tissue was processed for enumeration of *V. parahaemolyticus* via a three-tube MPN enrichment method following the FDA Bacteriological Analytical Manual coupled with culture-based and polymerase chain reaction (PCR) methods used to confirm the presence of *V. parahaemolyticus*. In the dataset COAST [60], the authors quantified *V. parahaemolyticus* from water samples by counting the total number of visible colonies using the CHROMagar Vibrio medium (CHROMagar, Paris, France).

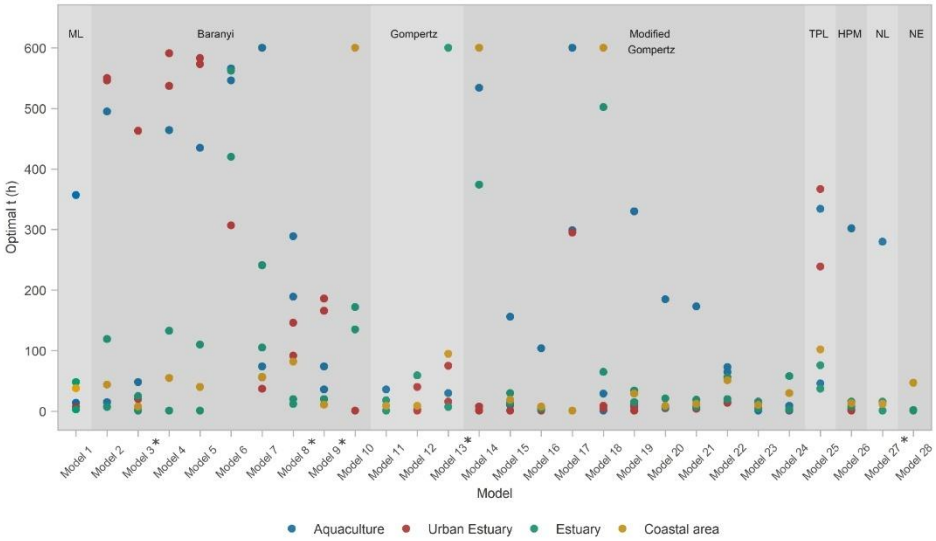

Figure A2. Optimal run time for the simulation duration that produces the best match between the model prediction and the observation (data point of a dataset) based on R^2 median value. Models based on Table 4 are specified on the x axis. Primary models are labeled as follows: ML—modified logistic, Baranyi, Gompertz, modified Gompertz, TPL—three-phase linear, HPM—Huang primary, NL—no-lag, and NE—net exponential. Star (*) signifies models that had an evaluation issue with some of the data points in some of the datasets (details in Appendix D).

Model performance and its applicability to specific datasets based on a comparison of R^2 median values ($R^2 > 0.13$) resulted in a total of 27 applicable models on datasets from four habitat types (aquaculture, urban estuary, estuary, and coastal area) (Figure 2). Model 6, based on the Baranyi equation and polinomial model for the effect of pH and NaCl, exhibited poor performance. Model performance and its applicability to specific datasets based on a comparison of Q1 (upper 25%) R^2 values ($R^2 > 0.22$) resulted in a total of 21 applicable models on datasets from three areas (urban estuary, estuary, and coastal area) (Figure 2). Here, Model 3, Model 6, Model 10, Model 12, Model 17, Model 22, and Model 23 displayed poor performance. Models 2, 6, and 10 are Baranyi's type with square root, polynomial, and modified Ratkowsky models for the effect of temperature or pH and NaCl, respectively. Model 12 is Gompertz's type with modified Ratkovsky and Arrhenius-based models, respectively. Model 22 and Model 23 were based on the modified Gompertz model and used square root and Arrhenius-based models for describing the effect of temperature.

Table A2. Results of robust ANOVA one-way test from the package WRS2 [61].

Function:	t1way (formula = max_r2~Habitat, data = as)
Test statistic:	F = 75.561
Degrees of freedom 1:	3
Degrees of freedom 2:	49.90
p-value:	0
Explanatory measure of effect size:	0.77
Bootstrap CI:	[0.68; 0.84]

Table A3. Results of post hoc lincon test from the package WRS2 [61].

Formula:		lincon (max_r2~Habitat, data = as)		
Habitat type	psihat	ci.lower	ci. upper	*p*-value
Aquaculture vs. Urban Estuary	−0.03910	−0.08237	0.00417	0.01688
Aquaculture vs. Estuary	−0.14203	−0.17331	−0.11075	0.00000
Aquaculture vs. Coastal Area	−0.21525	−0.27062	−0.15989	0.00000
Urban Estuary vs. Estuary	−0.10293	−0.14925	−0.05662	0.00000
Urban Estuary vs. Coastal Area	−0.17616	−0.23977	−0.11254	0.00000
Estuary vs. Coastal Area	−0.07322	−0.13069	−0.01575	0.00263

Appendix D. Model Evaluation Issues

Some models had evaluation issues, where a proportion of the data had to be disregarded. The issues for each model are described below, and Table A4 summarizes the issues and proportion of data disregarded at the optimum simulation time.

Model 3 had negative specific growth rates generated by the secondary model (the parameters used for a four-parameter polynomial model by [32]). This phenomenon was observed for temperatures below 15 °C. Model 3 also generated Inf values for higher simulation times and higher specific growth rates. Model 7 generated negative specific growth rates for temperatures between 11.1 and 11.7 °C. A secondary model was polynomial, as described in [34]. Additionally, in this secondary model, lag time had lower values at 17.6 °C and the highest at 27.5 °C. Model 8 resulted in Inf values in cases where maximum specific growth rate was between 1.630371 (12.3 °C) and 2.476982 (27.5 °C) and time range was between 289 and 600 h. This large value of the specific growth rate was generated by the response surface secondary model from [37]. Some of the predicted values were equal to the parameter maximum bacterial count (Y_{max}) (we obtained NA R^2). Model 9 resulted in Inf values in cases where mumax was between 1.491716 (15.9 °C) and 2.459676 (27.5 °C) and the time range was between 291 and 600. This large value of the specific growth rate was generated by Arrhenius–Davey secondary model from [37]. Some of the predicted values were equal to the parameter maximum bacterial count (Y_{max}) (we obtained NA R^2). Model 13 resulted in NAN predictions for values that had salinity, i.e., water activity higher than 0.998, and Model 27 had NAN predictions for temperatures below 10.5 °C.

Table A4. Models and their issues.

Model	Issue	Impacts % of Dataset	Dataset
Model 3	Negative specific growth rate	26/99 = 26.26%	AQC1 [54]
		16/81 = 19.75%	AQC2 [55]
		52/223 = 23.32%	EST1 [58]
		30/127 = 23.62%	EST2 [59]
		4/72 = 5.56%	COAST [60]
Model 8	High values of specific growth rate which generate Inf values	/	/
Model 9	High values of specific growth rate which generate Inf values	1/81 = 1.23%	AQC2 [55]
		6/223 = 2.69%	EST1 [58]
		4/127 = 3.15%	EST2 [59]
Model 13	Salinity, i.e., water activity >0.998	80/223 = 35.87%	EST1 [58]
		18/240 = 7.50%	URB2 [57]
Model 27	Temperature 10.5 °C	2/81 = 2.47%	AQC2 [55]
		29/223 = 13.00%	EST1 [58]
		16/127 = 12.60%	EST2 [59]
		1/72 = 1.39%	COAST [60]

References

1. Zhang, X.; Lin, H.; Wang, X.; Austin, B. Significance of Vibrio species in the marine organic carbon cycle—A review. *Sci. China Earth Sci.* **2018**, *61*, 1357–1368. [CrossRef]
2. Sampaio, A.; Silva, V.; Poeta, P.; Aonofriesei, F. *Vibrio* spp.: Life strategies, ecology, and risks in a changing environment. *Diversity* **2022**, *14*, 97. [CrossRef]
3. Thompson, J.R.; Polz, M.F. Dynamics of Vibrio populations and their role in environmental nutrient cycling. In *The Biology of Vibrios*; Wiley Online Library: Hoboken, NJ, USA, 2006; pp. 190–203.
4. Beardsley, C.; Pernthaler, J.; Wosniok, W.; Amann, R. Are readily culturable bacteria in coastal North Sea waters suppressed by selective grazing mortality? *Appl. Environ. Microbiol.* **2003**, *69*, 2624–2630. [CrossRef] [PubMed]
5. Ganesan, D.; Gupta, S.S.; Legros, D. Cholera surveillance and estimation of burden of cholera. *Vaccine* **2020**, *38*, A13–A17. [CrossRef]
6. Gilbert, J.A.; Steele, J.A.; Caporaso, J.G.; Steinbrück, L.; Reeder, J.; Temperton, B.; Huse, S.; McHardy, A.C.; Knight, R.; Joint, I.; et al. Defining seasonal marine microbial community dynamics. *ISME J.* **2012**, *6*, 298–308. [CrossRef]
7. Westrich, J.R.; Ebling, A.M.; Landing, W.M.; Joyner, J.L.; Kemp, K.M.; Griffin, D.W.; Lipp, E.K. Saharan dust nutrients promote Vibrio bloom formation in marine surface waters. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 5964–5969. [CrossRef]
8. Poloczanska, E.S.; Brown, C.J.; Sydeman, W.J.; Kiessling, W.; Schoeman, D.S.; Moore, P.J.; Brander, K.; Bruno, J.F.; Buckley, L.B.; Burrows, M.T.; et al. Global imprint of climate change on marine life. *Nat. Clim. Chang.* **2013**, *3*, 919–925. [CrossRef]
9. Dore, M.H. Climate change and changes in global precipitation patterns: What do we know? *Environ. Int.* **2005**, *31*, 1167–1181. [CrossRef]
10. Bijma, J.; Pörtner, H.O.; Yesson, C.; Rogers, A.D. Climate change and the oceans—What does the future hold? *Mar. Pollut. Bull.* **2013**, *74*, 495–505. [CrossRef]
11. Burge, C.A.; Hershberger, P.K. Climate change can drive marine diseases. In *Marine Disease Ecology*; Behringer, D.C., Silliman, B.R., Lafferty, K.D., Eds.; Oxford University Press: Oxford, UK , 2020; pp. 83–94.
12. Martinez-Urtaza, J.; Bowers, J.C.; Trinanes, J.; DePaola, A. Climate anomalies and the increasing risk of Vibrio parahaemolyticus and Vibrio vulnificus illnesses. *Food Res. Int.* **2010**, *43*, 1780–1790. [CrossRef]
13. da Cunha, D.T. Improving food safety practices in the foodservice industry. *Curr. Opin. Food Sci.* **2021**, *42*, 127–133. [CrossRef]
14. Levallois, P.; Villanueva, C.M. Drinking water quality and human health: An editorial. *Int. J. Environ. Res. Public Health* **2019**, *16*, 631. [CrossRef]
15. Li, H.; Xie, G.; Edmondson, A.S. Review of secondary mathematical models of predictive microbiology. *J. Food Prod. Mark.* **2008**, *14*, 57–74. [CrossRef]
16. Ferrer, J.; Prats, C.; López, D.; Vives-Rego, J. Mathematical modeling methodologies in predictive food microbiology: A SWOT analysis. *Int. J. Food Microbiol.* **2009**, *134*, 2–8. [CrossRef]
17. Peleg, M.; Corradini, M.G. Microbial growth curves: What the models tell us and what they cannot. *Crit. Rev. Food Sci. Nutr.* **2011**, *51*, 917–945. [CrossRef]
18. Perez-Rodriguez, F.; Valero, A. Predictive microbiology in foods. In *Predictive Microbiology in Foods*; Springer: Berlin/Heidelberg, Germany, 2013; pp. 1–10.

19. Oscar, T. Development and validation of primary, secondary, and tertiary models for growth of Salmonella Typhimurium on sterile chicken. *J. Food Prot.* **2005**, *68*, 2606–2613. [CrossRef]
20. Baranyi, J.; McClure, P.; Sutherland, J.; Roberts, T. Modeling bacterial growth responses. *J. Ind. Microbiol. Biotechnol.* **1993**, *12*, 190–194. [CrossRef]
21. Pirt, S.J. *Principles of Microbe and Cell Cultivation*; Blackwell Scientific Publications: Hoboken, NJ, USA , 1975.
22. Ratkowsky, D.; Lowry, R.; McMeekin, T.; Stokes, A.; Chandler, R. Model for bacterial culture growth rate throughout the entire biokinetic temperature range. *J. Bacteriol.* **1983**, *154*, 1222–1226. [CrossRef]
23. Davey, K. A predictive model for combined temperature and water activity on microbial growth during the growth phase. *J. Appl. Bacteriol.* **1989**, *67*, 483–488. [CrossRef]
24. Esser, D.S.; Leveau, J.H.; Meyer, K.M. Modeling microbial growth and dynamics. *Appl. Microbiol. Biotechnol.* **2015**, *99*, 8831–8846. [CrossRef]
25. Zwietering, M.; Jongenburger, I.; Rombouts, F.; Van't Riet, K. Modeling of the bacterial growth curve. *Appl. Environ. Microbiol.* **1990**, *56*, 1875–1881. [CrossRef] [PubMed]
26. Fu, S.; Shen, J.; Liu, Y.; Chen, K. A predictive model of V ibrio cholerae for combined temperature and organic nutrient in aquatic environments. *J. Appl. Microbiol.* **2013**, *114*, 574–585. [CrossRef]
27. Fu, S.; Liu, Y.; Li, X.; Tu, J.; Lan, R.; Tian, H. A preliminary stochastic model for managing microorganisms in a recirculating aquaculture system. *Ann. Microbiol.* **2015**, *65*, 1119–1129. [CrossRef]
28. Ma, F.; Liu, H.; Wang, J.; Zhang, Z.; Sun, X.; Pan, Y.; Zhao, Y. Behavior of Vibrio parahemolyticus cocktail including pathogenic and nonpathogenic strains on cooked shrimp. *Food Control* **2016**, *68*, 124–132. [CrossRef]
29. Zahaba, M.; Halmi, M.I.E.; Othman, A.R.; Shukor, M.Y. Mathematical modeling of the growth curve of Vibrio sp. isolate MZ grown in Seawater medium. *Bioremediation Sci. Technol. Res.* **2018**, *6*, 31–34. [CrossRef]
30. Baranyi, J.; Roberts, T.A. Mathematics of predictive food microbiology. *Int. J. Food Microbiol.* **1995**, *26*, 199–218. [CrossRef]
31. Boonyawantang, A.; Mahakarnchanakul, W.; Rachtanapun, C.; Boonsupthip, W. Behavior of pathogenic Vibrio parahaemolyticus in prawn in response to temperature in laboratory and factory. *Food Control* **2012**, *26*, 479–485. [CrossRef]
32. Chung, K.H.; Park, M.S.; Kim, H.Y.; Bahk, G.J. Growth prediction and time–temperature criteria model of Vibrio parahaemolyticus on traditional Korean raw crab marinated in soy sauce (ganjang-gejang) at different storage temperatures. *Food Control* **2019**, *98*, 187–193. [CrossRef]
33. Fernandez-Piquer, J.; Bowman, J.P.; Ross, T.; Tamplin, M.L. Predictive models for the effect of storage temperature on Vibrio parahaemolyticus viability and counts of total viable bacteria in Pacific oysters (Crassostrea gigas). *Appl. Environ. Microbiol.* **2011**, *77*, 8687–8695. [CrossRef] [PubMed]
34. Oh, H.; Yoon, Y.; Ha, J.; Lee, J.; Shin, I.S.; Kim, Y.M.; Park, K.S.; Kim, S. Risk assessment of vibriosis by Vibrio cholerae and Vibrio vulnificus in whip-arm octopus consumption in South Korea. *Fish. Aquat. Sci.* **2021**, *24*, 207–218. [CrossRef]
35. Parveen, S.; DaSilva, L.; DePaola, A.; Bowers, J.; White, C.; Munasinghe, K.A.; Brohawn, K.; Mudoh, M.; Tamplin, M. Development and validation of a predictive model for the growth of Vibrio parahaemolyticus in post-harvest shellstock oysters. *Int. J. Food Microbiol.* **2013**, *161*, 1–6. [CrossRef]
36. Posada-Izquierdo, G.D.; Valero, A.; Arroyo-López, F.N.; González-Serrano, M.; Ramos-Benítez, A.M.; Benítez-Cabello, A.; Rodríguez-Gómez, F.; Jimenez-Diaz, R.; García-Gimeno, R.M. Behavior of *Vibrio* spp. in table olives. *Front. Microbiol.* **2021**, *2021*, 1130. [CrossRef]
37. Zhou, K.; Gui, M.; Li, P.; Xing, S.; Cui, T.; Peng, Z. Effect of combined function of temperature and water activity on the growth of Vibrio harveyi. *Braz. J. Microbiol.* **2012**, *43*, 1365–1375. [CrossRef]
38. Miles, D.W.; Ross, T.; Olley, J.; McMeekin, T.A. Development and evaluation of a predictive model for the effect of temperature and water activity on the growth rate of Vibrio parahaemolyticus. *Int. J. Food Microbiol.* **1997**, *38*, 133–142. [CrossRef]
39. Nishina, T.; Wada, M.; Ozawa, H.; Hara-Kudo, Y.; Konuma, H.; Hasegawa, J.; Kumagi, S. Growth kinetics of Vibrio parahaemolyticus O3: K6 under varying conditions of pH, NaCl concentration and temperature. *Food Hyg. Saf. Sci.* **2004**, *45*, 35–37. [CrossRef]
40. Kim, Y.W.; Lee, S.H.; Hwang, I.G.; Yoon, K.S. Effect of temperature on growth of Vibrio paraphemolyticus and Vibrio vulnificus in flounder, salmon sashimi and oyster meat. *Int. J. Environ. Res. Public Health* **2012**, *9*, 4662–4675. [CrossRef]
41. Yang, Z.Q.; Jiao, X.A.; Li, P.; Pan, Z.M.; Huang, J.L.; Gu, R.X.; Fang, W.M.; Chao, G.X. Predictive model of Vibrio parahaemolyticus growth and survival on salmon meat as a function of temperature. *Food Microbiol.* **2009**, *26*, 606–614. [CrossRef]
42. Yoon, K.; Min, K.; Jung, Y.; Kwon, K.; Lee, J.; Oh, S. A model of the effect of temperature on the growth of pathogenic and nonpathogenic Vibrio parahaemolyticus isolated from oysters in Korea. *Food Microbiol.* **2008**, *25*, 635–641. [CrossRef]
43. van Boekel, M.A. On the use of the Weibull model to describe thermal inactivation of microbial vegetative cells. *Int. J. Food Microbiol.* **2002**, *74*, 139–159. [CrossRef]
44. Buchanan, R.; Whiting, R.; Damert, W. When is simple good enough: A comparison of the Gompertz, Baranyi, and three-phase linear models for fitting bacterial growth curves. *Food Microbiol.* **1997**, *14*, 313–326. [CrossRef]
45. Tang, X.; Zhao, Y.; Sun, X.; Xie, J.; Pan, Y.; Malakar, P.K. Predictive model of Vibrio parahaemolyticus O3: K6 growth on cooked Litopenaeus vannamei. *Ann. Microbiol.* **2015**, *65*, 487–493.
46. Huang, L. Growth kinetics of Listeria monocytogenes in broth and beef frankfurters—Determination of lag phase duration and exponential growth rate under isothermal conditions. *J. Food Sci.* **2008**, *73*, E235–E242. [CrossRef] [PubMed]

47. Chen, Y.R.; Hwang, C.A.; Huang, L.; Wu, V.C.; Hsiao, H.I. Kinetic analysis and dynamic prediction of growth of vibrio parahaemolyticus in raw white shrimp at refrigerated and abuse temperatures. *Food Control* **2019**, *100*, 204–211. [CrossRef]
48. Fang, T.; Gurtler, J.B.; Huang, L. Growth kinetics and model comparison of Cronobacter sakazakii in reconstituted powdered infant formula. *J. Food Sci.* **2012**, *77*, E247–E255. [CrossRef] [PubMed]
49. Froelich, B.; Bowen, J.; Gonzalez, R.; Snedeker, A.; Noble, R. Mechanistic and statistical models of total Vibrio abundance in the Neuse River Estuary. *Water Res.* **2013**, *47*, 5783–5793. [CrossRef]
50. Fujikawa, H.; Kai, A.; Morozumi, S. A new logistic model for bacterial growth. *Shokuhin Eiseigaku Zasshi. J. Food Hyg. Soc. Jpn.* **2003**, *44*, 155–160. [CrossRef] [PubMed]
51. Fujikawa, H. Application of the new logistic model to microbial growth prediction in food. *Biocontrol Sci.* **2011**, *16*, 47–54. [CrossRef]
52. Ratkowsky, D.A.; Olley, J.; McMeekin, T.; Ball, A. Relationship between temperature and growth rate of bacterial cultures. *J. Bacteriol.* **1982**, *149*, 1–5. [CrossRef]
53. Huang, L.; Hwang, C.A.; Phillips, J. Evaluating the Effect of Temperature on Microbial Growth Rate—The Ratkowsky and a Bělehrádek-Type Models. *J. Food Sci.* **2011**, *76*, M547–M557. [CrossRef]
54. Kapetanović, D.; Purgar, M.; Gavrilović, A., Hackenberger, B.K.; Kurtović, B.; Haberle, I.; Pečar Ilić, J.; Geček, S.; Hackenberger, D.K.; Djerdj, T.; et al. AqADAPT: Physiochemical parameters and *Vibrio* spp. abundance in water samples gathered in the Adriatic Sea. *PANGAEA* **2022**, *submitted*.
55. Jug-Dujaković, J.; Gavrilović, A.; Kolda, A.; Žunić, J.; Pikelj, K.; Listeš, E.; Vukić Lušić, D.; Vardić Smrzlić, L.; Perić, L.; El-Matbouli, M.; et al. AQUAHEALTH: Physiochemical parameters and *Vibrio* spp. abundance in water samples gathered in the Adriatic Sea. *PANGAEA* **2022**, *submitted*.
56. Bullington, J.A.; Golder, A.R.; Steward, G.F.; McManus, M.A.; Neuheimer, A.B.; Glazer, B.T.; Nigro, O.D.; Nelson, C.E. Refining real-time predictions of Vibrio vulnificus concentrations in a tropical urban estuary by incorporating dissolved organic matter dynamics. *Sci. Total Environ.* **2022**, *829*, 154075. [CrossRef] [PubMed]
57. Nigro, O.D.; James-Davis, L.I.; De Carlo, E.H.; Li, Y.H.; Steward, G.F. Variable freshwater influences on the abundance of Vibrio vulnificus in a tropical urban estuary. *Appl. Environ. Microbiol.* **2022**, *88*, e01884–21. [CrossRef] [PubMed]
58. Froelich, B.; Gonzalez, R.; Blackwood, D.; Lauer, K.; Noble, R. Decadal monitoring reveals an increase in *Vibrio* spp. concentrations in the Neuse River Estuary, North Carolina, USA. *PLoS ONE* **2019**, *14*, e0215254. [CrossRef]
59. Urquhart, E.A.; Jones, S.H.; Yu, J.W.; Schuster, B.M.; Marcinkiewicz, A.L.; Whistler, C.A.; Cooper, V.S. Environmental conditions associated with elevated Vibrio parahaemolyticus concentrations in Great Bay Estuary, New Hampshire. *PLoS ONE* **2016**, *11*, e0155018. [CrossRef]
60. Williams, T.C.; Froelich, B.A.; Phippen, B.; Fowler, P.; Noble, R.T.; Oliver, J.D. Different abundance and correlational patterns exist between total and presumed pathogenic Vibrio vulnificus and V. parahaemolyticus in shellfish and waters along the North Carolina coast. *FEMS Microbiol. Ecol.* **2017**, *93*, 71. [CrossRef] [PubMed]
61. Mair, P.; Wilcox, R. Robust statistical methods in R using the WRS2 package. *Behav. Res. Methods* **2020**, *52*, 464–488. [CrossRef]
62. Wilcox, R.R. *Introduction to Robust Estimation and Hypothesis Testing*, 3rd ed.; Elsevier: Amsterdam, The Netherlands, 2012.
63. RStudio Team. *RStudio: Integrated Development Environment for R*; RStudio, PBC: Boston, MA, USA, 2020.
64. Wickham, H.; Averick, M.; Bryan, J.; Chang, W.; McGowan, L.D.; François, R.; Grolemund, G.; Hayes, A.; Henry, L.; Hester, J.; et al. Welcome to the tidyverse. *J. Open Source Softw.* **2019**, *4*, 1686. [CrossRef]
65. Kuhn, M. Building Predictive Models in R Using the caret Package. *J. Stat. Softw. Artic.* **2008**, *28*, 1–26. [CrossRef]
66. Hamner, B.; Frasco, M.; LeDell, E. Evaluation Metrics for Machine Learning. 2018. Available online: https://cran.r-project.org/web/packages/Metrics/Metrics.pdf (accessed on 25 April 2022).
67. Grosjean, P. SciViews–Main Package, 2019. Available online: https://cran.r-project.org/web/packages/SciViews/SciViews.pdf (accessed on 25 April 2022).
68. Dowle, M. Package 'data.table': Extension of 'data.frame'. 2021. Available online: https://cran.r-project.org/web/packages/data.table/data.table.pdf (accessed on 25 April 2022).
69. Wickham, H. Read Excel Files. 2019. Available online: https://cran.microsoft.com/snapshot/2019-03-09/web/packages/readxl/readxl.pdf (accessed on 25 April 2022).
70. Dragulescu, A.; Arendt, C. Read, Write, Format Excel 2007 and Excel 97/2000/XP/2003 Files. 2020. Available online: https://cran.r-project.org/web/packages/xlsx/xlsx.pdf (accessed on 25 April 2022).
71. Purgar, M.; Kapetanović, D.; Geček, S.; Marn, N.; Haberle, I.; Hackenberger, K.B.; Gavrilović, A.; Pečar Ilić, J.; Hackenberger K., D.; Djerdj, T.; et al. Codes for Purgar et al. 2022: Investigating the ability of growth models to predict in situ *Vibrio* spp. abundances. *Zenodo* **2022**. [CrossRef]
72. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*; Springer: New York, NY, USA , 2016.
73. Martinez-Urtaza, J.; Lozano-Leon, A.; Varela-Pet, J.; Trinanes, J.; Pazos, Y.; Garcia-Martin, O. Environmental determinants of the occurrence and distribution of Vibrio parahaemolyticus in the rias of Galicia, Spain. *Appl. Environ. Microbiol.* **2008**, *74*, 265–274. [CrossRef]
74. Kirchman, D.L. Growth rates of microbes in the oceans. *Annu. Rev. Mar. Sci.* **2016**, *8*, 285–309. [CrossRef]
75. Norris, N.; Levine, N.M.; Fernandez, V.I.; Stocker, R. Mechanistic model of nutrient uptake explains dichotomy between marine oligotrophic and copiotrophic bacteria. *PLoS Comput. Biol.* **2021**, *17*, e1009023. [CrossRef]

76. Trinanes, J.; Martinez-Urtaza, J. Future scenarios of risk of Vibrio infections in a warming planet: A global mapping study. *Lancet Planet. Health* **2021**, *5*, e426–e435. [CrossRef]
77. Ouzzani, M.; Hammady, H.; Fedorowicz, Z.; Elmagarmid, A. Rayyan—A web and mobile app for systematic reviews. *Syst. Rev.* **2016**, *5*, 1–10. [CrossRef]
78. Gekas, V.; Gonzalez, C.; Sereno, A.; Chiralt, A.; Fito, P. Mass transfer properties of osmotic solutions. I. Water activity and osmotic pressure. *Int. J. Food Prop.* **1998**, *1*, 95–112. [CrossRef]

2.2. Publication II: Assessment of *Vibrio* spp. abundance as a water quality indicator: Insights from Mali Ston Bay in the Adriatic Sea

Purgar, M., Gavrilović, A., Kapetanović, D., Klanjšček, J., Jug-Dujaković, J., Kolda, A., Žunić, J., Kazazić, S., Vardić Smrzlić, I., Vukić Lušić, D., Pikelj, K., Listeš, E., El-Matbouli, M., Lillehaug, A., Lončarević, S., Knežević, D., Hengl, B., Geček, S., Klanjscek, T. (2023). Assessment of *Vibrio* spp. abundance as a water quality indicator: Insights from Mali Ston Bay in the Adriatic Sea. Estuarine, Coastal and Shelf Science, 295, 108558. https://doi.org/10.1016/j.ecss.2023.108558

Contributions: conceptualization, formal analysis, visualization, writing – original draft, and writing – review & editing.

Estuarine, Coastal and Shelf Science 295 (2023) 108558

Contents lists available at ScienceDirect

Estuarine, Coastal and Shelf Science

journal homepage: www.elsevier.com/locate/ecss

Assessment of *Vibrio* spp. abundance as a water quality indicator: Insights from Mali Ston Bay in the Adriatic Sea

Marija Purgar [a], Ana Gavrilović [b], Damir Kapetanović [a], Jasminka Klanjšček [a], Jurica Jug-Dujaković [c], Anamarija Kolda [a], Jakov Žunić [a], Snježana Kazazić [a], Irena Vardić Smrzlić [a], Darija Vukić Lušić [d], Kristina Pikelj [e], Eddy Listeš [f], Mansour El-Matbouli [g,h], Atle Lillehaug [i], Semir Lončarević [i], Dražen Knežević [j], Brigita Hengl [j], Sunčana Geček [a,*], Tin Klanjscek [a]

[a] *Ruđer Bošković Institute, Bijenička Cesta 54, Zagreb, 10000, Croatia*
[b] *University of Zagreb Faculty of Agriculture, Svetošimunska Cesta 25, Zagreb, 10000, Croatia*
[c] *Sustainable Aquaculture Systems, Inc., Frenchtown, 08825, New Jersey, USA*
[d] *University of Rijeka, Faculty of Medicine, Ul. Braće Branchetta 20/1, Rijeka, 51000, Croatia*
[e] *University of Zagreb, Faculty of Science, Horvatovac Ul. 102a, Zagreb, 10000, Croatia*
[f] *Croatian Veterinary Institute, Regional Veterinary Institute Split, Poljička Cesta 33, Split, 21000, Croatia*
[g] *University of Veterinary Medicine Vienna, Veterinärpl. 1, Wien, 1200, Austria*
[h] *School of Biotechnology, Badr University in Cairo, Cairo Governorate, Cairo, 11829, Egypt*
[i] *Norwegian Veterinary Institute, Arboretveien 57, Ås, 1433, Norway*
[j] *Croatian Agency for Agriculture and Food, Ul. Ivana Gundulića 36, Osijek, 31000, Croatia*

ARTICLE INFO

Keywords:
Aquaculture
Pathogens
Limit values
Regulations
Legislature

ABSTRACT

Due to high anthropogenic pressures, science-based coastal management required to ensure the sustainable use of coastal areas highly depends on environmental indicators used for decision-making. In this paper, we argue for the inclusion of *Vibrio* spp. abundance as a supplemental indicator of water quality for science-based coastal management by examining the environmental and bacterial indicators at a fish farm and a control site in Mali Ston Bay in the Adriatic Sea. Unexpectedly, heterotrophic bacteria, enterococci and *Vibrio* spp. were more abundant in the cold season, while *E. coli* and total coliforms, following a more traditional pattern, were more abundant in the warm season. Each of the currently used indicators has a specific purpose: heterotrophic bacteria indicate the presence of both nonpathogenic and pathogenic bacteria, while enterococci are pathogenic bacteria indicating fecal pollution. *Vibrio* spp. abundance additionally represents a non-fecal bacteria that can cause vibriosis in humans and aquatic organisms. Since vibriosis is the leading cause of disease-related fish mortality in aquaculture, pathogenic *Vibrio* spp. have large health and economic implications. These implications, as well as the added interpretative value when compared to other bacterial indicators, make *Vibrio* spp. abundance a good candidate as a water quality indicator. Significant dependence of the abundance on depth further differentiates *Vibrio* spp. from other indicators, thus bolstering the candidacy - especially in aquaculture areas. Before inclusion of any *Vibrio* spp. indicators into legislature, further research is needed particularly into (i) abundance thresholds characterizing water quality, and (ii) identification of species whose abundance should be monitored for best estimate of the disease risks.

1. Introduction

Coastal areas are home to more than 40% of human population and 60% of national economies; the related anthropogenic pressures endanger the continuous provision of the necessary ecosystem services (Maul and Duedall, 2019; Nobre, 2011; Martínez et al., 2007). Science-based integrated coastal management practices aim to ensure preservation of ecosystems and the related services, but the quality of

* Corresponding author.
E-mail address: suncana@irb.hr (S. Geček).

https://doi.org/10.1016/j.ecss.2023.108558
Received 28 March 2023; Received in revised form 30 October 2023; Accepted 3 November 2023
Available online 11 November 2023

management highly depends on indicators used for decision-making to determine the state of the environment and capture environmental trends (Elliot et al., 2017; Mazé et al., 2017; Atkins et al., 2015). Since water affects all coastal ecosystems, water quality indicators are of particular importance.

Typical water quality assessment includes indicators characterizing heterotrophic bacteria but is mostly focused on fecal bacteria (Cabral, 2010). Heterotrophic bacteria are a crucial element of the marine ecosystem nutrient cycling (Zhang et al., 2018; Moran et al., 2016; Benner, 2011; Fuhrman, 1992), however they can be pathogenic, causing severe diseases and economic losses (Feliatra et al., 2020; Bentzon-Tilia et al., 2016; Saxena et al., 2015). The most frequently used bacterial indicators for assessing water quality are heterotrophic plate counts (HPC), total coliforms (TC), and fecal bacterial indices (*E. coli* and enterococci) (Some et al., 2021; World Health Organization, 2017; Stewart et al., 2008).

The bacterial indicators account for specific anthropogenic pressures, and help estimate specific risks. HPC reflect the general load of different bacteria that need organic nutrients for growth in water bodies (Bartram et al., 2003). TC have been historically used as an indicator of human fecal influx, however TC are also commonly found in the environment and therefore are ineffective as a measure of fecal pathogens, and should primarily be used to gauge the efficacy of treatment and evaluate cleanliness and integrity of water distribution (World Health Organization, 2017). Specific fecal indicators, primarily *E. coli* and enterococci, are better indicators of fecal pollution than the TC (Some et al., 2021; Price and Wildeboer, 2017; Boehm and Sassoubre, 2014), and serve as contemporary indicators of water quality for recreational, industrial, agricultural, and water supply purposes. Other bacteria, such as *Vibrio* spp., have been largely ignored despite their potentially negative effect on human and marine organisms.

Vibrio spp., in particular, are ubiquitous heterotrophic bacteria present in marine environments (Ryder et al., 2014). Pathogenic *Vibrio* spp. (e.g. *V. parahaemolyticus*, *Listonella/V. alginolyticus*, *V. vulnificus*, *V. anguillarum*, *V. harveyi*, etc.) cause vibriosis, a potentially fatal disease in both humans and aquatic animals (Baker-Austin et al., 2018). Humans get infected through contact with seawater or eating raw or undercooked contaminated seafood, a problem especially prominent during heightened seafood consumption (Ryder et al., 2014). Additionally, *V. vulnificus* as an opportunistic pathogen, in humans causes wound infections that can develop into septicaemia. Marine organisms usually get vibriosis if their immune defenses are lowered due to stress (e.g. high temperature) or an immune deficiency (Manchanayake et al., 2023). The vibriosis mostly affects fish and shellfish (Sanches-Fernandes et al., 2022) and is caused by pathogens *V. parahaemolyticus*, *V. alginolyticus*, *V. harveyi*, *V. owensii* and *V. campbelli* that exploit the skin, gills, and gastrointestinal tract as portals for infection (Ina-Salwany et al., 2019). Despite the substantial body of research on *Vibrio* spp. (Froelich et al., 2013, 2019; Zadorozhnaya et al., 2015; Roux et al., 2015; Davis et al., 2019; Purgar et al., 2022) and the suggestion to use their abundance as a supplementary indicator of water quality as early as 1984 (Robertson, 1984) uniform guidelines, recommendations, and official regulations for monitoring *Vibrio* spp. are still lacking.

Vibrio spp. thrive in aquaculture environments due to favourable conditions (e.g. organic enrichment) (Toranzo and Barja, 1990; Chandrakala and Priya, 2017). There, vibriosis regularly causes significant mortality of fish and shellfish, resulting in environmental damage and significant economic losses (Sampaio et al., 2022). Furthermore, *Vibrio* spp. are persistent, with sediments serving as an important reservoir that can facilitate re-infection of the water column (Kapetanović et al., 2022). Presently, specific *Vibrio* spp. are only regulated for food products (e.g. in the U.S. (Center for Food Safety and Applied Nutrition, 2011), Australia (Food Standards Australia New Zealand, 2022)). Compared to the U.S. and Australia, the European Regulation (EC) No. 2073/2005 provides the microbiological standards for food products manufactured and exchanged within Europe; however, it does not include particular microbiological standards for *Vibrio* spp. (Hartnell et al., 2019). Given the risks to human health and the potential for environmental and economic damage, especially in aquaculture, suitability for water quality assessment of indicators characterizing *Vibrio* spp. should be investigated further.

Here we explore the suitability of *Vibrio* spp. abundance as a potential supplemental indicator of water quality for science-based coastal management. We investigated a broad set of indicators collected over three years in Mali Ston Bay - an intensive nearshore aquaculture site to ascertain whether *Vibrio* spp. abundance provides valuable information not captured by the other, more broadly used, indices. To understand general environmental and bacterial dynamics in the area, we investigated the seasonal, spatial and vertical (by depth) variability of environmental and bacterial indicators. In particular, we estimated: (i) reach of emissions from the fish farm, (ii) environmental factors affecting *Vibrio* spp. abundance, and (iii) variability of *Vibrio* spp. abundance with respect to other potential pathogens (*E. coli* and enterococci), and heterotrophic bacteria in general. Then, we show that *Vibrio* spp. abundance has patterns that differ from those of other bacteria in the same environment, and therefore provides additional information on the status of the environment. The information is especially relevant to aquaculture not only because of the potential environmental and economic damage due to vibriosis outbreaks, but also due to its relevance for vibrio-related foodborne diseases (Ryder et al., 2014).

2. Materials and methods

In this study, we conducted an analysis of the open dataset 'AQUAHEALTH' (PANGAEA repository, Jug-Dujaković et al., 2022) that includes measurements of two groups of indicators (environmental and bacterial) near a fish farm and a control site in the Mali Ston Bay (Adriatic Sea). The environmental indicators used in the analysis encompass temperature (Temp), salinity, total dissolved solids (TDS), pH, oxygen saturation (O2 (%)), total nitrogen (N), total phosphorus (P), particulate organic matter (POM), and particulate inorganic matter (PIM). The bacterial indicators include measurements of heterotrophic bacteria (HPC) and the potential indicator *Vibrio* spp. abundance. In addition to 'AQUAHEALTH' dataset, we analyzed measurements of total coliforms, *E. coli*, and enterococci, and all of the bacterial abundance in the sediment, that are available in the Zenodo repository as a part of this study's open analytical code (Purgar et al., 2023). The data was subjected to descriptive and inferential statistical analysis to reveal seasonal and spatial patterns within the dataset, as well as to explore potential interdependencies among the various indicators.

2.1. Sampling and in situ measurements

Sampling sites (Fig. 1) were the floating cage fish farm located near the island Maslinovac (Fish Farm) and a control site located near the island Pučenjak (Control). The observed area is under the freshwater influence of the Neretva river, submarine springs, coastal residential areas, and tourist and intensive fish and bivalve farming activities. The sampling period covers seasonal (warm and cold) measurements between 2016 and 2019.

Water sampling was conducted using a Niskin water sampler at four different depths (0.5 m below the surface, 5 m deep, 10 m deep, and 0.5 m above the bottom which is approximately 18 m). Samples were poured into sterile 0.5 L bottles. Sediment samples were collected using an Ekman grab (10 g of top sediment layer). In situ measurements of water temperature and oxygen saturation were taken with SevenGo pro/SG9 OptiOX (Mettler Toledo). Salinity and total dissolved solids (TDS) were determined with SevenGo pro/Conductivity (Mettler Toledo) and pH was measured with SevenGo pro/Ion (Mettler-Toledo).

Fig. 1. Sampling sites: Fish Farm near island Maslinovac (42°55.0719 N, 17°29.5401 E) and Control near island Pučenjak (42°55.7394 N, 17°29.7244 E) in the Southern part of the Adriatic Sea.

2.2. Seawater sample analyses

Total nitrogen was determined by the Oxidative digestion method (ISO 11905-1:1997) with peroxodisulfate using the Hach Lange, UV/VIS spectrophotometer DR/6000. The same spectrophotometer and the Hach method LCK348 for Water Analysis were used to measure the concentration of total phosphorus (P) (Hach, 2022).

Particulate matter analyses were performed in triplicates, where 1-L aliquot samples were filtered onto pre-ashed (3 h at 450 °C) 47-mm GF/C filters (Whatman). Total particulate matter (TPM), particulate organic matter (POM) and particulate inorganic matter (PIM) were determined according to the method described in Paterson et al. (2003).

2.3. Microbiological analysis

Vibrio spp. abundance from water samples was determined by counting the total number of visible colonies that exhibited relief from the plate surface from the Thiosulphate Citrate Bile Salt Sucrose (TCBS) (Difco™, BD) agar plates. TCBS agar plates were incubated at 22 °C for 3–5 days. Results were expressed as the mean number of colony forming units (CFU) in 1 mL of seawater or sediment. Similarly, heterotrophic marine bacteria (HPC) were enumerated by using the plate method on DifcoTM Marine Agar 2216 BD (BD, Sparks, MD, USA), and the plates were incubated at 22 °C for 3–5 days. TC and fecal indicators (*E. coli* and enterococci) were determined by defined substrate technology using Colilert-18 (IDEXX, Westbrook, ME, USA) for the total coliform bacteria and *E. coli*, and Enterolert-E (IDEXX) for enterococci. TC and fecal indicators (*E. coli* and enterococci) were enumerated using Quanti-tray2000 (IDEXX) which results in the most probable numbers (MPN/100 mL).

2.4. Data preparation

Data (N = 88 for water column, and N = 22 for sediment) preparation was conducted using RStudio Integrated Development Environment (RStudio Team, 2022), Version 4.2.2. Initially, 'pastecs' and 'psych' packages (Grosjean et al., 2018; Revelle, 2017) were used for descriptive statistics (SI Table 1 and SI Table 2). 'Base' packages (RStudio Team, 2022) were additionally used to assess normality (Shapiro-Wilks, QQplot), while the 'ggplot2 ' package was used for graphical display (Wickham, 2016). For a detailed summary please refer to SI Table 1B. Bacterial indicators *Vibrio*, HPC, and total coliforms were log-transformed using the 'decostand' function from the vegan package (Oksanen et al., 2013). Furthermore, extreme outliers were analyzed using the 'identify outliers' function from the 'rstatix' package (Kassambara, 2023). For a detailed summary of extreme outliers please refer to SI Table 1C. Ultimately, we replaced the detected extreme outliers for TDS, N, P and PIM and other missing values (Salinity, TDS, POM and PIM) with variable median values. Graphical display of variables used for further data analysis is displayed in Fig. 2.

2.5. Data analysis

Data analysis and graphical representation of results were performed in RStudio Integrated Development Environment (RStudio Team, 2022), Version 4.2.2. Results were visualised using the following packages: 'ggplot2' (Wickham, 2016) and 'corrplot' (Wei et al., 2017).

2.5.1. Environmental and bacterial water column indicators

To investigate the differences between (i) seasons (warm and cold), (ii) sites (fish farm and control), and (iii) depths (surface, 5m, 10m and bottom), in both groups of indicators (environmental and bacterial), we conducted a permutational multivariate analysis of variance (PERMANOVA) (Anderson, 2001). The original three-way multivariate PERMANOVA analysis (factors: season, site, depth) was replaced with a two-way multivariate PERMANOVA (factors: season, depth) after confirming that site and its interactions did not yield statistically significant change of indicator values in either the environmental indicator group or the bacterial indicator group. The multivariate two-way PERMANOVA analysis was followed by separate univariate two-way PERMANOVAs for each indicator variable to additionally assess the individual response of the indicator to changes in season and water layer depth. The R implementation of PERMANOVA analysis in the 'vegan' package (function 'adonis2') was applied to the Euclidean distance similarity matrices of environmental and bacterial indicators, respectively (Oksanen et al., 2013). The post-hoc test, i.e., pairwise multilevel comparison, was performed using the 'pariwise.adonis2' function from the 'pairwiseAdonis' package (Arbizu, 2020).

In addition, to determine the underlying relationships between environmental/bacterial indicators and changes in season, site, and water layer depth, a redundancy analysis (RDA) was performed (Legendre and Legendre, 2012). The RDA analysis was performed using the 'rda' function from the 'vegan' package (Oksanen et al., 2013). Both PERMANOVA and RDA were applied to data standardized using the 'decostand' function from the 'vegan' package.

Variables *E. coli* and enterococci were excluded from this analysis as their low abundance was expressed as categorical and not continuous variable. For categorical analysis please refer to: subchapter 2.5.4.

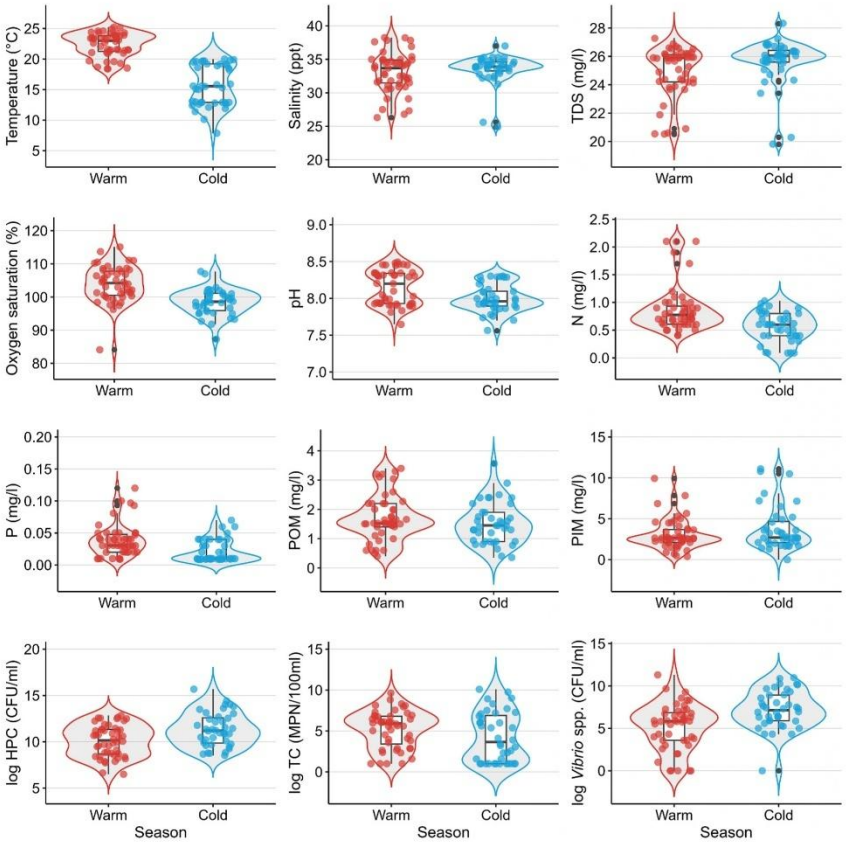

Fig. 2. Graphical display of AQUAHEALTH environmental and bacterial indicators used for analysis. The violin plots represent the relationship between each observed variable and the season. The box plot elements show the median value, the interquartile ranges, and potential outliers for each variable. On each side of the boxplot is a kernel density estimation to show the distribution shape of the data. HPC, TC, and *Vibrio* spp. were log transformed prior to analysis. Detailed descriptive statistics is reported in SI Table 2.

2.5.2. Interdependencies between indicators

Relationships between environmental indicators and the abundance of bacterial indicators were determined using Pearson's correlation coefficient (Benesty et al., 2009) and RDA analysis. Pearson's correlations were calculated using the 'cor' function from the 'stats' package (RStudio Team, 2022). Additionally, significance test and confidence intervals were quantified using the 'cor.mtest' function from the 'corrplot' package (Wei et al., 2017).

To assess the variation in the abundance of bacterial indicators affected by environmental indicators we used redundancy analysis (RDA) (Legendre and Legendre, 2012). RDA analysis was performed using the before mentioned 'rda' function from the 'vegan' package. The best exploratory variables were selected based on the results acquired by the function 'forward.sel' from the package 'adespatial' (Dray et al., 2022). This function performs a forward selection of variables by permutation of residuals under a reduced model.

2.5.3. Seasonal and spatial variation of environmental and bacterial indicators in sediment

In the sediment, differences in bacterial abundance between sites (fish farm and control) and seasons (warm and cold) were also examined by two-way PERMANOVA using the Euclidean distance similarity matrix (implementation: function 'adonis2' from the 'vegan' package in R, (Oksanen et al., 2013). RDA analysis was performed using the function 'rda', also from the 'vegan' package in R, to: (i) estimate the variation in bacterial abundance by site and season, and (ii) establish the relationship between bacteria in the sediment and in the lower layer of the water column.

2.5.4. Distribution of pathogenic bacteria in the water column

To assess the distribution of pathogenic bacteria (*E. coli*, enterococci and *Vibrio* spp.) in the water column we performed a comparison of the abundance categories (Table 1) based on the thresholds after each sampling campaign in Official Gazette of the Republic of Croatia 73/08 for monitoring and classification of bathing water quality (OGRC 73/08, 2008). We used the Croatian national regulation from 2008 which has more rigid limit values for microbiological indicators than those defined in the Bathing Water Directive 2006/07/EC (European Commission, 2006). Official standards for *Vibrio* spp. in assessing bathing water quality or near aquaculture sites do not yet exists, so we used standards

Table 1
Thresholds used for the assessment of coastal bathing water quality for *E. coli* and intestinal enterococci. Official standards for *Vibrio* spp. do not yet exist, so we used standards for enterococci (please see text for argumentation).

Indicator	Coastal bathing water quality			
	Excellent	Good	Sufficient	Poor
E. coli (MPN/100 ml)	<100	101–200	201–300	>300
intestinal enterococci (MPN/100 ml)	<60	61–100	101–200	>200
Vibrio spp. (CFU/ml)	Same as eneterococci as official standards for *Vibrio* spp. do not exist			

for enterococci. Abundance categories of individual pathogenic bacteria (*E. coli*, enterococci and *Vibrio* spp.) in the water column by depth, site and season were analyzed using Fisher's test (Field et al., 2012).

3. Results

Results presented here focus on variability of environmental and bacterial indicators in water column between seasons and depths, where the multivariate analysis identified a number of associations. The analysis on differences between sites, however, did not indicate significant associations. Hence, we assumed a horizontally uniform system, and relegated the analysis on differences between sites to the SI (SI Tables 3A and 4A).

Analysis of bacterial indicators in sediment neither found differences in the abundance of bacteria between sites or seasons, nor could link variability of bacterial indicators to season, depth, or site. Therefore, we considered the sediment as independent of site, depth and season, and relegated details into the SI (SI Tables 5–7, SI Figs. 1 and 2).

Therefore, here we only present results on (i) environmental indicators as functions of season and depth, (ii) bacterial indicators as functions of season and depth, and (iii) analysis of interdependencies between indicators. Finally, we present the results of categorical analysis describing distribution of pathogenic bacteria in the water column.

3.1. Environmental water column indicators

Season had a significant influence on the group of environmental indicators (Table 2A), as did the depth of the water layer in which the measurements were made (Table 2A). In particular, the surface layer showed different tendencies in environmental indicator values compared to other layers (SI Table 2C). However, the interaction between season and water layers did not significantly affect the environmental indicators (Table 2A).

Table 2
Two-way PERMANOVA analysis (factors: Season, Water layer depth) of environmental and bacterial indicators in the water column: (A/C) multivariate, (B/D) univariate. Statistically significant values are in bold. For a detailed statistical analysis, see SI Table 3B, SI Table 4B, and SI Table 8A-K.

	Season		Depth		Season:Depth	
	F(1, 87)	p	F(3, 87)	p	F(3, 87)	p
A Environmental indicators Multivariate analysis						
	16.186	**0.001**	3.054	**0.001**	0.738	0.823
B Environmental indicators Univariate analysis						
Temp	146.396	**0.001**	1.337	0.296	2.841	0.060
Salinity	0.799	0.376	11.618	**0.001**	0.667	0.601
TDS	5.024	**0.024**	9.222	**0.001**	0.086	0.957
pH	10.279	**0.003**	0.036	0.986	0.022	0.998
O_2 (%)	28.173	**0.001**	0.582	0.611	1.330	0.257
N	18.179	**0.001**	2.918	**0.045**	0.560	0.641
P	11.600	**0.002**	0.200	0.910	1.007	0.390
POM	1.467	0.217	1.441	0.218	0.935	0.431
PIM	2.096	0.151	1.660	0.168	0.459	0.707
C Bacterial indicators Multivariate analysis						
	9.00	0.001	1.17	0.333	0.88	0.565
D Bacterial indicators Univariate analysis						
HPC	10.593	**0.006**	0.517	0.679	1.081	0.384
Total coliforms	2.711	0.099	0.254	0.854	0.756	0.525
Vibrio spp.	15.032	**0.001**	3.036	**0.042**	0.7935	0.514

Additional univariate inferential assessment of each environmental indicator revealed a significant response of temperature, oxygen saturation, TDS, pH, N, and P to seasonal changes (Table 2B, SI Table 8). Higher TDS values were measured in the cold season (Fig. 2, SI Table 2, SI Table 8C), while higher values for temperature, oxygen saturation, pH, N, and P were measured in the warm season (Fig. 2, SI Table 2, SI Table 8ADEFG). These trends were also confirmed by the results of the RDA analysis (Fig. 3A), which indicated that the variables: temperature, oxygen saturation, pH, N, and P have similar directions of change, mainly related to the dominant RDA1 axis and seasonal warm-cold dynamics (SI Table 9FG).

Temperature distribution in the water column was consistent with summer stratification and winter mixing, with the largest atmospheric influence expressed on surface values (SI Table 5A). However, the significance of the interaction of depth and season for temperature values was only marginally confirmed ($p = 0.06$, SI Table 5A). Expectedly, in line with temperature, levels of inorganic nutrient salts (N, P) increased during the warm season.

Salinity and TDS differed between depths, with the surface layer statistically different from the deeper water layers (SI Table 8BC). These vertical trends are in agreement with the results of the RDA analysis (Fig. 3A), which indicated that salinity and TDS have similar directions of change mainly related to the RDA2 axis and the depth of the water layer (SI Table 9FG).

3.2. Bacterial water column indicators

Multivariate analysis of bacterial indicators revealed a significant influence of seasonal change on bacterial abundance (Table 2C). Increased abundance was more pronounced in the cold season for both HPC and *Vibrio* spp. (SI Table 8JL and Table 2D), consistent with the observed seasonal dynamics of oxygen saturation. RDA analysis indicated that *Vibrio* spp. and HPC have similar directions of seasonal change (mainly related to the RDA1 axis; Fig. 3B, SI Table 10), i.e. HPC and *Vibrio* spp. exhibit higher abundances in cold season, while Coliforms exhibit higher abundances in the warm season.

Depth was not a significant factor determining the vertical distribution of bacterial abundances, except in the case of *Vibrio* spp. where a marginally significant difference was observed between surface and other depths (SI Tables 8L and 2D). However, RDA clearly separated the surface layer from the others along the RDA2-axis suggesting that *Vibrio* spp. and Coliforms prefer deeper layers in respect to HPC (Fig. 3B, SI Table 10).

3.3. Interdependencies between indicators

Consistently with previous analysis, oxygen saturation, pH, N, and P show a significant positive correlation with temperature (Fig. 4A, SI Table 11A). An additional (positive) correlation is found for (i) inorganic nutrient salts (N and P), indicating coherent dynamics of their sources and ecosystem pathways, and (ii) salinity and TDS (Fig. 4A, SI Table 11A).

All bacterial abundances correlate with temperature (Fig. 4A, SI Table 11C). *Vibrio* spp. and HPC correlate negatively with temperature, consistent with our previous analyses that indicated a significant increase in their abundance during the cold season. Conversely, coliforms correlate positively with temperature. Significant negative correlation is obtained for: (i) *Vibrio* spp. and pH, N, P; and (ii) HPC and oxygen saturation, pH (Fig. 4A, SI Table 11C). Among bacterial indicators, only HPC and *Vibrio* spp. show a significant (positive) correlation (Fig. 4A, SI Table 11B). Innate positive correlations of oxygen saturation, pH, N, and P with temperature complicates interpretations of the observed

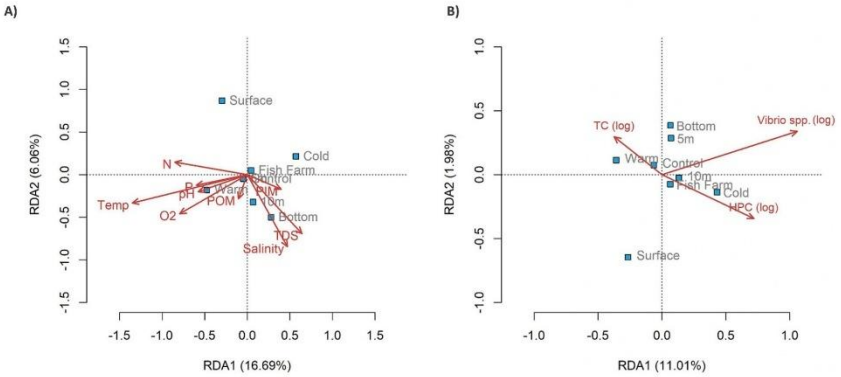

Fig. 3. Redundancy analysis (RDA) showing the dependence of environmental (Panel A), and bacterial (Panel B) indicators on explanatory variables: (i) site (Fish Farm, Control); (ii) water layer depth (Surface, 5m, 10m, bottom); and (iii) season (warm, cold). For details on RDA analysis, see SI Table 9 and SI Table 10.

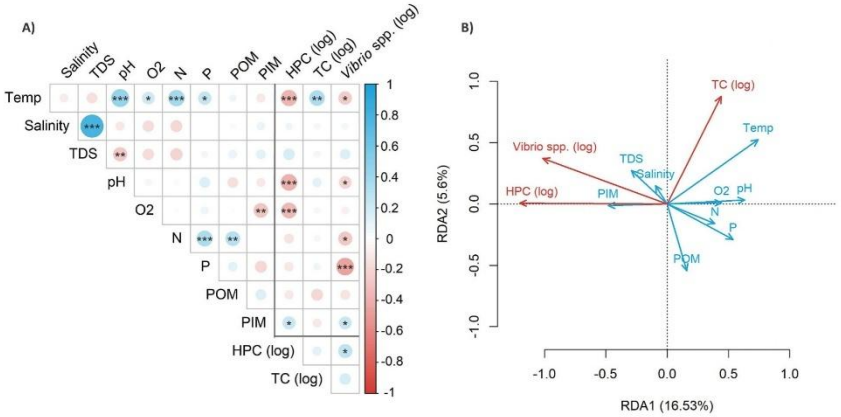

Fig. 4. Interdependencies between indicators. **Panel A**: Pearson's correlation matrix for environmental and bacterial variables (details in SI Table 8). Positive correlations are displayed in blue and negative correlations in red. The intensity of the color and the size of the circle are proportional to the correlation coefficient. **Panel B**: Redundancy analysis (RDA) showing the relationship between bacterial abundances and environmental explanatory variables (temperature, salinity, oxygen saturation, TDS, pH, N, P, POM and PIM). For details on RDA analysis, see SI Table 12. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

correlations. RDA analysis uses joint regression and ordination to account for common covariance on temperature, and explore explanatory power of environmental variables (Fig. 4B, SI Table 12). The environmental variables explain 26.12% (R_{adj} = 17.6%) of the variance in bacterial abundances (Fig. 4B; SI Table 12). When considering the correlation between the explanatory variables, temperature is the best predictor of bacterial abundances, followed by P (SI Table 13).

3.4. Distribution of pathogenic bacteria in the water column

Water quality, as defined by thresholds in Table 1, was generally excellent. *E. coli* (Fig. 5) had the best score: more than 95% of the samples collected at each station had excellent quality. Enterococci had excellent quality for more than 91% of the samples, while less than 64% of the samples tested for *Vibrio* spp. fell into this category. The significantly lower result for *Vibrio* spp. suggests that either (i) prevalence of potentially pathogenic *Vibrio* spp. is of real concern, or (ii) the thresholds originally adopted from the enterococci regulations do not correctly reflect the risk of disease.

Significant difference in water quality between seasons was detected for enterococci and *Vibrio* spp. (Fisher's tests, $p < 0.05$, SI Table 14, Fig. 5). Water samples analyzed for the presence of enterococci and *Vibrio* spp. were of lower quality during the cold season, which in the case of *Vibrio* spp. is consistent with PERMANOVA and RDA analyses (Subsection 3.2), which indicated greater abundance of *Vibrio* spp. during the cold season (Fig. 3, B; SI Table 8. L, SI Table 10 F, G). Fisher's test ($p > 0.05$, SI Table 14) did not confirm significantly worse water quality with respect to *E. coli* during the warm season, although consistently with the previous RDA analysis result (Subsection 3.2) there was a slight decrease in water quality during the warm season (Fig. 5). The decrease suggests an association between higher abundance of total coliform bacteria, and the warm season (Fig. 3, SI Table 10. F, G).

Water quality expressed by the presence of enterococci and *E. coli* did not vary significantly throughout the water column (Fisher's tests p >

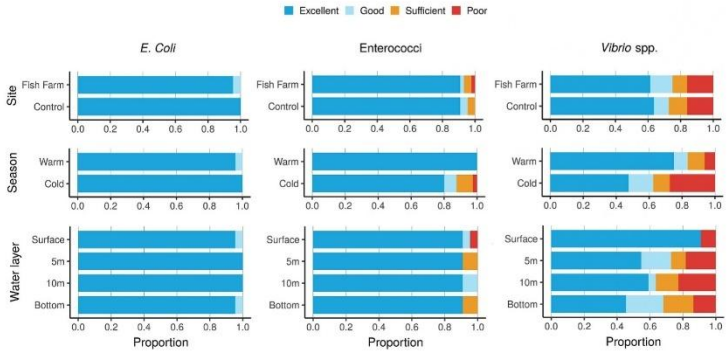

Fig. 5. Distribution patterns of water quality indicators for *E. coli*, enterococci, and *Vibrio* spp. by site, season and water layer. The water quality categories for *Vibrio* spp. are based on the thresholds originally adopted from the enterococci as the ones for potentially pathogenic *Vibrio* spp. do not yet exist.

0.05, SI Table 14, Fig. 5). Conversely, the quality expressed by the presence of *Vibrio* spp. had significantly better quality at the surface than at other depths (Fisher's test, $p < 0.05$, SI Table 14, Fig. 5). This result is consistent with the results of previous RDA analyses (Subsection 3.2), which indicated a greater preference of *Vibrio* spp. for deeper layers (Fig. 3, SI Table 8. L; SI Table 10. F, G).

4. Discussion

An intensive analysis of a broad set of indicators in Mali Ston Bay, as an example of a nearshore aquaculture site, generated four key findings. First, the environmental conditions, organic enrichment and bacterial abundances did not differ between the fish farm and the remote (control) site, thus indicating a uniform environment. Hence, either (i) the fish farm has no effect on the measured parameters even at the farming site, or (ii) the influence of the farm affects the whole area, or (iii) there is a strong alternative anthropogenic influence overriding effects of fish farming.

Second, bacterial abundances across seasons followed an unexpected pattern: HPC, enterococci, and *Vibrio* spp. thrived during the cold season despite optimal temperatures for *Vibrio* spp. growth reached during the warm season. Surprisingly, greater *Vibrio* spp. abundances were not related to temperature increase as expected (Sheikh et al., 2022; Froelich et al., 2019; Takemura et al., 2014; Pruzzo et al., 2005), and actually observed for the coliforms. The suppression of *Vibrio* spp. growth on both sites due to antibiotic use on the farm is unlikely to have caused the anomalous pattern as the use of antibiotics at the farm site is fairly limited. POM did not vary across seasons, and can therefore not be responsible for the anomalous patterns. TDS, however, was higher during the cold season, suggesting the organic component of TDS could be at least partly responsible for the anomalous patterns if used as a potential food source for the bacteria (Thickman and Gobler, 2017; Johnson et al., 2012).

Third, bacterial abundances observed in sediment support the thesis suggested by previous research that sediment is a reservoir for various bacterial species, especially *Vibrio* spp. (Kapetanović et al., 2022; Chase et al., 2015; Perkins et al., 2014; Vezzulli et al., 2009). A recent paper by Yang et al. (2022) on *V. parahaemolyticus* disease dynamics demonstrated how persistent reservoirs, such as sediment, potentially generate most of the outbreaks. Reinfection from sediment could be responsible for the observed uniformity of *Vibrio* spp. presence in our results.

Fourth, measurements of *Vibrio* spp. abundances provide additional information on water quality across seasons and depths:

- Traditionally used indicators cannot be used to estimate risks of vibriosis because *Vibrio* spp. abundance follows a different seasonal abundance pattern (Fig. 3, and Fig. 5),
- *Vibrio* spp. abundance frequently indicate poor water quality even when typically used microbial indicators indicate excellent and/or good water quality (Fig. 5).

Therefore, incorporating *Vibrio* species in water quality monitoring could broaden the perspective of the state of the environment and organisms in - and dependent on - coastal waters. Inclusion of *Vibrio* spp. into legislation has already been suggested in the 1980s (Robertson, 1984), but effective monitoring requires additional research e.g. on the appropriate thresholds for water quality categories.

Threshold values for *Vibrio* spp. abundance used in this paper are conservative. We set the threshold for 'sufficient' water quality to 100 CFU/ml, the value reported to result in bacterial transmission into organisms following prolonged exposure to *Vibrio* spp. (Kim and Lee, 2017). Therefore, even the 'sufficient' water quality category has known adverse effects on organisms which should arguably only appear in 'poor' category. Other categories were based on thresholds for enterococci (Table 1), who share the conservative threshold for 'sufficient' water quality. Hence, a scale tailored to *Vibrio* spp. would probably be even more stringent, with 100 CFU/ml probably already belonging to the 'poor' category further bolstering the case for *Vibrio* spp. monitoring.

4.1. Shortcomings in current Vibrio spp. monitoring practices

Monitoring *Vibrio* spp. abundance prior to potential hazards could help prevent vibriosis outbreaks, both in aquaculture and humans. This was recognized by the European Centre for Disease Prevention and Control (Levy, 2018). They have developed a *Vibrio* suitability tool, for informing ECDC and the public, which uses daily updated remote sensing data such as sea surface temperature and salinity to show the environmental suitability for *Vibrio* growth in the Baltic Sea during summer (https://geoportal.ecdc.europa.eu/vibriomapviewer/). While tracking limited to summer months may be suitable for the Baltic Sea, vibriosis outbreaks in the Adriatic (and, presumably other Mediterranean seas) have been observed during the spring (Veić, 2016; Zupičić et al., 2022). These, and potentially other areas, would then require broader monitoring than already exists in the Baltic region especially since existing fecal bacteria (fecal coliforms and/or *E. coli*), as seen in our results, cannot be used as indicators of risks from *Vibrio* spp. (Ryder et al., 2014).

As established disease agents (Baker-Austin et al., 2018; Brumfield et al., 2021), *Vibrio* spp. are primarily monitored in some countries e.g.

the U.S. (Center for Food Safety and Applied Nutrition, 2011) and Australia (Food Standards Australia New Zealand, 2022) to ensure food safety; and also identified after diagnosis of infections (Baker-Austin et al., 2018; Brumfield et al., 2021). The U.S. Center for Disease Control and Prevention runs a Cholera and Other *Vibrio* Illness Surveillance (COVIS) system that collects data on pathogenic *Vibrio* species: infection type and incidence, and geographic location of cases over time (Levy, 2018). In Europe, however, there's a significant lack of data concerning *Vibrio* spp. presence in the environment and its impact on human health. The European Union doesn't mandate *Vibrio* infection reporting, and laboratories test only for *Vibrio* infections in patients with post-travel diarrhea, primarily to exclude *Vibrio cholerae* (Semenza et al., 2017). Adequate *Vibrio* spp. data collecting and reporting would help enhance public health strategies, ensuring both food safety and timely medical interventions in affected regions.

4.2. Potential benefits of monitoring Vibrio spp. in marine aquaculture

Monitoring of *Vibrio* spp. would, even if not legislated, help investigate important issues in aquaculture. For example, the unusual *Vibrio* spp. abundance patterns could be caused by re-suspension from the sediment, or could be caused by an unobserved factor such as the influence of organic matter or nutrient influx. Regular long-term monitoring would help disentangle the two factors, and inform aquaculture facilities on risks of potential outbreaks in time to (i) prevent economic losses and environmental damage resulting from vibriosis outbreaks, and (ii) ensure seafood safety by preventing foodborne diseases. These issues are especially relevant for the Mediterranean region due to high levels of aquaculture production, sea-related tourism, and high levels of local seafood consumption.

4.3. Limitations to our approach

Limitations in our current approach are primarily related to data scarcity (Gorgoglione et al., 2020). Our analysis relied on the available dataset (Jug Dujaković et al., 2022; Purgar et al., 2023) collected at two stations in Mali Ston Bay, a coastal area of significant economic and recreational value. Hence, our results may be applicable to a limited region or types of environments only even though some studies suggest a wider applicability (Robertson, 1984; Levy, 2018).

While establishing appropriate water quality category thresholds is crucial for addressing *Vibrio* spp. abundance as a water quality indicator, several notable knowledge gaps exist. We cannot definitely assert that abundance is the ideal measure for monitoring *Vibrio* spp. To better address economic and health concerns, it might be more appropriate to focus on monitoring the pathogenic species only. Whether this is a viable (or even necessary) strategy depends on relative relationships between genus and (pathogenic) species-level trends (e.g. *V. parahaemolyticus*, *V. vulnificus*) an argument already made by Brumfield et al. (2021).

5. Conclusion

In conclusion, indices related to *Vibrio* spp. are a good candidate for inclusion in water quality monitoring strategies, especially in aquaculture, because they provide additional information on water quality and health risks for both humans and marine organisms. More research, however, is needed before the appropriate indices can be defined and included in legislature. The focus should be on identifying (i) thresholds for water quality categories, and (ii) appropriate species that capture disease risks for both humans and animals. Furthermore, the observed anomalous pattern of higher bacterial abundance during the cold season emphasizes the necessity for a holistic approach to *Vibrio* spp. research, and ecosystem studies in general.

Funding

This research was funded by the Croatian Science Foundation (HRZZ) project AQUAHEALTH no. IP-2014-09-3494 to D.K. and DOK-202102-6688 to T.K. for M.P.

CRediT authorship contribution

Conceptualization, S.G., J.K., D.K., M.P., and T.K.; methodology, D. K., A.G., S.G., and S.K.; writing—original draft preparation, M.P., S.G., J. K., D.K., and T.K.; data acquisition, J.J.D., A.K., D.K., A.G., and J.Ž.; investigation, S.K., I.V.S., D.V.L., K.P., E.L., D.K., and B.H.; validation, M.M., S.L., and A.L.; formal analysis, S.G., M.P., and J.K.; writing—review and editing, all authors; visualization, M.P., S.G., J.K., and T.K.; supervision, S.G., D.K., and T.K; funding acquisition, D.K. and T.K. All authors have read and agreed to the published version of the manuscript.

CRediT authorship contribution statement

Marija Purgar: Writing – review & editing, Writing – original draft, Visualization, Formal analysis, Conceptualization. **Ana Gavrilović:** Writing – review & editing, Methodology, Data curation. **Damir Kapetanović:** Writing – review & editing, Writing – original draft, Supervision, Methodology, Funding acquisition, Data curation, Conceptualization. **Jasminka Klanjšček:** Writing – review & editing, Writing – original draft, Visualization, Formal analysis, Conceptualization. **Jurica Jug-Dujaković:** Writing – review & editing, Data curation. **Anamarija Kolda:** Writing – review & editing, Data curation. **Jakov Žunić:** Writing – review & editing, Data curation. **Snježana Kazazić:** Writing – review & editing, Methodology, Investigation. **Irena Vardić Smrzlić:** Writing – review & editing, Investigation. **Darija Vukić Lušić:** Writing – review & editing, Investigation. **Kristina Pikelj:** Writing – review & editing, Investigation. **Eddy Listeš:** Writing – review & editing, Investigation. **Mansour El-Matbouli:** Writing – review & editing, Validation. **Atle Lillehaug:** Writing – review & editing, Validation. **Semir Lončarević:** Writing – review & editing, Validation. **Dražen Knežević:** Writing – review & editing, Investigation. **Brigita Hengl:** Writing – review & editing, Investigation. **Sunčana Geček:** Writing – review & editing, Writing – original draft, Visualization, Supervision, Methodology, Formal analysis, Conceptualization. **Tin Klanjscek:** Writing – review & editing, Writing – original draft, Visualization, Supervision, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data are available in: 1) PANGAEA repository (Jug-Dujaković et al., 2022) and 2) Zenodo repository as a part of this study's open analytical code (Purgar et al., 2023).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ecss.2023.108558.

References

Anderson, M.J., 2001. A new method for non-parametric multivariate analysis of variance. Austral Ecol. 26, 32–46.

Arbizu, P.M., 2020. pairwiseadonis: pairwise multilevel comparison using adonis r package version 0.4. Available online: https://github.com/pmartinezarbizu/pairwiseAdonis.
Atkins, J.P., Burdon, D., Elliott, M., 2015. Identification of a practicable set of ecosystem indicators for coastal and marine ecosystem services. In: Turner, R.K., Schaafsma, M. (Eds.), Coastal Zones Ecosystem Services. Springer, pp. 79–102.
Baker Austin, C., Oliver, J.D., Alam, M., Ali, A., Waldor, M.K., Qadri, F., Martinez Urtaza, J., 2018. *Vibrio* spp. infections. Nat. Rev. Dis. Prim. 4, 1–19. https://doi.org/10.1038/s41572-018-0005-8.
Bartram, J., Cotruvo, J., Exner, M., Fricker, C., Glasmacher, A., 2003. Heterotrophic Plate Counts and Drinking-Water Safety. IWA publishing.
Benesty, J., Chen, J., Huang, Y., Cohen, I., 2009. Pearson correlation coefficient. In: Noise Reduction in Speech Processing. Springer, pp. 1–4.
Benner, R., 2011. Biosequestration of carbon by heterotrophic microorganisms. Nat. Rev. Microbiol. 9, 75–75.
Bentzon Tilia, M., Sonnenschein, E.C., Gram, L., 2016. Monitoring and managing microbes in aquaculture–towards a sustainable industry. Microb. Biotechnol. 9, 576–584. https://doi.org/10.1111/1751-7915.12392.
Boehm, A.B., Sassoubre, L.M., 2014. Enterococci as indicators of environmental fecal contamination. In: Gilmore, M.S., Clewell, D.B., Ike, Y., Shankar, N. (Eds.), Enterococci from Commensals to Leading Causes of Drug Resistant Infection (Online), Massachusetts Eye and Ear Infirmary, Boston, MA.
Brumfield, K.D., Usmani, M., Chen, K.M., Gangwar, M., Jutla, A.S., Huq, A., Colwell, R. R., 2021. Environmental parameters associated with incidence and transmission of pathogenic vibrio spp. Environ. Microbiol. 23, 7314–7340. https://doi.org/10.1111/1462-2920.15716.
Cabral, J.P., 2010. Water microbiology. bacterial pathogens and water. Int. J. Environ. Res. Publ. Health 7, 3657–3703.
Center for Food Safety and Applied Nutrition, 2011. Fish and Fishery Products Hazards and Controls Guidance, fourth ed. Silver Spring, MD.
Chandrakala, N., Priya, S., 2017. Vibriosis in shrimp aquaculture a review. International Journal of Scientific Research in Science, Engineering and Technology 3, 27–33.
Chase, E., Young, S., Harwood, V.J., 2015. Sediment and vegetation as reservoirs of vibrio vulnificus in the tampa bay estuary and gulf of Mexico. Appl. Environ. Microbiol. 81, 2489–2494. https://doi.org/10.1128/AEM.03243-14.
Davis, B.J., Jacobs, J.M., Zaitchik, B., DePaola, A., Curriero, F.C., 2019. *Vibrio parahaemolyticus* in the chesapeake bay: operational in situ prediction and forecast models can benefit from inclusion of lagged water quality measurements. Appl. Environ. Microbiol. 85, e01007–e01019.
Dray, S., Blanchet, G., Borcard, D., Guenard, G., Jombart, T., Larocque, G., Legendre, P., Madi, N., Wagner, H.H., 2022. adespatial: multivariate multiscale spatial analysis 0.3-20. Available online: https://mirror.epn.edu.ec/CRAN/web/packages/adespatial/adespatial.pdf. .
Elliot, M., Burdon, D., Atkins, J.P., Borja, A., Cormier, R., de Jonge, V.N., Turner, R.K., 2017. "And DPSIR begat DAPSI(W)R(M)!" A unifying framework for marine environmental management. Mar. Pollut. Bull. 118, 27–40. https://doi.org/10.1016/j.marpolbul.2017.03.049.
European Commission, 2006. Directive 2006/7/EC of the European Parliament and of the Council of 15 February 2006 concerning the management of bathing water quality and repealing Directive 76/160/EEC. Off. J. Eur. Union 50.
Feliatra, F., Mardalisa, M., Setiadi, J., Lukistyowaty, I., Hutasoit, A., 2020. Potential of secondary metabolite from marine heterotrophic bacteria against pathogenic bacteria in aquaculture. In: Journal of Physics: Conference Series. IOP Publishing., 012044 https://doi.org/10.1088/1742-6596/1655/1/012044
Field, A., Miles, J., Field, Z., 2012. Discovering Statistics Using R. Sage publications.
Food Standards Australia New Zealand, 2022. Compendium of Microbiological Criteria for Food.
Froelich, B., Bowen, J., Gonzalez, R., Snedeker, A., Noble, R., 2013. Mechanistic and statistical models of total vibrio abundance in the neuse river estuary. Water Res. 47, 5783–5793.
Froelich, B., Gonzalez, R., Blackwood, D., Lauer, K., Noble, R., 2019. Decadal monitoring reveals an increase in vibrio spp. concentrations in the neuse river estuary, North Carolina, USA. PLoS One 14, e0215254.
Fuhrman, J., 1992. Bacterioplankton roles in cycling of organic matter: the microbial food web. In: Primary Productivity and Biogeochemical Cycles in the Sea. Springer, pp. 361–383.
Gorgoglione, A., Castro, A., Chreties, C., Etcheverry, L., 2020. Overcoming Data Scarcity in Earth Science.
Grosjean, P., Ibanez, F., Etienne, M., Grosjean, M.P., 2018. Package 'pastecs'. R Package Version 1.3.21. Available online: https://cran.r-project.org/web/packages/pastecs/index.html.
Hach, 2022. Water analysis handbook. Available online: https://www.hach.com/wah. (Accessed 19 September 2022).
Hartnell, R., Stockley, L., Keay, W., Rosec, J.P., Hervio-Heath, D., Van den Berg, H., Leoni, F., Ottaviani, D., Henigman, U., Denayer, S., et al., 2019. A paneuropean ring trial to validate an international standard for detection of vibrio cholerae, vibrio parahaemolyticus and vibrio vulnificus in seafoods. Int. J. Food Microbiol. 288, 58–65.
Ina-Salwany, M., Al-saari, N., Mohamad, A., Mursidi, F.A., Mohd-Aris, A., Amal, M., Kasai, H., Mino, S., Sawabe, T., Zamri Saad, M., 2019. Vibriosis in fish: a review on disease development and prevention. J. Aquat. Anim. Health 31, 3–22.
Johnson, C.N., Bowers, J.C., Griffitt, K.J., Molina, V., Clostio, R.W., Pei, S., Laws, E., Paranjpye, R.N., Strom, M.S., Chen, A., et al., 2012. Ecology of vibrio parahaemolyticus and vibrio vulnificus in the coastal and estuarine waters of Louisiana, Maryland, Mississippi, and Washington (United States). Appl. Environ. Microbiol. 78, 7249–7257. https://doi.org/10.1128/AEM.01296-12.
Jug Dujaković, Jurica, Gavrilović, Ana, Kolda, Anamarija, Žunić, Jakov, Pikelj, Kristina, Listeš, Eddy, Lušić, Vukić, Darija, Vardić Smrzlić, Irena, Perić, Lorena, El-Matbouli, Mansour, Lillehaug, Atle, Lončarević, Semir, Knežević, Dražen, Hengl, Brigita, Purgar, Marija, Kapetanović, Damir, 2022. AQUAHEALTH: physiochemical parameters and *Vibrio* spp. abundance in water samples gathered in the Adriatic Sea. PANGAEA. https://doi.org/10.1594/PANGAEA.951288.
Kapetanović, D., Vardić Smrzlić, I., Gavrilović, A., Jug Dujaković, J., Perić, L., Kazazić, S., Mišić Radić, T., Kolda, A., Čanković, M., Žunić, J., et al., 2022. Characterization of *Vibrio* populations from cultured European seabass and the surrounding marine environment with emphasis on *V. Anguillarum*. Microorganisms 10 (11), 2159. https://doi.org/10.3390/microorganisms10112159.
Kassambara, A., 2023. rstatix: pipe-friendly framework for basic statistical tests. R package, version 0.7.2. https://CRAN.R-project.org/package=rstatix.
Kim, J.Y., Lee, J.L., 2017. Correlation of total bacterial and vibrio spp. populations between fish and water in the aquaculture system. Front. Mar. Sci. 4, 147. https://doi.org/10.3389/fmars.2017.00147.
Legendre, P., Legendre, L., 2012. Numerical Ecology. Elsevier.
Levy, S., 2018. Ecdc vibrio map viewer: tracking the whereabouts of pathogenic species. Environ. Health Perspect. 126, 034003.
Manchanayake, T., Salleh, A., Amal, M.N.A., Yasin, I.S.M., Zamri-Saad, M., 2023. Pathology and pathogenesis of vibrio infection in fish: a review. Aquaculture Reports 28, 101459.
Martínez, M.L., Intralawan, A., Vázquez, G., Pérez-Maqueo, O., Sutton, P., Landgrave, R., 2007. The coasts of our world: ecological, economic and social importance. Ecol. Econ. 63 (2–3), 254–272.
Maul, G.A., Duedall, I.W., 2019. Demography of coastal populations. In: Finkl, C.W., Makowski, C. (Eds.), Encyclopedia of Coastal Science. Encyclopedia of Earth Sciences Series. Springer International Publishing, Cham, pp. 692–700.
Mazé, C., Dahou, T., Ragueneau, O., Danto, A., Mariat-Roy, E., Raimonet, M., Weisbein, J., 2017. Knowledge and power in integrated coastal management. For a political anthropology of the sea combined with the sciences of the marine environment. Compt. Rendus Geosci. 349 (6–7), 359–368.
Moran, M.A., Kujawinski, E.B., Stubbins, A., Fatland, R., Aluwihare, L.I., Buchan, A., Crump, B.C., Dorrestein, P.C., Dyhrman, S.T., Hess, N.J., et al., 2016. Deciphering ocean carbon in a changing world. Proc. Natl. Acad. Sci. USA 113, 3143–3151.
Nobre, A., 2011. Scientific approaches to address challenges in coastal management. Mar. Ecol. Prog. Ser. 434, 279–289. https://doi.org/10.3354/meps09250.
OGRC 73/08, 2008. Official gazette of the republic of Croatia 73/08. Regulation on sea bathing water quality. Ministry of environment and energy. Available online: https://narodne-novine.nn.hr/clanci/sluzbeni/2008_06_73_2426.html.
Oksanen, J., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R., O'hara, R., Simpson, G.L., Solymos, P., Stevens, M.H.H., Wagner, H., et al., 2013. Package 'vegan'. Community Ecology Package version 2, 1–295.
Paterson, K.J., Schreider, M.J., Zimmerman, K.D., 2003. Anthropogenic effects on seston quality and quantity and the growth and survival of sydney rock oyster (*Saccostrea glomerata*) in two estuaries in nsw, Australia. Aquaculture 221, 407–426.
Perkins, T.L., Clements, K., Baas, J.H., Jago, C.F., Jones, D.I., Malham, S.K., McDonald, J.E., 2014. Sediment composition influences spatial variation in the abundance of human pathogen indicator bacteria within an estuarine environment. PLoS One 9, e112951. https://doi.org/10.1371/journal.pone.0112951.
Price, R.G., Wildeboer, D., 2017. *E. coli* as an indicator of contamination and health risk in environmental waters. Escherichia coli-Recent Advances on Physiology. Pathogenesis and Biotechnological Applications 13doi. https://doi.org/10.5772/67330.
Pruzzo, C., Huq, A., Colwell, R.R., Donelli, G., 2005. Pathogenic vibrio species in the marine and estuarine environment. In: Belkin, S., Colwell, R.R. (Eds.), Oceans and Health: Pathogens in the Marine Environment. Springer, Boston, MA, pp. 217–252.
Purgar, M., Kapetanović, D., Geček, S., Mari, N., Haberle, I., Hackenberger, B.K., Gavrilović, A., Pećar Ilić, J., Hackenberger, D.K., Djerdj, T., Čaleta, B., Klanjscek, T., 2022. Investigating the ability of growth models to predict in situ *Vibrio* spp. abundances. Microorganisms 10 (9), 1765.
Purgar, M., Gavrilović, A., Kapetanović, D., Klanjšček, J., Jug-Dujaković, J., Kolda, A., Žunić, J., Kazazić, S., Vardić Smrzlić, I., Vukić Lušić, D., Pikelj, K., Listeš, E., El-Matbouli, M., Lillehaug, A., Lončarević, S., Knežević, D., Hengl, B., Geček, S., Klanjscek, T., 2023. Datasets and code for Purgar et al. 2023: Assessment of Vibrio spp. Abundance as a Water Quality Indicator: Insights from Mali Ston Bay in the Adriatic Sea. Zenodo. https://doi.org/10.5281/zenodo.10046184 [Data set].
Revelle, W.R., 2017. psych: Procedures for personality and psychological research. r package version 2.2.9. Available online: https://cran.r-project.org/web/packages/psych/index.html.
Robertson, W., 1984. Pollution indicators and potential pathogenic microorganisms in estuarine recreational waters. Canadian Journal of Public Health/Revue Canadienne de Sante'e Publique 19–24.
Roux, F.L., Wegner, K.M., Baker-Austin, C., Vezzulli, L., Osorio, C.R., Amaro, C., Ritchie, J.M., Defoirdt, T., Destoumieux-Garz'on, D., Blokesch, M., et al., 2015. The emergence of vibrio pathogens in europe: ecology, evolution, and pathogenesis (paris, 11–12th march 2015). Front. Microbiol. 6, 830.
RStudio Team, 2022. RStudio: Integrated Development Environment for R. RStudio. PBC, Boston, MA. URL: http://www.rstudio.com/.
Ryder, J., Iddya, K., Ababouch, L., 2014. Assessment and Management of Seafood Safety and Quality: Current Practices and Emerging Issues. FAO fisheries and aquaculture technical paper, 1.
Sampaio, A., Silva, V., Poeta, P., Aonofriesei, F., 2022. *Vibrio* spp.: life strategies, ecology, and risks in a changing environment. Diversity 14, 97.
Sanches-Fernandes, G.M., S'a-Correia, I., Costa, R., 2022. Vibriosis outbreaks in aquaculture: addressing environmental and public health concerns and preventive

therapies using gilthead seabream farming as a model system. Front. Microbiol. 13, 904815.
Saxena, G., Bharagava, R.N., Kaithwas, G., Raj, A., 2015. Microbial indicators, pathogens and methods for their monitoring in water environment. J. Water Health 13, 319–339. https://doi.org/10.2166/wh.2014.275.
Semenza, J.C., Trinanes, J., Lohr, W., Sudre, B., Löfdahl, M., Martinez-Urtaza, J., Nichols, G.L., Rocklöv, J., 2017. Environmental suitability of Vibrio infections in a warming climate: an early warning system. Environ. Health Perspect. 125 (10), 107004.
Sheikh, H., Musa, N., Fadhlina, A., Abdulrazzak, L.A., Nor, M.M., Kca, J., Kasan, N.A., 2022. Temperature upshift mostly but not always enhances the growth of vibrio species: a systematic review. Front. Mar. Sci. 1297doi https://doi.org/10.3389/fmars.2022.959830.
Some, S., Mondal, R., Mitra, D., Jain, D., Verma, D., Das, S., 2021. Microbial pollution of water with special reference to coliform bacteria and their nexus with environment. Energy Nexus 1, 100008. https://doi.org/10.1016/j.nexus.2021.100008.
Stewart, J.R., Gast, R.J., Fujioka, R.S., Solo-Gabriele, H.M., Meschke, J.S., Amaral-Zettler, L.A., Del Castillo, E., Polz, M.F., Collier, T.K., Strom, M.S., et al., 2008. The coastal environment and human health: microbial indicators, pathogens, sentinels and reservoirs. Environ. Health 7, 1–14. https://doi.org/10.1186/1476-069X-7-S2-S3.
Takemura, A.F., Chien, D.M., Polz, M.F., 2014. Associations and dynamics of vibrionaceae in the environment, from the genus to the population level. Front. Microbiol. 5, 38. https://doi.org/10.3389/fmicb.2014.00038.
Thickman, J.D., Gobler, C.J., 2017. The ability of algal organic matter and surface runoff to promote the abundance of pathogenic and non pathogenic strains of vibrio parahaemolyticus in long island sound, USA. PLoS One 12, e0185994. https://doi.org/10.1371/journal.pone.0185994.
Toranzo, A.E., Barja, J.L., 1990. A review of the taxonomy and seroepizootiology of vibrio anguillarum, with special reference to aquaculture in the northwest of Spain. Dis. Aquat. Org. 9, 73–82.
Veić, T., 2016. Characterization of *Vibrio* species isolated from european sea bass (*Dicentrarchus labrax* linnaeus, 1758) farmed on the east coast of the adriatic sea. Available at https://urn.nsk.hr/urn:nbn:hr:217740993. .
Vezzulli, L., Pezzati, E., Moreno, M., Fabiano, M., Pane, L., Pruzzo, C., Consortium, V., 2009. Benthic ecology of vibrio spp. and pathogenic vibrio species in a coastal mediterranean environment (la spezia gulf, Italy). Microb. Ecol. 58, 808–818. https://doi.org/10.1007/s00248-009-9542-8.
Wei, T., Simko, V., Levy, M., Xie, Y., Jin, Y., Zemla, J., et al., 2017. Package 'corrplot'. Statistician 56, e24.
Wickham, H., 2016. ggplot2: Elegant Graphics for Data Analysis.
World Health Organization, 2017. Guidelines for drinking-water quality: First addendum to the fourth edition.
Yang, C., Li, Y., Jiang, M., Wang, L., Jiang, Y., Hu, L., Shi, X., He, L., Cai, R., Wu, S., et al., 2022. Outbreak dynamics of foodborne pathogen vibrio parahaemolyticus over a seventeen year period implies hidden reservoirs. Nature Microbiology 7, 1221–1229. https://doi.org/10.1038/s41564-022-01182-0.
Zadorozhnaya, O., Kirsanov, D., Buzhinsky, I., Tsarev, F., Abramova, N., Bratov, A., Muñoz, F.J., Ribó, J., Bori, J., Riva, M.C., Legin, A., 2015. Water pollution monitoring by an artificial sensory system performing in terms of *Vibrio fischeri* bacteria. Sensor. Actuator. B Chem. 207, 1069–1075.
Zhang, X., Lin, H., Wang, X., Austin, B., 2018. Significance of vibrio species in the marine organic carbon cycle—a review. Sci. China Earth Sci. 61, 1357–1368. https://doi.org/10.1007/s11430-017-9229-x.
Zupičić, I.G., Pavlinec, Ž., Veić, T., Oraić, D., Zrnčić, S., 2022. Similarities and differences between vibriosis in European sea bass caused by *Vibrio anguillarum* and *Vibrio harveyi*. Vet. Arh. 92 (4), 509–524.

2.3. Publication III: Quantifying research waste in ecology

Purgar, M., Klanjscek, T., Culina, A. (2022b). Quantifying research waste in ecology. Nature Ecology & Evolution, 6(9), 1390-1397. https://doi.org/10.1038/s41559-022-01820-0

Contributions: data collection, data analysis, data interpretation, and writing - review and editing.

nature ecology & evolution

ANALYSIS

https://doi.org/10.1038/s41559-022-01820-0

Check for updates

Quantifying research waste in ecology

Marija Purgar[1,2], Tin Klanjscek[1] and Antica Culina[1,3] ✉

Research inefficiencies can generate huge waste: evidence from biomedical research has shown that most research is avoidably wasted and steps have been taken to tackle this costly problem. Although other scientific fields could also benefit from identifying and quantifying waste and acting to reduce it, no other estimates of research waste are available. Given that ecological issues interweave most of the United Nations Sustainable Development Goals, we argue that tackling research waste in ecology should be prioritized. Our study leads the way. We estimate components of waste in ecological research based on a literature review and a meta-analysis. Shockingly, our results suggest only 11–18% of conducted ecological research reaches its full informative value. All actors within the research system—including academic institutions, policymakers, funders and publishers—have a duty towards science, the environment, study organisms and the public, to urgently act and reduce this considerable yet preventable loss. We discuss potential ways forward and call for two major actions: (1) further research into waste in ecology (and beyond); (2) focused development and implementation of solutions to reduce unused potential of ecological research.

Research generates a wealth of output: datasets, workflows, analytical codes and—ultimately—derived results[1,2]. Only a small and likely biased subset of the output is published[3,4] and is thus available as information, which is often used in evidence synthesis[5,6]. Hence, much of potential knowledge stays hidden. More worryingly, when the 'publish or perish' research culture[7] couples with human cognitive biases[8] and lack of training[9], even data collection and analysis can be suboptimal and biased. These issues are becoming hard to ignore. Emerging evidence indicates that the problem could be relatively large across the sciences[10–12], including ecology[13–19], and is exacerbated by the failure to replicate results of previous studies across disciplines[10–12]. Some think we are facing a crisis[20]. Yet, to understand how much information we lose in the current research and publishing system, and how to best act to rectify the problem, we need a quantitative estimate of information loss (that is, research waste) over the research life cycle. Yet, research waste has been quantified only in medicine[21].

A highly influential seminal editorial by Altman[22], and follow-up work on research waste in medicine[21] (estimated 85% waste, globally equal to over US$170 billion annually[23]) triggered a series of seminars, meetings and the introduction of new policies that target reduction of the waste in medicine[24,25], thereby increasing the value of medicinal research. We want to start a comparable global and focused movement in ecology but also across the sciences to quantify the problem of research waste and facilitate a more serious and coordinated move towards changing standards for research and publishing. Identifying research waste is clearly the first step.

'Ignorance is expensive'[26]. This statement also applies to ignorance of research inefficiencies that can generate huge waste. The health of our environment, and thus of humans, and our ability to solve global challenges depends on robust and well-informed ecological research. As ecologists, as well as those who fund ecological research, we must aim to reduce the waste produced in our work. But how large is this waste and how big a problem is it?

Components of research waste

Research waste accumulates over the classical research life cycle (Fig. 1). The main stages of the research cycle for which we estimate research waste are: study planning (includes core study design, data collection and data analysis); results reporting; and publication. For our classification of waste components, we consider that research waste generated during data collection and data analysis is a problem of study planning. Well-planned studies should foresee, before data collection and analysis, the core study design (for example, experimental treatment allocation for the data collection set-up), exact data collection procedures (for example, blinding while collecting data) and statistical approaches that are appropriate given the core study design and the type of data collected (for example, controlling for covariates).

We distinguish two types of waste: core waste and exploitative waste. Core waste is all the conducted (and funded) work that never gets published. The causes of core waste are dual: low-quality studies and publication bias. Low-quality studies are unpublished because they are poorly planned or poorly conducted. Their publication would likely be detrimental. Publication bias, on the other hand, prevents publication of the research of adequate conceptual and methodological quality. This research is unpublished solely because its results are not considered 'interesting' (for example, null results). Exploitative waste represents a reduced potential of published work to inform the users (that is, to be exploited by the users). Exploitative waste is generated by all published studies with issues at the study planning stage[27] or result reporting stage[17]. Core and exploitative waste combine and lead to the overall waste that accumulates over the research life cycle.

How much research in ecology is avoidably wasted?

In this study, we provide a breakdown of the components of research waste based on a review of published literature (Methods and Supplementary Methods). We identified 34 meta-studies that estimated the components of research waste in ecology. We define a meta-study as a study that used published (and less often unpublished) studies to estimate different components of waste in ecology (at the study planning, at result reporting and at the publication stage). Only one of these meta-studies used an indirect estimation method (below and Supplementary Methods) and was thus excluded from the meta-analysis. Thus, our overall sample size was 33 meta-studies that, based on 10,464 studies, provided 43 estimates of research waste components. We summarized estimates of

[1]Rudjer Boskovic Institute, Zagreb, Croatia. [2]Department of Biology, University of Osijek, Osijek, Croatia. [3]Netherlands Institute of Ecology, NIOO-KNAW, Wageningen, the Netherlands. ✉e-mail: aculina@irb.hr

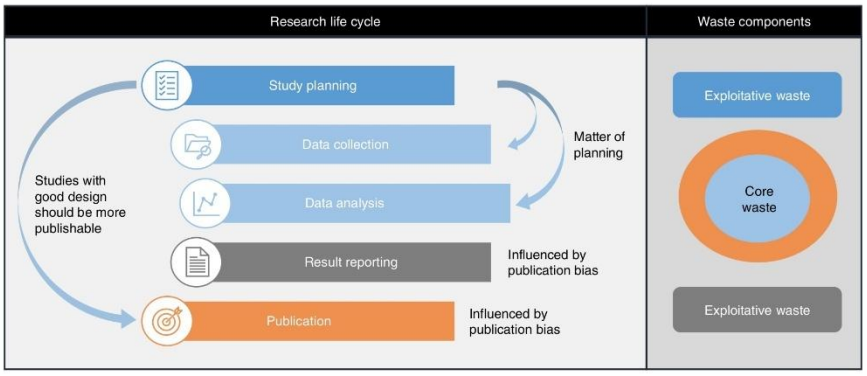

Fig. 1 | Stages of the classical research life cycle. We consider that any suboptimal study planning leads to waste in data collection and data analysis. This is because data collection and analysis should conceptually happen at the study planning stage even though physically conducted later. Further, the study planning stage influences the publication stage because badly planned studies are less likely to be published. The components of the research life cycle translate into components of research waste (right) where core waste represents all unpublished work (due to either low-quality study planning or publication bias) and exploitative waste represents the component of published research with a limited ability to inform future work (that is, to be exploited by the users) either because the study conducted (and later published) was of low quality (for example, issues with study design) or because the results of the study were reported in a way that prevents their use (for example, effect size or sample size not reported).

research waste that belong to the same waste component using a meta-analytical model (Methods). In this study, we weighted each effect size by the sample size of a meta-study. When combined, these meta-analytical estimates of the components of research waste led to the first estimate of the overall research waste in ecology.

We investigated two scenarios; both give worryingly high estimates of the overall research waste (Fig. 2). The best-case scenario assumes that waste components overlap, that is, that all under reporting appears in poorly planned studies, leading to 82% waste. In the worst-case scenario, poor planning and under reporting do not happen in the same studies, increasing the waste to 89%. Hence, between 82 and 89% of research appears to be avoidably wasted, or, in other words, unused. Interestingly, these numbers are very close to the only other existing estimate of 85% waste for medicine[31]. We provide the breakdown of the waste components below.

Core waste. Core waste is all the work that is unpublished due to either its low quality or publication bias. Meta-analysis of 10 direct estimates from 9 meta-studies (based on an overall sample size of 2,252 studies) estimated that core waste equals 44.7% (95% confidence interval (CI) 44.2–46.7%; Fig. 3a) of research. Estimates from the meta-studies included the percentage of unpublished projects (for example, projects collecting telemetry data that never published a single result[28]), unpublished theses chapters (for example, Koricheva[29]) or unpublished literature (for example, Bennett and Adams[30]). Only one of the meta-studies[15] provided an indirect estimate of unpublished research (using the trim and fill method[31]). We excluded this indirectly estimated value from the main meta-analysis (see the Supplementary Methods for the reasons) but we show the recalculated meta-analytical mean with this indirect estimate included (Supplementary Results and Supplementary Fig. 4). The meta-analytical estimates of core waste were similar for meta-studies that concern broader areas of ecology (for example, ecology, conservation ecology) and those with a narrower topic coverage (for example, facultative sex ratio adjustment in birds), as shown in Fig. 3a.

We lacked data to calculate the proportion of core waste caused by publication bias versus that caused by studies that are unpublished because of their low quality. Only one meta-study compared

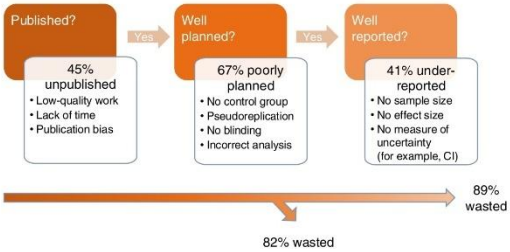

Fig. 2 | Overall estimate of research waste of ecological research based on a meta-analysis of waste at each stage (with examples of causes). In the best-case scenario, 82% of research is wasted and thus is unused because all under reporting is assumed to happen in poorly planned studies. In the worst-case scenario, 89% of the research is unused because all of the under reporting is assumed to happen in the otherwise well-planned research. Consequently, only 11–18% of conducted ecological research can inform users (other researchers, public, policymakers) fully.

the quality of study design between published and unpublished studies[32], finding that 13% of unpublished studies, and 25% of published studies, lacked a control group. Further, the study by Koricheva[29] broke down the reasons why some of the 187 doctoral theses chapters were never published. She found that 10.1% of these were never submitted for publication, largely due to lack of time (68%). Of 156 submitted chapters, 16.7% were rejected. Of these, 42.5% were rejected because of issues at the study planning stage (study design issues, data analysis issues, poor theoretical background), while around 14% were rejected because of the lack of novelty in the findings.

Exploitative waste. Exploitative waste represents the component of published research with a limited ability to inform future work either because the study conducted (and later published) was of

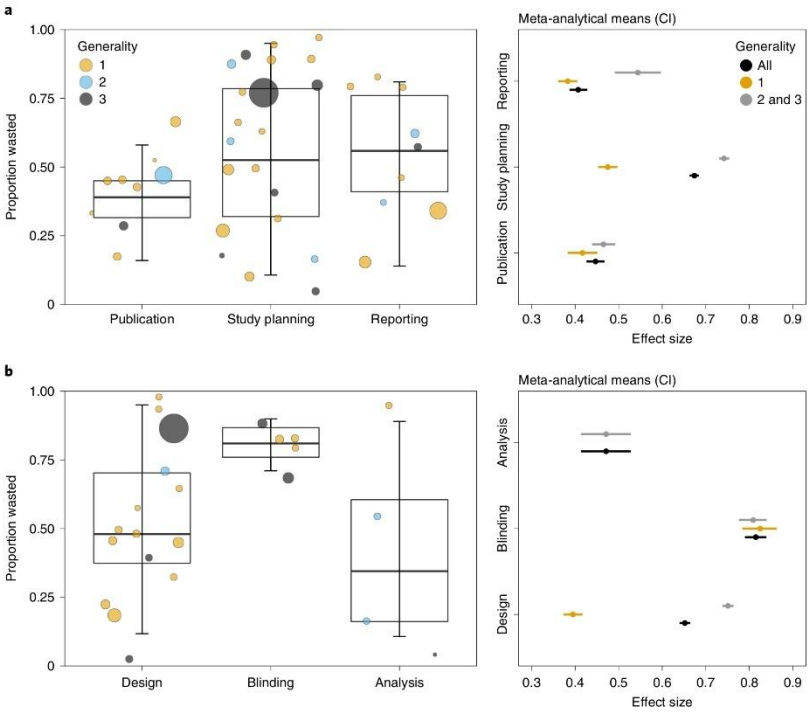

Fig. 3 | Estimates of the main components of research waste. a, Estimates of the main components of research waste, from each meta-study, and a boxplot of their distribution. **b**, Breakdown of research waste generated during the study planning stage, partitioned between different temporal stages of the research life cycle. Left panels: estimates of research waste (circles) as reported by each meta-study (whisker plot denotes their distribution). The circle size is proportional to the sample size used in each meta-study. The circles are coloured by the degree of generality, with 1 representing meta-studies covering narrow ecological subfields and 3 representing meta-studies not limited to a certain ecological subfield (that is, are broad). The boxplot central line represents the median of the estimates, the lower and upper edge of the boxplot represent the 25th and 75th percentiles of the distribution and the whiskers are the smallest and largest value within the 1.5 times interquartile range below and above the 25th and 75th percentiles. Right panels: meta-analytical mean of all effect sizes, i.e., proportion of research wasted (black circles), effect sizes coming from meta-studies with a narrow scope (generality 1, blue circles) and broad scope (generality of 2 and 3, grey circles), with a 95% CI.

low quality (for example, issues with study design) or because the results of the study were reported in a way that prevented their use (for example, effect size or sample size not reported). A shockingly high percentage of published research has issues at the level of study planning: the meta-analytical mean of 22 estimates from 21 meta-studies with an overall sample size of 7,505 studies, showed that 67.4% (95% CI 66.3–68.4%) of published studies in ecology have issues at the planning stage (Fig. 3a).

Conceptually, the core study design (for example, randomization of treatment units), data collection protocol (for example, blinded data collection) and analysis plan should be created at the study planning stage. Yet, timewise these happen sequentially and refer to different time steps of the classical research life cycle (Fig. 1). Thus, we broke down the study planning stage into estimates that correspond to these three different time steps of the research life cycle. The meta-analytical mean of 16 estimates from 15 meta-studies with an overall sample size of 6,606 studies, showed that 65.2% of studies (95% CI 64.0–66.4%) have core design issues (Fig. 3b). Most core design issues are caused by pseudo-replication (for example, Hurlbert[33]). At the data collection stage, the only available estimates were those for blinded versus non-blinded data collection: based on 5 estimates with a sample size of 981, it appears that most of the studies in ecology do not blind the observer to the data (81.5%, 95% CI 79.0–83.9%; Fig. 3b). Finally, at the statistical analysis stage, 4 estimates with a sample size of 288 showed that overall 47.1% (95% CI 41.3–52.8%) of analytical choices are suboptimal or incorrect. The severity of the problem is slightly worse when considering only the estimates from the meta-studies that capture the general field of ecology (Fig. 3b).

The results of the research will be used by different users (other researchers, policymakers, industry), commonly in the form of evidence synthesis[5,6]. The results can be well reported, reported incorrectly (misreported) or under-reported. Under-reporting is common, with 40.7% (95% CI 38.7–42.8%; Fig. 3a) of results being under-reported (based on 9 estimates with a sample size of 2,246). For example, a large proportion of results were reported without effect size, sample size or measure of uncertainty around the estimate. Our review did not identify any estimate of misreported results in ecology.

Core waste undoubtedly constitutes loss of knowledge. However, to determine how much exploitative waste contributes to information loss is difficult. Even non-rigorously conducted

and under-reported research can still have an informative value, albeit reduced compared to rigorous or well-reported research. For example, a study reporting a direction of an effect, without an effect size, will have a higher informative value than if the result was not reported at all. For a similar reason, we opted to exclude estimates of underpowered studies from our calculations of waste. Underpowered research can still lead to valid conclusions and can contribute to the overall evidence for a certain effect. Power is not only a statistical issue, but is limited by finances, time available and sometimes by the study system or organism (for example, rare species). However, we call for more consideration of sample size calculation in ecology and for study designs that are better adjusted to small sample sizes because our data suggest that almost all of the studies in ecology are underpowered (for example, Jennions and Møller[34]; also see the Dataset_starting data, available in the data package[35], for extracted estimates of underpowered research in ecology). Further, low-powered studies would benefit from being more straightforward about the implications that small sample sizes can have for the conclusions reached; they would benefit from coordination between groups that study the same phenomenon with the same methods.

Other factors that contribute to research waste

We estimate that a very high proportion of ecological research (82–89%) has limited information value because of the research waste accumulating over the research life cycle. Yet, other factors also contribute to the potential of research to inform future research, policy or interventions. These factors include (but are not limited to) access options (whether research has been published open access or with a paywall), the transparency and openness of the underlying research process and usability of codes and datasets. In this article, we develop on some if these factors.

Accessibility of publications. Published results are unfortunately not equally available to everyone. We estimated, based on the literature listed at the Europe PMC[36] (see Supplementary Methods for details) that 57.7% of 19,165 articles published in 94 ecological journals between 1957 and 2021 are open access. The situation changed for the better: among articles published after 2014 (11,980 articles), 73.0% are open access. This likely reflects overall trends in mandates by research funders to make funded research open access (ref. [37], also see ROARMAP https://roarmap.eprints.org/). Open access to published articles also exposes information to a higher number of users and thus has a higher potential to lead to discoveries, generate new ideas or spot errors. However, while open access to publications enables equality in access to information, it still creates inequality in who can publish open access[38,39] because open access fees are beyond reach for many researchers. Second, time between submission and publication (and thus its accessibility) can often be long, which can delay and even reduce the efficiency and impact of research[40]. Preprints might be a solution to both problems because they allow work to be visible before its official publication, while also making the preprint version available to anyone to read[41].

Unpublished data, methods and codes. Published results are only the tip of the iceberg, whose bulk consists of datasets, methods and data processing and analysis codes. These can be often more informative than the published results themselves, especially if the results are, as we have demonstrated in this work, under-reported. Additionally, having access to all research components helps the intended audience understand how published results were derived[42,43]. More importantly, reuse of data, methods and code can further accelerate scientific discovery and progress[14,44,45]. While the amount of open data is increasing in ecology[45], we lack a large-scale estimate of its quality and thus usability (for example, as done on a smaller sample by Roche et al.[46]), which seems rather low[46] (for example, lack of meta-data). A recent study[14] on code availability estimated that even among journals with a code policy, only around 27% of papers published also submitted their analytical codes, while only 21% papers were potentially computationally reproducible (that is, had data and code).

Reference to previous studies. Research waste is reduced when any new research is informed by past research[25,47] by, for example, conducting a systematic review of existing literature before starting new research. Such a practice has been encouraged (albeit still not widely adopted[48]) in medicine, especially since the 2014 *Lancet* series on 'Research: Increasing Value, Reducing Waste'. Ecology is lagging despite recent call for systematic review as a first stage of the research cycle[47]—probably because a lack of estimates (and therefore awareness) of the extent of the problem. When time or finances are limited, other types of review (for example, rapid evidence synthesis[49]) could be a solution. In this case, the costs and benefits of such an approach must be carefully considered[50,51].

Limitations of our approach

Our approach to calculating research waste components has a few limitations. First, like most literature reviews it is restricted to the literature published in English[52,53]. Thus, strictly speaking, we have estimated the research waste of research published in the English language. The evidence on whether research waste components differ between languages is limited and is non-conclusive in medical research[53,54]. Only one meta-study in our sample addressed the difference between English and non-English language literature: Vorobeichik and Kozlov[55] found that studies published in English tend to have a better quality of result reporting compared to studies published in Russian (68 versus 28% of results are well reported, respectively).

Second, we were not able to look into the trends because most of the meta-studies considered extended periods (for example, all the work published before a certain year). Based on several studies that reported separate values for different periods, it appears that there was no major shift in reducing waste components over time (see Dataset_MA_final data from the data package[35]).

Finally, our literature review did not retrieve any estimates of the prevalence of some of the questionable research practices[13]. Examples of these practices include optional stopping in data collection until a 'wanted' result is obtained[13,56] or taking advantage of the flexibility in the choice of analytical procedures (for example, including and excluding variables)[56] to obtain the desired result. One meta-study estimated the prevalence of questionable research practices in ecology but only based on surveys of researchers[13]. This study detected that among 807 ecologists and evolutionary biologists, 42% had collected more data after inspecting whether results were statistically significant, and 4.5% fabricated their data.

For the above reasons, we want to call for a community-wide discussion on the implications of different components of the research waste for knowledge generation and knowledge loss and for community-driven solutions to waste reduction. Further, we need to continue working on estimating the waste components on a larger set of ecological literature, including time trends.

Priority actions

Our results are plain—we have a huge knowledge loss from the onset of studies to the publication of results. In the twenty-first century and in line with meeting sustainable development goals[57], our priorities should be clear: reduce research waste and increase the knowledge gain from the rich ongoing ecological (and other) research. Responsibility to do this lies with funders, publishers, research institutions and researchers since all of them contribute to the research culture and research practices. The aim of our study was not to dissect all the possible ways for reducing research waste,

but start and facilitate a serious discussion and concrete actions on changing this alarming situation (as has happened in medicine). Thus, we provide only a brief outline of some potential solutions. These include changes in incentives and mandates, promotion of rigorous research practices and transparent research and better training of and support for scientists. Clearly, some solutions will differ among fields and subfields. Therefore, we strongly advocate further research into methods for quantifying the problem and finding optimal field-specific but also general solutions.

Some of the components of research waste, as detected by our study, should be easy to correct. For example, blinding leads to more robust results compared to unblinded research[18] and should not incur any additional study costs. Therefore, researchers should ideally blind themselves to data collection. However, in ecology, which is often based on field studies, blind data collection is often impossible. If so, researchers can blind themselves to data analysis. A nice overview on why blinding is important and how to do blind data analysis can be found in MacCoun and Perlmutter[58].

Quality of reporting can also be rapidly increased as high-quality result reporting should not be time-consuming or costly; many guidelines on result (and method) reporting are available[59,60]. Some changes, however, might require more effort and time. For example, preregistration of studies is still not widely adopted in ecology but it has been shown to reduce bias in research (in medicine[61]). Preregistration also enables detection of errors in study design before the study is conducted, thus reducing (or preventing) the main component of waste as detected in our study (study planning stage).

Funders and academic institutions have a primary responsibility for the reduction of waste. They shape the behaviour of researchers by deciding what research to fund and by setting the reward, promotion and mandate systems in science and academia. A long-set focus on journal publication (especially in high-impact factor journals) and an interconnected focus on securing competitive funding, were set up to select the best science and best scientists. However, it appears that this system is also good at selecting for questionable research practices and non-rigorous science and scientists, including low diversity of those selected[62]. For example, a recent large-scale study showed that over 50% of Dutch scientists engage in questionable research paractices[63].

The good news is that funders, institutions and publishers are becoming aware that incentives and mandates must change. Utrecht University has completely abandoned impact factor in hiring and promotion[64], while the European Commission is starting a reform of the research assessment[65]. In parallel, the European Commission has achieved a high level of open access publications (83%) under the Horizon 2020 programme[66], while the University of California leveraged its size and purchasing power to force open access concessions from Elsevier[67]. These are just some examples of changing incentives and mandates. Publishers can then build on the system by further regulating the type of research that gets published, and can set additional requirements. For example, an increase in the quantity of open data has been reported after many journals adopted open data policies[68]. Similarly, it has been recently shown that the introduction of *Nature*'s reproducibility checklist has improved the reporting standards of papers published by the *Nature* Publishing Group[69].

The bad news is that the incentives are shifting very slowly and in a non-synchronized way between countries and disciplines. Science is a global, cross-disciplinary endeavour. Thus, it is imperative to establish a global set of new incentives and rules. Further, new incentives should promote rigorous research even though such research takes longer, and might also be more likely to produce less 'exciting' but more robust findings. Consequences of notable international efforts to change the evaluation of researchers should be examined and, if successful, widely adopted (for example, the San Francisco Declaration on Research Assessment). Finally, funders need to become more transparent in their funding decisions, being mindful that the funded research is not only of high priority but also of high methodological quality[25,61].

Related to the above, funders and academic institutions should provide an adequate system to support scientists in conducting a more robust science. This support should include training of researchers and support from skilled personnel and infrastructures. Thus we join and substantiate calls for: (1) more courses on methodologically robust and transparent scientific research in student curricula and training of established researchers[9,61,70]; (2) increase in the involvement of experienced methodologists, statisticians and data stewards on projects[61,70] by, for example, securing funding for such personnel or establishing advisory bodies that would provide advice and guidance for funded projects; (3) better technical/infrastructural support[70] for enabling open science practices, rigorous reporting, archival of all elements of research and creating linkages among them. We especially call for support for preregistration of studies since many of the issues with study design and later appearing questionable research practices can be avoided this way.

The outlook

Apart from the immediate actions listed in the previous section, we also call for coordinated meta-scientific research and more funding for meta-science in ecology (as already done seven years ago in medicine[25]). Open science[3,71] and meta-science[1,72], two movements that span scientific disciplines, have emerged largely because of the need to reduce the impact of research biases on scientific knowledge. Open science aims to make all the components of the research cycle available to everyone. This generates higher knowledge gains based on the research conducted and increases trust in science[73]. Further, open science calls for changes in scientific incentives since these are likely at the root of research biases.

Meta-science goes in hand with open science as it investigates efficiency, quality and bias in the scientific ecosystem, and offers solutions to the challenges this system is facing[1,72]. Meta-science emerged as a discipline very recently, in parallel with a failure of several large-scale replication projects to replicate results of the previous studies[10–12]. However, meta-science is poorly integrated into most disciplines. In ecology, meta-science has not even emerged as a strong research line[74], although the number of meta-studies has been increasing (including this one).

With this work, we also introduce a new term—unused potential of research. Unused potential is likely much larger than waste but at the same time impossible to calculate (at present). For example, we cannot foresee what impact particular research would have had if its design had been better or if its results were fully rather than partially reported. Further, we believe that focusing on unused potential instead of waste better facilitates actionable recommendations for improvement and reduces resistance to adoption.

Our framework can be used (and potentially broadened) to identify and quantify waste components in other research fields or ecological subfields. Further, we should develop and apply methods to investigate additional unused potential that transcends pure waste. Given commonalities across research disciplines, we should then be able to arrive at a common set of policies that would decrease unused research potential in science. At the same time, and given specificity of each research field, we might need to be developing field-specific solutions. Further work should thus estimate (1) the exact costs of practices that contribute to research waste (for example, how much does non-blinding shift the estimates of an effect) and (2) the costs of different solutions to reducing waste (for example, the financial or time cost to apply blinding). In this way we could identify the best (that is, feasible and cost-effective) set of actions to reduce waste.

Conclusions

In this study, we arrived to a shockingly high estimate of research waste in ecological research. Thus, a large part of ecological research is unused. However, the overall unused potential of any research is impossible to calculate. This is because we cannot foresee the potential impact of any single result, dataset or method on knowledge development or applied solutions, especially since these are sometimes visible only in the far future. This is exactly why we need to urgently reduce the waste that accumulates over the research life cycle and open up all of the components of research. Only in this way we can enable the highest knowledge gain from past and ongoing research.

We hope our call will awaken funders, publishers, research institutions and researchers to the tremendous cost of ignoring unused potential in ecological research and research in general. 'Ignorance is expensive'[26] and we cannot allow this loss of knowledge to streamline and continue. Thus, in our conclusions we repeat the plain finding—due to suboptimal practices, only 11–18% of conducted ecological research reaches its full informative value.

Methods

Literature review and data extraction. In May 2021, we used the Web of Science Core Collection databases (please see Supplementary Methods for the exact content covered) to conduct a literature review to locate studies that have estimated one of the research waste components for ecological literature. We termed these meta-studies. We used the following search string: ((((unpublished OR unsubmitted OR 'non-published' OR 'not-published') NEAR/5 (thesis OR theses OR chapter* OR project* OR research OR studies OR study)) OR ((unpublished OR 'un-reported' OR 'under-reported' OR unsubmitted OR 'not published' OR 'non published' OR 'non reported' OR 'not submitted' OR 'non submitted') NEAR/5 (results OR effects OR effect OR result)) OR 'publication bias' OR 'confirmation bias' OR 'dissemination bias' OR 'small-study' OR 'selective reporting' OR 'incomplete reporting' OR 'biased reporting') OR ('research waste' OR 'wasted research effort' OR 'wasted research' OR 'unutilized research' OR 'non utilized research' OR 'wasted funds' OR 'funding waste' OR 'under-publication' OR 'file-drawer' OR 'low statistical power' OR underpowered OR 'cherry-picking' OR 'biased results' OR 'researcher degrees of freedom' OR 'research degrees of freedom' OR 'research bias' OR 'researcher bias' OR 'confirmation bias' OR 'p-hacking' OR 'observer bias' OR 'QRP' OR 'suboptimal research practices' OR 'sub-optimal research practices' OR 'questionable research practices' OR 'suboptimal research design' OR 'sub-optimal research design' OR 'questionable research design' OR 'suboptimal experimental design' OR 'sub-optimal experimental design' OR 'questionable experimental design')) AND (ecolog* OR evolution* OR biology* OR 'life sciences').

In this way, we obtained 474 studies that were screened independently by three reviewers (M.P, T.K. and A.C.) for eligibility. All the meta-studies deemed relevant after the full screening procedure (12 studies) were subjected to a backward and forward reference check to locate any additional relevant meta-studies. We repeated this until no new relevant meta-study was added to our list (four iterations). In this way, we obtained additional 23 studies. Five meta-studies were included from other sources based on the prior familiarity with the published literature. We excluded six meta-studies that only provided estimates of underpowered research (reasons for this decision can be found in the Supplementary Methods; see the data package for the references of the excluded studies[35]). Further, we excluded one meta-study that provided an indirect estimate of publication bias[15]. More details on the methods can be found in the Supplementary Methods. In this way, we obtained 33 meta-studies[17,18,27–30,32,33,55,75–98] with 43 estimates of research waste components and an overall sample size of 10,464. To each meta-study, we assigned a degree of generality from 1 to 3, depending on its literature coverage. The degree of generality describes whether a meta-study is concerned with a narrow research field within ecology (for example, facultative sex ratio adjustment in birds[17], coded with 1) or a broad area of ecological research (for example, literature from 9 prominent ecological journals[18], coded with 3). The final scores were derived based on scores given by all three reviewers (M.P., T.K., A.C.). Please see the details in the Supplementary Methods.

Meta-analyses. Nine meta-studies estimated the percentage of unpublished literature (either as unpublished project, thesis chapters or percentage of grey literature), based on an overall sample size of 2,252. There were 22 estimates on the study planning stage of research and 9 estimates of result reporting, based on an overall sample size of 7,505 and 2,246, respectively. To obtain the mean estimate of each waste component, we ran a weighted meta-analysis on the published estimates of the corresponding components (publication, study planning, result reporting). We also preformed meta-regressions to obtain mean estimates from the meta-studies (1) with a narrow coverage (degree of generality 1) and those with more general coverage (2 and 3 combined) and (2) for different subcomponents of the study planning stage (that is, core study design, data collection, data analysis). We performed the analysis in the RStudio integrated development environment v.1.4.1106 (ref. [99]) using the package Matafor v.2.4-0 (ref. [100]). Please see the details in the Supplementary Methods.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The data needed to reproduce the analyses and create the main text and supplementary figures have been deposited at Zenodo[35] https://doi.org/10.5281/zenodo.6566100. These include the original effect sizes as extracted from studies and the final set of the effect sizes used in the meta-analysis. Source data are provided with this paper.

Code availability

The codes/scripts needed to reproduce the analyses and create the main text and supplementary figures are deposited at Zenodo[35] https://doi.org/10.5281/zenodo.6566100.

Received: 17 November 2021; Accepted: 31 May 2022;
Published online: 21 July 2022

References

1. Ioannidis, J. P. A., Fanelli, D., Dunne, D. D. & Goodman, S. N. Meta-research: evaluation and improvement of research methods and practices. *PLoS Biol.* **13**, e1002264 (2015).
2. Hampton, S. E. et al. The Tao of open science for ecology. *Ecosphere* **6**, art120 (2015).
3. Rothstein, H. R., Sutton, A. J. & Borenstein, M. *Publication Bias in Meta-Analysis: Prevention, Assessment and Adjustments* (John Wiley & Sons, 2005).
4. Sutton, A. J. in *The Handbook of Research Synthesis and Meta-Analysis* (eds Cooper, H. et al.) 435–452 (Russell Sage Foundation, 2009).
5. Nakagawa, S., Koricheva, J., Macleod, M. & Viechtbauer, W. Introducing our series: research synthesis and meta-research in biology. *BMC Biol.* **18**, 20 (2020).
6. Nakagawa, S. et al. A new ecosystem for evidence synthesis. *Nat. Ecol. Evol.* **4**, 498–501 (2020).
7. Coolidge, H. J. & Lord, R. H. in *Archibald Cary Coolidge: Life and Letters* 308 (Houghton Mifflin Harcourt, 1932).
8. Nickerson, R. S. Confirmation bias: a ubiquitous phenomenon in many guises. *Rev. Gen. Psychol.* **2**, 175–220 (1998).
9. Touchon, J. C. & McCoy, M. W. The mismatch between current statistical practice and doctoral training in ecology. *Ecosphere* **7**, e01394 (2016).
10. Begley, C. G. & Ellis, L. M. Raise standards for preclinical cancer research. *Nature* **483**, 531–533 (2012).
11. Aarts, A. A. et al. Estimating the reproducibility of psychological science. *Science* **349**, aac4716 (2015).
12. Camerer, C. F. et al. Evaluating the replicability of social science experiments in Nature and Science between 2010 and 2015. *Nat. Hum. Behav.* **2**, 637–644 (2018).
13. Fraser, H., Parker, T., Nakagawa, S., Barnett, A. & Fidler, F. Questionable research practices in ecology and evolution. *PLoS ONE* **13**, e0200303 (2018).
14. Culina, A., van den Berg, I., Evans, S. & Sánchez-Tójar, A. Low availability of code in ecology: a call for urgent action. *PLoS Biol.* **18**, e3000763 (2020).
15. Jennions, M. D. & Møller, A. P. Publication bias in ecology and evolution: an empirical assessment using the 'trim and fill' method. *Biol. Rev. Camb. Philos. Soc.* **77**, 211–222 (2002).
16. Jennions, M. D. & Møller, A. P. A survey of the statistical power of research in behavioral ecology and animal behavior. *Behav. Ecol.* **14**, 438–445 (2003).
17. Cassey, P., Ewen, J. G., Blackburn, T. M. & Møller, A. P. A survey of publication bias within evolutionary ecology. *Proc. Biol. Sci.* **271**, S451–S454 (2004).
18. Kardish, M. R. et al. Blind trust in unblinded observation in ecology, evolution, and behavior. *Front. Ecol. Evol.* **3**, 51 (2015).
19. Jennions, M. D. & Møller, A. P. Relationships fade with time: a meta-analysis of temporal trends in publication in ecology and evolution. *Proc. Biol. Sci.* **269**, 43–48 (2002).
20. Baker, M. 1,500 scientists lift the lid on reproducibility. *Nature* **533**, 452–454 (2016).
21. Chalmers, I. & Glasziou, P. Avoidable waste in the production and reporting of research evidence. *Lancet* **374**, 86–89 (2009).
22. Altman, D. G. The scandal of poor medical research. *BMJ* **308**, 283–284 (1994).
23. Glasziou, P. & Chalmers, I. Is 85% of health research really 'wasted'? *BMJ Opinion* (14 January 2016).

24. Glasziou, P. & Chalmers, I. Research waste is still a scandal. *BMJ* **363**, k4645 (2018).
25. Chalmers, I. et al. How to increase value and reduce waste when research priorities are set. *Lancet* **383**, 156–165 (2014).
26. Kunin, W. E. Robust evidence of declines in insect abundance and biodiversity. *Nature* **574**, 641–642 (2019).
27. Christie, A. P. et al. Quantifying and addressing the prevalence and bias of study designs in the environmental and social sciences. *Nat. Commun.* **11**, 6377 (2020).
28. Campbell, H. A. et al. Finding our way: on the sharing and reuse of animal telemetry data in Australasia. *Sci. Total Environ.* **534**, 79–84 (2015).
29. Koricheva, J. Non-significant results in ecology: a burden or a blessing in disguise? *Oikos* **102**, 397–401 (2003).
30. Bennett, L. T. & Adams, M. A. Assessment of ecological effects due to forest harvesting: approaches and statistical issues. *J. Appl. Ecol.* **41**, 585–598 (2004).
31. Duval, S. & Tweedie, R. A nonparametric 'trim and fill' method of accounting for publication bias in meta-analysis. *J. Am. Stat. Assoc.* **95**, 89–98 (2012).
32. Brlík, V. et al. Weak effects of geolocators on small birds: a meta-analysis controlled for phylogeny and publication bias. *J. Anim. Ecol.* **89**, 207–220 (2020).
33. Hurlbert, S. H. Pseudoreplication and the design of ecological field experiments. *Ecol. Monogr.* **54**, 187–211 (1984).
34. Jennions, M. D. & Møller, A. P. A survey of the statistical power of research in behavioral ecology and animal behaviour. *Behav. Ecol.* **14**, 438–445 (2003).
35. Culina, A., Purgar, M. & Klanjscek, T. Datasets and codes for Purgar et al. 2022: quantifying research waste in ecology. *Zenodo* https://zenodo.org/record/6566100#.YrLWB-zMIqs (2022).
36. Ferguson, C. et al. Europe PMC in 2020. *Nucleic Acids Res.* **49**, D1507–D1514 (2021).
37. Huang, C.-K. et al. Meta-Research: Evaluating the impact of open access policies on research institutions. *eLife* **9**, e57067 (2020).
38. Ross-Hellauer, T. Open science, done wrong, will compound inequities. *Nature* **603**, 363 (2022).
39. Smith, A. C. et al. Assessing the effect of article processing charges on the geographic diversity of authors using Elsevier's 'Mirror journal' system. *Quant. Sci. Stud.* **2**, 1123–1143 (2021).
40. Christie, A. P. et al. Reducing publication delay to improve the efficiency and impact of conservation science. *PeerJ* **9**, e12245 (2021).
41. Desjardins-Proulx, P. et al. The case for open preprints in biology. *PLoS Biol.* **11**, e1001563 (2013).
42. Munafò, M. R. et al. A manifesto for reproducible science. *Nat. Hum. Behav.* **1**, 0021 (2017).
43. O'Dea, R. E. et al. Towards open, reliable, and transparent ecology and evolutionary biology. *BMC Biol.* **19**, 68 (2021).
44. Culina, A. et al. Navigating the unfolding open data landscape in ecology and evolution. *Nat. Ecol. Evol.* **2**, 420–426 (2018).
45. Culina, A., Crowther, T. W., Ramakers, J. J. C., Gienapp, P. & Visser, M. E. How to do meta-analysis of open datasets. *Nat. Ecol. Evol.* **2**, 1053–1056 (2018).
46. Roche, D. G., Kruuk, L. E. B., Lanfear, R. & Binning, S. A. Public data archiving in ecology and evolution: how well are we doing? *PLoS Biol.* **13**, e1002295 (2015).
47. Grainger, M. J., Bolam, F. C., Stewart, G. B. & Nilsen, E. B. Evidence synthesis for tackling research waste. *Nat. Ecol. Evol.* **4**, 495–497 (2020).
48. Nørgaard, B. et al. Systematic reviews are rarely used to inform study design—a systematic review and meta-analysis. *J. Clin. Epidemiol.* **145**, 1–13 (2022).
49. Webb, J. A. et al. Weaving common threads in environmental causal assessment methods: toward an ideal method for rapid evidence synthesis. *Freshw. Sci.* **36**, 250–256 (2017).
50. Collins, A., Coughlin, D., Miller, J. & Kirk, S. *The Production of Quick Scoping Reviews and Rapid Evidence Assessments: A How to Guide* (Joint Water Evidence Group, 2015).
51. Carrick, J. et al. Is planting trees the solution to reducing flood risks? *J. Flood Risk Manag.* **12**, e12484 (2019).
52. Nuñez, M. A. & Amano, T. Monolingual searches can limit and bias results in global literature reviews. *Nat. Ecol. Evol.* **5**, 264 (2021).
53. Morrison, A. et al. The effect of English-language restriction on systematic review-based meta-analyses: a systematic review of empirical studies. *Int. J. Technol. Assess. Health Care* **28**, 138–144 (2012).
54. Wu, T., Li, Y., Bian, Z., Liu, G. & Moher, D. Randomized trials published in some Chinese journals: how many are randomized? *Trials* **10**, 46 (2009).
55. Vorobeichik, E. L. & Kozlov, M. V. Impact of point polluters on terrestrial ecosystems: methodology of research, experimental design, and typical errors. *Russ. J. Ecol.* **43**, 89–96 (2012).
56. Simmons, J. P., Nelson, L. D. & Simonsohn, U. False-positive psychology: undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychol. Sci.* **22**, 1359–1366 (2011).
57. *Transforming Our World: the 2030 Agenda for Sustainable Development* (United Nations, 2015).
58. MacCoun, R. & Perlmutter, S. Blind analysis: hide results to seek the truth. *Nature* **526**, 187–189 (2015).
59. Parker, T. H. et al. Transparency in ecology and evolution: real problems, real solutions. *Trends Ecol. Evol.* **31**, 711–719 (2016).
60. Announcement: reducing our irreproducibility. *Nature* **496**, 398 (2013).
61. Moher, D. et al. Increasing value and reducing waste in biomedical research: who's listening? *Lancet* **387**, 1573–1586 (2016).
62. Smaldino, P. E. & McElreath, R. The natural selection of bad science. *R. Soc. Open Sci.* **3**, 160384 (2016).
63. Vrieze, J. Landmark research integrity survey finds questionable practices are surprisingly common. *ScienceInsider* https://www.sciencemag.org/news/2021/07/landmark-research-integrity-survey-finds-questionable-practices-are-surprisingly-common (2021).
64. Woolston, C. Impact factor abandoned by Dutch university in hiring and promotion decisions. *Nature* **595**, 462 (2021).
65. Directorate-General for Research and Innovation (European Commission). *Towards a Reform of the Research Assessment System*. Scoping Report (Publications Office, 2021).
66. Athena Research & Innovation Center, Directorate-General for Research and Innovation (European Commission), PPMI, UNU-MERIT. *Monitoring the Open Access Policy of Horizon 2020*. Final report (European Commission, 2021).
67. Kwon, D. University of California and Elsevier forge open-access deal. *TheScientist* https://www.the-scientist.com/news-opinion/university-of-california-and-elsevier-forge-open-access-deal–68557 (2021).
68. Vines, T. H. et al. Mandated data archiving greatly improves access to research data. *FASEB J.* **27**, 1304–1308 (2013).
69. NPQIP Collaborative Group. Did a change in *Nature* journals' editorial policy for life sciences research improve reporting? *BMJ Open Sci.* **3**, e000035 (2019).
70. Glasziou, P. et al. Reducing waste from incomplete or unusable reports of biomedical research. *Lancet* **383**, 267–276 (2014).
71. Fecher, B. & Friesike, S. in *Opening Science: the Evolving Guide on How the Internet is Changing Research, Collaboration and Scholarly Publishing* (eds Bartling, S. & Friesike, S.) 17–47 (Springer International Publishing, 2014).
72. Hardwicke, T. E. et al. Calibrating the scientific ecosystem through meta-research. *Annu. Rev. Stat. Appl.* **7**, 11–37 (2020).
73. McKiernan, E. C. et al. How open science helps researchers succeed. *eLife* **5**, e16800 (2016).
74. Fidler, F. et al. Metaresearch for evaluating reproducibility in ecology and evolution. *Bioscience* **67**, 282–289 (2017).
75. Cornwall, C. E. & Hurd, C. L. Experimental design in ocean acidification research: problems and solutions. *ICES J. Mar. Sci.* **73**, 572–581 (2016).
76. Fidler, F., Burgman, M. A., Cumming, G., Buttrose, R. & Thomason, N. Impact of criticism of null-hypothesis significance testing on statistical reporting practices in conservation biology. *Conserv. Biol.* **20**, 1539–1544 (2006).
77. Forstmeier, W. & Schielzeth, H. Cryptic multiple hypotheses testing in linear models: overestimated effect sizes and the winner's curse. *Behav. Ecol. Sociobiol.* **65**, 47–55 (2011).
78. Gillespie, B. R., Desmet, S., Kay, P., Tillotson, M. R. & Brown, L. E. A critical analysis of regulated river ecosystem responses to managed environmental flows from reservoirs. *Freshw. Biol.* **60**, 410–425 (2015).
79. Haddaway, N. R., Styles, D. & Pullin, A. S. Evidence on the environmental impacts of farm land abandonment in high altitude/mountain regions: a systematic map. *Environ. Evid.* **3**, 17 (2014).
80. Heffner, R. A., Butler, M. J. & Reilly, C. K. Pseudoreplication revisited. *Ecology* **77**, 2558–2562 (1996).
81. Holman, L., Head, M. L., Lanfear, R. & Jennions, M. D. Evidence of experimental bias in the life sciences: why we need blind data recording. *PLoS Biol.* **13**, e1002190 (2015).
82. Hurlbert, S. H. & White, M. D. Experiments with freshwater invertebrate zooplanktivores: quality of statistical analyses. *Bull. Mar. Sci.* **53**, 128–153 (1993).
83. Johnson, W. T.3rd & Freeberg, T. M. Pseudoreplication in use of predator stimuli in experiments on antipredator responses. *Anim. Behav.* **119**, 161–164 (2016).
84. Kozlov, M. V. Pseudoreplication in ecological research: the problem overlooked by Russian scientists. *Zh. Obshch. Biol.* **64**, 292–307 (2003).
85. Kozlov, M. V. Plant studies on fluctuating asymmetry in Russia: mythology and methodology. *Russ. J. Ecol.* **48**, 1–9 (2017).
86. McDonald, S., Cresswell, T., Hassell, K. & Keough, M. Experimental design and statistical analysis in aquatic live animal radiotracing studies: a systematic review. *Crit. Rev. Environ. Sci. Technol.* **52**, 2772–2801 (2021).
87. Møller, A. P., Thornhill, R. & Gangestad, S. W. Direct and indirect tests for publication bias: asymmetry and sexual selection. *Anim. Behav.* **70**, 497–506 (2005).

88. Mrosovsky, N. & Godfrey, M. H. The path from grey literature to Red Lists. *Endang. Species Res.* **6**, 185–191 (2008).
89. O'Brien, C., van Riper, C.3rd & Myers, D. E. Making reliable decisions in the study of wildlife diseases: using hypothesis tests, statistical power, and observed effects. *J. Wildl. Dis.* **45**, 700–712 (2009).
90. Parker, T. H. What do we really know about the signalling role of plumage colour in blue tits? A case study of impediments to progress in evolutionary biology. *Biol. Rev. Camb. Philos. Soc.* **88**, 511–536 (2013).
91. Ramage, B. S. et al. Pseudoreplication in tropical forests and the resulting effects on biodiversity conservation. *Conserv. Biol.* **27**, 364–372 (2013).
92. Sallabanks, R., Arnett, E. B. & Marzluff, J. M. An evaluation of research on the effects of timber harvest on bird populations. *Wildl. Soc. Bull.* **28**, 1144–1155 (2000).
93. Sánchez-Tójar, A. et al. Meta-analysis challenges a textbook example of status signalling and demonstrates publication bias. *eLife* **7**, e37385 (2018).
94. Waller, B., Warmelink, L., Liebal, K., Micheletta, J. & Slocombe, K. Pseudoreplication: a widespread problem in primate communication research. *Anim. Behav.* **86**, 483–488 (2013).
95. Van Wilgenburg, E. & Elgar, M. A. Confirmation bias in studies of nestmate recognition: a cautionary note for research into the behaviour of animals. *PLoS ONE* **8**, e53548 (2013).
96. Yoccoz, N. G. Use, overuse, and misuse of significance tests in evolutionary biology and ecology. *Bull. Ecol. Soc. Am.* **72**, 106–111 (1991).
97. Zaitsev, A. S., Gongalsky, K. B., Malmström, A., Persson, T. & Bengtsson, J. Why are forest fires generally neglected in soil fauna research? A mini-review. *Appl. Soil Ecol.* **98**, 261–271 (2016).
98. Zvereva, E. L. & Kozlov, M. V. Biases in studies of spatial patterns in insect herbivory. *Ecol. Monogr.* **89**, e01361 (2019).
99. RStudio Team. *RStudio: Integrated Development for R* (RStudio, 2020).
100. Viechtbauer, W. Conducting meta-analyses in R with the metafor package. *J. Stat. Softw.* **36**, 1–48 (2010).

Acknowledgements
We acknowledge funding from the Croatian Science Foundation project IP-2018-01-3150-AqADAPT that supported T.K. and M.P.

Author contributions
A.C. conceived the study and wrote the manuscript draft. A.C. and M.P. analysed the data. M.P., T.K. and A.C. designed the analysis, contributed to data collection, interpretation of the data and the manuscript revisions.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41559-022-01820-0.

Correspondence and requests for materials should be addressed to Antica Culina.

Peer review information *Nature Ecology & Evolution* thanks Matthew Grainger and Alec Christie for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

nature portfolio

Corresponding author(s): Antica Culina

Last updated by author(s): 23/05/2022

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☒	☐	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Data were collected via Web of Science search platform
Data analysis	Data analysis was done in R Studio Integrated Development Environment, Version 1.4.1106.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

The data needed to reproduce the analysis are deposited at Zenodo. https://doi.org/10.5281/zenodo.6566100. These include the original effect sizes as extracted from studies and the final set of the effect sizes used in the meta-analysis.

Human research participants

Policy information about studies involving human research participants and Sex and Gender in Research.

Reporting on sex and gender	*Use the terms sex (biological attribute) and gender (shaped by social and cultural circumstances) carefully in order to avoid confusing both terms. Indicate if findings apply to only one sex or gender; describe whether sex and gender were considered in study design whether sex and/or gender was determined based on self-reporting or assigned and methods used. Provide in the source data disaggregated sex and gender data where this information has been collected, and consent has been obtained for sharing of individual-level data; provide overall numbers in this Reporting Summary. Please state if this information has not been collected. Report sex- and gender-based analyses where performed, justify reasons for lack of sex- and gender-based analysis.*
Population characteristics	*Describe the covariate-relevant population characteristics of the human research participants (e.g. age, genotypic information, past and current diagnosis and treatment categories). If you filled out the behavioural & social sciences study design questions and have nothing to add here, write "See above."*
Recruitment	*Describe how participants were recruited. Outline any potential self-selection bias or other biases that may be present and how these are likely to impact results.*
Ethics oversight	*Identify the organization(s) that approved the study protocol.*

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☐ Behavioural & social sciences ☒ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	*Describe how sample size was determined, detailing any statistical methods used to predetermine sample size OR if no sample-size calculation was performed, describe how sample sizes were chosen and provide a rationale for why these sample sizes are sufficient.*
Data exclusions	*Describe any data exclusions. If no data were excluded from the analyses, state so OR if data were excluded, describe the exclusions and the rationale behind them, indicating whether exclusion criteria were pre-established.*
Replication	*Describe the measures taken to verify the reproducibility of the experimental findings. If all attempts at replication were successful, confirm this OR if there are any findings that were not replicated or cannot be reproduced, note this and describe why.*
Randomization	*Describe how samples/organisms/participants were allocated into experimental groups. If allocation was not random, describe how covariates were controlled OR if this is not relevant to your study, explain why.*
Blinding	*Describe whether the investigators were blinded to group allocation during data collection and/or analysis. If blinding was not possible, describe why OR explain why blinding was not relevant to your study.*

Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	*Briefly describe the study type including whether data are quantitative, qualitative, or mixed-methods (e.g. qualitative cross-sectional, quantitative experimental, mixed-methods case study).*
Research sample	*State the research sample (e.g. Harvard university undergraduates, villagers in rural India) and provide relevant demographic information (e.g. age, sex) and indicate whether the sample is representative. Provide a rationale for the study sample chosen. For studies involving existing datasets, please describe the dataset and source.*
Sampling strategy	*Describe the sampling procedure (e.g. random, snowball, stratified, convenience). Describe the statistical methods that were used to predetermine sample size OR if no sample-size calculation was performed, describe how sample sizes were chosen and provide a rationale for why these sample sizes are sufficient. For qualitative data, please indicate whether data saturation was considered, and what criteria were used to decide that no further sampling was needed.*

Data collection	*Provide details about the data collection procedure, including the instruments or devices used to record the data (e.g. pen and paper, computer, eye tracker, video or audio equipment) whether anyone was present besides the participant(s) and the researcher, and whether the researcher was blind to experimental condition and/or the study hypothesis during data collection.*
Timing	*Indicate the start and stop dates of data collection. If there is a gap between collection periods, state the dates for each sample cohort.*
Data exclusions	*If no data were excluded from the analyses, state so OR if data were excluded, provide the exact number of exclusions and the rationale behind them, indicating whether exclusion criteria were pre-established.*
Non-participation	*State how many participants dropped out/declined participation and the reason(s) given OR provide response rate OR state that no participants dropped out/declined participation.*
Randomization	*If participants were not allocated into experimental groups, state so OR describe how participants were allocated to groups, and if allocation was not random, describe how covariates were controlled.*

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	We conducted systematic review and meta-analysis of effects from published studies.
Research sample	We were targeting studies that have estimated one of the research waste components for ecological literature.
Sampling strategy	Literature review
Data collection	In May 2021, we used WoS Core Collection databases to search for relevant studies
Timing and spatial scale	Relevant literature published up to May 2021
Data exclusions	We excluded effect sizes that estimate power in ecology. We decided to do this after additinal discussion on the implicatin of low power in ecology.
Reproducibility	Coding of studies done by 3 reviewers
Randomization	NA
Blinding	Reviewers were blinded to each others decisions

Did the study involve field work? ☐ Yes ☒ No

Field work, collection and transport

Field conditions	*Describe the study conditions for field work, providing relevant parameters (e.g. temperature, rainfall).*
Location	*State the location of the sampling or experiment, providing relevant parameters (e.g. latitude and longitude, elevation, water depth).*
Access & import/export	*Describe the efforts you have made to access habitats and to collect and import/export your samples in a responsible manner and in compliance with local, national and international laws, noting any permits that were obtained (give the name of the issuing authority, the date of issue, and any identifying information).*
Disturbance	*Describe any disturbance caused by the study and how it was minimized.*

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems		Methods	
n/a	Involved in the study	n/a	Involved in the study
☐	☐ Antibodies	☐	☐ ChIP-seq
☐	☐ Eukaryotic cell lines	☐	☐ Flow cytometry
☐	☐ Palaeontology and archaeology	☐	☐ MRI-based neuroimaging
☐	☐ Animals and other organisms		
☐	☐ Clinical data		
☐	☐ Dual use research of concern		

Antibodies

Antibodies used — *Describe all antibodies used in the study; as applicable, provide supplier name, catalog number, clone name, and lot number.*

Validation — *Describe the validation of each primary antibody for the species and application, noting any validation statements on the manufacturer's website, relevant citations, antibody profiles in online databases, or data provided in the manuscript.*

Eukaryotic cell lines

Policy information about cell lines and Sex and Gender in Research

Cell line source(s) — *State the source of each cell line used and the sex of all primary cell lines and cells derived from human participants or vertebrate models.*

Authentication — *Describe the authentication procedures for each cell line used OR declare that none of the cell lines used were authenticated.*

Mycoplasma contamination — *Confirm that all cell lines tested negative for mycoplasma contamination OR describe the results of the testing for mycoplasma contamination OR declare that the cell lines were not tested for mycoplasma contamination.*

Commonly misidentified lines (See ICLAC register) — *Name any commonly misidentified cell lines used in the study and provide a rationale for their use.*

Palaeontology and Archaeology

Specimen provenance — *Provide provenance information for specimens and describe permits that were obtained for the work (including the name of the issuing authority, the date of issue, and any identifying information). Permits should encompass collection and, where applicable, export.*

Specimen deposition — *Indicate where the specimens have been deposited to permit free access by other researchers.*

Dating methods — *If new dates are provided, describe how they were obtained (e.g. collection, storage, sample pretreatment and measurement), where they were obtained (i.e. lab name), the calibration program and the protocol for quality assurance OR state that no new dates are provided.*

☐ Tick this box to confirm that the raw and calibrated dates are available in the paper or in Supplementary Information.

Ethics oversight — *Identify the organization(s) that approved or provided guidance on the study protocol, OR state that no ethical approval or guidance was required and explain why not.*

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Animals and other research organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research, and Sex and Gender in Research

Laboratory animals — *For laboratory animals, report species, strain and age OR state that the study did not involve laboratory animals.*

Wild animals — *Provide details on animals observed in or captured in the field; report species and age where possible. Describe how animals were caught and transported and what happened to captive animals after the study (if killed, explain why and describe method; if released, say where and when) OR state that the study did not involve wild animals.*

Reporting on sex — *Indicate if findings apply to only one sex; describe whether sex was considered in study design, methods used for assigning sex. Provide data disaggregated for sex where this information has been collected in the source data as appropriate; provide overall numbers in this Reporting Summary. Please state if this information has not been collected. Report sex-based analyses where performed, justify reasons for lack of sex-based analysis.*

Field-collected samples	*For laboratory work with field-collected samples, describe all relevant parameters such as housing, maintenance, temperature, photoperiod and end-of-experiment protocol OR state that the study did not involve samples collected from the field.*
Ethics oversight	*Identify the organization(s) that approved or provided guidance on the study protocol, OR state that no ethical approval or guidance was required and explain why not.*

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Clinical data

Policy information about clinical studies

All manuscripts should comply with the ICMJE guidelines for publication of clinical research and a completed CONSORT checklist must be included with all submissions.

Clinical trial registration	*Provide the trial registration number from ClinicalTrials.gov or an equivalent agency.*
Study protocol	*Note where the full trial protocol can be accessed OR if not available, explain why.*
Data collection	*Describe the settings and locales of data collection, noting the time periods of recruitment and data collection.*
Outcomes	*Describe how you pre-defined primary and secondary outcome measures and how you assessed these measures.*

Dual use research of concern

Policy information about dual use research of concern

Hazards

Could the accidental, deliberate or reckless misuse of agents or technologies generated in the work, or the application of information presented in the manuscript, pose a threat to:

No	Yes	
☐	☐	Public health
☐	☐	National security
☐	☐	Crops and/or livestock
☐	☐	Ecosystems
☐	☐	Any other significant area

Experiments of concern

Does the work involve any of these experiments of concern:

No	Yes	
☐	☐	Demonstrate how to render a vaccine ineffective
☐	☐	Confer resistance to therapeutically useful antibiotics or antiviral agents
☐	☐	Enhance the virulence of a pathogen or render a nonpathogen virulent
☐	☐	Increase transmissibility of a pathogen
☐	☐	Alter the host range of a pathogen
☐	☐	Enable evasion of diagnostic/detection modalities
☐	☐	Enable the weaponization of a biological agent or toxin
☐	☐	Any other potentially harmful combination of experiments and agents

ChIP-seq

Data deposition

☐ Confirm that both raw and final processed data have been deposited in a public database such as GEO.

☐ Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

Data access links *May remain private before publication.*	*For "Initial submission" or "Revised version" documents, provide reviewer access links. For your "Final submission" document, provide a link to the deposited data.*
Files in database submission	*Provide a list of all files available in the database submission.*
Genome browser session (e.g. UCSC)	

Provide a link to an anonymized genome browser session for "Initial submission" and "Revised version" documents only, to enable peer review. Write "no longer applicable" for "Final submission" documents.

Methodology

Replicates	Describe the experimental replicates, specifying number, type and replicate agreement.
Sequencing depth	Describe the sequencing depth for each experiment, providing the total number of reads, uniquely mapped reads, length of reads and whether they were paired- or single-end.
Antibodies	Describe the antibodies used for the ChIP-seq experiments; as applicable, provide supplier name, catalog number, clone name, and lot number.
Peak calling parameters	Specify the command line program and parameters used for read mapping and peak calling, including the ChIP, control and index files used.
Data quality	Describe the methods used to ensure data quality in full detail, including how many peaks are at FDR 5% and above 5-fold enrichment.
Software	Describe the software used to collect and analyze the ChIP-seq data. For custom code that has been deposited into a community repository, provide accession details.

Flow Cytometry

Plots

Confirm that:

☐ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☐ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☐ All plots are contour plots with outliers or pseudocolor plots.

☐ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	Describe the sample preparation, detailing the biological source of the cells and any tissue processing steps used.
Instrument	Identify the instrument used for data collection, specifying make and model number.
Software	Describe the software used to collect and analyze the flow cytometry data. For custom code that has been deposited into a community repository, provide accession details.
Cell population abundance	Describe the abundance of the relevant cell populations within post-sort fractions, providing details on the purity of the samples and how it was determined.
Gating strategy	Describe the gating strategy used for all relevant experiments, specifying the preliminary FSC/SSC gates of the starting cell population, indicating where boundaries between "positive" and "negative" staining cell populations are defined.

☐ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.

Magnetic resonance imaging

Experimental design

Design type	Indicate task or resting state; event-related or block design.
Design specifications	Specify the number of blocks, trials or experimental units per session and/or subject, and specify the length of each trial or block (if trials are blocked) and interval between trials.
Behavioral performance measures	State number and/or type of variables recorded (e.g. correct button press, response time) and what statistics were used to establish that the subjects were performing the task as expected (e.g. mean, range, and/or standard deviation across subjects).

Acquisition

Imaging type(s) — *Specify: functional, structural, diffusion, perfusion.*

Field strength — *Specify in Tesla*

Sequence & imaging parameters — *Specify the pulse sequence type (gradient echo, spin echo, etc.), imaging type (EPI, spiral, etc.), field of view, matrix size, slice thickness, orientation and TE/TR/flip angle.*

Area of acquisition — *State whether a whole brain scan was used OR define the area of acquisition, describing how the region was determined.*

Diffusion MRI ☐ Used ☐ Not used

Preprocessing

Preprocessing software — *Provide detail on software version and revision number and on specific parameters (model/functions, brain extraction, segmentation, smoothing kernel size, etc.).*

Normalization — *If data were normalized/standardized, describe the approach(es): specify linear or non-linear and define image types used for transformation OR indicate that data were not normalized and explain rationale for lack of normalization.*

Normalization template — *Describe the template used for normalization/transformation, specifying subject space or group standardized space (e.g. original Talairach, MNI305, ICBM152) OR indicate that the data were not normalized.*

Noise and artifact removal — *Describe your procedure(s) for artifact and structured noise removal, specifying motion parameters, tissue signals and physiological signals (heart rate, respiration).*

Volume censoring — *Define your software and/or method and criteria for volume censoring, and state the extent of such censoring.*

Statistical modeling & inference

Model type and settings — *Specify type (mass univariate, multivariate, RSA, predictive, etc.) and describe essential details of the model at the first and second levels (e.g. fixed, random or mixed effects; drift or auto-correlation).*

Effect(s) tested — *Define precise effect in terms of the task or stimulus conditions instead of psychological concepts and indicate whether ANOVA or factorial designs were used.*

Specify type of analysis: ☐ Whole brain ☐ ROI-based ☐ Both

Statistic type for inference (See Eklund et al. 2016) — *Specify voxel-wise or cluster-wise and report all relevant parameters for cluster-wise methods.*

Correction — *Describe the type of correction and how it is obtained for multiple comparisons (e.g. FWE, FDR, permutation or Monte Carlo).*

Models & analysis

n/a	Involved in the study
☐	☐ Functional and/or effective connectivity
☐	☐ Graph analysis
☐	☐ Multivariate modeling or predictive analysis

Functional and/or effective connectivity — *Report the measures of dependence used and the model details (e.g. Pearson correlation, partial correlation, mutual information).*

Graph analysis — *Report the dependent variable and connectivity measure, specifying weighted graph or binarized graph, subject- or group-level, and the global and/or node summaries used (e.g. clustering coefficient, efficiency, etc.).*

Multivariate modeling and predictive analysis — *Specify independent variables, features extraction and dimension reduction, model, training and evaluation metrics.*

3. DISCUSSION

This discussion chapter synthesizes key findings from the three peer-reviewed publications that comprise this doctoral dissertation. The research was centered around *Vibrio* spp. abundance, specifically, predictive modeling and statistical analysis, to enhance the understanding of applicability of existing models from the literature to predict *Vibrio* spp. abundance near mariculture and to assess its possible use as an additional indicator of marine water quality near mariculture. The dissertation also explored the informative value of ecological research, i.e., the extent of available information generated by ecological studies which can be reused. High informative value of research enables researchers, policymakers, and practitioners to reuse existing data and conclusions in designing more effective strategies for managing marine ecosystems, mitigating the impacts of climate change, and optimizing mariculture practices. Conversely, when studies remain unpublished or suffer from poor design and incomplete reporting, they contribute to research waste, a pervasive issue previously quantified primarily in medical sciences (Chalmers et al., 2014; Chalmers & Glasziou, 2009; Glasziou & Chalmers, 2016).

The rest of this chapter examines each hypothesis in turn, placing the findings in the context of previous research and discussing their implications for future studies.

3.1. *(H1) Existing models of Vibrio bacteria growth can be used to predict the abundance of Vibrio spp. in mariculture.*

Hypothesis H1 posited that existing *Vibrio* growth models from literature, many of which were originally developed under controlled laboratory conditions, can be applied to predict *Vibrio* spp. abundance in natural mariculture environments. The findings from *Publication I* (Purgar et al., 2022a) provide partial support for this hypothesis. While several models, particularly those based on the Baranyi framework, showed moderate predictive performance in specific datasets, none demonstrated consistently reliable predictions across all habitat types tested.

The study tested 28 standardized growth models on seven open datasets from diverse marine environments, including marine aquaculture, urban estuary, estuary, and coastal area. Predictive accuracy varied considerably depending on habitat conditions and the inclusion of key environmental drivers. For example, Baranyi models that incorporated both temperature and salinity performed better than those

with temperature alone. However, their overall performance was limited in ecosystems affected by strong anthropogenic pressures, such as around mariculture facilities and urban estuaries, where organic matter concentrations are elevated.

Elevated inputs of organic matter stem from anthropogenic activities such as fish feeding and feces, and terrestrial runoff. *Vibrio* spp., as prototypical copiotrophs, are adapted to thrive in nutrient-rich environments (Takemura et al., 2014; Thompson & Polz, 2006) and exhibit a feast-and-famine lifestyle and actively swim toward nutrient-enriched microzones, including organic particles and detritus (Azam & Malfatti, 2007; Stocker, 2012). Several studies have shown that *Vibrio* populations increase with rising concentrations of dissolved organic matter, particularly in coastal environments (Bullington et al., 2022; Eiler et al., 2007). Thus, the exclusion of organic matter from existing models likely undermines their predictive capacity, especially in habitats with fluctuating and elevated nutrient loads. This limitation aligns with previous discussions on factors that affect *Vibrio* spp. abundance beyond temperature and salinity (Brumfield et al., 2023; Oberbeckmann et al., 2012).

Future research should prioritize the development of more integrative and real-world data driven growth models. In particular, secondary models should be expanded to incorporate a broader and ecologically realistic range of environmental variables, especially temperature, salinity, and organic matter. Given the well-documented influence of organic matter on *Vibrio* proliferation, future field observations and datasets should include measurements of dissolved organic matter and related proxies. Mechanistic models may also benefit from synergy with machine learning techniques, which could offer improved adaptability and predictive accuracy in real-world conditions such as marine environments.

The study has several limitations. First, all *Vibrio* spp. were treated as a single group, despite known interspecies differences in environmental preferences and dynamics. While this simplification was appropriate for the study's broad objective, it limits the ability to detect species-specific patterns, especially under changing environmental conditions. Second, a single simulation time was applied to each dataset. Although varying simulation time per data point might improve fit, it would undermine the modeling objective by introducing overfitting. Exploring how simulation time could functionally depend on environmental variables (e.g., temperature) may offer value but requires further investigation. Third, the study focused exclusively on published models with well-defined functional forms. While this allowed for systematic

evaluation and application to both new and existing datasets, alternative parameterizations or combinations of models may offer improved predictive power. This study represents an initial step toward such efforts by organizing and standardizing available primary and secondary models. Alternatively, re-fitting statistical models may offer a better fit to the datasets and could be particularly useful in cases where the goal is to interpolate or extrapolate data within a specific geographic area.

Despite these limitations, the study offers a valuable foundation for adapting *Vibrio* growth models to mariculture settings. It highlights the importance of incorporating key environmental variables, especially organic matter, and demonstrates the challenges of applying laboratory-calibrated models to complex, real-world ecosystems. All datasets and code used in *Publication I* are openly available on Zenodo repository (Purgar et al., 2022c), ensuring the reproducibility of the analyses and allowing other researchers to further adapt, refine, or extend the evaluated models.

3.2. *(H2) Vibrio spp. abundance has significant potential to be included in the regular set of indicators for assessing the water quality in mariculture*

Hypothesis H2 proposed that indicators of *Vibrio* spp. abundance can serve as a meaningful supplementary indicator for water quality assessment in mariculture settings. Findings from *Publication II* (Purgar et al., 2023) lend support to this hypothesis by showing that *Vibrio* spp. abundance provides distinct and potentially earlier signals of microbial risk compared to traditional fecal indicators.

In colder months, *Vibrio* spp. abundance indicated microbial risk (i.e., poor water quality) when conventional fecal indicators signaled excellent or good water quality. Notably, *Vibrio* spp. more often exceeded 100 CFU/mL during winter months and across depths, a conservative threshold associated with risk of bacterial transmission to marine organisms following prolonged exposure (Kim & Lee, 2017). Although classified as ‘sufficient’ water quality, this level can still cause harm and may more appropriately fall under a ‘poor’ category. Since thresholds were based on enterococci, a scale tailored specifically to *Vibrio* spp. would likely be stricter. Furthermore, existing fecal indicators such as *E. coli* and coliforms, routinely used to assess water quality, do not correlate with non-fecal *Vibrio* spp. and thus fail to signal microbial risk

accurately (Ryder et al., 2014). These findings underscore *Vibrio* spp.'s potential as an early warning signal of microbial threat.

Previous studies have consistently shown that *Vibrio* spp. abundance typically increases with temperature (Sheikh et al., 2022; Froelich et al., 2019; Takemura et al., 2014; Pruzzo et al., 2005). The atypical *Vibrio* spp. abundance seasonal pattern observed in *Publication II*, where *Vibrio* spp. abundance was greater in colder months, is unlikely to be explained by antibiotic use at the farm, nor by variations in particulate organic matter (POM), which remained stable across seasons. However, total dissolved solids (TDS) were elevated during the cold season, suggesting that the organic component of TDS may have supported *Vibrio* proliferation by serving as a potential nutrient source (Johnson et al., 2012; Thickman & Gobler, 2017). This finding supports conclusions from *Publication I*, where organic matter emerged as a likely explanatory variable for poor model performance in aquaculture sites. Together, these results indicate that organic nutrient availability may exert a greater influence on *Vibrio* dynamics than temperature alone and should be incorporated into both monitoring programs and predictive models.

In the context of mariculture, monitoring *Vibrio* spp. remains crucial, even in the absence of formal regulatory requirements. Long-term surveillance can help identify drivers of anomalous abundance patterns, such as sediment resuspension or nutrient influx, enabling aquaculture operators to mitigate microbial risks proactively. Early detection reduces the likelihood of vibriosis-related losses, protecting both industry and consumers. Model-based predictions, if properly validated with monitoring data, could help regulatory agencies set evidence-based limits for *Vibrio* presence in coastal waters. These considerations are especially urgent in the Mediterranean, where intensive aquaculture, coastal tourism, and high local seafood consumption increase the importance of managing microbial water quality. In regions such as the Adriatic Sea, where vibriosis cases have been reported as early as spring (Veić, 2016; Zupičić et al., 2022), targeted microbial surveillance could strengthen early warning systems. Adapting these systems to reflect regional environmental dynamics would benefit from integration with model-based forecasting approaches, such as those explored in *Publication I*, particularly as models are refined to more accurately capture in situ variability.

Beyond their role in aquaculture, *Vibrio* spp. also present a public health concern. Despite the recognition as established human pathogens *Vibrio* spp.,

especially strains such as *V. parahaemolyticus*, *V. alginolyticus*, and *V. harveyi*, (Baker-Austin et al., 2018; Froelich et al., 2019; Ina-Salwany et al., 2019), *Vibrio* spp. are not monitored in most European countries. Countries such as the United States and Australia have implemented food safety regulations requiring *Vibrio* surveillance in seafood (Center for Food Safety and Applied Nutrition, 2011; Food Standards Australia New Zealand, 2022). The U.S. Centers for Disease Control and Prevention (CDC) operates the Cholera and Other *Vibrio* Illness Surveillance (COVIS) system, which tracks pathogenic *Vibrio* species, infection types, and geographic trends (Levy, 2018). Meanwhile, in the EU, *Vibrio* infections are underreported due to the absence of mandatory surveillance and reporting systems, with testing limited to travel-associated diarrhea to exclude *Vibrio cholerae* (Semenza et al., 2017). Closing this data gap would strengthen food safety, enable timely public health responses, and support the development of region-specific early warning systems.

The main limitation of the given study is primarily related to data scarcity (Gorgoglione et al., 2020). The analysis relied on the available dataset (Jug Dujaković et al., 2022) collected at two stations (fish farm and control site) in Mali Ston Bay, a coastal area of significant economic and recreational value. Hence, the results may be limited to a specific region even though some studies suggest a broader applicability (Levy, 2018; Robertson, 1984). The dataset covers a limited spatial range, which may affect generalizability. As such, broader sampling and cross-regional comparisons are needed to validate *Vibrio* spp. as a water quality indicator under diverse environmental conditions. The study used genus-level identification of *Vibrio* spp., which does not distinguish between pathogenic and non-pathogenic strains. Since only certain species pose health and ecological risks, species-level resolution is essential for regulatory use (Brumfield et al., 2021). Additionally, the absence of established threshold values for *Vibrio* spp. limits direct interpretation of results within current legislative frameworks, although the study conservatively applied thresholds designed for enterococci.

Generally, traditional indicators, which are primarily designed to assess fecal contamination, are inadequate for estimating vibriosis risk or detecting *Vibrio*-related microbial threats. Therefore, the findings support the hypothesis (H2) that *Vibrio* spp. abundance holds promise as a supplementary water quality indicator. However, further research is necessary to establish species-specific thresholds and to differentiate pathogenic from non-pathogenic strains for regulatory applications.

3.3. *(H3) The informative value of ecological research is similar to that estimated in medicine (about 15%)*

Publication III provided the first quantitative estimate of the informative value in ecology by synthesizing findings from 33 meta-research studies comprising 43 individual waste estimates. The results of meta-analysis showed that 44.7% of ecological studies remain unpublished (95% CI: 44.2-46.7%), 67.4% (95% CI: 66.3-68.4%) have methodological design flaws, and 40.7% (95% CI: 38.7-42.8%) incompletely report key results such as sample sizes, uncertainty measures, or effect sizes. Overall, only 11-18% of ecological research reaches its informative value, which is consistent with the estimate of 15% from medicine (Chalmers & Glasziou, 2009) and confirms H3. Low estimates of the informative value of ecological research, reduce potential for informing new studies, and practices such as evidence-based coastal management.

The informative value of research was estimated for broader field of ecology, rather than subfields such as, mariculture or aquatic ecology, for two main reasons. First, there is a limited number of available meta-research studies that quantify components of research waste across the research cycle within ecological subfields, which would constrain the robustness of subfield-level estimates. Second, scientific disciplines, and their subfields, often operate within similar structural frameworks, including shared research norms, evaluation criteria, and incentive systems. Therefore, it is reasonable to expect similar patterns of informative value across related ecological sub-fields. For example, comparable challenges were observed in the field of medicine (Chalmers & Glasziou, 2009). Nevertheless, future research should aim to disaggregate analyses by subfields to determine whether important differences exist and whether tailored, subfield-specific solutions are needed to effectively address inefficiencies in research cycle and improve scientific practice.

Several additional factors, though not quantified in *Publication III*, further undermine the informative value of ecological research. These include limited accessibility of publications, lack of open data, methods, and code, and insufficient engagement with prior research. As shown in *Publication III*, based on Europe PMC data from 94 ecological journals, 73.0% of articles published between 2014 and 2021 were open access, likely reflecting the growing influence of funder mandates and open science policies (Huang et al., 2020). Although, open access increases visibility,

promotes broader engagement, and enhances the potential for discovery and error detection, it also introduces inequities: high open access publication fees remain a barrier for many researchers, particularly those in low-income countries (Ross-Hellauer, 2022). These elements are essential for understanding how results were derived and for ensuring reproducibility and reuse of the existing data and code, for e.g. meta-analyses. Yet, many ecological studies fail to provide these components. For example, even in journals with code-sharing policies, only 27% of articles include code, and just 21% are computationally reproducible, often due to poor documentation or missing metadata (Culina et al., 2020). Finally, many ecological studies fail to build effectively on previous work by, for example, conducting systematic literature reviews before starting new research (Grainger et al., 2020). Any new research that is informed by past studies have the potential to reduce research waste (Chalmers et al., 2014; Grainger et al., 2020).

Publication III had several limitations. First, like most literature reviews, it focused only on English-language publications, which may not capture informative value in non-English research. Limited evidence suggests differences may exist; for example, one meta-study (Vorobeichik & Kozlov, 2012) found better result reporting in English than in Russian studies (68% versus 28% of results well reported, respectively). Second, it could not capture and analyze time trends, as most meta-studies spanned broad periods and did not report changes in research waste components over time. Finally, we found no empirical estimates quantifying the prevalence of certain questionable research practices, such as optional stopping or selective reporting, in the reviewed literature. While one survey-based study (Fraser et al., 2018) revealed that 42% of ecologists had collected more data after checking for statistical significance, and 4.5% admitted to fabricating data, systematic, evidence-based assessments of these practices in ecology are still lacking. These limitations highlight the need for broader community engagement to refine estimates of research waste and work on practices to improve informative value of research across ecological sub-fields.

Collectively, the findings from *Publication III* confirm H3 by demonstrating that the informative value of ecological research (11-18%) is indeed similar to that estimates from medicine (15%). Addressing this challenge calls for systemic reforms that promote methodological rigor, transparency, and inclusive publication practices. To truly harness the potential of ecological research, the field must shift toward open and

reproducible science that values the integrity and accessibility of all well-conducted studies, regardless of outcome (e.g., promote publication of non-significant research findings).

The findings from *Publication III* also provide broader context for understanding the challenges identified in *Publications I* and *II* related to *Vibrio* spp. abundance modeling and inclusion in water quality monitoring. For example, the limited predictive accuracy of existing *Vibrio* growth models (H1) and the atypical seasonal patterns observed in *Vibrio* abundance compared to conventional indicators (H2) may partly stem from the low informative value of ecological research. Methodological flaws, incomplete reporting, and limited access to underlying data of existing studies hinder model calibration, validation, and development, while also limiting the ability to explore datasets that could inform early warning signs of microbial threats in coastal environments and development of new indicators and setting threshold values. Thus, H3 highlights that advancing mariculture and evidence-based coastal management could benefit from greater availability, transparency, and methodological rigor in ecological research to effectively build on existing findings.

3.3.1. Pathways to increase the informative value and reusability of ecological research

Reducing avoidable research waste and maximizing the informative value of ecological research requires systemic changes, not only in incentives and mandates, but also in everyday research practices, training, and evaluation standards. Achieving this transformation requires collective responsibility and coordinated action from all key stakeholders, including funders, institutions, publishers, journals, and individual researchers.

Funders and institutions must move beyond quantity-based metrics (e.g., publication counts or journal impact factors) and should set up a reward system that focuses on methodological quality and reusability of research findings (Calster et al., 2021; Moher et al., 2018). This shift is particularly important in applied ecology, where studies often inform development of new models, conservation practices, environmental policies. Notably, Utrecht University has eliminated the use of journal impact factor in hiring and promotion decisions, serving as a model for progressive reform (Nakagawa et al., 2020). Funders should also balance research funding in all

types of research, including replication research (Bierer et al., 2018) and meta-research (Hardwicke et al., 2020). Replication research is essential for verifying novel claims and strengthening the evidence base, while meta-research, often described as *research on research*, examines the efficiency, quality, and potential biases within the scientific enterprise, and proposes evidence-based solutions to improve it (Hardwicke et al., 2020; Ioannidis et al., 2015).

To translate these systemic shifts into everyday research practice, funders and academic institutions must go beyond reforming evaluation criteria and actively invest in the infrastructure, training, and personnel needed to support robust and transparent science as described in *Publication III* (Purgar et al., 2022b). This includes integrating transparent research methods into academic training and student curricula (Glasziou et al., 2014; Moher et al., 2016; Touchon & McCoy, 2016), involving statisticians and data stewards in research projects through targeted funding or advisory boards (Glasziou et al., 2014; Moher et al., 2016), and building infrastructure for open science workflows such as preregistration, transparent reporting, and long-term archiving (Glasziou et al., 2014).

Improving the peer review process is another opportunity to enhance informative value of ecological research. Findings from *Publication III* suggest that peer review, in its current form, may not be effectively fulfilling its intended role in ensuring methodological rigor. Despite undergoing peer review, almost 70% of published studies exhibited poor study design that the review process is expected to detect. This indicates that current peer review practices may be insufficient for filtering out flawed research. At the same time, peer review process is often under-resourced, lacks transparency, and rarely incorporates specialist assessments of statistics, data availability, or study design (Bendiscioli, 2019; Tennant et al., 2017). Journals and publishers should try to assemble a cross-disciplinary review teams, including statisticians, data curators, and methodologists, to ensure more rigorous and comprehensive evaluations (Calster et al., 2021). In some journals, such as *The Royal Society* (*Data Sharing and Mining | Royal Society*, 2025), and *Behavioural Ecology and Sociobiology* (Bakker & Traniello, 2020) data and code are requested for review at the article submission stage. Another promising example is *Publons*, a platform designed to recognize and reward peer reviewers (Teixeira da Silva & Nazarovets, 2022). Integrated with systems such as Web of Science and ORCID, *Publons*

facilitates transparent and traceable review contributions, connecting researchers and publishers through a more accountable review process.

Implementing open science practices offers a tangible strategy for reducing knowledge loss in ecological research. Open science refers to a set of principles and practices that promote transparency, accessibility and reproducibility of scientific research, by making publications, data, code, and other resources freely available to the public (Bertram et al., 2023; Maedche et al., 2024). It emphasizes collaborative approaches, early sharing, and community engagement to accelerate knowledge transfer and reduce inefficiencies across the research lifecycle (Bertram et al., 2023; Besançon et al., 2021). Platforms such as the Open Science Framework (OSF) facilitate this process by allowing researchers to preregister study protocols, share code and data, and publicly archive entire research workflows, practices that have been shown to improve research credibility and reuse (Nosek et al., 2018; Kidwell et al., 2016; Nuzzo, 2015).

Open science also plays a transformative role in democratizing access to ecological knowledge. Publishing preprints (Berg et al., 2016; N. Fraser et al., 2021; Noble et al., 2024) and sharing open datasets can help reduce global inequities in access to research outputs, especially for researchers in underfunded or resource-constrained settings (Baker, 2023; Chan et al., 2009; Petersen, 2021). Making data openly available under the FAIR principles (Findable, Accessible, Interoperable, and Reusable) (Wilkinson et al., 2016), can foster collaborative projects, facilitate data synthesis, and enable secondary analyses that extend the value of the original work (Culina et al., 2018). Although legal, ethical, or national constraints may restrict full data openness in some cases, most ecological data can be shared responsibly when appropriate licenses, data anonymization, and proper crediting are in place (Culina et al., 2018).

As climate change accelerates and global demand for aquatic food continues to grow, the production of robust, transparent, and fully informative research becomes not only a scientific priority but a societal necessity. High-quality, reusable evidence is critical for guiding policy, reducing uncertainty in environmental decision-making, and ensuring that mariculture practices are both sustainable and resilient. Advancing sustainable and healthy mariculture is a key step toward achieving Sustainable Development Goal 14 (Life Below Water), which promotes the conservation and responsible use of oceans, seas, and marine resources (Stead, 2019; Troell et al.,

2023). With its considerable growth potential, marine aquaculture is well positioned to positively influence livelihoods, employment, and local economic development in coastal communities around the world. However, this can only happen if supported by a reliable and informative scientific foundation, which can be achieved through joint efforts between different stakeholders.

4. CONCLUSIONS

Through three peer-reviewed scientific publications, this dissertation integrated predictive modeling, statistical analysis, and meta-analysis to address critical knowledge gaps concerning the predictive performance of existing *Vibrio* growth models, the suitability of *Vibrio* spp. abundance as a supplementary indicator of water quality, and the overall informative value of ecological research. The key findings and contributions are summarized as follows:

1. A total of 28 functional growth models for *Vibrio* spp. were extracted and standardized using a unified nomenclature, providing a structured overview of existing *Vibrio* modeling efforts to date.
2. Baranyi models demonstrated the greatest applicability in mariculture, however, no model offered reliable predictions across all coastal habitats, underscoring the importance of tailoring models to site-specific conditions, particularly in areas under strong anthropogenic pressures.
3. *Vibrio* spp. abundance has potential to be included as a supplementary microbial indicator of water quality near mariculture as it provides unique, seasonally relevant information that traditional fecal indicators may overlook. Further research is required to (i) establish appropriate threshold values for water quality classification and (ii) identify key pathogenic species relevant to both human and marine organism health.
4. Only 11–18% of ecological research reaches full informative value, which is the first quantitative estimate of informative value of research in ecology and second in any field of science. Specifically, 45% of ecological studies are never published, and out of those that are published 67% suffer from suboptimal study planning and 41% are under-reported.
5. The low informative value of research underscores the need for systemic reform in science and provides compelling evidence for funders, institutions, publishers, and researchers to improve how studies are planned, reported, and published.
6. All three studies included in this thesis are grounded in open science principles, with datasets and analytical code publicly available via Zenodo, promoting transparency, reproducibility, and reuse.

5. REFERENCES

1. Alver, M. O., & Føre, M. (2023). Mathematical Modelling Applied to Aquaculture. In Q. Zhang (Ed.), Encyclopedia of Digital Agricultural Technologies (pp. 791–807). Springer International Publishing. https://doi.org/10.1007/978-3-031-24861-0_197
2. Azam, F., & Malfatti, F. (2007). Microbial structuring of marine ecosystems. Nature Reviews Microbiology, 5(10), 782–791. https://doi.org/10.1038/nrmicro1747
3. Baker, S. (2023). North–south publishing data show stark inequities in global research. Nature, 624(7991), S1–S1. https://doi.org/10.1038/d41586-023-03901-x
4. Baker-Austin, C., Oliver, J. D., Alam, M., Ali, A., Waldor, M. K., Qadri, F., & Martinez-Urtaza, J. (2018). *Vibrio* spp. Infections. Nature Reviews Disease Primers, 4(1), 1–19. https://doi.org/10.1038/s41572-018-0005-8
5. Baker-Austin, C., Trinanes, J. A., Taylor, N. G. H., Hartnell, R., Siitonen, A., & Martinez-Urtaza, J. (2013). Emerging *Vibrio* risk at high latitudes in response to ocean warming. Nature Climate Change, 3(1), 73–77. https://doi.org/10.1038/nclimate1628
6. Bakker, T. C. M., & Traniello, J. F. A. (2020). Ensuring data access, transparency, and preservation: Mandatory data deposition for Behavioral Ecology and Sociobiology. Behavioral Ecology and Sociobiology, 74(11), 132. https://doi.org/10.1007/s00265-020-02910-4
7. Baranyi, J., McClure, P. J., Sutherland, J. P., & Roberts, T. A. (1993). Modeling bacterial growth responses. Journal of Industrial Microbiology, 12(3), 190–194. https://doi.org/10.1007/BF01584189
8. Bartram, J., Cotruvo, J., Exner, M., Fricker, C., & Glasmacher, A. (Eds.). (2013). Heterotrophic plate counts and drinking-water safety: The significance of HPCs for water quality and the human health. IWA Publishing. https://www.who.int/publications/i/item/9241562269
9. Begley, C. G., & Ellis, L. M. (2012). Raise standards for preclinical cancer research. Nature, 483(7391), 531–533. https://doi.org/10.1038/483531a

10. Bendiscioli, S. (2019). The troubles with peer review for allocating research funding. EMBO Reports, 20(12), e49472. https://doi.org/10.15252/embr.201949472
11. Berg, J. M., Bhalla, N., Bourne, P. E., Chalfie, M., Drubin, D. G., Fraser, J. S., Greider, C. W., Hendricks, M., Jones, C., Kiley, R., King, S., Kirschner, M. W., Krumholz, H. M., Lehmann, R., Leptin, M., Pulverer, B., Rosenzweig, B., Spiro, J. E., Stebbins, M., … Wolberger, C. (2016). Preprints for the life sciences. Science, 352(6288), 899–901. https://doi.org/10.1126/science.aaf9133
12. Bertram, M. G., Sundin, J., Roche, D. G., Sánchez-Tójar, A., Thoré, E. S. J., & Brodin, T. (2023). Open science. Current Biology, 33(15), R792–R797. https://doi.org/10.1016/j.cub.2023.05.036
13. Besançon, L., Peiffer-Smadja, N., Segalas, C., Jiang, H., Masuzzo, P., Smout, C., Billy, E., Deforet, M., & Leyrat, C. (2021). Open science saves lives: Lessons from the COVID-19 pandemic. BMC Medical Research Methodology, 21(1), 117. https://doi.org/10.1186/s12874-021-01304-y
14. Bierer, B. E., Strauss ,David H., White ,Sarah A., & and Zarin, D. A. (2018). Universal Funder Responsibilities That Advance Social Value. The American Journal of Bioethics, 18(11), 30–32. https://doi.org/10.1080/15265161.2018.1523498
15. Brumfield, K. D., Chen, A. J., Gangwar, M., Usmani, M., Hasan, N. A., Jutla, A. S., Huq, A., & Colwell, R. R. (2023). Environmental Factors Influencing Occurrence of *Vibrio* parahaemolyticus and *Vibrio* vulnificus. Applied and Environmental Microbiology, 89(6), e00307-23. https://doi.org/10.1128/aem.00307-23
16. Brumfield, K. D., Usmani, M., Chen, K. M., Gangwar, M., Jutla, A. S., Huq, A., & Colwell, R. R. (2021). Environmental parameters associated with incidence and transmission of pathogenic *Vibrio* spp. Environmental Microbiology, 23(12), 7314–7340. https://doi.org/10.1111/1462-2920.15716
17. Bullington, J. A., Golder, A. R., Steward, G. F., McManus, M. A., Neuheimer, A. B., Glazer, B. T., Nigro, O. D., & Nelson, C. E. (2022). Refining real-time predictions of *Vibrio* vulnificus concentrations in a tropical urban estuary by incorporating dissolved organic matter dynamics. Science of The Total Environment, 829, 154075. https://doi.org/10.1016/j.scitotenv.2022.154075

18. Calster, B. V., Wynants, L., Riley, R. D., Smeden, M. van, & Collins, G. S. (2021). Methodology over metrics: Current scientific standards are a disservice to patients and society. Journal of Clinical Epidemiology, 138, 219–226. https://doi.org/10.1016/j.jclinepi.2021.05.018
19. Camerer, C. F., Dreber, A., Holzmeister, F., Ho, T.-H., Huber, J., Johannesson, M., Kirchler, M., Nave, G., Nosek, B. A., Pfeiffer, T., Altmejd, A., Buttrick, N., Chan, T., Chen, Y., Forsell, E., Gampa, A., Heikensten, E., Hummer, L., Imai, T., … Wu, H. (2018). Evaluating the replicability of social science experiments in Nature and Science between 2010 and 2015. Nature Human Behaviour, 2(9), 637–644. https://doi.org/10.1038/s41562-018-0399-z
20. Chalmers, I., Bracken, M. B., Djulbegovic, B., Garattini, S., Grant, J., Gülmezoglu, A. M., Howells, D. W., Ioannidis, J. P. A., & Oliver, S. (2014). How to increase value and reduce waste when research priorities are set. The Lancet, 383(9912), 156–165. https://doi.org/10.1016/S0140-6736(13)62229-1
21. Chalmers, I., & Glasziou, P. (2009). Avoidable waste in the production and reporting of research evidence. Lancet (London, England), 374(9683), 86–89. https://doi.org/10.1016/S0140-6736(09)60329-9
22. Chan, L., Arunachalam, S., & Kirsop, B. (2009). Open access: A giant leap towards bridging health inequities. Bulletin of the World Health Organization, 87(8), 631–635. https://doi.org/10.2471/blt.09.064659
23. Culina, A., Baglioni, M., Crowther, T. W., Visser, M. E., Woutersen-Windhouwer, S., & Manghi, P. (2018). Navigating the unfolding open data landscape in ecology and evolution. Nature Ecology & Evolution, 2(3), 420–426.
24. Culina, A., Berg, I. van den, Evans, S., & Sánchez-Tójar, A. (2020). Low availability of code in ecology: A call for urgent action. PLOS Biology, 18(7), e3000763. https://doi.org/10.1371/journal.pbio.3000763
25. Data sharing and mining | Royal Society. (2025, May 1). https://royalsociety.org/journals/ethics-policies/data-sharing-mining/
26. Davey, K. R. (1989). A predictive model for combined temperature and water activity on microbial growth during the growth phase. The Journal of Applied Bacteriology, 67(5), 483–488. https://doi.org/10.1111/j.1365-2672.1989.tb02519.x
27. Eiler, A., Gonzalez-Rey, C., Allen, S., & Bertilsson, S. (2007). Growth response of *Vibrio* cholerae and other *Vibrio* spp. To cyanobacterial dissolved organic

matter and temperature in brackish water. FEMS Microbiology Ecology, 60(3), 411–418.

28. Esser, D. S., Leveau, J. H. J., & Meyer, K. M. (2015). Modeling microbial growth and dynamics. Applied Microbiology and Biotechnology, 99(21), 8831–8846. https://doi.org/10.1007/s00253-015-6877-6
29. FAO. (2024). The State of World Fisheries and Aquaculture 2024 – Blue Transformation in action. Food and Agriculture Organization of the United Nations. https://openknowledge.fao.org/handle/20.500.14283/cd0683en
30. Fraser, H., Parker, T., Nakagawa, S., Barnett, A., & Fidler, F. (2018). Questionable research practices in ecology and evolution. PloS One, 13(7), e0200303. https://doi.org/10.1371/journal.pone.0200303
31. Fraser, N., Brierley, L., Dey, G., Polka, J. K., Pálfy, M., Nanni, F., & Coates, J. A. (2021). The evolving role of preprints in the dissemination of COVID-19 research and their impact on the science communication landscape. PLOS Biology, 19(4), e3000959. https://doi.org/10.1371/journal.pbio.3000959
32. Froelich, B., Gonzalez, R., Blackwood, D., Lauer, K., & Noble, R. (2019). Decadal monitoring reveals an increase in *Vibrio* spp. Concentrations in the Neuse River Estuary, North Carolina, USA. PLOS ONE, 14(4), e0215254. https://doi.org/10.1371/journal.pone.0215254
33. Glasziou, P., Altman, D. G., Bossuyt, P., Boutron, I., Clarke, M., Julious, S., Michie, S., Moher, D., & Wager, E. (2014). Reducing waste from incomplete or unusable reports of biomedical research. The Lancet, 383(9913), 267–276. https://doi.org/10.1016/S0140-6736(13)62228-X
34. Glasziou, P., & Chalmers, I. (2016). Is 85% of health research really “wasted.” The BMJ.
35. Gorgoglione, A., Castro, A., Chreties, C., & Etcheverry, L. (2020). Overcoming Data Scarcity in Earth Science. Data, 5(1), Article 1. https://doi.org/10.3390/data5010005
36. Grainger, M. J., Bolam, F. C., Stewart, G. B., & Nilsen, E. B. (2020). Evidence synthesis for tackling research waste. Nature Ecology & Evolution, 4(4), 495–497. https://doi.org/10.1038/s41559-020-1141-6
37. Hardwicke, T. E., Serghiou, S., Janiaud, P., Danchev, V., Crüwell, S., Goodman, S. N., & Ioannidis, J. P. A. (2020). Calibrating the Scientific Ecosystem Through

Meta-Research. Annual Review of Statistics and Its Application, 7(Volume 7, 2020), 11–37. https://doi.org/10.1146/annurev-statistics-031219-041104

38. Huang, C.-K. (Karl), Neylon, C., Hosking, R., Montgomery, L., Wilson, K. S., Ozaygen, A., & Brookes-Kenworthy, C. (2020). Evaluating the impact of open access policies on research institutions. eLife, 9, e57067. https://doi.org/10.7554/eLife.57067
39. Huss, H. H., Ababouch, L., & Gram, L. (2004). Assessment and management of seafood safety and quality. No.444, iv + 230 pp.
40. Ina-Salwany, M. Y., Al-saari, N., Mohamad, A., Mursidi, F.-A., Mohd-Aris, A., Amal, M. N. A., Kasai, H., Mino, S., Sawabe, T., & Zamri-Saad, M. (2019). *Vibrio*sis in Fish: A Review on Disease Development and Prevention. Journal of Aquatic Animal Health, 31(1), 3–22.
41. Ioannidis, J. P. A., Fanelli, D., Dunne, D. D., & Goodman, S. N. (2015). Meta-research: Evaluation and Improvement of Research Methods and Practices. PLOS Biology, 13(10), e1002264. https://doi.org/10.1371/journal.pbio.1002264
42. Johnson, C. N., Bowers, J. C., Griffitt, K. J., Molina, V., Clostio, R. W., Pei, S., Laws, E., Paranjpye, R. N., Strom, M. S., Chen, A., Hasan, N. A., Huq, A., Noriea, N. F., Grimes, D. J., & Colwell, R. R. (2012). Ecology of *Vibrio* parahaemolyticus and *Vibrio* vulnificus in the Coastal and Estuarine Waters of Louisiana, Maryland, Mississippi, and Washington (United States). Applied and Environmental Microbiology, 78(20), 7249–7257. https://doi.org/10.1128/AEM.01296-12
43. Kidwell, M. C., Lazarević, L. B., Baranski, E., Hardwicke, T. E., Piechowski, S., Falkenberg, L.-S., Kennett, C., Slowik, A., Sonnleitner, C., Hess-Holden, C., Errington, T. M., Fiedler, S., & Nosek, B. A. (2016). Badges to Acknowledge Open Practices: A Simple, Low-Cost, Effective Method for Increasing Transparency. PLOS Biology, 14(5), e1002456. https://doi.org/10.1371/journal.pbio.1002456
44. Kim, J. Y., & Lee, J.-L. (2017). Correlation of Total Bacterial and *Vibrio* spp. Populations between Fish and Water in the Aquaculture System. Frontiers in Marine Science, 4. https://doi.org/10.3389/fmars.2017.00147
45. Leung, H. M., Leung, S. K. S., Au, C. K., Cheung, K. C., Wong, Y. K., Leung, A. O. W., & Yung, K. K. L. (2015). Comparative assessment of water quality parameters of mariculture for fish production in Hong Kong Waters. Marine

Pollution Bulletin, 94(1), 318–322. https://doi.org/10.1016/j.marpolbul.2015.01.028

46. Levy, S. (2018). ECDC *Vibrio* Map Viewer: Tracking the Whereabouts of Pathogenic Species. Environmental Health Perspectives, 126(3), 034003. https://doi.org/10.1289/EHP2904
47. Liu, Y., Wang, Z., Yang, X., Wang, S., Liu, X., Liu, B., Zhang, J., Meng, D., Ding, K., Gao, K., Zeng, X., & Ding, Y. (2023). Changes in mariculture and offshore seawater quality in China during the past 20 years. Ecological Indicators, 157, 111220. https://doi.org/10.1016/j.ecolind.2023.111220
48. Maedche, A., Elshan, E., Höhle, H., Lehrer, C., Recker, J., Sunyaev, A., Sturm, B., & Werth, O. (2024). Open Science. Business & Information Systems Engineering, 66(4), 517–532. https://doi.org/10.1007/s12599-024-00858-7
49. Manchanayake, T., Salleh, A., Amal, M. N. A., Yasin, I. S. M., & Zamri-Saad, M. (2023). Pathology and pathogenesis of *Vibrio* infection in fish: A review. Aquaculture Reports, 28, 101459. https://doi.org/10.1016/j.aqrep.2022.101459
50. Martinez-Urtaza, J., Bowers, J. C., Trinanes, J., & DePaola, A. (2010). Climate anomalies and the increasing risk of *Vibrio* parahaemolyticus and *Vibrio* vulnificus illnesses. Food Research International, 43(7), 1780–1790. https://doi.org/10.1016/j.foodres.2010.04.001
51. Moher, D., Glasziou, P., Chalmers, I., Nasser, M., Bossuyt, P. M. M., Korevaar, D. A., Graham, I. D., Ravaud, P., & Boutron, I. (2016). Increasing value and reducing waste in biomedical research: Who's listening? The Lancet, 387(10027), 1573–1586. https://doi.org/10.1016/S0140-6736(15)00307-4
52. Moher, D., Naudet, F., Cristea, I. A., Miedema, F., Ioannidis, J. P. A., & Goodman, S. N. (2018). Assessing scientists for hiring, promotion, and tenure. PLOS Biology, 16(3), e2004089. https://doi.org/10.1371/journal.pbio.2004089
53. Nakagawa, S., Dunn, A. G., Lagisz, M., Bannach-Brown, A., Grames, E. M., Sánchez-Tójar, A., O'Dea, R. E., Noble, D. W. A., Westgate, M. J., Arnold, P. A., Barrow, S., Bethel, A., Cooper, E., Foo, Y. Z., Geange, S. R., Hennessy, E., Mapanga, W., Mengersen, K., Munera, C., … Haddaway, N. R. (2020). A new ecosystem for evidence synthesis. Nature Ecology & Evolution, 4(4), 498–501. https://doi.org/10.1038/s41559-020-1153-2

54. Nakagawa, S., Koricheva, J., Macleod, M., & Viechtbauer, W. (2020). Introducing our series: Research synthesis and meta-research in biology. BMC Biology, 18(1), 20. https://doi.org/10.1186/s12915-020-0755-0
55. Noble, D., Xirocostas, Z., Wu, N., Martinig, A., Almeida, R., Bairos-Novak, K., Balti, H., Bertram, M., Bliard, L., Brand, J., Byrne, I., Chan, Y.-C., Clink, D., Corbel, Q., Correia, R., Crawford-Ash, J., Culina, A., D'Bastiani, E., Gywa, G., … Lagisz, M. (2024). The promise of community-driven preprints in ecology and evolution. https://doi.org/10.32942/X2SS46
56. Nogales, B., Lanfranconi, M. P., Piña-Villalonga, J. M., & Bosch, R. (2011). Anthropogenic perturbations in marine microbial communities. FEMS Microbiology Reviews, 35(2), 275–298. https://doi.org/10.1111/j.1574-6976.2010.00248.x
57. Nosek, B. A., Ebersole, C. R., DeHaven, A. C., & Mellor, D. T. (2018). The preregistration revolution. Proceedings of the National Academy of Sciences, 115(11), 2600–2606.
58. Nuzzo, R. (2015). How scientists fool themselves – and how they can stop. Nature, 526(7572), 182–185. https://doi.org/10.1038/526182a
59. Oberbeckmann, S., Fuchs, B. M., Meiners, M., Wichels, A., Wiltshire, K. H., & Gerdts, G. (2012). Seasonal Dynamics and Modeling of a *Vibrio* Community in Coastal Waters of the North Sea. Microbial Ecology, 63(3), 543–551. https://doi.org/10.1007/s00248-011-9990-9
60. Onohuean, H., Agwu, E., & Nwodo, U. (2022). A Global Perspective of *Vibrio* Species and Associated Diseases: Three-Decade Meta-Synthesis of Research Advancement. Environmental Health Insights, 16, 11786302221099406. https://doi.org/10.1177/11786302221099406
61. Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. Science (New York, N.Y.), 349(6251), aac4716. https://doi.org/10.1126/science.aac4716
62. Peleg, M., & Corradini, M. G. (2011). Microbial growth curves: What the models tell us and what they cannot. Critical Reviews in Food Science and Nutrition, 51(10), 917–945. https://doi.org/10.1080/10408398.2011.570463
63. Petersen, O. H. (2021). Inequality of Research Funding between Different Countries and Regions is a Serious Problem for Global Science. Function, 2(6), zqab060. https://doi.org/10.1093/function/zqab060

64. Price, R. G., Wildeboer, D., Price, R. G., & Wildeboer, D. (2017). E. coli as an Indicator of Contamination and Health Risk in Environmental Waters. In Escherichia coli—Recent Advances on Physiology, Pathogenesis and Biotechnological Applications. IntechOpen. https://doi.org/10.5772/67330
65. Pruzzo, C., Huq, A., Colwell, R. R., & Donelli, G. (2005). Pathogenic *Vibrio* Species in the Marine and Estuarine Environment. In S. Belkin & R. R. Colwell (Eds.), Oceans and Health: Pathogens in the Marine Environment (pp. 217–252). Springer US. https://doi.org/10.1007/0-387-23709-7_9
66. Purgar, M., Gavrilović, A., Kapetanović, D., Klanjšček, J., Jug-Dujaković, J., Kolda, A., Žunić, J., Kazazić, S., Vardić Smrzlić, I., Vukić Lušić, D., Pikelj, K., Listeš, E., El-Matbouli, M., Lillehaug, A., Lončarević, S., Knežević, D., Hengl, B., Geček, S., & Klanjscek, T. (2023). Assessment of *Vibrio* spp. abundance as a water quality indicator: Insights from Mali Ston Bay in the Adriatic Sea. Estuarine, Coastal and Shelf Science, 295, 108558. https://doi.org/10.1016/j.ecss.2023.108558
67. Purgar, M., Kapetanović, D., Geček, S., Marn, N., Haberle, I., Hackenberger, B. K., Gavrilović, A., Pečar Ilić, J., Hackenberger, D. K., Djerdj, T., Ćaleta, B., & Klanjscek, T. (2022a). Investigating the Ability of Growth Models to Predict In Situ *Vibrio* spp. Abundances. Microorganisms, 10(9), Article 9. https://doi.org/10.3390/microorganisms10091765
68. Purgar, M., Kapetanović, D., Geček, S., Marn, N., Haberle, I., Hackenberger K., B., Gavrilović, A., Pečar Ilić, J., Hackenberger K., D., Djerdj, T., Ćaleta, B., & Klanjšček, T. (2022c). Codes for Purgar et al. 2022: Investigating the ability of growth models to predict in situ *Vibrio* spp. abundances (Version 1) [Data set]. Zenodo. https://doi.org/10.5281/zenodo.7013394
69. Purgar, M., Klanjscek, T., & Culina, A. (2022b). Quantifying research waste in ecology. Nature Ecology & Evolution, 6(9), Article 9. https://doi.org/10.1038/s41559-022-01820-0
70. Ramsey, M. H. (2021). Challenges for the estimation of uncertainty of measurements made in situ. Accreditation and Quality Assurance, 26(4), 183–192. https://doi.org/10.1007/s00769-020-01446-4
71. Ratkowsky, D. A., Lowry, R. K., McMeekin, T. A., Stokes, A. N., & Chandler, R. E. (1983). Model for bacterial culture growth rate throughout the entire biokinetic temperature range. Journal of Bacteriology, 154(3), 1222–1226.

72. Riedinger, D. J., Hassenrück, C., Herlemann, D., & Labrenz, M. (2025). Global distribution and predictive modeling of *Vibrio* vulnificus abundance. Communications Earth & Environment, 6(1), 1–8. https://doi.org/10.1038/s43247-025-02182-8
73. Robertson, W. J. (1984). Pollution indicators and potential pathogenic microorganisms in estuarine recreational waters. Canadian Journal of Public Health = Revue Canadienne De Sante Publique, 75(1), 19–24.
74. Ross-Hellauer, T. (2022). Open science, done wrong, will compound inequities. Nature, 603(7901), 363–363. https://doi.org/10.1038/d41586-022-00724-0
75. Rothstein, H. R., Sutton, A. J., & Borenstein, M. (2005). Publication Bias in Meta-Analysis. In Publication Bias in Meta-Analysis (pp. 1–7). John Wiley & Sons, Ltd. https://doi.org/10.1002/0470870168.ch1
76. Ryder, J., Karunasagar, I., Ababouch, L., & Division, F. and A. E. and P. (2014). Assessment and management of seafood safety and quality: Current practices and emerging issues. https://openknowledge.fao.org/handle/20.500.14283/i3215e
77. Sanches-Fernandes, G. M. M., Sá-Correia, I., & Costa, R. (2022). *Vibrio*sis Outbreaks in Aquaculture: Addressing Environmental and Public Health Concerns and Preventive Therapies Using Gilthead Seabream Farming as a Model System. Frontiers in Microbiology, 13. https://doi.org/10.3389/fmicb.2022.904815
78. Semenza, J. C., Trinanes, J., Lohr, W., Sudre, B., Löfdahl, M., Martinez-Urtaza, J., Nichols, G. L., & Rocklöv, J. (2017). Environmental Suitability of *Vibrio* Infections in a Warming Climate: An Early Warning System. Environmental Health Perspectives, 125(10), 107004. https://doi.org/10.1289/EHP2198
79. Shafiee, N. S., Mazlan, N., Abd Rahman, N. N., Shapawi, R., & Shah, M. D. (2024). *Vibrio*sis in Aquaculture: Pathogenic *Vibrio* Species, Affected Animals, Diagnosis and Adverse Effects. In N. M. Faudzi, M. D. Shah, N. Mazlan, & S. Raehanah Muhamad Shaleh (Eds.), Essentials of Aquaculture Practices (pp. 113–138). Springer Nature. https://doi.org/10.1007/978-981-97-6699-4_7
80. Sheikh, H. I., Najiah, M., Fadhlina, A., Laith, A. A., Nor, M. M., Jalal, K. C. A., & Kasan, N. A. (2022). Temperature Upshift Mostly but not Always Enhances the Growth of *Vibrio* Species: A Systematic Review. Frontiers in Marine Science, 9. https://doi.org/10.3389/fmars.2022.959830

81. Some, S., Mondal, R., Mitra, D., Jain, D., Verma, D., & Das, S. (2021). Microbial pollution of water with special reference to coliform bacteria and their nexus with environment. Energy Nexus, 1, 100008. https://doi.org/10.1016/j.nexus.2021.100008
82. Stead, S. M. (2019). Using systems thinking and open innovation to strengthen aquaculture policy for the United Nations Sustainable Development Goals. Journal of Fish Biology, 94(6), 837–844. https://doi.org/10.1111/jfb.13970
83. Stein, A., Riley, J., & Halberg, N. (2001). Issues of scale for environmental indicators. Agriculture, Ecosystems & Environment, 87(2), 215–232. https://doi.org/10.1016/S0167-8809(01)00280-8
84. Stewart, J. R., Gast, R. J., Fujioka, R. S., Solo-Gabriele, H. M., Meschke, J. S., Amaral-Zettler, L. A., del Castillo, E., Polz, M. F., Collier, T. K., Strom, M. S., Sinigalliano, C. D., Moeller, P. D., & Holland, A. F. (2008). The coastal environment and human health: Microbial indicators, pathogens, sentinels and reservoirs. Environmental Health, 7(2), S3. https://doi.org/10.1186/1476-069X-7-S2-S3
85. Stocker, R. (2012). Marine Microbes See a Sea of Gradients. Science, 338(6107), 628–633. https://doi.org/10.1126/science.1208929
86. Takemura, A. F., Chien, D. M., & Polz, M. F. (2014). Associations and dynamics of *Vibrio*naceae in the environment, from the genus to the population level. Frontiers in Microbiology, 5. https://doi.org/10.3389/fmicb.2014.00038
87. Teixeira da Silva, J. A., & Nazarovets, S. (2022). The Role of Publons in the Context of Open Peer Review. Publishing Research Quarterly, 38(4), 760–781. https://doi.org/10.1007/s12109-022-09914-0
88. Tennant, J. P., Dugan, J. M., Graziotin, D., Jacques, D. C., Waldner, F., Mietchen, D., Elkhatib, Y., Collister, L. B., Pikas, C. K., Crick, T., Masuzzo, P., Caravaggi, A., Berg, D. R., Niemeyer, K. E., Ross-Hellauer, T., Mannheimer, S., Rigling, L., Katz, D. S., Tzovaras, B. G., … Colomb, J. (2017). A multi-disciplinary perspective on emergent and future innovations in peer review (6:1151). F1000Research. https://doi.org/10.12688/f1000research.12037.3
89. Thickman, J. D., & Gobler, C. J. (2017). The ability of algal organic matter and surface runoff to promote the abundance of pathogenic and non-pathogenic strains of *Vibrio* parahaemolyticus in Long Island Sound, USA. PLOS ONE, 12(10), e0185994. https://doi.org/10.1371/journal.pone.0185994

90. Thompson, J. R., & Polz, M. F. (2006). Dynamics of *Vibrio* Populations and Their Role in Environmental Nutrient Cycling. In The Biology of *Vibrio*s (pp. 190–203). John Wiley & Sons, Ltd. https://doi.org/10.1128/9781555815714.ch13
91. Touchon, J. C., & McCoy, M. W. (2016). The mismatch between current statistical practice and doctoral training in ecology. Ecosphere, 7(8), e01394. https://doi.org/10.1002/ecs2.1394
92. Troell, M., Costa-Pierce, B., Stead, S., Cottrell, R. S., Brugere, C., Farmery, A. K., Little, D. C., Strand, Å., Pullin, R., Soto, D., Beveridge, M., Salie, K., Dresdner, J., Moraes-Valenti, P., Blanchard, J., James, P., Yossa, R., Allison, E., Devaney, C., & Barg, U. (2023). Perspectives on aquaculture's contribution to the Sustainable Development Goals for improved human and planetary health. Journal of the World Aquaculture Society, 54(2), 251–342. https://doi.org/10.1111/jwas.12946
93. Udesky, L. (2025). 'Publish or perish' culture blamed for reproducibility crisis. Nature. https://doi.org/10.1038/d41586-024-04253-w
94. van den Berg, N. I., Machado, D., Santos, S., Rocha, I., Chacon, J., Harcombe, W., Mitri, S., & Patil, K. R. (2022). Ecological modelling approaches for predicting emergent properties in microbial communities. Nature Ecology & Evolution, 6(7), 855–865. https://doi.org/10.1038/s41559-022-01746-7
95. Veić, T., 2016. Characterization of *Vibrio* species isolated from european sea bass (Dicentrarchus labrax linnaeus, 1758) farmed on the east coast of the adriatic sea. Available at https://urn.nsk.hr/urn:nbn:hr:217740993.
96. Vorobeichik, E. L., & Kozlov, M. V. (2012). Impact of point polluters on terrestrial ecosystems: Methodology of research, experimental design, and typical errors. Russian Journal of Ecology, 43(2), 89–96. https://doi.org/10.1134/S1067413612020166
97. Webber, J. L., Tyler, C. R., Carless, D., Jackson, B., Tingley, D., Stewart-Sinclair, P., Artioli, Y., Torres, R., Galli, G., Miller, P. I., Land, P., Zonneveld, S., Austen, M. C., & Brown, A. R. (2021). Impacts of land use on water quality and the viability of bivalve shellfish mariculture in the UK: A case study and review for SW England. Environmental Science & Policy, 126, 122–131. https://doi.org/10.1016/j.envsci.2021.09.027
98. Wilkinson, M. D., Dumontier, M., Aalbersberg, Ij. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E.,

Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., … Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data, 3(1), 160018. https://doi.org/10.1038/sdata.2016.18

99. Zakaria, N. H., Abd Rahim, N. D. E., Rosilan, N. F., Sung, Y. Y., Waiho, K., Harun, S., Zainal Abidin, R. A., & Afiqah-Aleng, N. (2025). Global landscape of *Vibrio* parahaemolyticus research: A bibliometric analysis. World Journal of Microbiology and Biotechnology, 41(2), 45. https://doi.org/10.1007/s11274-025-04262-5
100. Zupičić, I.G., Pavlinec, Ž., Veić, T., Oraić, D., Zrnčić, S., 2022. Similarities and differences between vibriosis in European sea bass caused by *Vibrio* anguillarum and *Vibrio* harveyi. Vet. Arh. 92 (4), 509–524.

6. AUTHOR RESUME

Marija Purgar Filjak was born on November 19, 1996. She completed her secondary education in Našice, Croatia, and earned a bachelor's degree in Environmental Engineering from the Faculty of Geotechnical Engineering, University of Zagreb, in 2019. She continued her studies at the J. J. Strossmayer University of Osijek, where she completed a master's degree in Nature and Environmental Protection in 2021. That same year, she joined the Ruđer Bošković Institute as a research assistant in the Laboratory for Informatics and Environmental Modelling and enrolled in the Interdisciplinary PhD Program in Oceanology at the Faculty of Science, University of Zagreb.

During her PhD, Marija published 12 scientific papers, including five as first author. Three of these appeared in Q1 journals, with two published in Nature Portfolio journals. Marija was awarded the Fulbright Scholarship for the 2024/25 academic year, which she spent at Emory University’s Rollins School of Public Health, Department of Epidemiology. Beyond her research, Marija advocates for open science. She serves on the Board of Directors of the *Society for Open, Reliable, and Transparent Ecology and Evolutionary Biology* (SORTEE) and is President-elect for 2026. The society gathers more than 500 members from over 50 countries to enhance research transparency, accessibility, and reproducibility in ecology and evolution.

She is also a Croatia Field Representative for *Seacology*, a nonprofit organization based in Berkeley, California, dedicated to preserving island ecosystems and cultures worldwide.