Integrating Knowledge about Complex Adaptive Systems - Insights from Modelling the Eastern Baltic Cod.

Dissertation

to

obtain the degree

Doctor of Engineering (Dr.-Ing.)

from the Faculty of Computer Science and Electrical Engineering

of the University of Rostock.

Submitted by:

Maria Elena Pierce,
born on December 31st 1981,
in Goch, Germany

Rostock, June 7th 2022

https://doi.org/10.18453/rosdok_id00004121

Dieses Werk ist lizenziert unter einer Creative Commons Namensnennung - Weitergabe unter gleichen Bedingungen 4.0 International Lizenz.

COLOPHON

This document was typeset in LaTeX using the `tuft-book` package. Thousands of digital cod were badly hurt and killed during the completion of this thesis

Integrating Knowledge about Complex Adaptive Systems,
Insights from Modelling the Eastern Baltic Cod.
© June 7^{th} 2022, Maria E. Pierce

Gutachter:

1. Gutachterin:
Prof. Dr. Adelinde M. Uhrmacher,
Universität Rostock

2. Gutachter:
Prof. Dr. Volker Grimmm,
Universität Potsdam und
Helmholtz Zentrum für Umweltforschung - UFZ

3. Gutachter:
Dr. Uwe Krumme,
Thünen Institut für Ostseefischerei

Datum der Einreichung: 07. Juni 2022
Datum der Verteidigung: 27. September 2022

Abstract

Currently, the Eastern Baltic cod (EBC) is in continuing decline regarding both the productivity of the stock and the condition of its individuals. Supporting management efforts to assist the stock in its recovery will require a functional understanding of the new dynamics of the EBC stock and the Baltic ecosystem. However, aquatic environments are challenging to research because they are not directly accessible to humans, encompass many scientific disciplines, and are complex adaptive systems (CAS). This thesis explores how modelling and simulation methods can be applied and adapted to meet the specific needs of fisheries biologies' current challenges regarding the management of the EBC and potentially other stocks facing similar challenges.

To effectively incorporate modelling and simulation into the workflow of departmental research, specific key requirements and matching solutions were identified: functionally integrating knowledge across scientific domains, which can be achieved through multi-faceted modelling; providing the required accessibility, which can be supported by suitable domain-specific languages (DSLs) and thorough documentation; modelling CAS to a high degree of verisimilitude of functionality, which requires a careful selection of type of model based on the respective system of interest and research question; and continued validity, which can be streamlined by iterative validation.

As the general approach, multi-faceted modelling was adapted to the given requirements by using white-box integration of submodels to provide insight into mechanisms and interaction of processes in the modelled system. These models were specified in the DSL ML-Rules, a language for multi-level modelling and simulation of cell-biological systems which was previously tested for its suitability for ecological models. The systematic reuse of simulation experiments during the iterative integration of models helped to maintain the continued validity of the model. Additionally, a focus was placed on documentation to further ensure accessibility and invite scrutinisation of the simulation study. Both structured natural language documentation using the TRACE protocol and formal provenance documentation based on the provenance data model (ProvDM) were applied here. Provenance metadata was then used to link the two documentations, which provides increased navigability by supporting queries of the provenance graph.

The simulation study currently covers the facets of physiology, reproduction, behaviour, environment, prey and parasitation and iteratively worked towards understanding the decline in growth and condition. Although the processes formalised to this point have not uncovered the mechanisms of the decline, several suspected causes could be dismissed. Notwithstanding the domain results, the study illustrates the merits and viability of the multi-faceted approach for the functional integration of knowledge about CAS.

Zusammenfassung

Gegenwärtig ist beim Ostdorsch (EBC - "Eastern Baltic cod") eine anhaltende Verschlechterung, sowohl hinsichtlich der Produktivität des Bestands als auch des Zustands seiner Individuen, festzustellen. Um Bewirtschaftungsmaßnahmen, welche den Bestand bei seiner Erholung unterstützen würden, zu identifizieren, wird ein funktionales Verständnis der neuen Dynamik des EBC-Bestands und des Ökosystems Ostsee benötigt. Die Erforschung aquatischer Systeme birgt jedoch eine Reihe an Herausforderungen, sie sind für den Menschen nicht direkt zugänglich, umfassen eine Vielzahl an wissenschaftlichen Disziplinen und fallen in die Kategorie komplexer adaptiver Systeme (CAS - 'complex adaptive systems'). In Rahmen dieser Arbeit wurde untersucht, wie Modellierungs- und Simulationsmethoden angewendet und angepasst werden können, um den spezifischen Bedürfnisse gerecht zu werden, die sich aus den aktuellen Herausforderungen der Fischereibiologie in Bezug auf die Bewirtschaftung des EBC und möglicherweise anderer Bestände, welche vor ähnlichen Herausforderungen stehen, ergeben.

Um Modellierung und Simulation effektiv in den Arbeitsablauf der Ressortforschung zu integrieren, wurden spezifische Schlüsselanforderungen und passende Lösungsansätze identifiziert: funktionale Integration von Wissen über wissenschaftliche Domänen hinweg kann durch die 'Modellierung unterschiedlichen Facetten' (multi-faceted modelling) erreicht werden; die erforderlichen Zugänglichkeit kann durch geeignete domänenspezifische Sprachen (DSLs - 'domain specific languages') und eine gründliche Dokumentation unterstützt werden; Modellierung von CAS mit einem hohen Maß an Realitätsnähe in Bezug auf Funktionalität benötigt eine sorgfältige Auswahl des Modelltyps basierend auf dem spezifischen System und der jeweiligen Forschungsfrage; und fortgesetzte Validität kann durch iterative Validierung effizient gestaltet werden.

Als übergreifender Ansatz wurde die Methode der 'Modellierung unterschiedlichen Facetten' an die gegebenen Anforderungen angepasst, indem eine White-Box-Integration von Teilmodellen angewandt wurde, um Zugang zu Mechanismen und Wechselwirkungen von Prozessen im modellierten System zu gewährleisten. Zur Umsetzung des Ansatzes wurde die DSL ML-Rules verwendet, welche für die Mehrebenenmodellierung zellbiologischer Systeme konzipiert und bereits auf ihre Eignung für ökologische Modelle getestet wurde. Die systematische Wiederverwendung von Simulationsexperimenten während der iterativen Integration von Modellen trug dazu bei, die Validität des Modells aufrechtzuerhalten. Darüber

hinaus wurde ein Fokus auf die Dokumentation gelegt, um stets Zugänglichkeit zu gewährleisten und zur Prüfung der Simulationsstudie einzuladen. Dabei kamen sowohl die strukturierte Dokumentation nach dem TRACE-Protokoll als auch die formale Provenienzdokumentation nach dem Provenance Data Model (ProvDM) zum Einsatz. Provenienz-Metadaten wurden anschließend verwendet, um die beiden Dokumentationen zu verknüpfen, was eine verbesserte Navigierbarkeit bietet, indem Abfragen des Provenienzgraphen unterstützt werden.

Die Simulationsstudie deckt derzeit die Facetten Physiologie, Reproduktion, Verhalten, Umwelt, Beute und Parasitierung ab und arbeitete während der Entwicklung iterativ auf das Verständnis der Zustandsverschlechterung und des Wachstumsrückgangs hin. Obwohl die bis zu diesem Punkt formalisierten Prozesse die Mechanismen des Rückgangs nicht aufgedeckt haben, konnten mehrere vermutete Ursachen verworfen werden. Ungeachtet der fischereibiologischen Ergebnisse veranschaulicht die Studie die Vorzüge und Realisierbarkeit den Ansatz der ‘Modellierung unterschiedlichen Facetten’ für die funktionale Integration von Wissen über CAS.

The difference between science and screwing around is writing it down.

- Adam Savage

Acknowledgements

There are many People I would like to thank for helping me complete this thesis, here are some of them:

- Adelinde 'Lin' Uhrmacher has had a the most formative impact on my growth and development as a scientist. I am truly grateful for all the opportunities she has given me access to and the enjoyment and excitement of conducting research with her.
- Uwe Krumme was my domain supervisor during my masters-thesis and got me back from the other side of the country to work on his team. This is the greatest vote of confidence I have ever received in my scientific career, and it has changed its direction permanently.
- Christopher Zimmermann found the funds and room to send me on this journey and supported me through all manner of challenges. He has been the type of mentor usually only found in stories and helped me to become a confident scientist.
- Oliver Reinhardt for his correctness on every level.
- Kate McQueen for her well-founded enthusiasm and friendship.
- Tom Warnke for being my idol, voice of reason and calm.

Florian Peters *who never placed his perfect style above substance*
Isabella Kratzer *who upkept the physicist spirits for all of us*
Meike Klettke *who turned up the light*
Nadja Schlungbaum *who always provided*
Pia Wilsdorf *who shared her clear head*
Tobias Helms *who knows what is important and was always kind*

I did all of this standing on your shoulders.

There is also the infrastructure in which this project existed: The Thünen-Institute for Baltic Sea Fisheries, the University of Rostock, the German state and the European Union - I thank those who built them and those who fill them with life.

Maria Elena Pierce
Rostock,
June 7th 2022

Contents

Chapter I

Chapter II

Chapter III

Chapter IV

Chapter V

Part I

PROLOGUE

Introduction

To whomever much is given, of him will much be required; and to whom much was entrusted, of him more will be asked.

Luke 12:48

Managing Natural Resources

Humans rely heavily on natural resources for food and raw materials. Managing when, how and how much of these resources we extract is important to control our impact on the environment, and, in the case of ecosystem goods, such as fish, necessary to be able to continue to do so. The former reflects the responsibility we have for the impact of our actions. The latter results from the fact that maximising the use of any renewable resource long term requires the amount extracted to be managed according to its subsequent renewal rate: if it is over-extracted, future yield is diminished, if it is under-extracted, current yield is lost. The management of a renewable resource, therefore, is both a question of stewardship and economics [1].

Management is consequently based on reliable knowledge about the renewal rate of a given resource and the impact of extraction on this renewal rate. When proper management is implemented and the system from which the resources is taken is productive and resilient, renewal rates tend to be high and stable, allowing for consistent management and stability for the depending industries. However, in recent years increasing anthropogenic pressures have moved some systems past their respective points of resilience and renewal rates have decreased and become more erratic.

In these cases it appears that knowledge about renewal rates can no longer be based on past observations but has to be based on a deep understanding of the newly evolved properties of the system [2]. As a consequence the consistency and subsequent trust in management to be responsible and provide stability can only be upheld if it is equipped with such deepened understanding.

The Framework of Public R&D - Departmental Research

The German word "Ressortforschung" refers to research and development efforts coordinated and financed by federal and state ministries [3]. This structure is meant to serve as a conduit between science, society and politics. Part of "departmental research" or "public research and development" (R&D) is providing scientific services (which might even be statutorily determined), such as testing and regulation. The aim is to gain scientific knowledge of importance to a field of activ-

[1] Garrett Hardin. "The Tragedy of the Commons: The population problem has no technical solution; it requires a fundamental extension in morality." In: *Science* 162.3859 (1968), pp. 1243–1248. ISSN: 0036-8075, 1095-9203. DOI: `10.1126/science.162.3859.1243`

[2] Francois Bastardie et al. "A Review Characterizing 25 Ecosystem Challenges to Be Addressed by an Ecosystem Approach to Fisheries Management in Europe". In: *Frontiers in Marine Science* 7 (2021), p. 629186. ISSN: 2296-7745. DOI: `10.3389/fmars.2020.629186`

[3] *Empfehlungen zur Profilierung der Einrichtungen mit Ressortforschungsaufgaben des Bundes*. Wissenschaftsrat. 2010

ity or ministry and, most importantly, provide a scientific basis for administrative or political decisions.

This placement results in typical characteristics of Public R&D: its motivation always begins with challenges of society, economy, politics and often all three, it is practical problem-oriented and usually draws a number of different scientific disciplines. It can, when the need is determined and not covered elsewhere, also include basic research.

Crucially, it ensures knowledge-based policy advice, thereby securing a sound basis for legislation and enforcement. This is also the point at which the tie-in with management takes place. Departmental research provides the knowledge base for management decisions concerning natural resources. Making this a high-stakes endeavour as there are usually very direct real-world implications based on its input.

Some management decisions have to be made on an annual basis which has an impact on the priorities in the institutes conducting this research: a large proportion of resources has to be allocated to the collection and processing of the latest data for which there is often an established institutional workflow. This results in constraints for the individual scientists involved concerning time and resources as well as constraints concerning the deviation from the regular workflow.

As a result, meeting new challenges (like the arising need for a deeper understanding of ecosystem behaviour) will also need to take into account the realities of departmental research by considering how to make the best use of limited resources and practical integration into an established workflow.

Understanding the Behaviour of Ecosystems

The renewal rate of a managed resource is one aspect of the behaviour of the ecosystem of which it is a part and by which it is sustained. As ecosystems fall into the category of complex adaptive systems (CAS) predicting their behaviour is not a straightforward task: in a simple system, one set of inputs leads to one set of behaviours; in a complicated system there might be a less direct link between a set of inputs and a resulting set of behaviours but their connection is determinable and can be reproducibly predicted. However, in a complex system, short-term behaviour is (at least at this point in history) a much more elaborate thing to predict, and long-term behaviour might evade our grasp completely. Complex adaptive systems are distinguished by consisting of large dynamic networks of numerous interactions and the development of these interactions over time results in properties such as emergence (i.e. the behaviour of the whole can not be derived from the behaviour of its parts) and self-organisation (i.e. the development of an overall order from local interactions) [4].

On a practical level, central to investigating the behaviour of a complex adaptive system, therefore, is the analysis of interactions, first within the manageable scope of its respective subsystems, and sub-

[4] Madhur Anand et al. "Ecological Systems as Complex Systems: Challenges for an Emerging Science". In: *Diversity* 2.3 (2010), pp. 395–410. ISSN: 1424-2818. DOI: `10.3390/d2030395`

sequently investigation of behaviour resulting from the interactions of these subsystems. The former is reflected in the many established scientific disciplines and methodological approaches

Applying Modelling and Simulation

Modelling and simulation is a well established tool in many scientific endeavours [5]. It is also a very broad umbrella term as models range from the straightforward (e.g.: Newton's laws of motion) to the sprawling (e.g.: climate models) and simulation studies can have a lot of different purposes (e.g.: formalising a relation, predicting a behaviour and many more). However, here the ambition is to meet the challenges that the investigation of complex adaptive systems pose, while remaining within the established framework specified by the realities of departmental research.

Exploring the specifics and details of these requirements and constraints, and matching them with the tools and approaches best suited to the task and, where required, suitably customising them is at the heart of this thesis. In this endeavour modelling and simulation studies shall also not be viewed in isolation but rather as being part of a specific workflow (e.g.: stock assessment) and supported by auxiliary infrastructure (e.g. documentation). As including these aspects at the point of conceptualisation, rather than treating them as optional add-ons, forms a more comprehensive understanding of how modelling and simulation will be functionally embedded to support reaching the overarching goal.

These objectives are put to the test by an extensive modelling and simulation study of the Eastern Baltic cod which illustrates how the many theoretical considerations can be translated into practical implementation.

[5] Eric B. Winsberg. *Science in the age of computer simulation*. Chicago: The University of Chicago Press, 2010. 152 pp. isbn: 978-0-226-90202-9 978-0-226-90204-3

Prior Work

Parts of this thesis contain work published, submitted or in preparation for publication as, articles, posters and abstracts:

• Statements on the use of domain specific languages to investigate Eastern Baltic Cod ecology were first presented in:

> Maria E. Pierce et al. "Individual-based cod simulation with ML-Rules". In: Proceedings of the Winter Simulation Conference, Huntington Beach, CA, USA; Poster. 2015, pp. 3192-3193. doi: `10.1109/WSC.2015.7408465`

• The first iteration of the Cod model and the subsequent evaluations of domain specific languages for investigative ecological models was published as:

> Maria E. Pierce, Tom Warnke, Uwe Krumme, Tobias Helms, Cornelius Hammer, and Adelinde M. Uhrmacher. "Developing and validating a multi-level ecological model of eastern Baltic cod (Gadus morhua) in the Bornholm Basin - A case for domain-specific languages". In: Ecological Modelling 361 (2017), pp. 49-65. issn: 03043800. doi: `10.1016/j.ecolmodel.2017.07.012`.

• The first conceptualizations of the challenges and opportunities of tailoring tools to the emerging requirements of fisheries management were presented in:

> Maria E. Pierce, Methods, Modellers and Fisheries Scientists – Building Bridges. In: Winter Simulation Conference, 09-12 Dec 2018, Gothenburg, Sweden. Poster. 2018, pp. 4168–4169.

• The chapter on validation (Chapter IV) is based on:

> Maria E. Pierce, Uwe Krumme, and Adelinde M. Uhrmacher. "Building Simulation Models of Complex Ecological Systems by Successive Composition and Reusing Simulation Experiments". In: 2018 Winter Simulation Conference (WSC). Gothenburg, Sweden: IEEE, 2018 , pp. 2363-2374 . isbn: 978-1-5386-6572-5 . doi: `10.1109/WSC.2018.8632262`.

Part II

Chapter I
Case in Point – The Eastern Baltic Cod

> *Cod – It is amazingly prolific. Leewenhoek counted 9384000 eggs in a cod-fish of middling size – a number that will baffle all the efforts of man to exterminate.*
>
> J. Smith Homans Sr. & Jr. (1858)

Figure 1: Cod Fish (*Gadus morhua*) from Ichtylogie, ou Histoire naturelle: génerale et particuliére des poissons (1785–1797) by Marcus Elieser Bloch. (New York Public Library cc)

Cod - *Gadus morhua*

Cod are a species of carnivorous, marine fish found along the continental shelves (with a depth range form 0 - 300m) of the North Atlantic, the North Sea and the Baltic Sea. They utilise open waters and the seabed (i.e. are benthopelagic) and individuals can reach well above 1m total length, and can weigh more than 30kg. They usually reach sexual maturity between the ages of two and four years but can take as few as one and as many as eight years. This reflects the broad range of habitats they populate, the remarkable plasticity of their behaviour and food sources, and their active and wide-ranging migrations, which all contribute to their success. Cod are top-level predators and thereby critical to the ecosystems they are part of by exerting top-down control of the respective population dynamics. Cod are also very valuable to humans as a source of (once extremely abundant) protein, they are the object of reverence[6] and have been the cause of wars [7].

The Baltic Sea

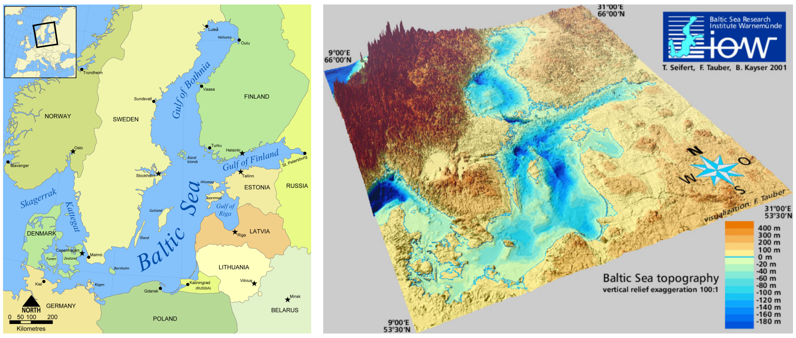


Figure 2:
Left: Map of the Baltic Sea (by NormanEinstein cc);
Right: Digital topography of the Baltic and Baltic Sea showing the overall shallow waters and deep basins (by The Leibniz Institute for Baltic Sea Research, Warnemünde IOW)

The Baltic Sea is a marginal sea of the North Sea and, by extension, the North Atlantic. Its only connection to the North Sea are three shallow straights: the Øresund (max. depth 40m, max. width 28km), the Great Belt (max. depth 60m, max. width 32km) and the Little Belt (max. depth 81m, max. width 28km). With a surface area of 337000km^2, the Baltic Sea is too small to develop significant tides

---

[6] An effigy of a cod named 'The Sacred Cod' hangs in the House of Representatives chamber of Boston Massachusetts

[7] GuÐmundur J. GuÐmundsson. "THE COD AND THE COLD WAR". in: *Scandinavian Journal of History* 31.2 (2006), pp. 97–118. ISSN: 0346-8755, 1502-7716. DOI: `10.1080/03468750600604184`

and is generally considered micro-tidal (with a few centimetres of tidal range in the Southern Baltic) or non-tidal (which is a sound assumption for most intents and purposes). In combination with its humid climate (more freshwater is introduced than evaporates) and the limited exchange with the North Sea, this has led to a salinity gradient from West (at near 20 PSU) to North (at near 2 PSU) with brackish water conditions throughout, making it a rare type of habitat. Another distinguishing property of the Baltic Sea is its bathymetry, it is overall relatively shallow, with an average depth of only 55m but a maximum depth of 459m (the Landsort Deep), see Figure 2. It can be effectively viewed as a shallow, stratified sea containing consecutive deep basins. The deeper basins, when seen from west to east, are the Arkona Basin (max. depth 48m), followed by the Bornholm Basin (max. depth 92m), the Gotland Deep (max. depth 249m) and the Western Gotland Basin in parallel containing the Landsort Deep (max. depth 459m) and the Eastern Gotland Basin (max. depth 231m).

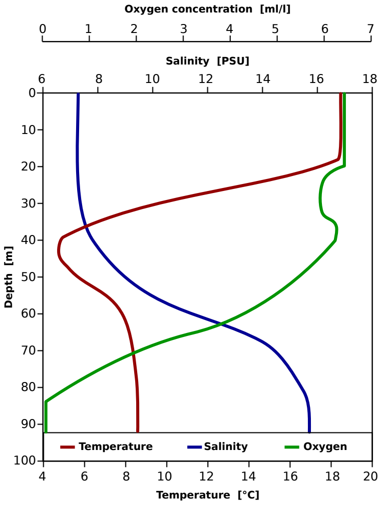


Figure 3: Example of typical stratification of temperature, salinity and oxygen in the Bornholm Basin in early summer.

The Baltic Sea's unique salinity, bathymetric conditions and lack of tides have resulted in an exceptionally stable stratification in its basins. It not only displays the typical thermocline of lakes and oceans, separating the upper mixed layer from less turbulent deeper layers but in addition develops a halocline and one to two oxyclines (see Figure 3). The halocline separates the lower density brackish waters at the top from the saltier, denser waters at the bottom. The stratifications resulting from the salinity gradient are remarkably stable, and the lower strata are generally only mixed and aerated through Major Baltic Inflow events (MBIs) [8], [9]. These events transport large amounts (in the order of magnitude of $500km^3$) of oxygen-rich, saline water from the Kattegat through the shallow straights into the Baltic. Since the inflowing water has a higher specific gravity than the resident water, it moves along the seabed (i.e., underneath the resident water) and ventilates the Baltic Basins if sufficient in volume. The combination of bathymetry, MBIs and oxygen consumption then accounts for the formation of oxyclines. The first oxycline develops when deeper waters are not supplied with oxygen due to the stable halocline and subsequent lack of mixing, and the oxygen that is present is depleted through biological activity. When an MBI occurs, its oxygenated waters can move underneath these hypoxic or anoxic waters forming a second oxycline with the lowest strata now oxygen-rich. However, the second oxycline is not stable and will, over time, dissolve and aerate the oxygen-poor waters above it. If the amount of oxygen is sufficient, mixing can also lead to the dissolution of the first oxycline.

Furthermore, the 'current' Baltic Sea is, geologically speaking, very young, having begun its formation about 10000 years ago as the Baltic Ice Lake and reaching its current form as the Modern Baltic Sea roughly 4000 years ago [10]. This time span does not allow for the evolution of fully adapted (macroscopic) species, and organisms in the Baltic Sea mostly operate on the margins of their respective osmoregulatory spectrum as its fauna is recruited either from marine

[8] Peter L. Holtermann et al. "Deepwater dynamics and mixing processes during a major inflow event in the central Baltic Sea: DEEPWATER DYNAMICS DURING A MBI". in: *Journal of Geophysical Research: Oceans* 122.8 (2017), pp. 6648–6667. ISSN: 21699275. DOI: `10.1002/2017JC013050`

[9] Volker Mohrholz. "Major Baltic Inflow Statistics – Revised". In: *Frontiers in Marine Science* 5 (2018). ISSN: 2296-7745. DOI: `10.3389/fmars.2018.00384`

[10] Jan Harff, Svante Björck, and Peer Hoth, eds. *The Baltic Sea Basin*. Central and Eastern European Development Studies (CEEDES. Heidelberg ; New York: Springer, 2011. 449 pp. ISBN: 978-3-642-17219-9

or freshwater environments.

In comparison to other bodies of water with comparable size, the Baltic Sea is, due to all of its characteristics mentioned above, particularly species-poor (less than 100 marine species of animals in the central Baltic, compared to the North Sea with more than 1 500), decreasing the potential resilience of its ecosystem.

Eastern Baltic Cod Ecology

In the Baltic Sea, cod is the apex predator among fish and has been commercially highly relevant prior to its decline. There are two 'stocks' (subpopulations considered a unit by management) of cod in the Baltic: the Western Baltic cod (WBC) and the Eastern Baltic cod (EBC). The stock of interest in this thesis is the EBC.

As do all marine species in the Baltic Sea, cod operate on the margins of their physiological capabilities when dwelling in its brackish waters. The EBC, residing in the less brackish end of the Baltic Sea, has become genetically distinct under adaptation to these conditions and is considered an example of ongoing speciation [11], [12]. Subsequently, there is also an increasing genetic distance between EBC and WBC.

The main prey species in terms of fish for the EBC are herring (*Clupea harengus*) and sprat (*Sprattus sprattus*), which, as is the cod, are originally marine species. The herring and sprat stocks naturally display dynamics of their own, with the Central Baltic herring in steady decline[13] and the Baltic sprat exhibiting continuing fluctuation[14]. Other fish cod feed on in the Baltic sea are flatfish, gobies and, given a sufficient difference in size, other cod. As a benthopelagic species, cod are equally adept at foraging benthic prey and stomach sampling of Baltic cod reveals that they feed on crustaceans in general, including large amounts of brown shrimp and crabs as well as benthic isopods (which could be of particular interest in terms of essential nutrients), ringworms and molluscs [15].

The Baltic Sea is also home to four species of marine mammals: the grey seal (*Halichoerus grypus*), harbour seal (*Phoca vitulina*), ringed seal (*Pusa hispida*) and the harbour porpoise (*Phocoena phocoena*) which are all potential predators of cod in the Baltic Sea.

In order to reproduce in EBC depend on what is termed a 'Reproductive Volume'. Such volumes are defined by being sufficiently saline to allow the cods eggs to float while simultaneously being sufficiently oxygenated to prevent suffocation [16], [17]. This volume typically fluctuates between 200 and 400 km^3 and is limited to the sufficiently deep and oxygenated basins. Ongoing oxygen depletion has moved the cut-off for sufficient oxygenation in the deep basins further south, and currently, the primary recruitment area is the Bornholm Basin.

[11] Paul R. Berg et al. "Adaptation to Low Salinity Promotes Genomic Divergence in Atlantic Cod (*Gadus morhua* L.". In: *Genome Biology and Evolution* 7.6 (2015), pp. 1644–1663. ISSN: 1759-6653. DOI: `10.1093/gbe/evv093`

[12] A. Poćwierz-Kotus et al. "Genetic differentiation of brackish water populations of cod *Gadus morhua* in the southern Baltic, inferred from genotyping using SNP-arrays". In: *Marine Genomics* 19 (2015), pp. 17–22. ISSN: 18747787. DOI: `10.1016/j.margen.2014.05.010`

[13] ICES. "Herring (Clupea harengus) in subdivisions 25-29 and 32, excluding the Gulf of Riga (central Baltic Sea)". In: (2021). Publisher: ICES. DOI: `10.17895/ICES.ADVICE.7767`

[14] ICES. "Sprat (Sprattus sprattus) in subdivisions 22-32 (Baltic Sea)". In: (2021). Publisher: ICES. DOI: `10.17895/ICES.ADVICE.7867`

[15] Peter Hornetz. "Spatio-temporal distribution and food intake of cod (*Gadus morhua*) along a depth gradient in the western Bornholm Sea". MA thesis. University of Hamburg, 2020

[16] A. Nissling and L. Westin. "Egg mortality and hatching rate of Baltic cod (*Gadus morhua*) in different salinities". In: *Marine Biology* 111.1 (1991), pp. 29–32. DOI: `10.1007/BF01986341`

[17] Uwe Waller and Dietrich Schnack. "Development of Baltic cod eggs at different levels of temperature and oxygen content". In: *Dana* 10 (1994), pp. 163–177

The Challenges of the Eastern Baltic Cod

Some of the unique conditions the EBC has to navigate have already been broached, such as the Baltic's brackish and stratified water conditions, leading to an increased osmoregulatory effort and reduced reproductive volume. These conditions and ongoing changes in the Baltic ecosystem have resulted in several, likely interconnected, challenges for the EBC.

Eastern Baltic Cod Fisheries Decline

One source of data on the development of the EBC are its fisheries. Fisheries science aims to record (at the very least) data about the spawning stock biomass (SSB - biomass of sexually mature fish), the landings, the fishing mortality (F - all fish removed from the population by the fisheries) and the recruitment (the amount of fish entering into the population) for each fish stock and year (see Figure 4 for this data on the EBC).

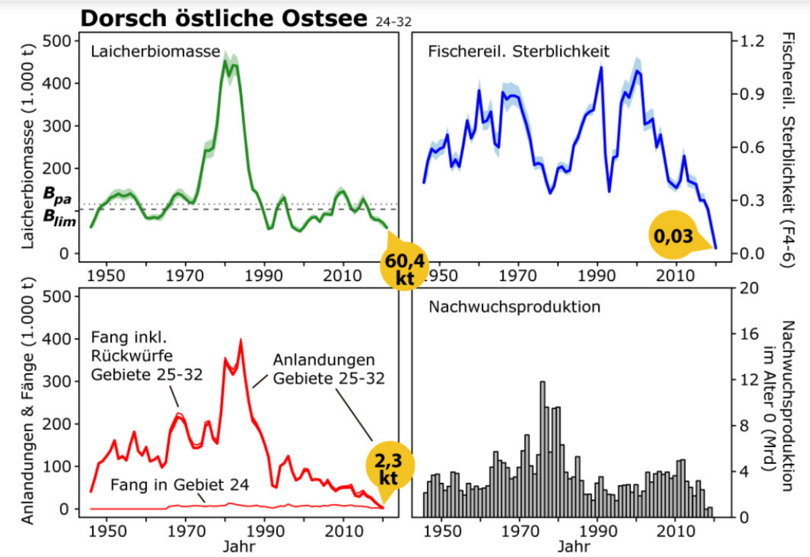


Figure 4: Bestandsentwicklung (engl.: stock development) of the Eastern Baltic cod (by Fischbestände Online, 2022)

Laicherbiomasse →
Spawning stock biomass SSB
B_{pa} → Biomass precautionary approach
B_{lim} → Biomass limit reference point

Fischereil. Sterblichkeit →
Fishing mortality F

Anlandung und Fänge →
Landings and Catches

Nachwuchsproduktion →
Recruitment

For the EBC this data shows a highly variable development during the last 70 years. Throughout the 1950s and 60s biomass was mostly stable and catches slowly trended upwards. After a strong increase in biomass in the mid-1970s, maximum catches of almost 400 000 t were achieved (see Figure 5 for illustration of the condition of fish caught during this period). From the mid-1980s, however, the population quickly decreased again. With high fishing mortality, the stock became severely overfished, mainly due to excessive legal catches and, in addition, considerable illegal catches.

After the stock exhibited some trends between 2008 and 2015, which were interpreted as a sign of recovery, a steady decrease in biomass has been observed ever since. As of 2017, it has been below the limit value of the precautionary approach (B_{lim} - limit reference

Figure 5: The fisheries scientist Eero Aro and a big catch of cod in the southern Baltic Sea in 1987.

point for SSB). Fishing mortality has been reduced since 2012, with the lowest value in the time series being reached in 2021. However, it appears that natural mortality has increased significantly in recent years and is now eight times higher than the fishing rate. Additionally, recruitment has been weak for almost two decades and was the lowest in the time series in 2018 [18] (compare Figure 4).

Important Drivers of EBC Stock Development

The current decline of the Eastern Baltic cod stock is not limited to the significant increase in natural mortality and decrease in spawning stock biomass and recruitment. There has also been an alarming decrease in the body condition (see Figure 6 for illustration of condition and comparison to Figure 5). Several possible drivers of these developments have been identified and are under ongoing scientific investigation. A short introduction of those drivers currently considered likely to contribute significantly and of related phenomena will be given in this section.

Figure 6: Eastern Baltic Cod in poor condition by Peter Ljungberg

Hypoxia and Anoxia in the Baltic Sea

One of the main distinguishing traits of the Baltic Sea is that its stable stratification makes it dependent on infrequent events, MBIs, for the aeration of its deeper layers. Additionally, its almost landlocked position limits water exchange. At the same time, an increase in the use of fertilisers by several countries whose agricultural run-off ultimately drains into the Baltic Sea has led to significant eutrophication. This abundance in nutrients supports a notable increase in microorganisms and change in the composition of planktivorous organisms, which is a drain on the already limited oxygen supply of its deeper waters [19].

This affects the cod in multiple ways: it decreases the Reproductive Volume impacting recruitment, it impacts the Benthos (i.e., the community of organisms that populate the seabed), limiting or possibly even removing a source of high-quality prey in deeper basins and it causes physiological stress when cod are forced to reside in hypoxic waters. Limited oxygen in deeper waters is a defining characteristic of the Baltic Sea ecosystem, and it can be considered a likely primary driver of current developments.

Shift in Diet – Decrease in Food Quality

The shift of dominant pelagic prey species from herring to sprat means there has also likely been a corresponding shift in the diet of EBC [20]. Since the two species have different nutrient profiles, this could have had a detrimental effect on the fitness of individual cod [21,22].

The benthos provides several varied prey species, especially for smaller cod. However, the depletion of oxygen near the seabed has led to a substantial decrease in species variation and the abundance of benthic organisms, particularly in deeper and more northern parts

[18] ICES (2019). Benchmark Workshop on Baltic Cod Stocks (WKBALTCOD2). ICES Scientific Reports. 1:9. 310 pp. DOI: `10.17895/ices.pub.4984`

[19] Natalie Loick-Wilde et al. "Stratification, nitrogen fixation, and cyanobacterial bloom stage regulate the planktonic food web structure". In: *Global Change Biology* 25.3 (2019), pp. 794–810. ISSN: 1354-1013, 1365-2486. DOI: `10.1111/gcb.14546`

[20] K Haase et al. "Diet of dominant demersal fish species in the Baltic Sea: Is flounder stealing benthic food from cod?" In: *Marine Ecology Progress Series* 645 (2020), pp. 159–170. ISSN: 0171-8630, 1616-1599. DOI: `10.3354/meps13360`

[21] M. C. Rojbek et al. "Forage fish quality: seasonal lipid dynamics of herring (Clupea harengus L.) and sprat (Sprattus sprattus L.) in the Baltic Sea". In: *ICES Journal of Marine Science* 71.1 (2014), pp. 56–71. ISSN: 1054-3139, 1095-9289. DOI: `10.1093/icesjms/fst106`

[22] Henrik Svedäng et al. "Compensatory Feeding in Eastern Baltic Cod (*Gadus morhua*): Recent Shifts in Otolith Growth and Nitrogen Content Suggest Unprecedented Metabolic Changes". In: *Frontiers in Marine Science* 7 (2020), p. 565. ISSN: 2296-7745. DOI: `10.3389/fmars.2020.00565`

of the Baltic Sea. Long-term hypoxia and anoxia may have eradicated the benthic fauna in those more affected areas. This reduces the variety of prey for the EBC and it remains to be determined how this impacts its access to essential nutrients.

Density

Density effects are also considered a factor that limits the amount of prey available, particularly for younger cod with specific habitat requirements and exhibiting relatively localised distributions after settlement. This effect assumes that given the unfavourable spatio-temporal distribution of year-classes and a reduced suitable habitat, the density of cod could exceed the carrying capacity at these more minor points in time and space, resulting in a devastating increase in competition. This reduced availability of resources would impact both the overall as well as the individual state of cod [23].

Thiamine and Other Micronutrients

Assuming a eutrophication-driven reorganisation of all trophic levels in the Baltic Sea, this would also include those trophic levels responsible for the primary production of micronutrients. A micronutrient that is under current investigation for possible impact on the condition of the EBC is thiamine [24].

Thiamine, also known as vitamin B_1, is essential to higher organisms, as they lack the ability to synthesise it yet rely on it, as it is required for many cellular processes. A deficiency in thiamine for fish results in neurological disorder, a decreased functioning of the immune system and lower reproductive viability.

Increased Prevalence of Parasites

Parasites are prevalent in all ecosystems and the Baltic Sea is no exception. For the EBC the seal worm (*Pseudoterranova decipiens*), the liver worm (*Contracaecum osculatum*) [25] and the herring worm (*Anisakis simplex*) [26] are currently being investigated as possible drivers of its decline. These parasites infest the liver or other internal organs. The infestation of the liver is a major stressor for cod as it is the organ used for energy storage in the form of lipids. Infestation of the cod liver thus directly impacts the condition of individuals and a subsequent second-order impact on the resilience of the stock as low condition could reduce fecundity (the number of viable eggs per female).

The life cycle of many parasites requires being eaten by their host (trophically transmitted parasites). Likewise, many parasites have a designated animal species in which they reproduce (final host). For the seal worm and the liver worm, the end host is the seal. After they reproduce, their eggs are spread with the seals' feces. These eggs are then ingested by benthic invertebrates, which are in turn ingested by cod and subsequently infest their flesh and liver, respectively. It has been hypothesised that the increase of seals in the Baltic Sea creates an increase of the final host of cod parasites, and it will

---

[23] H.-H. Hinrichsen et al. "Spatio-temporal dynamics of cod nursery areas in the Baltic Sea". In: *Progress in Oceanography* 155 (2017), pp. 28–40. ISSN: 00796611. DOI: `10.1016/j.pocean.2017.05.007`

[24] Josefin Engelhardt et al. "Severe thiamine deficiency in eastern Baltic cod (*Gadus morhua*". In: *PLOS ONE* 15.1 (2020). Ed. by Ilaria Corsi, e0227201. ISSN: 1932-6203. DOI: `10.1371/journal.pone.0227201`

[25] M Sokolova et al. "Spatial patterns in infection of cod *Gadus morhua* with the seal-associated liver worm Contracaecum osculatum from the Skagerrak to the central Baltic Sea". In: *Marine Ecology Progress Series* 606 (2018), pp. 105–118. ISSN: 0171-8630, 1616-1599. DOI: `10.3354/meps12773`

[26] S Zuo et al. "Host size-dependent anisakid infection in Baltic cod *Gadus morhua* associated with differential food preferences". In: *Diseases of Aquatic Organisms* 120.1 (2016), pp. 69–75. ISSN: 0177-5103, 1616-1580. DOI: `10.3354/dao03002`

have to be investigated if there is a connection with the increase in poor-condition cod [27].

Recruitment

Recruitment of the EBC depends on several factors [28]. Important parameters are the quality of the eggs and the suitability of the environment in which they are laid. As introduced above, cod eggs require sufficiently dense water to keep them from sinking and sufficiently oxygenated to keep them healthy. Cod eggs vary to quite an extent in their size depending on the size and health of the female that produced them. Therefore cod that are small, young, or in less than optimal condition both in body mass or in health might produce eggs not able to survive to hatching, and if the condition of the female is extremely poor, these females might even skip spawning altogether [29].

Behaviour

One of the least explored aspects of Eastern Baltic Cod ecology is its behaviour both in terms of migration and in terms of collective memory or knowledgeable mature individuals, sometimes termed scouts [30]. While the question of the existence of such knowledgeable individuals is quite elusive, more ground is being gained in the investigation of migrations. The cod conduct both yearly and daily migrations. Yearly migrations are likely motivated by the movement of prey and the urge to reproduce. Daily migrations are typically vertical or up- and downshore (which involves a vertical component) and likely serve to optimize several factors such as prey availability, abiotic optimisation for oxygen and temperature and possibly avoidance of predators.

Lines of Investigation Relevant to Management: Ageing and Stock Mixing

Depending on the particular assessment model used, management requires information on age and, for adjacent stocks of the same species how the stocks overlap spatially. Both of these requirements apply to the Eastern Baltic cod.

Ageing of cod is accomplished by otolith reading. An otolith (oto-lith: ear-stone) is found in the inner ear, mainly composed of calcium carbonate. When extracted, it is sown in half or into a very slim segment it shows rings. These rings are then interpreted and used to age fish. In many temperate fish, the calcium produces two types of rings, a translucent and an opaque ring corresponding to periods of favourable and less favourable conditions. When stock assessment relies on a known age-length key, proper ageing of fish is vital. However, age reading of EBC otolith is notoriously uncertain, and direct age information is no longer used in the assessment.

The habitat of Western and Eastern Baltic cod overlap in the region of the Arcona Basin or SD24 (see Figure 7). Since this is also a

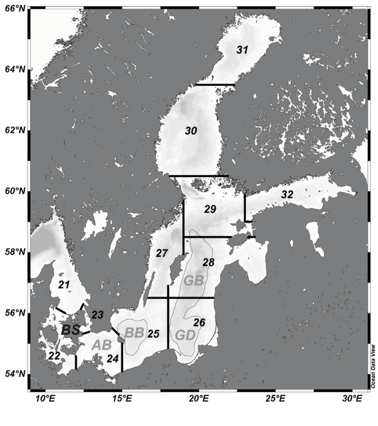


Figure 7: ICES subdivisions (SD) in the Baltic Sea. BS: Belt Sea (SD 22), AB: Arkona Basin (SD 24), BB: Bornholm Basin (SD 25), GD: Gdansk Deep (SD 26), GB: Gotland Basin (SD 28).

[27] A.C. Setyawan et al. "Baltic cod endohelminths reflect recent ecological changes". In: *Journal of Helminthology* 94 (2020), e155. ISSN: 0022-149X, 1475-2697. DOI: `10.1017/S0022149X20000176`

[28] Friedrich W. Köster et al. "Eastern Baltic cod recruitment revisited—dynamics and impacting factors". In: *ICES Journal of Marine Science* 74.1 (2017). Ed. by Howard Browman, pp. 3–19. ISSN: 1054-3139, 1095-9289. DOI: `10.1093/icesjms/fsw172`

[29] M. Mion et al. "Effect of fish length and nutritional condition on the fecundity of distressed Atlantic cod *Gadus morhua* from the Baltic Sea: POTENTIAL FECUNDITY OF BALTIC *G. MORHUA*". in: *Journal of Fish Biology* 92.4 (2018), pp. 1016–1034. ISSN: 00221112. DOI: `10.1111/jfb.13563`

[30] G. De Luca et al. "Fishing out collective memory of migratory schools". In: *Journal of The Royal Society Interface* 11 (2014), pp. 20140043–20140043. DOI: `10.1098/rsif.2014.0043`

region where a considerable amount of fishing takes place, the ratio of the two stocks in this area has to be considered when assigning quotas. Stock assignment is achieved by either direct genetic analysis or otolith shape analysis against a genetically validated baseline of each stock[31].

It is currently under investigation whether the overlap of the Eastern and the Western Baltic cod stocks might be only present in 2D, whereas in actuality, the two stocks occupy different depth strata, with the Western Baltic cod swimming and foraging in shallower waters and the Eastern Baltic cod using mostly waters of greater depth.

Multi-Driver Investigations

Although the prominent lines of investigations concerning the decline of the EBC have been introduced individually, no explanation makes sense in isolation. This emphasises a point of tension when working within a system as interconnected as an ecosystem: scientific investigation is often conducted with a single focus, while the issues of ecosystems are typically inextricably interconnected. Single focus investigations are more likely to make and clearly communicate significant progress. However, the relative impact, the interdependencies, and most importantly, the underlying mechanics of the interaction of drivers need to be uncovered to support a thorough understanding of the causes and mechanisms driving stock dynamics and possible means to ensure recovery of the stock.

Some combination of drivers have already been investigated, such as the combined effect of climate change and eutrophication[32] or the reduced abundance of benthic prey due to an increase in hypoxic areas combined with a lack of spatio-temporal overlap with pelagic prey [33]. However, there continues to be a lack of a sufficiently robust understanding to support the stock in its recovery, which emphasizes the need for a structured approach [34].

Eastern Baltic Cod as an Example of a Complex System

Making explicit the abstract properties of the Eastern Baltic cod and the Baltic Sea, which supports it and of which it is a part, can be considered a requirement to investigate its decline and forge a path to its recovery. Only if the abstract nature of its properties are understood a formal representation can be soundly judged for its suitability. With such suitability confirmed, investigating the specific properties can become the main focus of the required scientific efforts.

"What Is in a System?"

"What's in a name? That which we call a rose by any other name would smell as sweet" (from Romeo and Juliet by William Shakespeare)

A deceptively simple definition of system is:

> "A system is what is distinguished as a system".[35]

The candid nature of this statement reflects that 'system' is a *concept*

---

[31] Fm Schade, P Weist, and U Krumme. "Evaluation of four stock discrimination methods to assign individuals from mixed-stock fisheries using genetically validated baseline samples". In: *Marine Ecology Progress Series* 627 (2019), pp. 125–139. ISSN: 0171-8630, 1616-1599. DOI: `10.3354/meps13061`

[32] Sieme Bossier et al. "Integrated ecosystem impacts of climate change and eutrophication on main Baltic fishery resources". In: *Ecological Modelling* 453 (2021), p. 109609. ISSN: 03043800. DOI: `10.1016/j.ecolmodel.2021.109609`

[33] Stefan Neuenfeldt et al. "Feeding and growth of Atlantic cod (*Gadus morhua* L.) in the eastern Baltic Sea under environmental change". In: *ICES Journal of Marine Science* 77.2 (2020). Ed. by Henn Ojaveer, pp. 624–632. ISSN: 1054-3139, 1095-9289. DOI: `10.1093/icesjms/fsz224`

[34] Margit Eero, Massimiliano Cardinale, and Marie Storr-Paulsen. "Emerging challenges for resource management under ecosystem change: Example of cod in the Baltic Sea". In: *Ocean & Coastal Management* 198 (2020), p. 105314. ISSN: 09645691. DOI: `10.1016/j.ocecoaman.2020.105314`

[35] Brian R Gaines. "General systems research: quo vadis?" In: *General Systems Yearbook* 24 (1979), p. 15

rather than a *thing* and that a choice is being made as to what is part of any given system and what is not .

Translated into more tangible qualities is the dictionary definition of a system:

> "a group or set of related or associated things perceived or thought of as a unity or complex whole".[36]

Again, there is an aspect of choice, made either subconsciously by our perception or consciously by what we choose to group together.

Both definitions have in common that choices are being made. It could therefore be argued that what defines a system is neither intrinsic nor objective. This leads to the conclusion that it is a not only pragmatic but sound to resize any given system, when investigating it, to the expanse most useful to the respective investigation.

Properties of Complex Adaptive Systems

As stated in the introduction of this section, investigating specific properties of a system can profit by first explicitly stating the abstract properties of the system under investigation. Ecosystems are an example of complex adaptive systems (CAS), therefore, some of the relevant properties of complex adaptive systems will be introduced here.

Complexity

When the behaviour of a system can not easily be inferred from its properties, it is understood to be complex. However, currently, there is no agreed-upon technical definition of complexity but rather the assumption that if the properties introduced next (non-linearity, emergence, self-organisation) apply, the system in question would be deemed complex. Therefore complexity can be understood as not being straightforward and exceeding complicated (i.e., involving many different parts and challenging, but possible, to understand) in its elusiveness.

Non-linearity [37]

A non-linear system is one where the change in input is not proportional to the change in output and includes a wide variety of relationships. Mathematically, every polynomial, including an indeterminate with an exponent different from 1, represents a non-linear relationship. Phenomenological, which is much more illustrative, non-linear relationships can reveal themselves in the form of intervals (i.e., holding your breath ten seconds out of every sixty seconds is fine, but holding your breath ten minutes out of every sixty minutes might not be good), growth curves (i.e., the same capacity for reproduction will yield quite different proportions in growth dependent on the supporting system and the absolute number of individuals) and many more.

[36] *system, n.* In: *OED Online*. Oxford University Press, 2021

[37] "Using a term like non-linear science is like referring to the bulk of zoology as the study of non-elephant animals." - Stansilaw Ulam

Nested Sets, Nested Hierarchies or Multi-Levelness

The components of the complex system might themselves be complex systems (although not always for ecological systems). The overall ecosystem is a subsystem of the earth's system, a stock of fish is the subsystem of the ecosystem, a school of its individuals is again a complex system, and so is each individual of this school, as are their organs and the cells constituting these organs.

'Nested sets' are not to be confused with 'nestedness' which has a specific definition in ecology and refers to a well-defined mesure of sturcture for the relations of species and sites or species and species

In the investigation of ecological systems, this has the advantage that concentrating on subsystems limits the complexity of the system under investigation but has the disadvantage that upward and downward causations might not be taken into consideration in cases where they are vital for a thorough understanding.

Self-Organisation

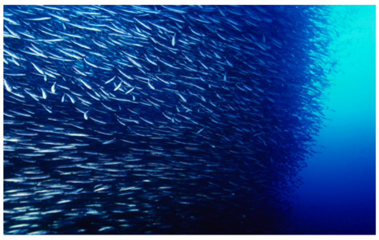


Figure 8: A school of fish: the shape of the school is a result of self-organisation resulting from the interactions of the individuals and results in the emergent property of its particular shape and top-down impact of (limited) protection (by OpenStax College cc)

Self-organisation is the process of acquiring and maintaining a structure (spatial, temporal or functional - see Figure 8) without external control. In biology and ecology, self-organisation often occurs together with emergence (see below). Self-organisation without emergence would require the maintaining of a structure without the emergence of a pattern or property on the next-higher system-level [38].

Emergence

Emergence is a process whereby larger entities (i.e., things with distinct and independent existence), patterns and regularities arise through interactions among smaller or simpler entities that themselves do not exhibit such properties. In this sense, the shape of a school of fish is an emergent property of the interactions of its constituting individuals (see Figure 8), and life is an emergent property of chemistry.

It has been suggested to differentiate between 'strong' and 'weak' emergence. It has also been argued that this is merely a question of the current depth of understanding. Nevertheless, in this line of argumentation, the goal of the scientific investigation of a complex system is to turn 'strong' emergence into 'weak' emergence.

Adaption

The 'adaptive' of a complex adaptive system means that such systems possess and implement the capability to respond to change, both to changes in their respective environment and changes in their constituting parts.

For ecosystems, this means they will adapt to changes in their environment, such as climate change and will adapt to changes in their constituting parts, such as an increase in available nutrients due to eutrophication.

[38] Tom De Wolf and Tom Holvoet. "Emergence Versus Self-Organisation: Different Concepts but Promising When Combined". In: *Engineering Self-Organising Systems*. Ed. by Sven A. Brueckner et al. Vol. 3464. Berlin, Heidelberg: Springer Berlin Heidelberg, 2005, pp. 1–15. ISBN: 978-3-540-26180-3 978-3-540-31901-6. DOI: `10.1007/11494676\_1`

Eastern Baltic Cod Management

Management of the Eastern Baltic Cod takes place within the Frameworks of ICES, the European Union and the nation-states bordering the Baltic Sea. Here some of the principles, institutions and procedures relevant to the process of past and future management of the Eastern Baltic Cod stock will be introduced.

The Basis for Management: Tragedy of the Commons and MSY

When a resource is available to all community members but access is not regulated, this can lead to overexploitation due to a lack of coordination. This phenomenon is known as 'the tragedy of the commons' [39]. Historically 'the common' refers to common land on which members of the associated community could let their animals graze. When too much and/or uncoordinated grazing occurred, this practice became unsustainable and the resource would be lost to all.

Because marine fish do not have an owner in any practical sense, overexploitation, or rather overfishing, will also occur with this common resource as soon as technological advancements make it possible, but fishing remains unregulated. As described in the introduction, there are two main drawbacks to overfishing: it has an extreme impact on the respective ecosystem and it reduces overall revenue. The impact on the ecosystem is usually a destabilising one, as the excessive reduction of the number of individuals of a stock not only reduces the resilience of this particular stock but also reduces the resilience of the supporting ecosystem by drastically impacting one of its key components. The loss of revenue is understood in a time frame spanning several harvest seasons: though maximum extraction might produce one year of high revenue, the impact on growth potential reduces the revenue of future years. Therefore common resources, such as fisheries, have to be managed if the goal is to maximise and sustain their use.

Assuming stable environmental conditions, the growth of fish stocks approximately follows a logistic growth curve (see Figure 9), where at the lowest population size, growth is slow (due to the small numbers of individuals reproducing) and subsequently increases with its population until a point of inflexion is reached, from whereon available resources cap growth, growth then continues to slow until the carrying capacity of the respective system entirely prevents any further increase of the population.

To ensure the highest overall revenue, the goal is to take the largest harvest which can be taken from a fish stock over an indefinite period: the maximum sustainable yield (MSY) [40]. Assuming logistic growth, the goal would be to keep the population at the point of inflexion where the number of individuals added is the largest. Therefore the maximum of the derivative of the growth function is the (theoretical) optimum at which to keep the population of a stock to ensure the maximum sustainable yield. Marine fish generally reproduce on a

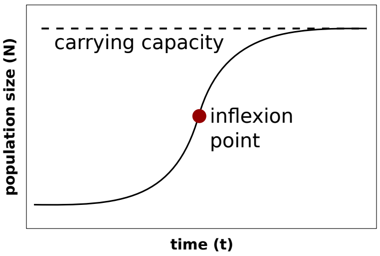


Figure 9: Logistic/Sigmoidal growth curve for a population.

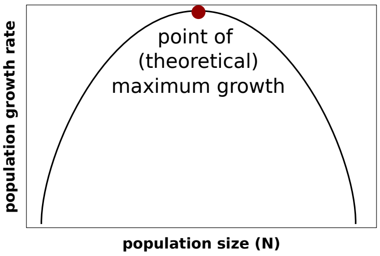


Figure 10: Derivative of Logistic/Sigmoidal growth curve for a population.

[39] Garrett Hardin. "The Tragedy of the Commons: The population problem has no technical solution; it requires a fundamental extension in morality." In: *Science* 162.3859 (1968), pp. 1243–1248. ISSN: 0036-8075, 1095-9203. DOI: `10.1126/science.162.3859.1243`

[40] Milner B. Schaefer. "Some aspects of the dynamics of populations important to the management of the commercial marine fisheries". In: *Inter-American Tropical Tuna Commission Bulletin* 1.2 (1954), pp. 23–56. DOI: `10.1016/S0092-8240(05)80049-7`

yearly cycle, therefore, the abstract growth rate can be interpreted as a yearly growth rate and be translated into the amount of 'surplus' population suitable to be extracted. In practice, the goal is to keep the population slightly ahead of the maximum to account for ecological variability and margins of error in assessing the stock.

ICES, EU and Bordering Nation-States

Concurring with the implications of the tragedy of the commons, the international community is in agreement that management of the shared fisheries resources is required (although there can be, at times, considerable disagreement about what the specifics of this management should be). Therefore several institutions and political units work together and fulfil different functions to promote a sustainable, stable and fair allocation of marine fish.

ICES – Scientific Services

The International Council for the Exploration of the Sea (ICES[41]) was founded in 1902 to encourage an international and scientific approach to the harvesting of fish stocks. It currently has 20 member nations (Belgium, Canada, Denmark, Estonia, Finland, France, Germany, Iceland, Ireland, Latvia, Lithuania, the Netherlands, Norway, Poland, Portugal, the Russian Federation, Spain, Sweden, the United Kingdom and the United States) and thereby covers a substantial area of the northern hemisphere.

ICES is responsible for coordinating the scientific assessment of the state of fish stocks and ecosystems and provides scientific advice to legislators. It has two main branches, one for scientific work (SCICOM - the scientific committee) and one for providing scientific advice (ACOM - the advisory committee).

EU Fisheries Ministry and Nation-States - Legislation

The EU takes part in the legislative process of fisheries management. Scientific advice provided by ICES is considered by the European Commission and forwarded to the council of ministers who agree on TACs (total allowable catches). Employing a stable international allocation key, these catches are then assigned to the respective nation-states, which in turn provide quotas to their fishing licensees.

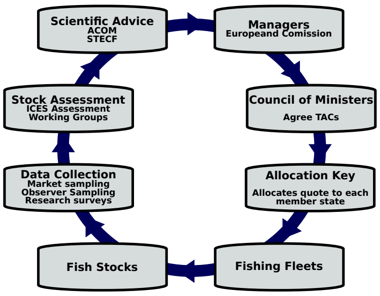


Figure 11: The annual management cycle of fish stock within Europe as an ICES areas of competence.

Public Fisheries Research in Germany

Each nation-state sharing fisheries resources is tasked with collecting and providing both the required data and the scientific input, in the form of research and scientific personnel, to create a good scientific basis for management. In Germany, the responsible research institutes to tie-in with the development of scientific advice for catch-quotas are the Thünen-Institutes: The Thünen-Institute for Baltic Sea Fisheries in Rostock, mainly responsible for surveying the Baltic stocks and the Thünen-Institute for Sea Fisheries in Bremerhaven, mainly responsible

[41] French: CIEM, *Conseil International de l'Exploration de la Mer*

for surveying the North Sea stocks (please note that responsibilities are partly shared and overlap).

Advice: A Scientific Product

The basis for the legislative process of allocating catch quotas is scientific Advice. The 'Advice' produced by ICES is not merely scientific input, but an established scientific 'product' that results from a highly refined and scrutinized workflow (e.g., 2021 Advice for the EBC[42]).

The Advisory Process

Advice is typically requested from public authorities such as the governments of ICES member countries or the European Commission (EC). For most commercially harvested stocks, Advice is produced yearly as a 'standing order' (single species assessment). Nation-states that claim fishing rights typically maintain dedicated research institutes that survey data and conduct required research (e.g., the Thünen Institutes). Data is collected from the sampling of commercial fisheries and conducting fisheries-independent surveys and further analysis of the data collected, such as age readings. These institutes subsequently furnish the respective ICES working groups with national data and scientists.

For the single species assessment, working groups call for and accumulate data, typically for each stock. This data is then evaluated often by use of stock assessment models to make predictions about the 'surplus production' (according to the MSY approach). Stock assessment models are typically data-driven mathematical models, often a combination of a population model and a data model combining functional knowledge about the population (e.g., biomass dynamics, age structure, length structure, etc.) with current data on the stock (e.g., catches, abundance, recruitment, etc.).

Results from the stock assessment models and other relevant output from the respective working group are then synthesized into Advice.

[42] ICES. "Cod (*Gadus morhua*) in subdivisions 24-32, eastern Baltic stock (eastern Baltic Sea". In: (2021). Publisher: ICES. DOI: `10.17895/ICES.ADVICE.7745`

Chapter II
Supporting Management of Complex Systems by Modelling and Simulation

Science is built of facts the same way a house is built of bricks; but an accumulation of facts is no more science than a pile of bricks is a house.

Henri Poincare

Concepts and Terminology of Modelling and Simulation

As do all scientific disciplines, modelling and simulation has a number of core concepts and accompanying definitions. These will be introduced first to provide the necessary foundations for its application to the Eastern Baltic cod (EBC). Most of the following citations are taken from CELLIER[43].

System

In Chapter I several definitions were given for the term 'system' to introduce its relation to fisheries research and the concept and properties of complex adaptive systems (CAS). To make the present section more self-contained, the most crucial definition will be repeated here:

"A system is anything that can be distinguished as a system." [44].

To which CELLIER[43] adds that systems have 'inputs' (variables which influence the behaviour of a given system) and 'outputs' (variable which are determined by the behaviour of a given system) which sets up the additional definition by ZEIGLER that:

'A system is a potential source of data' [45].

Experiment

This second definition of system then leads CELLIER directly to his definition of an experiment:

"An experiment is the process of extracting data from a system by exerting it through its inputs."[43]

It is worth noting that this is true for all experiments, not only those conducted employing computers and/or models. Wet-lab experiments conducted in the context of fisheries research extract data from the well-defined system of a fish by exerting control over its environment.

[43] François E. Cellier. *Continuous System Modeling*. New York, NY: Springer New York, 1991. ISBN: 978-1-4757-3924-4 978-1-4757-3922-0. DOI: `10.1007/978-1-4757-3922-0`

[44] Brian R Gaines. "General systems research: quo vadis?" In: *General Systems Yearbook* 24 (1979), p. 15

[45] Bernard P. Zeigler. *Theory of modelling and simulation*. New York: Wiley, 1976. 435 pp. ISBN: 978-0-471-98152-7

Model

Relying on the definition of 'system' and 'experiment' the following definition of 'model' is attributed to Minsky[46]:

"A model (M) for a system (S) and an experiment (E) is anything to which E can be applied in order to answer questions about S"

Again, models do not have to be computer models, and they can range from the abstract to the physical. The critical property is that they aim to supply the same output as the system they are modelling for a given input. Thereby they allow for experimentation regarding the original system at reduced effort.

Simulation

Korn and Wait[47] define a simulation simply as:

"A simulation is an experiment performed on a model."

Note that even for simulations computers are not a requirement, simulations can easily be performed with mental or physical models [48]. The critical requirement remains that the relationship between input and output is maintained.

In the context of this thesis the focus will be on formal, executable models and if not stated otherwise from this point on 'model' and 'simulation' will refer to computer models (implemented in a programming language) and computer simulations (executed by a computer).

Modelling and Simulation Study

A modelling and simulation study then puts the concepts and artefacts defined above into practice, using them to work towards a scientific objective of a particular application domain. Now the use of modelling and simulation becomes more elaborate. Rather than building one model once and conducting one experiment, these steps are undertaken iteratively and in identifiable phases. Sometimes these phases can be sorted into the so-called 'modelling and simulation life-cycle' which has many forms, a very basic version of which is:

1.) There exists a real-world (problem) - undertake conceptual modelling 2.) There exists a conceptual model - implement computer model 3.) There exists a computer model - undertake experiments with the computer model 4.) There exist solutions/improved understanding - implement into real-world (see Figure 12).

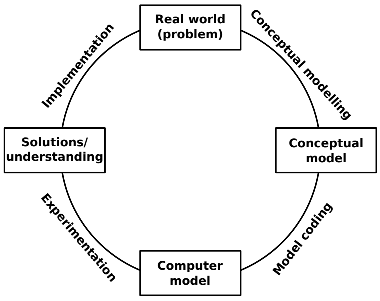


Figure 12: A basic modelling and simulation life cycle based on Brooks and Robinson (2000).

The process can be repeated at this point, as it, for example, is in the management of fish stocks. However, each process, such as the implementation of models, can be iterative in and of itself. Within these processes, additional steps are subsumed, such as calibration and validation, and additional infrastructure is advisable, such as documentation.

More elaborate versions of the modelling and simulation life-cycle

---

[46] Marvin Minsky. *Models, Minds, Machines*. Proceedings IFIP Congress, pp 45-49. 1965
Note that the orginial source could not be retrieved succesullfy and this reference is taken from Cellier[43]

[47] Granino A. Korn and John V. Wait. *Digital continuous-system simulation*. Englewood Cliffs, N.J: Prentice-Hall, 1978. 212 pp. isbn: 978-0-13-212274-0

[48] One example are the Waterloopbos, a facility that has been using large scale physical models long before the computer age to simulate water management scenarios in the Netherlands and globally

taking these additional aspects into account of course exist, notable examples are presented in [49] and [50].

Applying Modelling and Simulation to the EBC

As described in Chapter I the workflow of management mostly (excluding updates such as Benchmarks) has a repeating annual cycle consisting of: legislative tasks, which regulate human activity, the resulting actions then, in part, influences the new state of a fish stock, which then is scientifically evaluated to produce Advice, completing the cycle by informing the legislative tasks.

While it is important not to lose track of the requirements and constraints resulting from the legislative part of this cycle, from here on the focus will be on the scientific remits: understanding the state and dynamics of the Eastern Baltic cod stock, which should be understood to include the state and dynamics of the Baltic ecosystem of which it is a part and by which it is supported. It should be appreciated that this understanding relies to significant extent on the task of data collection and is the prerequisite for sound stock assessment and producing Advice[51] (see Figure 13).

Specifically, it shall be evaluated how and in which manner modelling and simulation can aid in the continued scientific support of consistent management of the Eastern Baltic cod stock and by extension other stocks where management faces comparable challenges. To this end, the many requirements resulting from the characteristics of management, fisheries science, complex adaptive systems and departmental research need to be identified and presented in detail so they can subsequently be matched with suitable approaches, methods and tools both existing and to be developed.

To provide a structured delivery the proposed general approach will be covered first in order to establish its operating principle and outline its placement within the fisheries science workflow. Based on the then established framework the second half of this chapter will address the necessary specific tools, methods and supporting infrastructure to successfully put the general approach into practice.

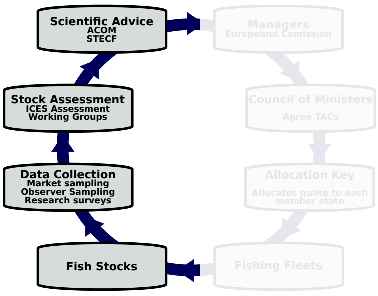


Figure 13: The annual management cycle of fish stock within Europe as an ICES area of competence: legislative half greyed out highlighting the state of the fish stocks and subsequent scientific parts of the annual cycle.

Fundamental Objectives and Requirements

Modelling and simulation is a well established tool to produce requisite results, for example the modelling and simulation of a bridge to gauge the viability of an engineering project or, as is the case in fisheries management, the modelling and simulation of a fish stock to gauge the optimal removal rate for its respective fisheries. However, the act of modelling and simulation in and of itself, i.e., conducting a modelling and simulation study can also be considered a tool. Not (only) to provide requisite results (although theses are also quite valuable) but rather to provide insight about the system with which it is concerned [52]. This aspect of modelling and simulation is what is meant by 'general approach' in this thesis and which will be the

"Modelling means the process of organizing knowledge about a system."

[49] R G Sargent. "Verification and validation of simulation models". In: *Journal of Simulation* 7.1 (2013), pp. 12–24. ISSN: 1747-7778, 1747-7786. DOI: `10.1057/jos.2012.20`

[50] Osman Balci. "Verification, Validation and Accrediation of Simulation Models". In: *Proceedings of the 1997 Winter Simulation Conference*. Ed. by S. Andradóttir et al. IEEE. Piscataway, New Jersey, 1997, pp. 135–141. DOI: `10.1109/IEEESTD.1997.8685803`

[51] Advice, the scientific product as defined and implemented by ICES

[52] Bernard P Zeigler. *Multifacetted modelling and discrete event simulation*. Academic Press Professional, Inc., 1984. ISBN: 978-0127784502

subject matter of this section.

Arising Challenges and the Resulting Need to Go Deeper

Chapter I introduced the current developments of the Eastern Baltic cod stock and subsequent challenges for management. Namely there has been a significant decline in the productivity of the stock and the health of its individuals without a clear understanding as to the exact causes of theses developments. This lack of understanding means that there is no basis for the implementation of strategies which could aid in the recovery of the stock [53].

As long as the EBC was embedded in a productive system, it was sufficient to observe it predominately on the 'level of interest', i.e., at the level of the stocks' productivity, without incorporating much information from other system levels. This implicitly assumed that the many underlying and interacting mechanisms, which together resulted in the stocks capacity for recruitment, growth and the carrying capacity of the Eastern Baltic ecosystem, was stable (or varied within observed limits) and that, consequently, the dynamic of the productivity would remain predictable with the perturbation of fishing pressure. Thus, the MSY approach was a reliable metric for management decisions [54].

However, since the Baltic ecosystem is no longer in a stable state (and considerably less productive), likely top-level drivers being eutrophication and/or climate change, the assumption of an underlying equilibrium no longer holds.[55] As a result, system levels can no longer be understood in isolation, especially when they need to be understood well enough to support future management. At this point, there is a need to take into account phenomena across the different levels of the entire ecosystem to establish new, reliable metrics to inform management decisions [56,57].

Consequently, facets of Baltic cod ecology which are typically investigated independently from each other (see section on 'Important Drivers of EBC Stock development' in Chapter I), separated by scientific (sub)-domain (e.g. genetics, parasitology, fishery biology, etc.), line of investigation (e.g. impact of hypoxia, impact of stock mixing, density-dependence, etc.) or both now need to be integrated into a cohesive understanding.

As a result, the pressure on fisheries science to establish the understanding necessary to support the recovery of the stock is high. Simultaneously, this does not only require more intense efforts but likely also an expansion of approach as the regular workflow is geared at assessing an abundant stock in a stable ecosystem [10] whereas now the objective has shifted to assessing a stock with decreasing productivity in a rapidly changing ecosystem [58,59].

It has been relayed how each respective line of investigation concerning likely drivers of the EBC decline has made some and in many cases, significant progress (again, see section on 'Important Drivers of EBC Stock development' in Chapter I). However, the relative impact, the interdependencies and most importantly the underlying mechan-

---

[53] Margit Eero et al. "Eastern Baltic cod in distress: biological changes and challenges for stock assessment". In: *ICES J Mar Sci* 72 (2015), pp. 2180–2186. ISSN: 1095-9289. DOI: `10.1093/icesjms/fsv109`

[54] M Schaefer. "Some aspects of the dynamics of populations important to the management of the commercial marine fisheries". In: *Bulletin of Mathematical Biology* 53.1 (1991), pp. 253–279. ISSN: 00928240. DOI: `10.1016/S0092-8240(05)80049-7`

[55] It is worth noting that the increase in productivity of stock during the 1980s was and is also not understood but was simply accepted.

[56] L. W. Botsford. "The Management of Fisheries and Marine Ecosystems". In: *Science* 277.5325 (1997), pp. 509–515. ISSN: 00368075, 10959203. DOI: `10.1126/science.277.5325.509`

[57] Francois Bastardie et al. "A Review Characterizing 25 Ecosystem Challenges to Be Addressed by an Ecosystem Approach to Fisheries Management in Europe". In: *Frontiers in Marine Science* 7 (2021), p. 629186. ISSN: 2296-7745. DOI: `10.3389/fmars.2020.629186`

[58] Margit Eero, Massimiliano Cardinale, and Marie Storr-Paulsen. "Emerging challenges for resource management under ecosystem change: Example of cod in the Baltic Sea". In: *Ocean & Coastal Management* 198 (2020), p. 105314. ISSN: 09645691. DOI: `10.1016/j.ocecoaman.2020.105314`

[59] Francois Bastardie et al. "A Review Characterizing 25 Ecosystem Challenges to Be Addressed by an Ecosystem Approach to Fisheries Management in Europe". In: *Frontiers in Marine Science* 7 (2021), p. 629186. ISSN: 2296-7745. DOI: `10.3389/fmars.2020.629186`

ics of these drivers need to be uncovered without delay, as the stock shows no sign of recovery even with fisheries being closed [60]. In agreement with the urgency of establishing a robust understanding of EBC dynamics are publications concerned with investigating these relationships (see section on 'Multi-Driver Investigations' in Chapter I).

As the Baltic ecosystem continues to evolve at a fast pace, this effort will also have to go beyond describing 'the new system' assuming just a different steady state, but will rather require gaining a sufficiently deep understanding to allow reliable predictions during ongoing change [61]. Meaning, not only extrapolations of states (e.g.: spawning stock biomass x will now lead to recruitment value y: $\mathrm{f}(x) = y$) but also an extrapolation in the change of mechanisms will be necessary (e.g.: nutrient input will impact the mechanisms $\alpha, \beta, \gamma, \delta, \epsilon$, to be applied, in these various manners and in consequence this will impact the relationship between spawning stock biomass and recruitment thusly: $\mathrm{g}(\{\alpha, \beta, \gamma, \delta, \epsilon\}, x) = y$).

How to Move Forward

While it is unlikely that long-term predictions (i.e. spanning decades or more) of the behaviour of changing ecosystems will be feasible in the near future [62] management of fisheries could already benefit from improved 'directionality' over the span of a few years, identifying key points of intervention while avoiding erratic Advice[63] and thereby maintaining trust. Currently Advice is almost exclusively generated from data about past states of the system, excluding the possibility to predict system states outside of the space of system states that have already been observed (i.e. non-ergodicity of the system can not be grasped). However, by investigating how the many interacting mechanisms produce (novel) system behaviour, likely trajectories of the systems' dynamics, moving it outside of previously observed states, can be uncovered.

In summary the fundamental objective is to reach a depth of understanding about the mechanics of the EBC stock and its supporting ecosystem which allows for pre-emptive rather than reactive management. The requirement is that this goal be achieved as swiftly as possible with a realistic scope of resources.

This aspiration poses several considerable challenges: how to provide an efficient path for integrating mechanisms across scientific domains (facets) and system levels and how to keep this process of integration transparent and easy to access for the different scientists that need to be involved?

General Approach: Multi-faceted Modelling

Considering the complexity of the EBC ecosystem and the multiple and wide-ranging questions pertaining to its recovery, an approach is required that allows to functionally integrate the various individual scientists' partial knowledge of this system into a comprehensive

[60] ICES. "Cod (*Gadus morhua*) in subdivisions 24-32, eastern Baltic stock (eastern Baltic Sea". In: (2021). Publisher: ICES. DOI: `10.17895/ICES.ADVICE.7745`

[61] Matthew R. Evans. "Modelling ecological systems in a changing world". In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 367.1586 (2012), pp. 181–190. ISSN: 0962-8436, 1471-2970. DOI: `10.1098/rstb.2011.0172`

[62] Benjamin Planque. "Projecting the future state of marine ecosystems, "la grande illusion"?" In: *ICES Journal of Marine Science: Journal du Conseil* 73.2 (2016), pp. 204–208. ISSN: 1054-3139, 1095-9289. DOI: `10.1093/icesjms/fsv155`

[63] Advice, the scientific product as defined and implemented by ICES

understanding. This will not only consolidate what is known but also reveal gaps in knowledge that need to be filled. "Multi-faceted" modelling was developed precisely to meet this type of challenge: by using partial models (representing partial knowledge and solutions obtained under constraints imposed by disciplinary or problem-oriented perspectives) with which to capture facets of understanding and to subsequently integrate this formalised understanding [64].

Augmenting Multi-faceted Modelling for Fisheries Science

However, using the approach of multi-faceted modelling in this manner, i.e., to not only model a complex system but to uncover its underlying mechanisms, requires adjustments be made to the practical implementation of the approach. Where the original approach relied on a modular composition of models, coupling outputs of one model component by using them as inputs of another model component, investigating ecological systems requires the understanding of interactions in the form of mechanisms. This requirement is fulfilled by integrating models in a white box manner resulting in larger models holding several, transparently and functionally integrated facets of understanding.

White-box Models

Here 'white-box' refers to models which allow full mechanistic insight as opposed to 'black-box' models which allows none. In reality of course this is seldom a dichotomy but rather a spectrum, where in reality most models are 'grey-box' models. More importantly the level of mechanistic insight is also dependent on factors such as documentation, skill set and familiarity with the type of model.

We Are All Modellers: Mental Models

There is an argument to be made that, at least in regards to mental models, all scientists are modellers. As a pragmatic approach a mental model shall be defined as: 'A mental representation of a system, made up of concepts, connections and interactions which people use to know, understand, and mentally simulate the system they are thinking about'. [65] Based on this definition a (sub)-domain expert has, and is continuously updating, a mental model of 'their system', focusing on what they have deep knowledge on and including what is required from other (sub)-domains to support the structure of their understanding.

Currently these different mental models have to be exchanged and communicated, almost exclusively, through natural language by means of peer-reviewed publications, reports, working groups and conversations. This has three main disadvantages: 1) it does not aim to consolidate the knowledge they contain (the aforementioned modes are all geared more towards scientific discourse and less towards collaborative integration), 2) it is often ambiguous (speaking of 'low oxygen conditions' might mean below 77% (threshold for reduced

[64] Bernard P Zeigler. *Multifacetted modelling and discrete event simulation*. Academic Press Professional, Inc., 1984. ISBN: 978-0127784502

[65] Jay Wright Forrester stated in 1971 that: "The image of the world around us, which we carry in our head, is just a model. Nobody in his head imagines all the world, government or country. He has only selected concepts, and relationships between them, and uses those to represent the real system"

growth) to one scientist while it means below 37% (threshold for increased mortality) to another) and it is not always efficient (reading and writing reports simply takes time, resulting in not everyone being able to be updated on all relevant mental models).

While the workflow of stock assessment models needs to be upheld as a proven approach to deliver the predictions required to produce Advice[66] in a timely fashion I suggest to expand this type of modelling and simulation (which will be referred to as a 'prediction tool') by a separate workflow which uses modelling and simulation as 'thinking tool'. This distinction mirrors the ones made by Evans et al.[67] between a focus on prediction ('tactical models') versus a focus on communication and clarification ('synthetic models for a system'). The overall approach of transparent multi-faceted modelling and its placement within the fisheries science workflow is depicted in Figure 14. As illustrated in the Figure, the emphasis lies on different aspects of the process. Where 'prediction tool' is aimed at providing forecasts based on established understanding of mechanisms, a 'thinking tool' model is aimed at including the perspective of different scientists, integrating their knowledge and thereby updating understanding of mechanisms. Thus the latter supplements the former in the combined goal of producing dependable Advice.

In detail the process can be seen to begin with with the collection of data about the ecosystem, which of course includes surveys but also experiments conducted inside the ecosystem and conducted in wet-labs as well as any activity that leads to insight, we arrive at the aforementioned multitude of mental models. This is the starting point of the 'thinking tool' track.

The nature of the aforementioned mental models naturally results in a certain amount of overlap between the mental models of the different experts on 'their' respective (sub)-systems (e.g.: diet, growth, reproduction, etc.). I propose that integrating these different individual mental models, in a concerted effort, into executable formalisations (simulation models) will necessarily bring the different mental models more into alignment, as ambiguity will have to be resolved and vague concepts will have to be made explicit. Subsequently using the resulting formalisation for hypothesis testing, etc. will bring - the now shared - understanding closer to reality. Both processes can be iteratively repeated until understanding is consolidated on a new level.

At this point questions about what to survey and which newly uncovered mechanisms could be of use to the stock assessment, including predictions of further changes in dynamics, will have new and updated answers and will thereby deliver presorted suggestions for surveys and wet-lab experiments and, given the appropriate path, improved management support (see bold arrows in Figure 14).

Although models always omit large amounts of detail, in this approach simplifications are made by each expert with regards to their respective area of expertise, thus no facet in its entirety, possibly vital to a better understanding, will be excluded.

[66] Advice, the scientific product as defined and implemented by ICES

[67] Matthew R. Evans et al. "Do simple models lead to generality in ecology?" In: *Trends in Ecology & Evolution* 28.10 (2013), pp. 578–583. ISSN: 01695347. DOI: `10.1016/j.tree.2013.05.022`

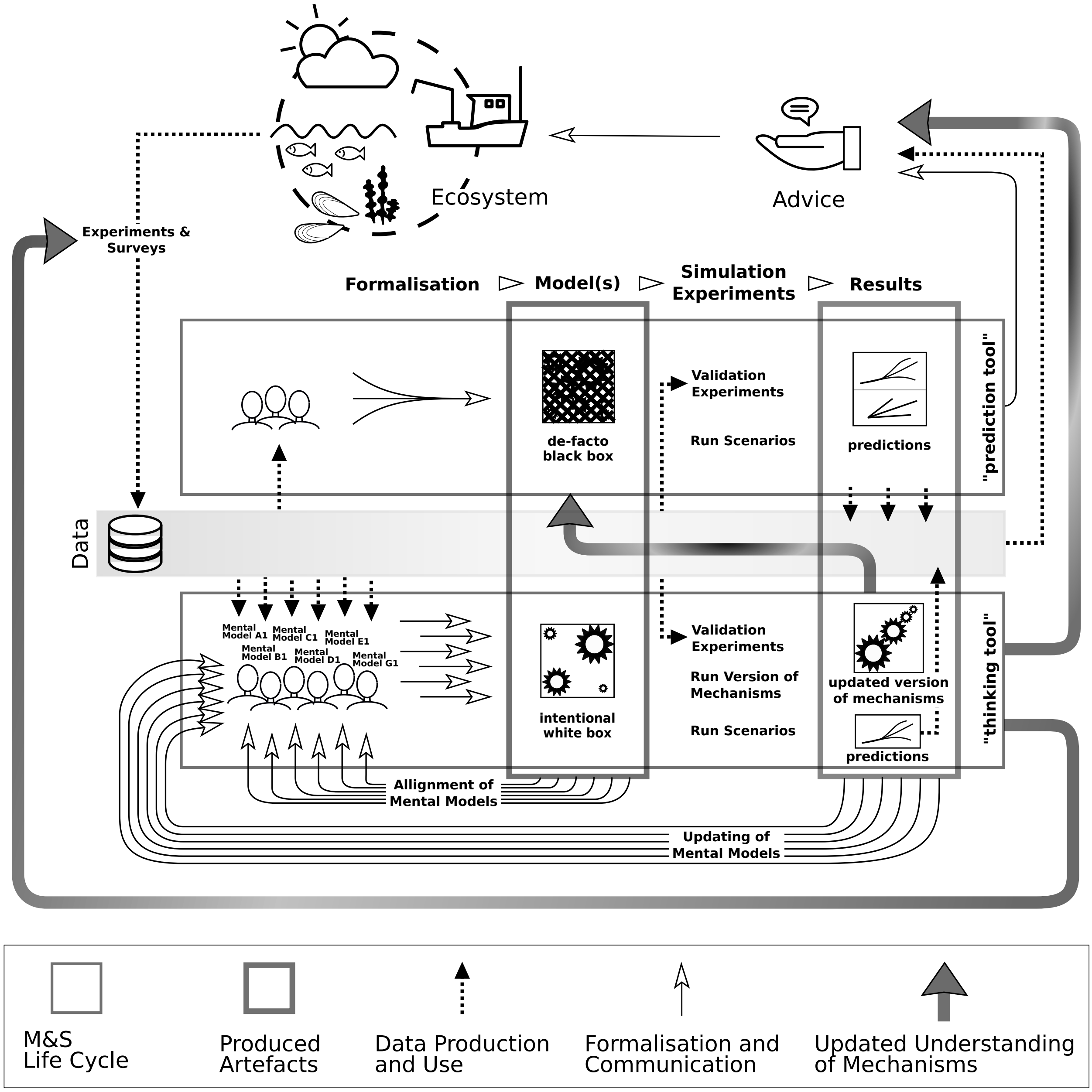


Figure 14: Comparison of Modelling and Simulation life cycles of 'prediction tools' and 'thinking tools'. The real-world problem is posed by the desire to predict the development of stocks and adjust fishing pressure accordingly. In the prediction track, formalisation will typically be focused on producing a working model based on current knowledge and understanding. In the complementary 'thinking tool' track formalisation will be focused on communication and clarification. As discrepancies between predictions and real-world outcomes emerge, submodels will reflect the detailed and deep knowledge and understanding of different facets of the system by the respective experts, and these will subsequently be composed into integrated model(s). Validation and running scenarios takes place in both tracks but the latter is aimed at the investigation of mechanisms. The strength of the prediction tool lies in providing reliable predictions based on the current understanding of the system; the strength of the thinking tool lies in removing ambiguity in the communication of experts, subsequently integrating their knowledge and ultimately updating understanding of the system. The results can then inform the process of producing advice, amend what data to survey and revise the mechanisms and assumption of the next generation of prediction models.

At this point the question arises how each scientist can be asked to participate in *actual* modelling when it has already been made clear that their time is quite limited an their areas of expertise are quite diverse. This will be addressed in the second half of this chapter after a discussion of the general approach.

Discussion: Simulation Models as a Focal Point

The results presented so far aim to meat the arising challenges of managing fish stocks in changing ecosystems by providing a tailor made approach for the functional integration of the distributed knowledge about the stock and its supporting ecosystem.

It can be argued that grasping all relevant aspects of an ecosystem, even if only those pertaining to one component (for example, a particular fish stock), exceeds the capacity of one or a small group of scientist to comfortably handle [68]. Especially keeping in mind that this task has to be completed regularly with changing participants and with limited time as all participants have other, albeit related, duties to fulfil. As input from different scientist is essential the group of participants will have to be expanded.

This larger and heterogeneous group then requires and unambiguous and efficient means of communication. Though anecdotal, the unlikeness of natural language correspondence being able to fulfil this requirements shall still be stated as an issue here [69], if only because it has been the means of communication available in the form of reports and scientific publications but has so far failed to sustain the required insights.

Therefore means of communication, such as formalised simulation models which are more straightforward and inhibit ambiguity are a promising next candidate. In addition executable simulation models not only provide the required lack of ambiguity but also allow for the communication of progressions, i.e., the interactions of different mechanisms their change over time and their shifting dynamics become tangible through the execution of the appropriate simulation experiments. Thereby the means of communication also directly provide the means of testing the formalised understanding, providing the ability to compare simulation results with field observations.

Unlike wet-lab experiments and surveying, simulation experiments are also highly controllable and the number of 'alternative treatments' that can be conducted is exceedingly higher, while less time and facilities are required, and no animals have to die for scientific purposes. For both animals at liberty and hindcasting, controlled experiments can only be conducted with simulation experiments and, additionally, data that would not be surveyable in the field (e.g., how much did they eat during which period, who did they mate with, etc.) can easily be recorded and evaluated.

Having stated these advantages, it is important to point out that modelling and simulation is not a panacea and requires diligent application. Models require validation and scrutiny. They are also

[68] George A. Miller. "The magical number seven, plus or minus two: some limits on our capacity for processing information." In: *Psychological Review* 63.2 (19569), pp. 81–97. ISSN: 1939-1471, 0033-295X. DOI: `10.1037/h0043158`

[69] more than once in seminars and working groups it has been made clear that extensive reports are being produced with great enthusiasm but rarely read

simply abstractions for a specific research question about a particular system [70] and results of simulation experiments should always be interpreted only within this scope of applicability.

Although it has been shown how and where the methodology presented here fits into the life-cycle of fisheries research it has yet to be clarified why more established methods currently available might not be sufficient to achieve the same goal of providing continued robust Advice[71].

Statistical analysis is one of the oldest and most powerful tools in biology, ecology and consequently fisheries science and, therefore, a natural candidate. It allows us to unearth correlations and thereby deepen our understanding of patterns and probable relationships in the systems we study. But as the complexity of the matter under investigation escalates, this method faces diminishing returns, meaning that with the necessity of increasingly statistically complex research, there is a decrease in explanatory power [72]. For the experiment design in wet-lab settings, which depend on replicating treatments for controlling variables, there are additional practical limitations to support sound statistical analysis. With an increasing number of variables to account for, the number of required replicating treatments can quickly become impractically high. Consequently, investigations across different facets and several levels of a complex system, as will be required for a sufficiently deep understanding, are currently out of the realm of this particular tool.

There can also be a temptation, especially if there is abundance of historical data (as is the case for the Baltic Sea) and with the appeal of ever more sophisticated tools available for analysis, to demand or even claim inclusion of functional environmental changes in predictions based on historic behaviour of the ecosystem. But if changes have led to truly novel dynamics these can, by definition, not be revealed using historic data, which makes this approach precarious.

Further limiting the study of complex adaptive systems is the realisation that non-linearity is not intuitive to humans and the number of interdependent phenomena we are able to keep under consideration at one time is limited [73]. Although it has been implicitly argued at the beginning of this chapter, it should be stated explicitly that expecting individual or small groups of scientists to uncover sets of interdependent mechanisms spanning a range of scientific domains by sheer accumulation of knowledge and insight appears overly optimistic. Going down the path of a 'community approach', aims to facilitate the most effective use of research resources (fisheries biology is highly resource intensive, requiring research vessels, etc.) both in terms of effort and of the time it will take us to sufficiently deepen our understanding.

Faced with ecological systems evolving at unprecedented pace under several different anthropogenic pressures, the objective of sustainably managing embedded living resources has shifted: we are no longer able to predict the future based on past observed behaviour, but now have to predict novel behaviour based on a functional under-

[70] As stated in the definition given for models: *A model (M) for a system (S) and an experiment (E) is anything to which E can be applied in order answer questions about S*

[71] Advice, the scientific product as defined and implemented by ICES

[72] Etienne Low-Décarie, Corey Chivers, and Monica Granados. "Rising complexity and falling explanatory power in ecology". In: *Frontiers in Ecology and the Environment* 12.7 (2014), pp. 412–418. ISSN: 1540-9295. DOI: `10.1890/130230`

[73] George A. Miller. "The magical number seven, plus or minus two: some limits on our capacity for processing information." In: *Psychological Review* 63.2 (19569), pp. 81–97. ISSN: 1939-1471, 0033-295X. DOI: `10.1037/h0043158`

standing of the mechanics of the respective system.

Taking advantage of modelling and simulation tools able to provide a focal point for transparent and efficient integration of the many facets of the supporting ecological systems, will greatly benefit sustainable management of stocks embedded in quickly evolving ecosystems [74].

As long as methodologies are limited to those which only allow to understand high-impact shifts in ecosystems in hindsight, management will remain 'behind the crest of the wave'. The modelling and simulation approach suggested here, has the potential to place fisheries science ahead of the crest and thus continue to support reliable Advice[75].

Methods, Tools and Supporting Infrastructure

Having identified the overarching requirements and subsequently settled upon the matching general approach of multi-faceted modelling, this framework now needs to be outfitted with specific modelling methods, tools and supporting infrastructure meeting the particular requirements of this task as exhaustively as possible.

Formalising a (or parts of a) fish stock and the relevant parts of the ecosystem in which it is embedded to communicate about and deepen understanding of its functionality makes several different demands on the tools, approaches, and methods used. Therefore, it is vital to be aware of the specific requirements before attempting to derive the desired properties of these components suited for this particular purpose. To this end, the basis for the requirements, the resulting criteria and the specific properties need to be made explicit.

The requirements resulting from the given task can be sorted into four groups, reflecting the different aspects of the endeavour:
a) stemming from the higher-order motivation of the creation of management Advice[76], are the requirements of trust and transparency
b) based on the composition of the intended users, i.e., the quite heterogeneous and busy fisheries science community, there is a requirement for ease of participation facilitated through accessibility
c) following from the properties of complex adaptive systems is the requirement to enable the representation of the system's functionality with a high degree of verisimilitude and without unintentional omissions

From these requirements and criteria, then follow the specific properties for evaluation of suitability for the proposed task:
a: Trust and transparency: any simulations conducted have to be without ambiguity as to how a given model relates to "its" simulation results and need to be sufficiently transparent to be subject to a high degree of scrutiny.
b: Ease of participation: as fisheries science deals with entire ecosystems, its contributors are a highly diverse group of scientists in terms of field of specialisation, expertise and methodological training. They

[74] Ferdi L. Hellweger. "75 years since Monod: It is time to increase the complexity of our predictive ecosystem models (opinion". In: *Ecological Modelling* 346 (2017), pp. 77–87. ISSN: 03043800. DOI: `10.1016/j.ecolmodel.2016.12.001`

[75] Advice, the scientific product as defined and implemented by ICES

[76] Advice, the scientific product as defined and implemented by ICES

are also, more often than not, committed to a time-sensitive workflow that requires contributions on a yearly cycle, leaving little time and few resources for additional training. This demands that the formalisations can be understood by all parties involved, which can be facilitated by the use of metaphors intuitive to all participants resulting in minimal requirements in terms of formal training and a high degree of accessibility.
c: Verisimilitude and no unintentional omissions: ecological systems span a broad range of dimensional scales as they are comprised of processes ranging from the fast and small-scale (e.g., oxidising a metabolite) to the slow and large-scale (e.g., yearly migrations) as well as the interactions across these scales (e.g., having access to sufficient metabolites for oxidising might require a yearly migration). Depending on a given mental model, these scales can also correspond to different system levels. Ecological systems also fall into the category of complex adaptive systems and share their properties of emergence, path-dependence, stochasticity and more. Therefore a suitable modelling language needs to be able to handle a broad range of temporal and spatial scales, the interactions across these scales, the option to interpret them as system levels, stochasticity and emergence.

Characteristics of Suitable Modelling Languages

The limits of my language are the limits of my thoughts - Wittegenstein

The first half of this chapter introduced the notion that formalising executable models as a way to record mental models unambiguously is a viable path to efficient and functional knowledge integration. Now the most suitable means with which these formalisations would be implemented have to be identified.

Typically, the type of executable models referred to here are implemented in a computer language. In general, a 'language' can be defined as a set of words and symbols (usually a subset of all possible combinations in a given alphabet) [77]. This subset becomes a means of structured communication by possessing a syntax, which defines the combination of permitted words and symbols (i.e., are grammatical), and semantics, which define what these combinations mean. This is true for both natural and computer languages, although there is far greater leniency in natural languages. This leniency allows for robustness in their use and supports their evolution over time, but it also results in ambiguity.

A 'programming language' is then defined as a formal language that is used to communicate with a computer, generally to convey instructions. Subsequently, a 'modelling language' can then be defined as a computer language that is used for modelling. However, in the field of modelling and simulation, as in some other fields, a distinction can be made between general-purpose and domain-specific languages. The former are languages such as C, Python and Java, which, as the labelling implies, can be used for all purposes and are not customised. Domain-specific modelling languages, on the other

[77] John E. Hopcroft and Jeffrey D. Ullman. *Introduction to automata theory, languages, and computation*. Addison-Wesley series in computer science. Reading, Mass: Addison-Wesley, 1979. 418 pp. ISBN: 978-0-201-02988-8

hand, are specifically customised to meet the needs and requirements of a specific domain [78].

As Henziger, Jobstman and Wolf[79] stated, properties of interest of modelling languages are:
"*compositionality* (how does the formalism support the construction of complex models from simpler parts?), *expressiveness and succinctness* (which systems can be specified in the formalism and how large are the specifications?), *executability* (how easy is it to compute the possible direct successor states of a given state?), and *well-formedness* (how easy is it to check if a model has a unique solution?)."

The following sections will explore the properties of particular interest to modelling the EBC with the general objective of supporting management and using the approach outlined in the first half of this chapter in more detail. In doing so, the four requirements and associated properties established in the previous section will be referred to as a means of categorisation.

Separation of Model and Simulator – Trust and Transparency

To acquire simulation results, two things are needed: firstly, a formalised system, which is the model, and secondly, the trajectory (series of states) of the system produced by an algorithm and given an initial state, which is the simulation. Those parts of code representing the algorithm and producing the progression of states are not the model but the simulator. When simulation models are built as one unit, the code that is the model and the code that is the simulator can be highly interconnected [80]. However, model and simulator can also be purposefully separated, which has several advantages:

During the implementation stage, the two tasks can be matched with the required skill sets: before any model is built those educated in programming can design and implement a simulator ensuring well-defined processes. When models are built those educated in the content domain can concentrate all their resources on formalising the system they are knowledgeable about. Therefore separating the task of providing a simulator from the task of formalising the given system will likely result in a higher quality of both parts [81].

Throughout the modelling and simulation study, more resources will be available for the domain expert now able to remain on the level of the conceptual model concentrating only on the system they are investigating. If required, the simulator can be optimised for properties like performance without interfering in the model code. This allows concentrating on one aspect without having to use (in this case cognitive) resources for the other [82].

Upon completion of the simulation study, there is greater transparency of both model and simulator as they can each be scrutinized independently. Stakeholders relying on simulation results can scrutinise the simulator for the exact nature of its processes (independently of the models it is meant to simulate) and can invite programming experts to participate since no domain knowledge is required. Also, the

---

[78] Tom Warnke. "Domain-specific languages for modeling and simulation". Dissertation. Rostock: University of Rostock, 2020

[79] Barbara Jobstmann1 Thomas A. Henzinger and Verena Wolf. "Formalisms for Specifying Markovian Population Models". In: *Reachability problems: third international workshop, RP 2009, Palaiseau, France, September 23-25, 2009: proceedings*. Ed. by Olivier Bournez and Igor Potapov. Lecture notes in computer science 5797. Springer, 2009. isbn: 978-3-642-04419-9

Section 2.1.4 of:
[80] Bernard P. Zeigler, Alexandre Muzy, and Ernesto Kofman. *Theory of modeling and simulation: discrete event and iterative system computational foundations*. Third edition. OCLC: on1020030023. London San Diego Cambridge, MA Oxford, UK: Academic Press, an imprint of Elsevier, 2019. isbn: 978-0-12-813370-5

Section 2.2.2 of:
[81] Bernard P. Zeigler, Alexandre Muzy, and Ernesto Kofman. *Theory of modeling and simulation: discrete event and iterative system computational foundations*. Third edition. OCLC: on1020030023. London San Diego Cambridge, MA Oxford, UK: Academic Press, an imprint of Elsevier, 2019. isbn: 978-0-12-813370-5

[82] Tobias Helms et al. "Semantics and Efficient Simulation Algorithms of an Expressive Multilevel Modeling Language". In: *ACM Transactions on Modeling and Computer Simulation* 27.2 (2017), pp. 1–25. issn: 1049-3301, 1558-1195. doi: `10.1145/2998499`

review/examination of the simulator will only need to be done *once* as there is no need to write 'simulator code' for each model. Equally, models can be judged purely on the merit of how well they represent a given system, in this case inviting domain experts to participate, as there is no interference from code only necessary for simulating.

Therefore the separation of model and simulator supports transparency of the model and the simulation process, as each can be investigated as a stand-alone artefact and with assistance from (readily available) experts. This level of transparency then translates into trustworthiness as those asked to give trust have direct access to well-defined parts which can be inspected thoroughly, with a single objective and, when required, paired with suitable experts [83].

Semantic Domain - Verisimilitude of Functionality

The separation of model and simulator allows defining exactly which *processes* move a modelled system from one state to the next and, by extension, which states the model can occupy. The sum of these possible states then defines the 'state space'. Together these well-defined processes and the corresponding state space make up the 'semantic domain' of a given modelling language. From the domain perspective, this can be interpreted as 'defining the manner of things which can happen'. This explicit definition then is conducive to clarity and focus.

As Harel and Rumpe [84] state:
"The semantic domain is not to be taken lightly: It specifies the very concepts that exist in the universe of discourse. As such, it serves as an abstraction of reality, capturing decisions about the kinds of things the language should express. The domain is also a prerequisite to comparing semantic definitions."

Modelling and simulation as a field of computer science conducts research regarding the implementation of different semantic domains such as stochasticity and continuous time. However, the distinctions made from the perspective of realisation (i.e., actually implementing a simulator) are less relevant here than their suitability for modelling the EBC. Ideally, from the perspective of application (i.e., conducting modelling and simulation studies), the choice of (or development of a new) modelling language should base decisions about the semantic domain only on the ability to reflect relevant properties of a system (i.e., provide verisimilitude), regardless of the technical challenges or ease. This requires an understanding of how the desired properties of the formalised system map onto properties of the semantic domain. To this end, some of the more common decisions regarding the best-suited state space for a given model will be discussed in the section on 'Choice of Modelling Approach'.

Domain Specific Languages - Accessibility

As has been stated in the section on the suitability of modelling

[83] Roland Ewald et al. "Flexible experimentation in the modeling and simulation framework JAMES II - implications for computational systems biology". In: *Briefings in Bioinformatics* 11 (2010), pp. 290–300. DOI: `10.1093/bib/bbp067`

[84] D. Harel and B. Rumpe. "Meaningful modeling: what's the semantics of "semantics"?" In: *Computer* 37.10 (2004), pp. 64–72. ISSN: 0018-9162. DOI: `10.1109/MC.2004.172`

languages, a 'domain specific language' (DSL) is a language tailored to the requirements of a specific domain. This means that rather than requiring translation of the common concepts of a domain into the syntax of a general-purpose programming language, a DSL accommodates these common concepts by providing a suitably tailored syntax [85].

In modelling and simulation, this can be achieved by offering a syntax that mirrors well-established concepts of the respective domain and is therefore intuitive to domain experts. Such DSLs, which use familiar notations, are highly legible and thus provide an accessible formalisation of the respective system [86]. Since the general approach, outlined in the first half of this chapter, requires the participation of a range of domain scientists, this legibility and accessibility are vital for it to succeed.

However, DSLs currently available for fisheries science were not suitable for the work presented in this thesis for different reasons. For example, the Sprat Ecosystem DSL[86] has a focus on 3D spatial distribution, which was not the focus of the model developed here. At this point, Ecopath with Ecosim (EwE)[87] shall also be mentioned, which is not a DSL but an established ecosystem modelling software suite for fisheries science and aquatic ecosystems. It is a pioneer in this field which regrettably has led to it becoming the kind of legacy software that is less than transparent[88]. It also focuses more on the mass-balance of food webs and spatial aspects rather than the mechanisms of physiology and behaviour, which are relevant to the work presented in this thesis.

At present, there appears to be no established domain-specific modelling language for fisheries research that supports the approach of the work presented in this thesis (the modelling language used in this thesis is ML-Rule and will be introduced later in this chapter). Developing such a language can therefore be regarded as a requirement for implementing the augmented version of multi-faceted modelling introduced earlier.

Developing a DSL is a considerable undertaking that poses the challenges of proper implementation and deciding on its semantic domain. Already with a very narrow audience for its use, a DSL requires their involvement in choosing the best-suited concepts, which is an iterative and elaborate process. The DSL required for the approach presented here will have to provide ease of participation to a broad range of domain experts with an equally broad range of well-established concepts. In this case, optimising the necessary trade-offs will be difficult. Nevertheless, developing such a DSL would provide a helpful tool and possibly even insight since the process of development (mirroring the integration of submodels) demands unambiguous communication about the different mental models.

[85] Tom Warnke. "Domain-specific languages for modeling and simulation". Dissertation. Rostock: University of Rostock, 2020

[86] Arne N. Johanson and Wilhelm Hasselbring. "Effectiveness and efficiency of a domain-specific language for high-performance marine ecosystem simulation: a controlled experiment". In: *Empirical Software Engineering* 22.4 (2017), pp. 2206–2236. ISSN: 1382-3256, 1573-7616. DOI: `10.1007/s10664-016-9483-z`

[87] https://ecopath.org/

[88] É. E. Plagányi and D. S. Butterworth. "A critical look at the potential of Ecopath with ecosim to assist in practical fisheries management". In: *African Journal of Marine Science* 26.1 (2004), pp. 261–287. ISSN: 1814-232X, 1814-2338. DOI: `10.2989/18142320409504061`

Choice of Modelling Approach

A number of requirements has so far been matched with suitable approaches and tools. The general approach of multi-faceted modelling has been determined to suit the proposed goal of knowledge integration given the requirements of a diverse community of scientist, each holding valuable knowledge about the system in question. Separating model and simulator allows for trust and transparency both through the abundant scrutiny separation of model and simulator can support and the transparency which it can provide. Using a tailored DSL to formalise the required models provides accessibility by inviting participation as it can be tailored to the customs of the community. Furthermore, the control of the semantic domain supports verisimilitude of functionality.

However, given the possibility for a number of design choices of the semantic domain these also have to be made, here the different options and likely candidates will be evaluated particularity with regard to the requirements posed by the properties of complex adaptive systems and the multi-faceted approach.

Multi-level Models

'Multi-level models' are not to be confused with 'Multilevel models' which are often used in ecology and are defined as statistical models which vary at several levels. The former refers to simulation models which take into account the different levels of a modelled system [89].

As discussed in Chapter I ecosystems function as a whole and consist of many subsystems. These systems can conceptually be understood as multiple levels. For example organs → organisms → stocks → ecosystem. Within these levels, upward and downward causation occurs, thriving individuals lead to high recruitment and a dense stock (upward causation), and a dense stock increases competition and subsequent stress on its individuals (downward causation).

Providing explicit multileveldness by a modelling language is therefore conducive to accessibility by mirroring the established metaphor of nested subsystems (the parasite is inside the liver which is inside the cod). More importantly it is vital to avoid unintentional omissions of causation by providing the framework to emulate the corresponding real-world structure[90] thus ensuring verisimilitude.

As Noble[91] argued:

"... these modelling approaches are still doomed to fail, if they do not consider the inherent multi-levelness of biological (and by extension ecological) systems to which upward causation, as well as downward causation, belong, i.e., that 'all processes at the lower levels of a hierarchy are restrained by and act in conformity to the laws of the higher levels.[92]' "

Discrete and Continuous Time Base

Depending on the required resolution different types of time bases

---

[89] Carsten Maus, Stefan Rybacki, and Adelinde M. Uhrmacher. "Rule-based multi-level modeling of cell biological systems". In: *BMC Syst Biol* 5 (2011), p. 166. DOI: `10.1186/1752-0509-5-166`

[90] Naturally, there will be many intentional omissions as it the nature of modelling and simulation to define the system of interest by omitting those parts outside of the respective definition.

[91] Denis Noble. "Claude Bernard, the first systems biologist, and the future of physiology: Systems biology and the future of physiology". In: *Experimental Physiology* 93.1 (2008), pp. 16–26. ISSN: 09580670. DOI: `10.1113/expphysiol.2007.038695`

[92] D.t. Campbell. "Downward Causation". In: *Studies in the philosophy of biology*. University of California Press, 1974

can be used in modelling and simulation. The fundamental distinction being between discrete time, where transitions can occur at a countable number of points in time and continuous time, where transitions can occur at any point in time[93]. In individual-based models, which are of particular interest here (see section on macro- vs. micro-models below) both are used. Most agent-based modelling and simulation frameworks focus on discrete time, a prominent example is NetLogo[94] Another class of population based models are Markov population models[95]. They are based on continuous time markov chains (CTMCs).

CTMCs are random processes that have the property of memorylessness. Therefore, the transition from one state of the system to another is only dependent on the system's present state and does not take prior states or dwelling times into account. In mathematical terms P_{ij} (t) is the probability that the model will be in state j, t time units from now, given it is in state i now. Once it has entered a state it will remain in that state, independent of the past, for an exponentially distributed amount of time before changing state again.

By using continuous time in modelling and simulation unwanted artefacts resulting from unclear scheduling during execution can be avoided, for this reason already it is to be preferred [96]. In terms of verisimilitude continuous time simply resembles the characteristic of the real world system more closely, allowing "to produce a virtual population that closely resembles a real population" [97]. Furthermore, as already stated in the previous section the many sub-systems of interest when investigating the EBC span a broad range of scales, this includes spatial- and time-scales. If this broad range of scales is to be functionally integrated as the general approach envisages, continuous time provides the basis for a clean integration and interaction of processes of vastly different scales.

However, although it has been stated that the focus here is only on the perspective of application, it is worth noting that from a technical perspective continuous time can be more challenging to realise.

Deterministic vs. Stochastic Models

Deterministic models will always produce the same output given the same parameter input, whereas with a stochastic model, identical parameter inputs will lead to a range of outputs (see Figure 15 for an illustration of a stochastic simulation). If we apply this to the question of growth of individual EBC we find that fish of the same cohort living in close proximity will invariably exhibit a variance in growth due to slight random variation (e.g., variations in speed, and thus efficiency of food conversion) which are specific to their experiences and to each individual. To begin with, these random variations are likely to have minute impacts, but over time they will lead to cod being sufficiently different in size to accelerate their diverging paths (e.g., larger cod can ingest larger prey item, reach more favourable conditions faster leading to more growth). This property of path-dependence exhibited by the real-world system can therefore only be

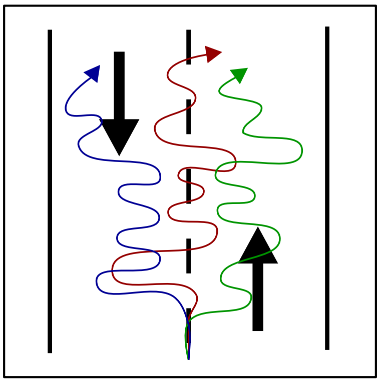


Figure 15: Random walk on a busy road: beginning at the same point stochasticity will lead to varying outcomes.

---

[93] Karl Johan Åström, Hilding Elmqvist, and Sven Erik Mattsson. "Evolution of Continuous-Time Modeling and Simulation". In: The 12th European Simulation Multiconference. ESM'98. Manchester, UK, 1998

[94] http://www.netlogoweb.org/

[95] Barbara Jobstmann1 Thomas A. Henzinger and Verena Wolf. "Formalisms for Specifying Markovian Population Models". In: *Reachability problems: third international workshop, RP 2009, Palaiseau, France, September 23-25, 2009: proceedings*. Ed. by Olivier Bournez and Igor Potapov. Lecture notes in computer science 5797. Springer, 2009. ISBN: 978-3-642-04419-9

Appendix 1A of:

[96] Averill M Law. *Simulation modeling and analysis*. OCLC: 1200558671. Boston: McGraw-Hill, 2007. ISBN: 978-0-07-132169-3

[97] Frans Willekens. "Continuous-time Microsimulation in Longitudinal Analysis". In: *New Frontiers in Microsimulation Modelling*. Ed. by Asghar Zaidi, Ann Harding, and Paul Williamson. 1st ed. Ashgate: Routledge, 2009, pp. 413–436. ISBN: 978-1-315-24806-6. DOI: `10.4324/9781315248066`

replicated with stochastic simulations thus ensuring a high degree of verisimilitude.

It is important to note that stochastic simulations are typically more expensive, due to the number of replications required and averaging the many results will be necessary for interpretation.

Deceptive Averages

While stochastic models allow for path-dependency and thus distributed results, path-dependency also plays a role in the matter of averages.

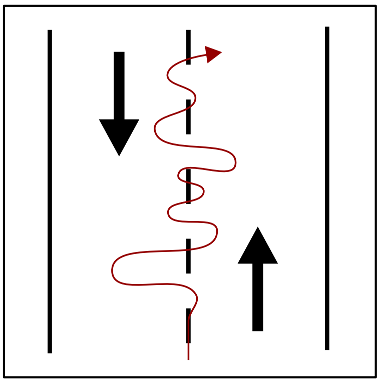


Figure 16: Random walk on a busy road: averaging before evaluation will leave you unharmed but averaging after evaluation will leave you run over.

This example is dapted from an example from Sagave (2015) *The flaw of averages*

An illustration of this is an inebriated person walking along the middle of a busy two-way street, see Figure 16. Due to their state, the person swerves left and right and the distribution of their position ξ gives the positions along their path (Equation 1). The assumption is that they remain unharmed if they are exactly in the middle of the road and get run over when they are not (Equation 2). If we average this path before we evaluate the model (e.g., we calculate where they are on average first and then let the cars drive on the road), the person is determined to be in the middle of the road and thus unharmed (Equation 3). If we, however, refer to multiple samples of the exact position when we evaluate the outcome (e.g., we let the cars drive on the road while they swerve left and right), they will, at some point, not be in the middle of the road and therefore run over (Equation 4). In summary the state of the inebriated person at their average position is unharmed, but the average state of the inebriated person is run over.

Difference between averagging before and after evaluation. ξ is uniform distibuten of the position of the inebriated person along their path. E denotes the expected value of the evaluation of the model.

$$\text{Random Position}: \xi \sim \text{Uniform}[0,1] \tag{1}$$

$$\text{Vital Status}: f(y) = \begin{cases} \text{run over} & \text{if } y \neq \frac{1}{2} \\ \text{unharmed} & \text{if } y = \frac{1}{2} \end{cases} \tag{2}$$

$$f(E\xi) \Rightarrow \text{unharmed} \tag{3}$$

$$Ef(\xi) \Rightarrow \text{run over} \tag{4}$$

This means that averaging at different points in the implementation and execution of a model can produce markedly different results.

Resolution

This also relates to the question of resolution as low resolution models require averaging before implementation (the degree of detail cannot fall below the size of the steps or grid) and high resolution models allow the heterogeneity of a given property to be upheld. Depending on the properties of the modelled system, this can be a critical distinction as illustrated above. If the resolution is low, simplification using averages is required during model implementation, potentially omitting crucial interactions. Conversely, a high resolutions allows for more detail, thereby including a greater number of these potentially crucial interactions.

Concerning the EBC resolution pertains more often to processes (e.g., the resolution of the process of respiration) than to a predetermined step or grid size, this lies more in the realm of model implementation than of semantic domain.

Macro vs. Micro Models

The distinction between micro- and macro-models is one of perspective, where a macro-model uses an aggregation of constituting entities, a micro-model presumes the constituting entities to be distinct from each other [98].

Individual- or agent-based models (IBMs/ABMs) are the most common form of micro-models and the one of interest here. They were first developed in the late 1960s [99] for the economic and social sciences, which was followed by IBMs for ecology in the early 1970s [100]. Rather than describing systems from a macro-perspective (e.g., MSY or the original Lotka-Volterra equations), the approach wanted to understand systems from a micro-perspective, as the interactions of its constituting individuals.

For the EBC, the micro perspective is required for three reasons. Firstly, the investigation across multiple system layers does not lend itself to the macro-perspective as the aggregation of the constituting entities of the respective lower system level could conceal mechanisms of interest. This begins at the level of stocks, where using the concept of SSB (spawning stock biomass) does not allow for the investigation of variations in the fecundity of its individuals. But this also extends all the way down the food chain, where variations in the phyto- and zooplankton population could impact the sprat and herring population in a manner relevant to the EBC.

Secondly, the detail oriented investigation of mechanism requires the ability to formalise heterogeneity. For example finding the mechanisms responsible for some, but not other, individuals reaching a particularly poor condition or mature females not spawning while others do can only be formalised if this heterogeneity can be. Although it is possible to formalise heterogeneity in macro-models this is not practical past a small number of variation as a large number of variations will produce a combinatorial explosion (see also the next section but one).

Thirdly, there is the requirement to grasp emergence as the EBC is part of a CAS. At least in the case of 'weak emergence'[101] this can be captured by computer models. The requirement for capturing emergence, however, is that these models formalise system 'from the bottom up' (e.g., the interactions of the fish in a school with each other and their environment), thereby modelling the *process* of emergence, rather than modelling the emergent property (e.g., the observed movement of a school of fish).

For these reasons, of all the distinctions, made so far, the one of a micro perspective and an individual-based approach could be considered indispensable to investigating the EBC.

[98] Leif Gustafsson and Mikael Sternad. "Consistent micro, macro and state-based population modelling". In: *Mathematical Biosciences* 225.2 (2010), pp. 94–107. ISSN: 0025-5564. DOI: `10.1016/j.mbs.2010.02.003`

[99] Thomas C. Schelling. "Models of Segregation". In: *The American Economic Review* 59.2 (1969), pp. 488–493. DOI: `10.1080/0022250x.1971.9989794`

[100] Daniel B. Botkin, James F. Janak, and James R. Wallis. "Some Ecological Consequences of a Computer Model of Forest Growth". In: *The Journal of Ecology* 60.3 (1972), p. 849. ISSN: 00220477. DOI: `10.2307/2258570`

[101] The distinction between 'weak' and 'strong' emergence is highly debated and this debate would exceed the scope of this thesis

ML-Rules

The DSL used for the modelling and simulation study presented in this thesis is ML-Rules, which is a multi-level rule-based modelling language originally developed for modelling cell biological systems at different levels of organization [102,103]. ML-Rules supports nested rule schemata, the hierarchical dynamic nesting of entities, the assignment of attributes and solutions to entities at each level, and a flexible definition of reaction rate kinetics. Therefore, ML-Rules exhibits many features essential for the work presented in this thesis. Entities and attributes allow to distinguish and conveniently model individual cod. Nested hierarchies are an intuitive representation of a hierarchically structured ecological system and dynamic nesting allows modelled cod to move between layers of their habitat or become adults by adding reproductive organs.

In addition, its rule-based syntax (inspired by reactions) appears rather suitable for the accessible description of processes. Since it supports arbitrary attributes and functions on these attributes and the content of entities to constrain the kinetics of the systems, it is a very expressive approach and thus, poses few constraints to the modelling of the different submodels.

ML-Rules also adopts the "don't care, don't write" approach: a rule refers only to those entities and only those attributes that are of importance for a reaction to take place. Thereby, the combinatorial complexity of other, e.g., equation-based, approaches, where all possible combinations have to be enumerated, is avoided.

A rule-based syntax and the "don't care, don't write" approach also lend themselves to the transparent and sound integration of processes by remaining succinct and legible.

Given these properties ML-Rules appears to be a good starting point for implementing the general approach and was used in the modelling and simulation study presented in the following chapter. However, as stated in the section on DSLs, a DSL augmented or developed for and thereby tuned to the specific requirements of the fisheries science community, accommodating their established metaphors, would be optimal.

An ML-Rules model consists of five sections: parameters, functions, definition of entities, initial solution and the rule schemata. The first segment is optional and allows for the definition of parameters used within the model. These can be numerical, boolean, strings or expressions. The second segment is also optional and allows for the definition of arbitrary functions, satisfying the demand for a high degree of control over the description of processes. The third segment is mandatory and requires the definition of the entities within the model. An ML-Rules entity can have many attributes and these can again be numerical, boolean or strings. The fourth segment is also mandatory and here, the model's initial state is defined. Again, the fifth segment is mandatory and requires the definition of rules that have the appearance of chemical reactions. First, the reactants are stated, then the products are defined and finally, a rate is given. There

[102] Carsten Maus, Stefan Rybacki, and Adelinde M. Uhrmacher. "Rule-based multi-level modeling of cell biological systems". In: *BMC Syst Biol* 5 (2011), p. 166. DOI: `10.1186/1752-0509-5-166`

[103] Tobias Helms et al. "Semantics and Efficient Simulation Algorithms of an Expressive Multilevel Modeling Language". In: *ACM Transactions on Modeling and Computer Simulation* 27.2 (2017), pp. 1–25. ISSN: 1049-3301, 1558-1195. DOI: `10.1145/2998499`

is also the option to define guards which are conditions under which the rule will be executed.

Lotka-Volterra: comparing ODE and IBM

This section will illustrate the difference between deterministic macro models and stochastic individual-based models. The example shall also introduce the modelling language used for the work presented in this thesis, ML-Rules.

Coupled ordinary differential equations (ODEs) are a prominent example of a formal modelling language used in ecology. They are typically used to describe macro-models, are based on continuous time and are deterministic. A prominent example from theoretical ecology are the Lotka-Volterra equations (see Equation 5). This system of two coupled ODEs describes the idealised interactions of a predator and a prey population. They were derived from theoretical consideration and to explain data sets independently by mathematician, chemist and statistician Lotka and mathematician and physicist Volterra[104,105,106]. If its parameters are set to specific values it can be evaluated and shows two staggered oscillations of population numbers (see Figure 17).

$$\frac{\mathrm{d}N_1}{\mathrm{d}t} = N_1(\alpha - \beta N_2), \qquad \frac{\mathrm{d}N_2}{\mathrm{d}t} = -N_2(\gamma - \delta N_1) \tag{5}$$

Lotka-Volterra equations, description of variables is given in Table 1

Table 1: Description of parameters in Lotka-Volterra equations, see Equation 5

$N_1 = N_1(t)$	number of prey organisms	time dependent
$\alpha > 0$	reproductive rate of prey without disruption and with sufficient food supply	constant
$\beta > 0$	consumption rate of predators per prey individual = mortality rate of prey per predator	constant
$N_2 = N_2(t)$	number of predators	time dependent
$\gamma > 0$	mortality rate of predators when there is no prey	constant
$\delta > 0$	reproductive rate of predators per prey individuals	constant

The idealising assumptions made for the ODE model are:
- there is always sufficient food for the prey population
- predator food supply consists entirely of the prey population
- the rate at which a population changes is proportional to its size
- there is no adaptation
- there are no environmental change
- the appetite of predators is limitless

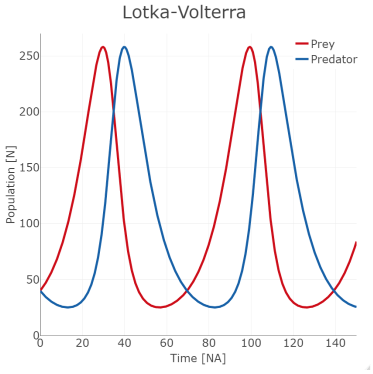


Figure 17: Visual representation of the Lotka-Volterra equations for α=0.1, β=0.001, γ=0.1, δ=0.001 and both populations with N_0 of 40.

As stated this is an idealised formalisation of the interaction of the predator and prey population. Nonetheless it was able to provide insight and a theoretical base which could be further investigated to

---

[104] Lotka worked with data sets on lynx and snowshoe hare as recorded by the Hudson's Bay Company and Volterra worked with fisheries data from the Mediterranean Sea during WWI primarily on sharks

[105] Alfred James Lotka. *Elements of Physical Biology*. Williams & Wilkins Company, 1925. 514 pp.

[106] Vito Volterra. "Fluctuations in the Abundance of a Species considered Mathematically". In: *Nature* 118.2972 (1926), pp. 558–560. ISSN: 0028-0836, 1476-4687. DOI: `10.1038/118558a0`

reach insight such as realising that, in a wide range of ecological scenarios, prey populations control corresponding predator populations more than predator populations control the populations of their prey.

The system modelled by the ODE can also be modelled as an IBM by translating the processes to the individual level. The code below shows an IBM implementation of the Lotka-Volterra model in ML-Rules. It is based on the same parameter values as were used for the simulation of the ODE as shown in Figure 17.

ML-Rules code of Lotka-Volterra implemented as an IBM in ML-Rules: First, the type of entities are specified. Then the initial model state for the subsequent simulation is given here there are 40 predator and prey entities each. Next the model is defined by four rules corresponding to the four processes described by the Lotka-volterra equations.
1.) 'Reproduction of Prey': one prey entity becomes two prey entites, the rate of this process (which sets the parameter for the exponentially distributed amount of time it will remain in its current sate - see section above on continous time bases) is determined by the parameter α (0.1) and the current number of prey entities which is recorded in the local variable a
2.) 'Predation': Prey and Predator entity interact and the prey entity is removed, the rate is determined, again, by the number of involved entities, in this case the prey (a) and predator (b) and β (0.001)
3.) 'Death of Predators': a predator entity is removed, the rate is determined by the number of predators (b) and γ (0.1)
4.) 'Reproduction of Predators': one predator entitiy transforms into two predator entities, the rate is determined by the number of prey entities (a), number of predator entities (b) and δ (0.001)

```
//ENTITIES
Predator();             //Predator
Prey();                 //Prey


//INITIAL STATE
>>INIT[40 Predator + 40 Prey];


//RULES
//Reproduction of Prey
Prey:a ->
2 Prey
@ 0.1 * #a;             //alpha

//Predation
Prey:a + Predator:b ->
Predator
@ 0.001 * #a * #b;      //beta

//Death of Predators
Predator:b ->

@ 0.1 * #b;             //gamma

//Reproduction of Predators
Prey:a +   Predator:b ->
Prey   + 2 Predator +
@ 0.001 * #a * #b;      //delta
```

Simulation results of the ML-Rules implementation of the Lotka-Volterra model exhibit the same general behaviour as the ODE version: two staggered oscillations of population numbers. However, stemming from a stochastic simulation, the results are irregular (see Figure 18). To properly evaluate the behaviour of a stochastic model, several simulations must be run and interpreted as a whole (see Figure 19). This evaluation shows that although their general behaviour is persistent, the stochasticity is sufficient for the length of the oscillations to grow significantly apart over time, which is illustrated by increasing variance and confidence band over time.

```
// PARAMETER
e       :2.71828; //Euler’s number

//FUNCTION
//sigmoidal increase of skill
ssi :: num -> num;
ssi y  =
    20/(1+(e)^(-0.1*(((10*log(-y/(y-20))+50)+1)-50)));

//ENTITIES
Predator(num,num);      //Predator
Prey();                 //Prey


//INITIAL STATE
>>INIT[20 Predator(0,0.14783) +
20 Predator(0,10)      + 40 Prey];


//RULES
//Reproduction of Prey
Prey:a ->
2 Prey
@ 0.05 * #a;                    //alpha

//Predation
Prey:a + Predator(kills,skill):b ->
Predator(kills+1,ssi(skill))
@ 0.001 * #a * #b * (skill/10); //beta

//Death of Predators
Predator(kills,skill):b ->

@ 0.1 * #b;                     //gamma

//Reproduction of Predator
Predator(kills,skill):b + Prey:a ->
Predator(kills,skill) + Predator(0,unif(0.014783,10)) +
    Prey
@ 0.001 * #a * #b;              //delta
```

ML-Rules code of Lotka-Volterra implemented as an IBM in ML-Rules expanded for a heterogenous predator population whose individuals vary according to their skills: First, a parameter (Euler’s number) is defined for subsequent use in a sigmoidal function that models skill (ssi - sigmoidal skill increase). Next, the type of entities are specified. Then the simulation’s initial state is given: there are 20 predators with minimal skill (0.14783), 20 predators with median skill (10) and 40 prey entities. Next the model is defined by four rules corresponding to the four processes described by the Lotka-Volterra equations.

1.) ‘Reproduction of Prey’: one prey entity transforms into two prey entites, the rate of this process is determined by the parameter α (0.1) and the current number of prey entities which is recorded in the local variable a

2.) ‘Predation’: Prey and Predator entity interact and the prey entity is removed, this increases the kill count by one and the hunting skill of the individual predator according to the evaluated sigmoidal function, which increases skill platouing at a value of 20 over a number of 100 kills, the rate is determined by the number of prey entities (a), number of predator entities (b), β (0.001) and the skill of the specific predator involved

3.) ‘Death of Predators’: a predator entity is removed, the rate is determined by the number of predators (b) and γ (0.1)

4.) ‘Reproduction of Predators’: one predator entity transforms into two predator entities, the rate is determined by the number of prey entities (a), number of predator entities (b) and δ (0.001)

To realise the example given above as an ODE model would require a separate population for each skill level, and even if the continuous value were suitably discretised, the resulting combinatorial complexity would prevent modelling of agent-based models as ODEs. Rule-based modelling is an established means to deal with the combinatorial complexity of biological systems [107]. The above example illustrates that in ML-Rules, attributes can easily be added to the entities and arbitrary functions can be defined, in this case, to determine attribute values, but equally to determine the content of entities as well as the rates (see also Figure 20).

The examples above illustrate the practical expressiveness of ML-Rules as a language to legibly describe processes. The distinction

---

[107] James R. Faeder et al. “Rule-based modeling of biochemical networks”. In: *Complexity* 10.4 (2005), pp. 22–41. ISSN: 1076-2787, 1099-0526. DOI: `10.1002/cplx.20074`

between theoretical and practical expressiveness is introduced in the next section.

Discussion: Theoretical and Practical Expressiveness

The expressive power of a language is a measure of the span of concepts that can be formalised and communicated by it. This measure can be further differentiated into *theoretical* and *practical* expressiveness, where the former requires that the span is broad [108] and the latter demands that it also be concise and accessible [109]. Theoretical expressiveness can be precisely defined and is a statement about weather or not one given language can express more, fewer or the same amount of concepts than another language.

However, given that most computer languages are Turing-complete (i.e., are able to simulate a universal Turing machine) theoretical expressiveness is rarely a limiting factor. As Perlis stated: "Beware of the Turing tar-pit in which everything is possible but nothing of interest is easy" [110]. A Turing tar-pit (coined after Perlis' quote) therefore is any interaction with a computer (an interface or a language) that is difficult to use for its intended task, despite having great expressive power. Consequently, a more relevant, albeit less quantifiable, measure of a languages usefulness for a given task is its practical expressiveness.

For the purposes of using modelling and simulation to investigate the EBC in the manner suggested only practical expressiveness is relevant. Although no user study will be performed, the modelling efforts have shown limitations but also promises of ML-Rules as a DSL for modelling a complex adaptive system. This will be illustrated further in the following chapter.

Validation and Documentation

As stated above a model is designed to answer a specific question of interest, whether it provides a valid answer is subject to validation, and how it does so is subject of documentation.

Validation

Balci[111] states that:
"*Model Validation* is substantiating that the model, within its domain of applicability, behaves with satisfactory accuracy consistent with the M&S objectives. Model validation deals with building the *right* model."

This means that validations aims to show that, within given boundaries, a validated model produces the same output as the model system under the same exertion within its scope of applicability, meaning that under the definition given at the beginning of this chapter a valid model can be used for experimentation instead of the real world system for those experiments it has been intended for. This

---

[108] Matthias Felleisen. "On the expressive power of programming languages". In: *Science of Computer Programming* 17.1 (1991), pp. 35–75. ISSN: 01676423. DOI: 10.1016/0167-6423(91)90036-W

[109] Tom Warnke and Adelinde M. Uhrmacher. "Practical expressiveness of internal and external domain-specific modeling languages". In: *2017 Winter Simulation Conference (WSC*. 2017 Winter Simulation Conference (WSC. Las Vegas, NV: IEEE, 2017, pp. 4566–4567. ISBN: 978-1-5386-3428-8. DOI: 10.1109/WSC.2017.8248207

[110] Alan J. Perlis. "Special Feature: Epigrams on programming". In: *ACM SIGPLAN Notices* 17.9 (1982), pp. 7–13. ISSN: 0362-1340, 1558-1160. DOI: 10.1145/947955.1083808

[111] Osman Balci. "Verification, Validation and Accrediation of Simulation Models". In: *Proceedings of the 1997 Winter Simulation Conference*. Ed. by S. Andradöttir et al. IEEE. Piscataway, New Jersey, 1997, pp. 135–141. DOI: 10.1109/IEEESTD.1997.8685803

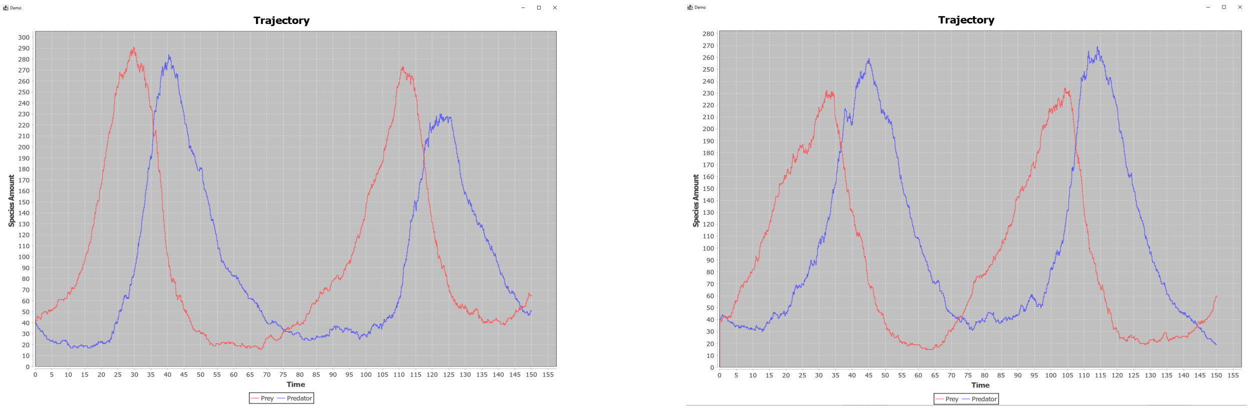
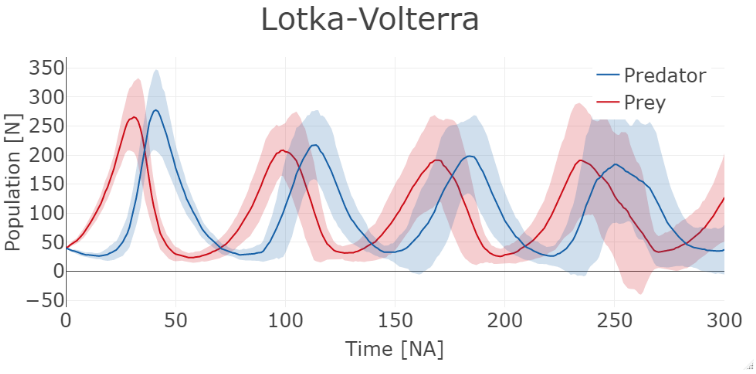


Figure 18: Simulation of Lotka-Volterra as IBM in ML-Rules - see code above. Parameters values are equivalent ot visualisation of coupled DE in Figure 17. Note the stochasticity and

Figure 19: Rule-based simulation of Lotka-Volterra with 20 replications: mean and confidence band of sd given.



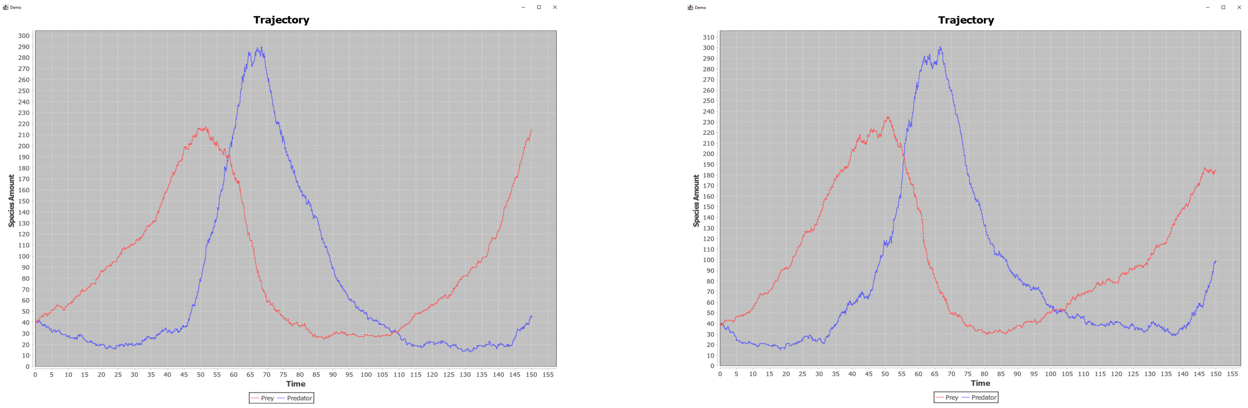


Figure 20: Simulation of Lotka-Volterra as IBM in ML-Rules with heterogeneous entities - see code above. Parameters values are equivalent ot visualisation of coupled ODE in Figure 17. The original code has been expanded to attributed predators which have a kill-count and corresponding hunting skill which modulates the rate of predation.

satisfies an essential requirement of the general approach.

A more extensive introduction to validation and an investigation of a suitable workflow to provide continued validity of iteratively expanded models will be provided in Chapter IV.

Documentation

As it is the goal of the multi-faceted approach to accumulate and functionally integrate a broad and diverse span of understanding, not only the models but also the knowledge, information and decisions (e.g. omission, simplifications, data, narratives, etc.) they contain will become quite extensive. Only based on this information it is possible to interpret the results achieved with a simulation model. Therefore this knowledge is vital and will only be accessible if it has been documented. However, simply requiring documentation does not match the importance and expanse of that which need to be documented.

This is not a new insight and proper, documentation of ecological models is a field of study in and of itself. Great progress in particular with the intention of informing management decisions has been made and brought forth the TRACE documentation [112]. This manner of documentation has been used to document the model of the Eastern Baltic cod which was developed in the scope of this thesis and will be presented in the next chapter.

But even with accessible formalisations, accumulating insights through an expansive modelling and simulation study means the models will contain ever more, densely packed, information. At the same time it is important that new participants can join with little effort, as their insights could be required to make progress. Providing comprehensive access to the model in an autodidactic manner, will allow collaborators to work independently without the requirement to coordinate schedules. This type of access can be provided in the form of appropriate documentation. The TRACE protocol [113,114] has been developed specifically to provide structured and accessible documentation of ecological modelling studies conducted with the goal to support management. We therefore propose using this already available and established tool for documentation.

However, the TRACE documentation can be further expanded by coupling it with a provenance approach, this will be covered in Chapter V of this thesis.

Conclusion about Approaches and Tools

Managing the EBC in its current situation requires a deep understanding so far not provided by the established tools and approaches. I argue that a viable new approach is to comprehensively integrate the expert knowledge on the different facets of EBC ecology, currently spread across the many scientists investigating them, by means more efficient than natural language communications. To this end, the

[112] Volker Grimm et al. "Towards better modelling and decision support: Documenting model development, testing, and analysis using TRACE". en. In: *Ecological Modelling* 280 (2014), pp. 129–139. ISSN: 03043800. DOI: `10.1016/j.ecolmodel.2014.01.018`

[113] Amelie Schmolke et al. "Ecological models supporting environmental decision making: a strategy for the future". In: *Trends in Ecology & Evolution* 25.8 (2010), pp. 479–486. ISSN: 01695347. DOI: `10.1016/j.tree.2010.05.001`

[114] Volker Grimm et al. "Towards better modelling and decision support: Documenting model development, testing, and analysis using TRACE". en. In: *Ecological Modelling* 280 (2014), pp. 129–139. ISSN: 03043800. DOI: `10.1016/j.ecolmodel.2014.01.018`

multi-faceted modelling approach can be adapted by augmenting it to use functional integration of white-box models to align and refine the mental models of the scientific community surrounding the EBC. This could provide the necessary insight the emerging challenges of managing ecosystem goods such as the EBC require [115]. In addition the approach will also require adapting or developing a suitable DSL that provides the accessibility required by the general approach and the semantic domain required to avoid unintentional omissions and provide the necessary verisimilitude. To be truly functional these approaches and tools have to be accompanied by, preferably low effort, validation and documentation.

Although the complete set of tools required to execute the general approach is not yet available, fisheries science has the advantage that it has a 'central body'. ICES already requires intense (sub)-domain spanning communication and the integration of hypotheses in statistical stock assessment models. This existing infrastructure could be built upon to accelerate the establishment of new approaches such as the one presented here.

[115] Nicolas Mouquet et al. "REVIEW: Predictive ecology in a changing world". In: *Journal of Applied Ecology* 52.5 (2015). Ed. by Marc Cadotte, pp. 1293–1310. ISSN: 00218901. DOI: `10.1111/1365-2664.12482`

Chapter III
Modelling and Simulating the Eastern Baltic Cod

In theory, theory and practice are the same. In practice, they are not.
The Yale Literary Magazine, 1882

Modelling and Simulation Results

This chapter will illustrate the application of multi-faceted modelling to the Eastern Baltic cod (EBC). Excerpts of the modelling and simulation results will be introduced facet by facet to discuss model development from both a modelling and simulation and a fisheries science perspective.

The facets which are included in the current iteration of the study are physiology, reproduction, behaviour, environment, prey and parasitation (see Figure 21).

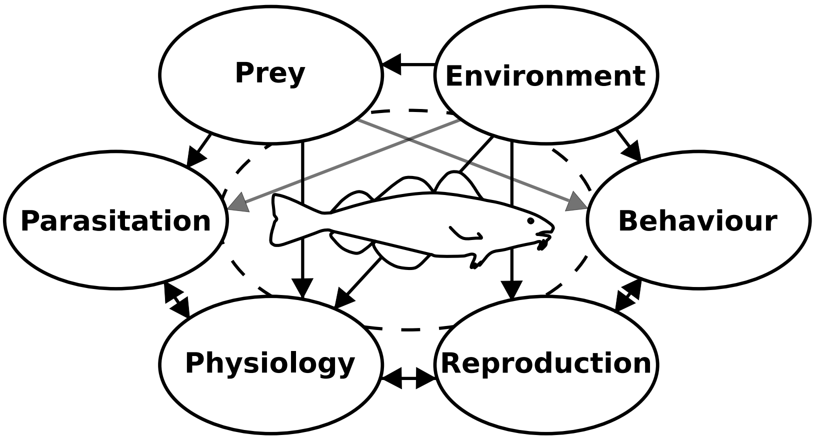


Figure 21: Schematic overview of the facets of Eastern Baltic Cod taken into account in the modelling and simulation study presented here. Black arrows mark integration of facets which have already been implemented, grey arrow mark integration of facets for which the necessary infrastructure is already implemented but formalising of mechanisms has not yet taken place.

Chapter II discussed the objective of supporting management and the requirements stemming from working within departmental research and investigating CAS. These requirements were met by a multi-faceted approach, the use of DSLs, and a suitable choice of semantic domain. The focus here will be on gauging the following aspects as a measure of how well the results from the previous chapter could be put into practice:

a.) Verisimilitude: how well does the implementation compare to the desired simplified version of the system intended by the conceptual model
b.) Comprehensiveness: how well could the different aspects of a given mechanism be integrated
c.) Complexity: how well could more elaborate interdependencies be

formalised
d.) Experimentation: how do experiments conducted with the model allow to compare results with results of wet-lab experiments and field observations, or how do they allow for investigation of competing hypotheses (to a certain extent, this also touches on the subject of validation, however validation will be covered more extensively in Chapter IV)

As introduced in the previous chapter the modelling language used for this implementation is ML-Rules. To enhance legibility code will not be presented in the exact ML-Rules syntax:

The rule which formalises the uptake of oxygen in ML-Rules syntax, with the relevant attributes underlined and functions denoted in red

```
//Oxygen Intake
GM(l,bm,ox  ,jpo,s,rc,hs,sta,p,v,ui)[gm?] ->
GM(l,bm,ox+1,jpo,s,rc,hs,sta,p,v,ui)[gm?]
@ if (ox < 30)
then amr(o,t) * lbmf(bm) else 0;
```

but rather in a pseudocode more suited to discuss content related details without requiring the reader to make themselves completely familiar with all aspect of the model:

The rule which formalises the uptake of oxygen expressed in pseudocode with the guards and rate given as natural language descriptions

```
Cod(oxygen) ->
Cod(oxygen + 1)
IF the maximum oxygen content is not exceeded
@ a rate of the active metabolic rate (dependent on
    oxygen and temperature) * scaled for body mass
```

The pseudocode will differ from the ML-Rules syntax by being drastically simplified to enhance legibility and alleviate formatting issues. This includes reducing code to its relevant parts (e.g.: omission of attributes and entire mechanisms), substituting mathematical functions with natural language descriptions of their in- and output, presenting incomplete rules, replacing abbreviations with long form descriptions and the omission of symbols (e.g.: semicolons at the end of rules, etc.). This also points to one of the two main shortcomings of using ML-Rules for its application to aquatic ecology, although its semantic domain, as introduced and discussed in the previous chapter, meets all of the crucial requirements, it was designed for the domain of cell biology. This domain requires fewer on often no attributes thus, the syntax is not designed to keep models legible when a larger number of them is in use. However the full (executable) code and comprehensive supporting documentation are publicly available[116].

Starting Point: the Juvenile Physiology

An early version of the EBC model focused on oxygen and had been developed as part of a master's thesis. A subsequent version, highlighting the merits of DSLs was also published[117]. It has since been significantly updated, particularly concerning how oxygen is handled conceptually, shifting oxygen from entity to attribute. This updated version is the central facet of the EBC simulation study

[116] `https://github.com/Baltic-Cod/EBC_IBM`

[117] Maria E. Pierce et al. "Developing and validating a multi-level ecological model of eastern Baltic cod (*Gadus morhua*) in the Bornholm Basin - A case for domain-specific languages". In: *Ecological Modelling* 361 (2017), pp. 49–65. ISSN: 03043800. DOI: `10.1016/j.ecolmodel.2017.07.012`

presented here and will therefore be introduced first.

Juvenile Physiology

This model aims to formalise how hypoxic conditions affect the physiology of cod and how the impact of hypoxia is related to reduced growth and condition observed in the field. Therefore, it provides a set of rules formalising aspects of metabolism and growth, which are reactive to oxygen saturation, temperature and available food, and together represent a basic physiology of a juvenile cod. The model was validated for growth and sensitivity to hypoxia by reproducing wet-lab experiments with the modelled cod and comparing results. However, while the model was successfully validated, experimentation with the model did not provide a viable explanation for the diminishing state of the EBC stock.

At this stage, the model consists of 11 rules, which formalise the processes of:

1. Oxygen Intake - increasing the cod's oxygen content if this does not exceed its maximum at a rate dependent on ambient oxygen, temperature, and body mass.

2. Standard Metabolic Rate - using oxygen and available metabolites (to sustain life), if sufficient metabolites are available, at a rate dependent on temperature and body mass.

3. Starvation - using oxygen and reducing body mass (to sustain life), if sufficient metabolites are not available, a rate dependent on temperature and body mass.

4. Normal Growth - using oxygen and available metabolites to increase body mass and length if available oxygen, free metabolites, and condition of the cod are sufficiently high and a rate dependent on temperature and body mass.

5. Growth at Low CI - using oxygen and available metabolites to increase body mass and length if available oxygen and free metabolites of the cod are sufficiently high but the condition is not at a rate dependent on temperature and body mass.

6. Stomach Evacuation - transforming prey pieces in the stomach into available metabolites using oxygen quantified by the properties of the prey pieces if available oxygen and appetite are sufficient and at a rate dependent on temperature, body mass, stomach content, and type of prey.

7. Ingestion - transferring prey within the same compartment into the stomach if the cod has previously successfully hunted (note that the process of ingestion and hunting are kept separate as they are conceptually separate and can consequently be augmented and expanded independently) and the cod is hungry, and the prey is not too large subsequently resetting the boolean tracking hunting success.

8. Hunting - setting the boolean tracking hunting success to successful at a rate reflecting field data on stomach sampling.

9. Activity - using oxygen and metabolites (for implied movement) if the cod has sufficient amounts of both at a rate dependent on ambient oxygen and body mass.

Table 2: Attributes of GM (*Gadus morhua*), Stomach, AJPO (food pieces) and Prey. Attributes and abbreviations used in the model are given, ranges refer to those inside the scope of applicability of the current iteration of the model. Attributes held by two entities are generally used to transfer information. Please note that the table includes attributes added during the inclusion of additional facets.

Attribute		**Use**	**Range**	**Unit**
		GM - *Gadus morhua*		
length	(l)	GM length	292-721	[mm]
body mass	(bm)	GM mass (excluding gonads)	150 - 4500	[g]
oxygen content	(ox)	tracks oxygen	0 - 30	0.1 [mg] O_2
energy content	(jpo)	tracks metabolites	0 - 675000	$\frac{2.7\ [J]}{0.2\ [mg] O_2}$
sex	(s)	assign sex	0,1 - female, male	
reproductive cycle	(rc)	track reproductive cycle	1,2,3,4 - spent, preparing, ready, spawning	
hunting success	(hs)	control feeding	0,1	
stock affiliation	(sta)	assign Eastern, Western, Hybrid	'E','W','H'	
particular	(p)	force properties	as needed	
volume	(v)	track position	1 - 48	1
Identifier	(ui)	track individuals	0 - ∞	1
		Stc - Stomach		
prey consumption	(pc)	tracks consumed prey	0 - max. stc. cont.	[g]
variability of food intake	(va)	allows to control appetite	0-1	0.01
identifier	(uis)	track Stc and match with GM	0 - ∞	1
		AJPO - Food Pieces		0.013 g
oxygen demand	(od)	set oxygen demand for digestion		0.1 [mg] O_2
energy transfer	(et)	set energy transfer	0-20	0.0001
digestive factor	(df)	speed of digestion	1 - 3	1
		Prey		
name	(pn)	readability		
size	(ps)	define size	0 - 10000	0.0013 [g]
oxygen demand	(pod)	set oxygen demand for digestion		0.1 [mg] O_2
energy transfer	(pet)	set energy transfer	0 - 20	0.0001
digestive factor	(pdf)	speed of digestion	1-3	1
Anisakis simplex	(ast)	track infection with parasite		
Contracaecum osculatum	(cot)	track infection with parasite	0 - 20	0.0001

10. Death from Starvation - removing the cod if its condition is too low to survive.
11. Death from Asphyxiation - removing the cod if it has no more oxygen available.

The central entity, of course, is the cod which is equipped with several attributes (see Table 2, please note that the table includes attributes added during the integration of additional facets) which define the specific state of each individual. At this stage they would include length, body mass, oxygen content, energy content and hunting success.

Additional entities at this stage are the stomach (`Stc`), prey (`Prey`) and "food pieces" (`AJPO`). The latter acquire their acronym from the perspective of the metabolism. Currently metabolites are summarised under their property to provide a given amount of energy for a given amount of oxygen ($[\frac{2.7\,[\mathrm{J}]}{0.2\,[\mathrm{mg}]\mathrm{O}_2}]$), resulting in the acronym joule-per-oxygen → jpo. As food pieces come before metabolites, they were named *ante*-joule-per-oxygen → AJPO.

Making food pieces entities results from the requirement of heterogeneity. Different types of prey act differently in various processes, and relevant information needs to remain with the prey pieces during digestion. For example, the speed of digestion, the oxygen required, and the amount of energy transferred all vary depending on the specific type of prey. In later stages of the simulation study, having food pieces as entities will also support implementing the transfer of essential micro-nutrient such as thiamine and corresponding detrimental enzymes such as thiaminase based on available prey.

Guards

Rule	Rank	Guard value
Oxygen Intake	0	-
Standard Metabolic Rate	0	-
Starvation	0	-
Normal Growth	3	>28
Growth at low CF	3	>28
Stomach Evacuation	1	>25
Ingestion	0	-
Hunting	0	-
Activity	2	>26
Death from starvation	0	-
Death from asphyxiation	-1	<1
Gonad energy allocation	1	>25

Table 3: Ranking of functions through guards evaluating available oxygen: Rank indicating order of priority for processes when oxygen is limited.
(-: process is executed regardless of oxygen status)

One manner in which the model formalises the mental models of domain experts is through the use of guards. For example, physiological priorities are formalised by parametrisation of the guards setting oxygen requirements for the execution of the respective rule, see Table 3, (please note that the table includes for a process added during the integration of the reproduction facet). This approach was also used for the hierarchy of processes with regards to available metabolites, where guards could also require a minimal amount (e.g., the minimum requirement not to go into starvation mode to satisfy the requirements of the basic metabolic rate), or a maximum amount (e.g., the maximum permissible amount of metabolites for the cod to still 'be hungry' and 'want to hunt').

Rates

Rates, where applicable, were either taken directly or derived from the results of wet-lab experiments. For example, the rates for the processes of oxygen intake, the standard metabolic rate and starvation were implemented by translating the results of Claireaux[118] into functions. These implement the equations for the active metabolic rate (amr):

$$\mathrm{AMR}(T, S_{O_2}) = (17.29T^{-0.015T+1.062} + 30.01) \cdot (1 - e^{-0.035S_{O_2}+0.34}) \; \frac{\mathrm{mg\ O_2}}{\mathrm{h\ kg}}, \tag{6}$$

where T is the temperature in °C and S_{O_2} is the oxygen saturation given in %,
and the standard metabolic rate (smr):

$$\mathrm{SMR} = 80.1(1 - e^{-0.185T^{0.79}}) \; \frac{\mathrm{mg\ O_2}}{\mathrm{h\ kg}}, \tag{7}$$

where T is the temperature in °C.

In the model, these functions are then further modulated for body mass by an additional function that completes the rate. Although these equations are shown here in part to illustrate that the model uses the best available input, regardless of how mathematical simple or complicated, there is a more critical point to be made about trade-offs not only for rates but also for the mathematical relationships used for guards and transfers. Chapter II introduced the importance of accessibility for the general approach[119] and subsequently suggested using a DSL and providing documentation to meet this requirement. However, no DSL can make complicated equations more accessible (in ML-Rules, the syntax for functions is one of its weaknesses in terms of legibility), and regarding this one particular point, the requirements introduced at the beginning of this chapter, of verisimilitude, comprehensiveness, and complexity, in my opinion, override the requirement of accessibility. Everything should cater to the ease of participation unless this entails the risk of overlooking a significant relationship or interdependency.

Validation Experiments

Using a DSL for simulation experiment specification, SESSL (Simulation Experiment Specification via a Scala Layer), which will be introduced in Chapter IV experiments to validate aspects of the model. These simulation experiments reproduce two wet-lab experiments,

---

[118] G. Claireaux et al. "Influence of water temperature and oxygenation on the aerobic metabolic scope of Atlantic cod (*Gadus morhua*". In: *Journal of Sea Research* 44.3 (2000), pp. 257–265. ISSN: 13851101. DOI: `10.1016/S1385-1101(00)00053-8`

[119] referring to the use of modelling and simulation as a means of increasing knowledge about a system, here in particular through multi-faceted modelling

one investigating asphyxiation [120] and one investigating oxygen-dependent growth [121]. They will be thoroughly introduced in the next section and results on the early version of the model will also be thoroughly discussed in Chapter IV, relevant at this point is that the model performed well, or rather too well, for the purposes of the simulation study. In particular, the lack of variability in growth, as shown by a minimal standard deviation, pointed toward missing processes.

Expanded Juvenile Physiology

Based on the aforementioned consideration, the model was subsequently expanded to control the variability of food intake levels (including a slightly larger margin for stochasticity).

This is an example of introducing mathematical control of a process (in this case, ingestion of food) as a placeholder for the large number of effects resulting from mechanisms at different systems levels not yet available for formalisation (e.g., foraging success, stress, individual variance in metabolic performance, etc.) as there were not yet sufficient data and/or hypotheses available to soundly implement these mechanisms. Even though the details of these latter processes are ‘missing’, their quantitative impact can be controlled through the proxy of controlling food intake levels, and subsequent simulation experiments can be face-validated (meaning persons knowledgeable about the aspect of interest will compare the simulation results to what they would expect in the field or lab based on their experience) (see Figure 22). This proxy will continue to impact quantitative model behaviour as required throughout the modelling and simulation cycles until it can be retired, bit by bit, in favour of the actual mechanisms producing the effect observed in the field or wet-lab. The rule for ingestion, reduced to the relevant parts, now reads as:

```
GM(hunting success = ‘TRUE‘)[Stc[stc?]] +
    Prey(preysize,oxygen demand,energy transfer,speed)
    ->
GM(fil(preycount,variability,bodymass,oxygen,temperature))[
Stc[preysize * Foodpieces(oxygen demand,energy
    transfer,speed) + stc?]]
IF ’the cod is hungry’ AND ’ingesting the prey will not
    exceed its size dependent stomach volume and oxygen
    dependent appetite and temperature and body mass
    dependent appetite’
@ a rate of 1.0
```

Pseducode for the ‘Ingestion’ rule from the ‘Expanded Juvenile Physiology’ model.

Where GM is the cod (*Gadus morhua*) and its attribute is a boolean determining the ability to ingest food: only if the boolean is TRUE prey can be ingested. Stc is its stomach and stc? a variable that holds the information on the amount of ‘prey pieces’ it contains, and Prey is a prey organism with its respective size, the oxygen demand required for its digestion, and the energy transferred during its digestion and its ‘digestive speed’ (a factor that reflects the variation in the speed at which different prey pieces are digested) as its attributes. This is

---

[120] S. Plante, D. Chabot, and J.-D. Dutil. “Hypoxia tolerance in Atlantic cod”. In: *J Fish Biol* 53 (1998), pp. 1342–1356. DOI: `10.1111/j.1095-8649.1998.tb00253.x`

[121] D. Chabot and J.-D. Dutil. “Reduced growth of Atlantic cod in non-lethal hypoxic conditions”. In: *Journal of Fish Biology* 55 (1999), pp. 472–491. DOI: `10.1111/j.1095-8649.1999.tb00693.x`

the state before the execution of the rule. After the rule is executed (after the arrow), the prey is transformed into Foodpieces (these are normed prey pieces of 0.013g mass) of an amount corresponding to its size and equipped with the attributes relevant for digestion. In the previous version, the boolean was always set to FALSE directly after ingestion. In this updated version, the fil-function (fil: food intake level) is evaluated to set the boolean either to FALSE, ending this foraging session after the last hunting success, or to remain at TRUE allowing the cod to continue to ingest prey without requiring a new successful hunt (hunting is formalised in a separate rule). By having separate rules for hunting and ingestion mechanisms pertaining to the former (energy expenditure, success rate, etc.) could remain conceptually separate from the mechanisms pertaining to the latter (transfer for information about prey, etc.). The fil-function factors in the amount of food in the cods stomach, a calibrating attribute to be used for experimentation named variability, as described above, its body mass, the ambient oxygen and ambient temperature. This function thereby formalises a combination of mechanisms, based on the quantity of stomach content, body mass, ambient oxygen and temperature and of calibration which can be experimented with and is the aforementioned proxy, i.e., 'variability'. If any prey is ingested at all is not only determined by the required successful hunting but also by two guards (to reduce uninformative maths, these are given in plain text - the actual guards can be found in the models' documentation): the cod must be hungry which is modelled based on free metabolites relative to body mass AND (which is a logical 'AND') several equations which, as does the fil-function, factor in the given variables. If all conditions, implicit and explicit, are met, ingestion takes place at a rate of 1.

This example of one aspect of one rule illustrates how this approach can satisfy the need for verisimilitude, comprehensiveness, and complexity, even though ML-Rules was not designed for the domain of fisheries ecology but 'merely' for the domain of cell biology. A very specific process has been formalised, providing verisimilitude, it is reactive to several variables and adjustable in fine detail, providing increasing complexity of interdependencies, while these interdependencies cover a range of different aspects making the process comprehensive.

This updated version of the physiology was subsequently revalidated using simulation experiments introduced earlier on asphyxiation [122] and on oxygen-dependent growth [123] (see Figure 22). The experiments where implemented by modelling experimental tanks to replicate the different treatments. For the example during the growth experiment the temperature was kept at 10°C and the modelled cod were exposed to the same six oxygen saturations as where the cod in the wet-lab. The feeding regime was also transferred by feeding the modelled cod three times a week with prey holding properties modelled on the capelin fillet with which the experimental animals had been fed and all food not ingested after one hour of simulated

[122] S. Plante, D. Chabot, and J.-D. Dutil. "Hypoxia tolerance in Atlantic cod". In: *J Fish Biol* 53 (1998), pp. 1342–1356. DOI: `10.1111/j.1095-8649.1998.tb00253.x`

[123] D. Chabot and J.-D. Dutil. "Reduced growth of Atlantic cod in non-lethal hypoxic conditions". In: *Journal of Fish Biology* 55 (1999), pp. 472–491. DOI: `10.1111/j.1095-8649.1999.tb00693.x`

time was then removed from the modelled tank. The simulated cod are sensitive to oxygen saturations in their growth, although for the highest food intake level (1.0) the plateau in growth observed in the wet-lab for the three higher oxygen saturations could not be reproduced. The results for the high food intake level (0.75) nevertheless, are congruent with the wet-lab results.

This illustrates how simulation experiments can be kept comparable to wet-lab experimentation throughout the expansion of the model, crucially providing accessibility to fisheries scientists who spent their careers in wet-labs and subsequently hold invaluable knowledge.

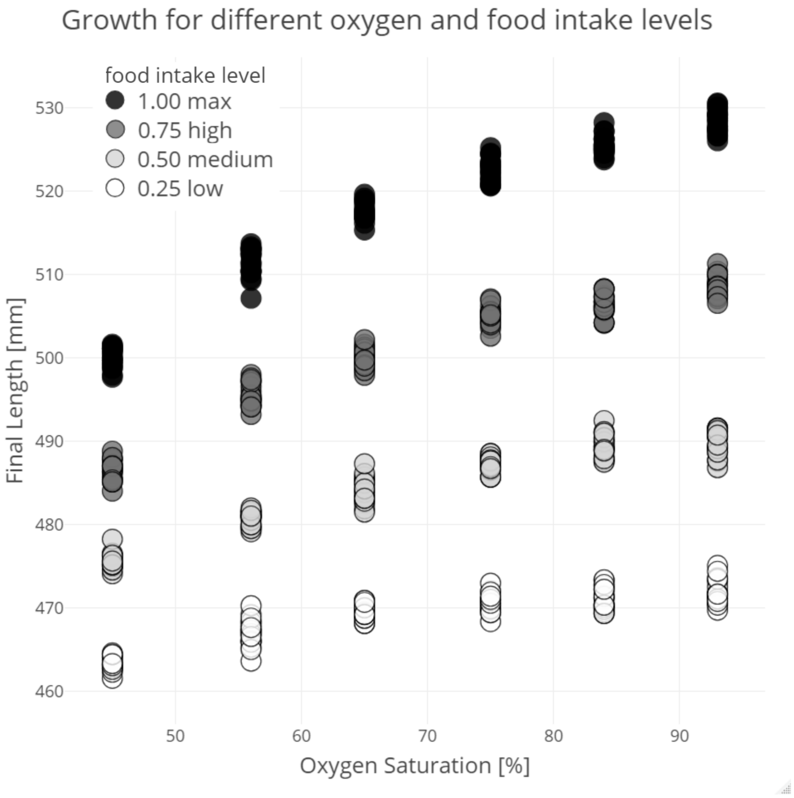


Figure 22: Quantifying the impact of food intake level (variability) control on EBC growth exposed to six different oxygen saturations. Simulation experiment based on twelve week growth experiment by Chabot[153]. Length at beginning of experiment is 442mm with 20 replications per treatment. The simulation results for 'high variability' (0.75) are congruent with the wet-lab results.

However, neither before nor after the expansion of the juvenile physiology presented here, simulation results were able to support the hypotheses that the low condition of EBC stock observed in the field can be explained by the hypoxic conditions found in its habitat. In fact, in this scenario, cod would be required to dwell in sufficiently low oxygen saturations for prolonged amounts of time, which can not be supported by observations of cod behaviour in the field [124]. Therefore further development of the model was required to take into account new hypothesis which in turn required the implementation and functional integration of additional facets of EBC ecology.

[124] Matthias Schaber, Hans-Harald Hinrichsen, and Joachim Gröger. "Seasonal changes in vertical distribution patterns of cod (*Gadus morhua*) in the Bornholm Basin, central Baltic Sea". In: *Fish Oceanogr* 21 (2012), pp. 33–43

Integration of Reproduction

The original model had always been developed with the intention of investigating different interactions within the EBCs environment, however, these were first limited to abiotic and behavioural aspects. As these factors turned out to be insufficient to reproduce the low conditions observed in the field, other facets required for different lines of investigation were formalised and subsequently integrated with the existing model, both by expanding the existing rules (functional integration of the additional facet, or a part of that facet, into an already formalised process) and by adding new rules (adding new processes). The next facet required for further investigation is reproduction allowing to experiment with the physiology and behaviour of adult cod.

Reproduction

Formalising reproduction touches (at least) two fields of research: physiology and ethology and links between the two. The physiological aspects can be further subdivided into bioenergetics (e.g., formalising how reproduction uses physiological resources) and reproductive biology (e.g., formalising stages of gonadal development, etc.). This facet of cod ecology was kept distinct by introducing gonads as an entity and focal point for reproduction. Modelled gonads contain information about: sex, number of eggs, maturity of these eggs and their weight:

Pseudocode of attribute added to the cod and new entity representing gonads to implement reproduction.

```
GM(reproductive stage)[
    Gon(sex,number,maturity,weight)]
```

Information about the stage in the reproductive cycle of each cod is attributed to the cod entity, however, this is a question of style and could be easily changed to also be an attribute of the gonads, if desired. This decision is an example of how the mental model of the scientist developing a model will translate into the concrete formalisation. In the desired setting of including several scientists, a consensus on the most suitable entity to carry the attribute would be found, which, moving further, will provide an alignment of mental models, thus supporting clear communication. An example for an alternate version would be:

Pseudocde of alternate manner in which to distribute attributes added for reproduction.

```
GM[Gon(sex,reproductive stage,number,maturity,energy
    counter,weight)]
```

The drain of building gonads on physiological resources (oxygen and metabolites) is implemented by three rules added to the juvenile physiology, the first rule formalises energy allocation:

```
GM(oxygen,metabolites,reproductive stage)[Gon(weight,
    energy counter)] ->
GM(oxygen-4,metabolites-12,reproductive stage)
    [Gon(weight+growth increment, energy counter +
    growth increment)]
IF reproductive stage = preparing AND 'oxygen is
    sufficient' AND 'metabolites are sufficient for cod
    mass'
@ a rate of 'dependent on temperature and body mass' *
    'exponential body mass relation'
```

Pseudocode for the 'Gonad energy allocation' rule from the 'Adult' model.

Here the oxygen and metabolite reserve of the cod is decreased by a set amount, and the weight of the gonads and the 'energy counter' is increased by a corresponding amount, i.e., the growth increment (gi=0.00587[g]), upon execution of the rule (the energy counter is used to track the different stage of gonad maturity its use is explained in the following example). Energy-mass-balance considerations currently determine the value of the growth increment. The guards for the execution of the rule are that the cod (GM) is at a point in its reproductive cycle where the growth of gonads is indicated (preparation stage) AND that the GM has sufficient oxygen to spend on gonad development, AND that it has sufficient metabolites in relation to its size to spend on gonad growth. The rate of execution of this process is, as are most physiological processes, a function of its temperature optimum, determined by its mass and the ambient temperature and its size accounting for declining length growth with increased mass.

This example illustrates how new rules functionally integrate with the existing model and thereby satisfy the need for comprehensiveness: simply by making use of the established resources (oxygen and metabolites), the rule ties in with the existing formalisation of the juvenile physiology and represents the added expenditures the adult physiology coordinates. It also illustrates how very specific relations can be individually investigated in a more extensive context/formalised system: by setting guards (here referring to the second and third guard) for the use of these resources, it allows to experiment with the physiological priorities of the different processes. Currently, gonad development is prioritised over the length and mass growth of the parent cod in terms of lower cut-off values for available oxygen and metabolites than growth. Experimenting with different rankings, therefore, merely requires changing two values (see also Table 3).

The second and third rules added to the juvenile physiology control the different stages of growth/maturation of gonads (the two rules are identical except for the percentage of body mass, which is 8% for females and only 5% for males). These rules are an example of accommodating comparisons with field data where stages of gonadal development are frequently used for classification. This represents a vital link to making the results of simulation experiments and field data comparable, thus providing access to fisheries scientists who have acquired invaluable knowledge through time spent at sea conducting surveys and gathering data. They also represent the link between bioenergetics and reproductive biology to be further refined to account for mechanisms regulating gonadal development (e.g., re-absorption of eggs, skipped spawning, etc.).

Pseudocode of a ‘Gonad growth’ rule, moving the gonads through eight stages of maturity.

```
GM()[Gon(maturity, energy counter)] ->
GM()[Gon(maturity+1, 0)]
IF ’full maturity is not reached’ AND ’cod is female’
    AND ’energy content has reached one eighth of 8\%
    body mass’
@ a rate of 1.0
```

Where the energy (oxygen and metabolites) invested in gonad growth has been tracked by the ‘energy counter’, and upon execution of the rule, this is reset to zero while the maturation stage is increased by 1. The guard for the execution of the rule is that the gonads are not fully mature, AND the energy counter has reached one eight of the total gonad mass set in relation to the body mass of the cod (8% for females and 5% for males). The rule is executed if the guards are met with a rate of 1.

The introduction of the reproductive cycle, currently formalised by the four stages of: ‘spent’, ‘preparing’, ‘ready to spawn’ and ‘spawning’ also provides a tangible link between physiology (bioenergetics and reproductive biology), environment and behaviour. Now, independent of perspective (e.g.: bioenergetics, behaviour, etc.) naming the ‘state’ of the cod offers control over proposed causality and a clear and efficient communication.

For example: energy allocation to the gonads begins as soon as the respective individual transitions to a state of ‘preparing’, this transition is controlled by daylength, stating a clear chain of causality: daylength triggers onset of gonadal development; the ‘behavioural decision’ of movement to the Bornholm Basin deep is triggered by completion of growing mature gonads: mature gonads cause movement to spawning grounds (this example is shown in the ‘environment’ section).

As a ‘proof of concept’ the full life cycle from eggs produced during spawning, through their maturation (or failure to mature), to a larval stage was implemented thus showing that multi-generational investigation (for example the question of genetic mixing with other stocks) is also feasible. The integration of reproduction into the existing physiology and the integration of new facets into the model is partially illustrated in Figure 27.

Now, while continuing to provide full access to all mechanisms, the impact of the facets of reproduction on physiology and behaviour and the impact of environmental factors on reproduction have been functionally integrated (see Figure 21). Nonetheless simulation experiments with adult cod would also not reproduce the low conditions observed in the field as the expenditure of resources for reproduction is not sufficiently high (and was not expected to be), rather they needed to be in place for the following investigations. Subsequently the next step was to introduce further facets to provide a 'feature complete' set-up, consisting of realistic types and distribution of prey and a functionally structured environment to take into account yearly cycles of abiotic conditions and behaviour in order to investigate hypothesis which require a more intricate set of mechanisms.

Prey design

Formalising **`Prey`** entities inherently allowed for the straightforward formalisation of both quantity and especially the quality of prey while maintaining a structured and thereby accessible approach, as any desired properties are simply tied to the entity. Additional factors such as infection of prey with parasites were also included allowing them to be used as a vector and thereby preparing for the formalisation of highly entangled phenomena (high quality, highly infected prey vs. low-quality non-infected prey or infection gradients across the Baltic, etc.).

It is also an example of methodically (and tediously) translating known factors (energy values and speed of digestion correlating with oxygen demand and efficiency of digestion) into a model to benefit as much as possible from known factors (here the data was taken from [125]), thus grounding as many quantitative aspects as possible to be able to speculate on others.

```
Prey(species, size, oxygen demand, energy transfer,
    digestive speed, A. simplex, C. osculatum))
```

Pseudocode of the entitiy 'Prey' and its attributes.

Prey is also an instance of an extensive proxy, as prey attributes represent the cumulative result of the impact of a vast range of mechanisms: the quantity of different types of prey can be a result of their own, respective, stock dynamics and their distribution can be a result of factors ranging from abiotic conditions to interactions with the predator population and the quality of prey items can be a result of mechanisms at the system level of plankton production (and below) and translate through the food web with all the interactions occurring on this journey (resulting, for example, in a reduced prevalence of essential nutrients).

As mentioned earlier in regard to food pieces this also illustrates the capacity of individual based modelling to support the investigation of hypothesis that require pronounced heterogeneity. While equation-based models typically rely on averages of available nutrients when simulating growth (for both quantity and quality), IBMs are able to introduce specific prey items that, for example, carry a particular essential nutrient and subsequently experiment with their frequency at exceptionally low levels, hence allowing for investigation of hypotheses that consider nutrient depletion scenarios.

For simulation experiments, knowledge of prey characteristics was implemented, and the composition of prey in different depth strata, surveyed in the form of stomach samples, was used to populate the modelled environmental compartments (which will be introduced in

---

[125] Peter Hornetz. "Spatio-temporal distribution and food intake of cod (*Gadus morhua*) along a depth gradient in the western Bornholm Sea". MA thesis. University of Hamburg, 2020

the following section) with the appropriate prey, thus providing food for the modelled cod during at liberty simulations (these reproduce field experiments where animals are set free to be recaptured at a later date aka spending time at liberty). For experiments containing the facet of parasitation, the prey was infested at the different desired levels. Additionally, the 'infrastructure' for the effect of prey on behaviour has also been implemented, although rules formalising specific processes have not yet been included (see Figure 21).

Environment and Behaviour

To provide an environment sufficient to formalise interactions between physiology, abiotic and biotic environmental conditions and to tie these in with behaviour in at liberty simulations, a series of 'environmental boxes' were built. This will allow for more detailed and elaborate narratives about the impact of different facets in the field (type of prey, the structure of habitat, etc.) to be explored. At a minimum, these are four boxes representing the 'functional habitats' found in and around the Bornholm Basin: the Bornholm basin Deep (for spawning), the Bornholm Basin Pelagic (for transition and sometimes hunting), the Bornholm Basin Slope (for hunting on occasion) and the Bornholm Island Slope (for hunting most of the year). Since the underlying mechanisms controlling the movement of cod throughout the Baltic Sea are all but unknown, these implementations are, of course, highly speculative and more of a blanc but a willing canvas to pose and explore various hypotheses and further clarify and investigate these. Environments contain information about: temperature, oxygen saturation, salinity, day of the year, zone, substrate, prey availability (which is part of the 'infrastructure' mentioned in the previous section), surface area and maximum and minimum depth, although not all of this information is utilised yet.

To include more variation in abiotic conditions, the enormous volumes of the environmental boxes were each subdivided/fragmented into sub-volumes with slight variations in temperature and oxygen saturations. To make two spreads of conditions available, a version with a fragmentation into six and a version with a fragmentation into twelve environmental boxes were developed. The space is discretised into volumes (represented by ML-Rules compartments), each of which contains an index (which allows to determine its position) and a depth range. Rules take this index into account when determining the movement of EBC between neighbouring volumes, implicitly modelling a spatial grid of adjacent compartments.

Formalising movement takes place foremost through guards while the content of the respective rule is trivial: the cod leaves one compartment and enters an adjacent one. However, under which conditions it will do this represents its behaviour, in the current version of the model, this is influenced by day length (representing the time of year) and stage of the reproductive cycle. Although, including the influence of other factors such as prey availability, salinity, and many others would require little effort.

```
Box('bornhom slope',day)[Cod] + Box('bornholm
    pelagic',day)[] ->
Box('bornhom slope',day)[] + Box('bornholm
    pelagic',day)[Cod]
IF 'the cods reproductive stage is `ready'' OR ('the
    cods reproductive stage is `spent'' AND 'daylenth
    is shorter than 7 hours')
@ a rate of 1.0
```

Pseudocode for the 'bornholm slope -> bornholm pelagic' role guiding movement of the cod through the habitat compartments.

More importantly, movement is now tied to explicit reasons which can be further investigated, refined or rejected. From an ethological perspective, movement of EBC through their habitat currently has a very speculative basis for hypothesis since experimentation would require not only monitoring but also manipulation of (at least) an entire basin. Having a model which allows to transparently run alternative hypotheses of the underlying mechanisms that guide movement provides the opportunity to hone in on the most likely candidates before returning to the field to seek corroboration (this is one way in which this type of approach provides feedback for surveys and experimentation, as stated in Chapter II). At this point integration of all facets, bar parasitation, indicated in Figure 21 have been successfully implemented an could be used for simulation experiments. For these simulation experiments twenty-five pairs of a male and a female cod, which are now equipped with behaviour to move through their environment and reproduce where 'set free' to be at liberty for one year and were 'monitored' weekly. Since the simulation provides stochasticity for all process the simulated individuals display a variance in growth as do the animals in the field. At this point experimentation with the model produces data which in its scope of available parameters not only covers what fisheries scientists are accustomed to receive from surveys (i.e., length and weight distributions) but exceeds this by data they can not easily survey (i.e., movement throughout the year, length and weight at previous points in time or the amount of food an individual has ingested).

Movement was also visualised (as a 2D animation) to allow for face validation and compare modelled behaviour with hypotheses and/or insights from tagging data [126] (see Figure 23). This is an additional advantage of the high degree of detail this type of modelling and simulation can provide: by being able to log all attributes of all entities, detailed visualisations are supported, in this case illustrating 'narratives' of the mechanisms that have been formalised, supplying a spatio-temporal visualisation for spatio-temporal relationships. This type of visualisation (e.g., showing the movement, growth, and reproductive cycle of individual cod as a complex whole) then also aids in the alignment of mental models as different participating scientists can use this artefact as a focal point for discussion. Additionally, as spatio-temporal visualisation of field data becomes a more fully developed and established analytical tool, the comparison of modelled and the corresponding real world systems will benefit from the

[126] Stefanie Haase et al. "Validation approaches of a geolocation framework to reconstruct movements of demersal fish equipped with data storage tags in a stratified environment". In: *Fisheries Research* 237 (2021), p. 105884. ISSN: 01657836. DOI: `10.1016/j.fishres.2021.105884`

spatio-temporal resolutions of simulations matching field data. Building on the availability of detailed visualisation for modelled systems and real-world data can then make optimal use of the considerable capability of humans to process visually displayed narratives [127].

However, even with unrealistically high expenditure for swimming/activity and the most adverse yet realistic combinations of oxygen saturations and temperature (low oxygen and high temperatures - data on observed abiotic conditions in the corresponding areas of the Baltic Sea was supplied by the IOW) low growth and low condition observed in the field could not be reproduced. Therefore additional facets impacting the cod needed to be implemented for further investigation.

Parasitation

Not only the next prominent candidate for negatively impacting the EBC but also an example of a current 'chicken or the egg problem' is parasitation of cod livers with *Contracaecum osculatum*[128]. Regarding causality, it is not yet clear whether parasitation causes the low condition or low condition encourages high levels of parasitation. As with reproduction, an entity, in this case, a liver, was introduced as the focal point for formalising the process of uptake of liver parasites by logging their amount.

Pseudocode of entity 'Liver' and its attributes.

```
Liver(number Anisakis simplex, number Contracaecum
    osculatum)
```

Although mechanisms for uptake were only formalised for *C. osculatum*, the 'infrastructure' for *Anisakis simplex* was also provided in anticipation of likely interactions. This demonstrates the relative ease of adding structurally similar factors. To investigate the two alternative hypotheses regarding uptake of *C. osculatum*, two separate mechanisms were implemented.

For both mechanisms, the progression of infection was based on a sigmoidal curve (see Equation 8), this implicitly postulates that the first stages of infection are met with some immune defence, the further infection has a fast progression, and ultimately there is a form of 'carrying capacity' of the liver. This is an example of how clearly non-linear relationships can be implemented and consequently tested in broader settings. Also, for both cases, the infestation was integrated into the ingestion rule so that increased parasitation of the liver requires the ingestion of infected prey.

Equation 8 describes the sigmoidal curve modelling the continous increase of parasitation with discrete ingestion of infested prey. I is the infection with parasites and 20 is the assumed carrying capacity of the liver.

$$I_{parasite} = \frac{20}{1 + \mathrm{e}^{-ln - \frac{I_{parasite-1}}{I_{parasite-1}-20} + 1}} \tag{8}$$

The ingestion rules was already introduced in the section on the expanded juvenile physiology, the following example shows the expansion which integrates the mechanism of ingestion based parasite increase:

---

[127] Gennady Andrienko et al. "Space, time and visual analytics". In: *International Journal of Geographical Information Science* 24.10 (2010), pp. 1577–1600. ISSN: 1365-8816, 1362-3087. DOI: `10.1080/13658816.2010.508043`

[128] Marie Plambech Ryberg et al. "Physiological condition of Eastern Baltic cod, *Gadus morhua*, infected with the parasitic nematode *Contracaecum osculatum*". In: *Conservation Physiology* 8.1 (2020). Ed. by Steven Cooke, coaa093. ISSN: 2051-1434. DOI: `10.1093/conphys/coaa093`

Pseudocode of the ‘Ingestion’ rule for the ‘Liver’ model. Implementing the ingestion based hypothesis H1.

```
GM()[Stc[stc? ] +
Liver(C.o.Liver)] + Prey(preysize, C.o.Prey) ->
GM()[Stc[preysize * AJPO] +
Liver (C.o.Liver + spi(C.o.Liver,C.o.Prey))]
IF 'the cod is hungry' AND 'ingesting the prey will not
    exceed its size dependent stomach volume and oxygen
    dependent appetite and temperature and body mass
    dependent appetite and parasite dependent apetite'
@ a rate of 1.0
```

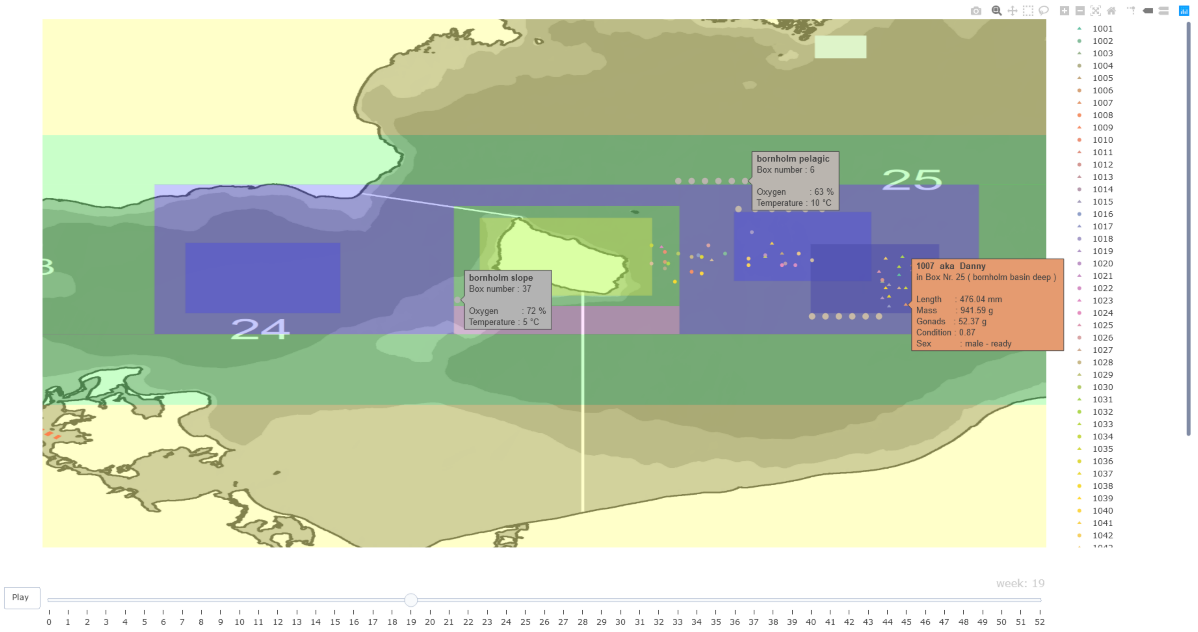
Here the guards remain unchanged from the original rule (see Expanded Juvenile Physiology example) as ingestion occurs regardless of the parasitation of the prey, which implies that cod can not identify or do not care if the prey is infested. The spi function (sigmoidal parasite increase) formalises the described progression of infection, taking into account only the current amount of parasites in the cod’s liver and if the ingested prey contains parasites (see Figure 24 left). The alternate expansion of the ingestion rules merely requires using a different function, albeit a more elaborate one. The cfbpi function (condition factor based parasite increase) factors in not only the current parasite load of the liver and if the ingested prey contains parasites but also the condition of the ingesting cod, calculated from its mass and length:

Pseudocode of the ‘Ingestion’ rule for the ‘Condition’ model. Implementing the condition based hypothesis H2.

```
GM(body mass,length)[Stc[stc? ] +
Liver(C.o.Liver)] + Prey(preysize, C.o.Prey) ->
GM(body mass,length)[Stc[preysize * AJPO] +
Liver (C.o.Liver + cfbpi(body mass,
    length,C.o.Liver,C.o.Prey))]
IF 'the cod is hungry' AND 'ingesting the prey will not
    exceed its size dependent stomach volume and oxygen
    dependent appetite and temperature and body mass
    dependent appetite and parasite dependent apetite'
@ a rate of 1.0
```

The distinction between the two mechanisms is mathematically implemented by considering the condition of the individual fish for the second hypothesis: where in the first case, any ingestion of infested prey will cause an increase in *C. osculatum* load, in the second case, the original sigmoidal function is modulated with an additional factor. This factor [1,0.1] is calculated by a negative exponential function over the range of cod condition, which can vary between poor (0.6) and good (1.2), meaning for a poor condition, the increase per ingestion of infected prey is identical to the first mechanism but is only one-tenth for cod in good condition (see Figure 24 right).

This implementation is also an example of a first stage of iteratively closing in on more refined differentiation between hypothesised mechanisms: the assumption was made that for both causal directions, the increase of parasites progresses along a sigmoidal curve with the same shape. In a subsequent iteration, this assumption could be rejected in favour of using two differently shaped (possibly not even sigmoidal) curves. Rather than having to ‘get it right on the



bornholm pelagic
Box number : 6

Oxygen : 63 %
Temperature : 10 °C

bornholm slope
Box number : 37

Oxygen : 72 %
Temperature : 5 °C

1007 aka Danny
in Box Nr. 25 (bornholm basin deep)

Length : 476.04 mm
Mass : 941.59 g
Gonads : 52.37 g
Condition : 0.87
Sex : male - ready

Figure 23: Screenshot of visualisation in R for 'at liberty' simulation experiments including visible mouse-over for two environmental compartments and one individual fish. Compartments are overlaid on a map of the southern central Baltic Sea featuring the Bornholm Basin (east of the Bornholm Island) as a spawning ground. Each environmental compartment is fragmented into six versions with slight variations in temperature and oxygen saturations. Slider tracks weeks of simulation experiment.

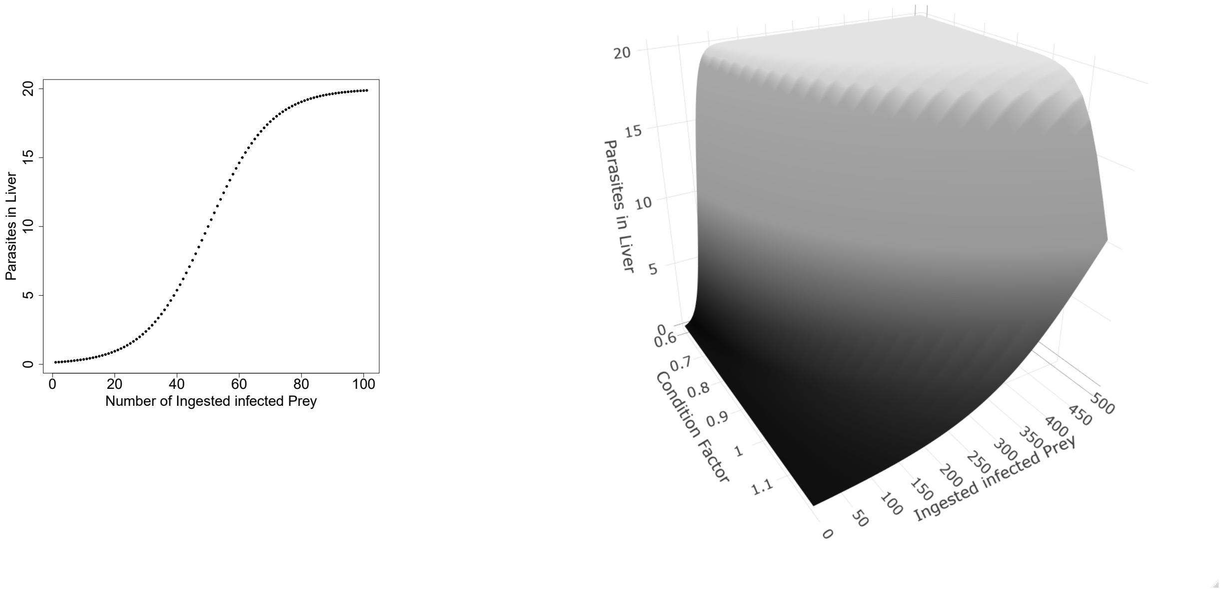


Figure 24:
Left: Progression of increase in the parasite *C. osculatum* with number of ingested infected prey.
Right: Progression of increase in the parasite *C. osculatum* with number of ingested infected prey and dependent on condition of cod at ingestion.

first try', there is a starting point from which iteratively to hone in on the correct relationship, which could be exceedingly mathematically complex.

As with reproduction, the impact (of parasitation) on bioenergetics was integrated into the physiological sub-model: wet-lab experiments found a quantifiable impact on basal metabolic rate and 'appetite'[129] and an estimate of both effects was made part of the respective rules in the physiological model. The impact on appetite is part of the guards of the ingestion rules, see examples above, and the integrated standard metabolic rate rules now reads as:

```
GM(oxygen  ,metabolites  )[Liver(C.o.Liver)] ->
GM(oxygen-1,metabolites-1)[Liver(C.o.Liver)]
IF 'oxygen is greater zero' AND 'metabolites are
    sufficient not to go into starvation mode'
@ rate of (smr(temperature) - dsmr(C.o.Liver)) *
    'linear body mass relation'
```

Pseudocode for the 'Standard metabolic rate' for the 'Liver' and the 'Condition' model.

Where oxygen and metabolites are removed at an experimentally determined rate[130] translated into the smr function (standard metabolic rate) to cover basic metabolic functions and is scaled to account for the respective body mass of the individual cod. The rule is executed if the cod has a minimum of oxygen and enough metabolites, if it does not have any oxygen, the guard of the asphyxiation rule is satisfied and the cod dies, and if it does not have enough metabolites, the basic metabolic functions are covered at the expense of its body mass which is handled by the starvation rule. The integration of the experimentally observed impact of parasitation lowering the standard metabolic rate is implemented by lowering the rate by an amount calculated using the dsmr function (decrease of standard metabolic) rate, which takes into account the infestation level of the liver and was derived from experimental findings [131].

With the two mechanisms (H1 and H2) implemented, they can be experimented with. First, the feeding regime of the recreated wet-lab growth experiment, as introduced in the section on the expanded juvenile physiology, was altered to include various proportions of infested prey (see Figure 25 - H1). Then the same experiment was set up to include cod of a range of different conditions by changing the initial weight of the simulated cod to investigate the second mechanism (see Figure 26 - H2). It is interesting to note that these experiments reveal open questions concerning the impact on bioenergetics since simulated cod with low levels of parasitation out-perform those with no parasitation Figure 25. This particular result can now be speculated on: naturally, the first line of investigation is to double-check if this is an artefact resulting from improper implementation of known mechanisms in the model. If this can be excluded, it could be discussed if further wet-lab experimentation is required to uncover additional impacts of parasitation as the recorded ones lead to these unintuitive results, or if wet-lab experiments formed a complete picture and the implementation was sufficient, this might support the (improbable) hypothesis that the EBC as adapted to low levels of parasitation to

---

[129] Marie Plambech Ryberg et al. "Physiological condition of Eastern Baltic cod, *Gadus morhua*, infected with the parasitic nematode *Contracaecum osculatum*". In: *Conservation Physiology* 8.1 (2020). Ed. by Steven Cooke, coaa093. ISSN: 2051-1434. DOI: `10.1093/conphys/coaa093`

[130] G. Claireaux et al. "Influence of water temperature and oxygenation on the aerobic metabolic scope of Atlantic cod (*Gadus morhua*". In: *Journal of Sea Research* 44.3 (2000), pp. 257–265. ISSN: 13851101. DOI: `10.1016/S1385-1101(00)00053-8`

[131] Marie Plambech Ryberg et al. "Physiological condition of Eastern Baltic cod, *Gadus morhua*, infected with the parasitic nematode *Contracaecum osculatum*". In: *Conservation Physiology* 8.1 (2020). Ed. by Steven Cooke, coaa093. ISSN: 2051-1434. DOI: `10.1093/conphys/coaa093`

the point of thriving with them. Please note that this speculation is meant to illustrate how the formalisation of intricate processes can support a workflow of structured investigation of interacting mechanisms. For the second mechanism, the experiment only used all or no infected prey, and therefore variation of infection resulted from initial conditions subsequently, the mentioned open question presents very differently as parasitation of the low condition individuals rises quickly and the advantage of a decreased resource requirement for the standard metabolism sets in earlier (see Figure 26).

As an example, this process shows interactions of numerous factors (ambient oxygen, condition, physiological performance and parasitation) through several mechanisms (condition-based increase of parasite load and its effect on physiological performance) while remaining accessible. As a result, the implementations demonstrate verisimilitude by formalising the different hypotheses as intended by the conceptual model, comprehensiveness by functionally integrating yet another facet into the model, and complexity by providing tie-ins into several processes without the need to eliminate any previously implemented interactions. It can also be used for simulation experiments to reproduce wet-lab experiments and investigate competing hypotheses.

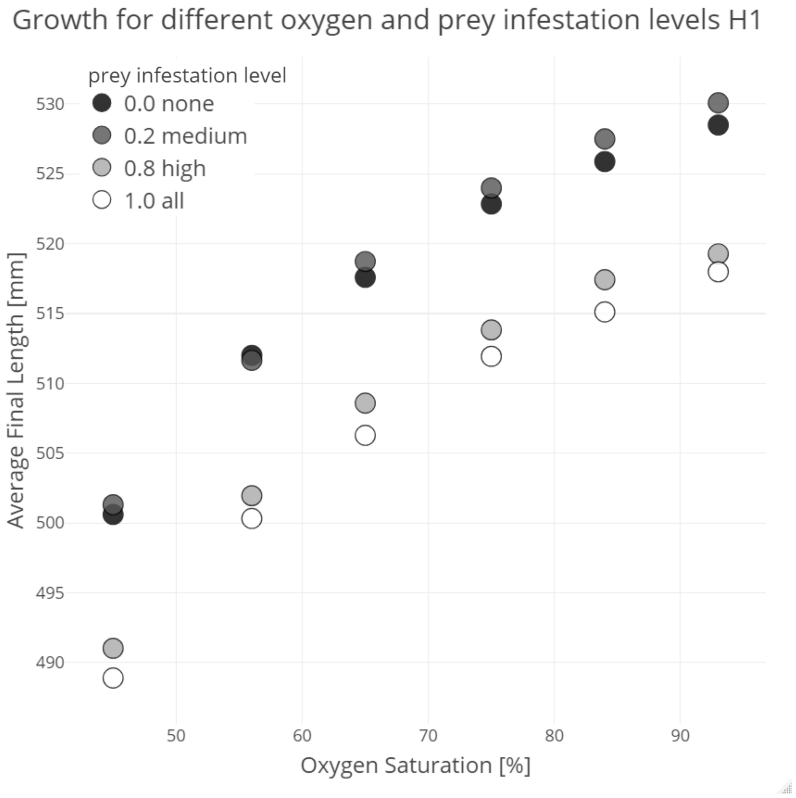


Figure 25: Simulation experiment based on Chabot et al. (1999) twelve week growth experiment at different oxygen levels. Impact of parasitation on growth for four levels of prey infestation [none, medium, high, all] with *C. osculatum*. Parasite uptake based on ingestion of infected prey (H1).

With the functionality of the integration of the two parasitation hypothesis established, these two types of modelled cod were also subjected to the at liberty experiments in the modelled environment.

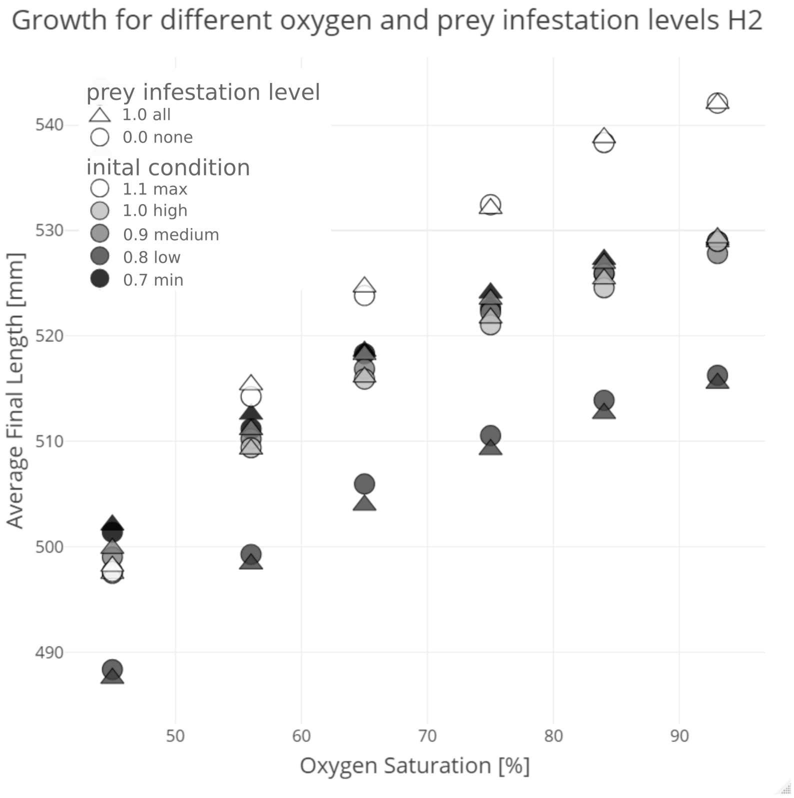


Figure 26: Simulation experiment based on Chabot et al. (1999) twelve week growth experiment at different oxygen levels. Impact of parasitation on growth for two levels of prey infestation [none, all] with *C. osculatum* and five levels of initial condition factor [min, low, medium, high, max]. Parasite uptake based on condition at time of ingestion of infected prey (H2).

Simulations were run for both mechanisms with varying amounts of infected prey and for the second mechanism (H2) with broad ranges of initial condition factors. However, again the reduced growth observed in the field could not be reproduced experimentally.

At this point, the only factor that was able to impact growth sufficiently was to severely restrict the modelled cod's 'appetite', which has already been introduced as an extensive proxy and therefore results produced by its manipulation do not provide definite answers but nonetheless a rough direction. Therefore, the Baltic ecosystem changes that have the most significant impact on the EBC are either not yet included or not yet pieced together correctly in the model.

Discussion

As illustrated by Figure 27 and the examples in this chapter, the model has been successfully developed to a high degree of comprehensiveness and complexity. For several processes, at least two facets could be functionally integrated, e.g., standard metabolic function is impacted by environmental factors and parasitation, and the movement through the habitat draws on input from the reproductive cycle (which in turn is impacted by physiology) and abiotic conditions. During these integrations, no simplifications were required in terms of how interdependencies were formalised neither regarding maths (i.e., by using products of functions) nor regarding interaction

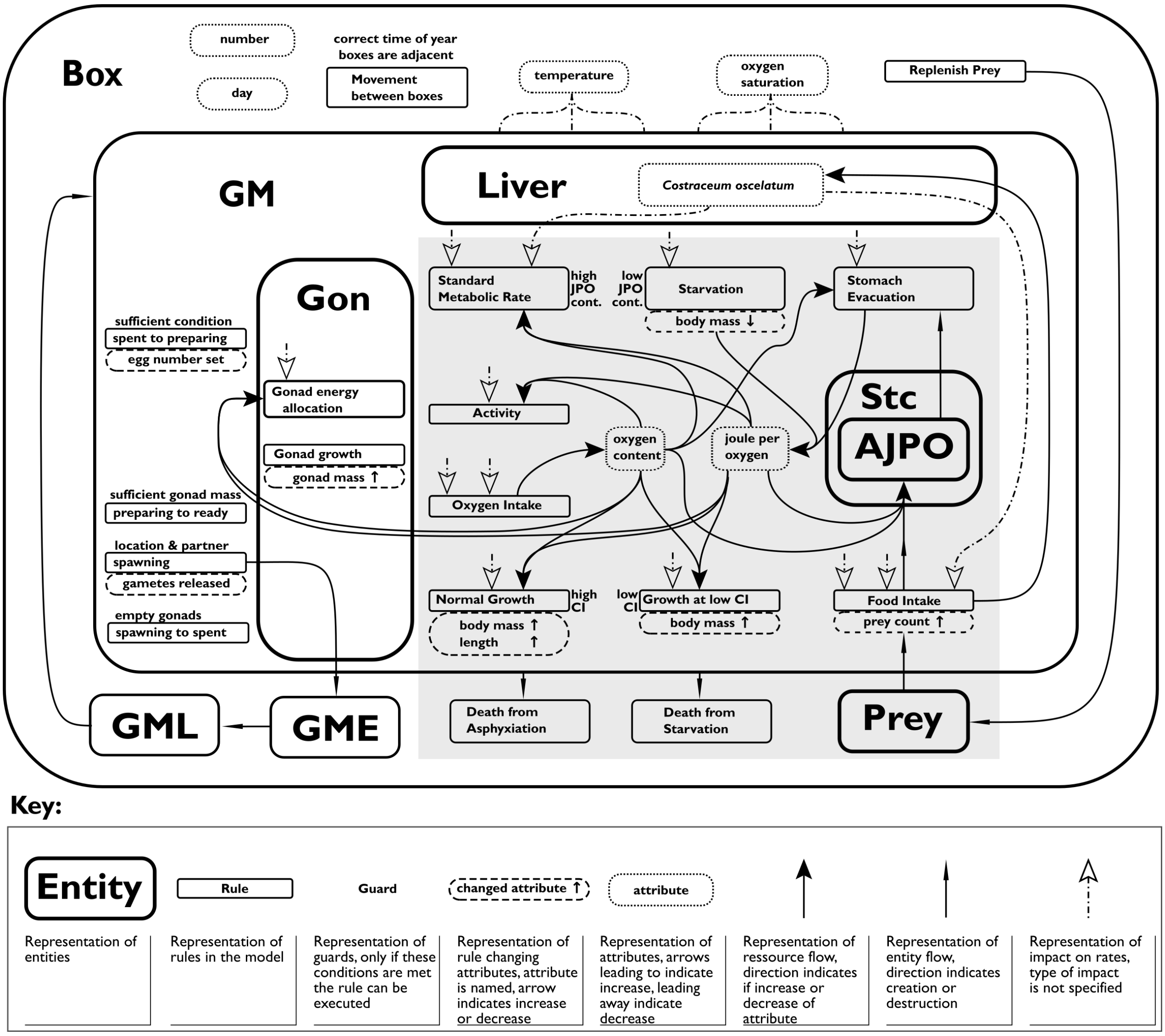


Figure 27: Representation of the major aspects of the expanded Eastern Baltic cod model - the grey background marks the physiology of the juvenile cod. The central elements are the different entities defined by their attributes. Entities can be nested, here the Box entity provides an environment for the cod (GM) which in turn contains organs like a stomach (Stc), gonads (Gon) and a liver (Liver). Entities can be created, for example, eggs (GME) are created (narrow arrows pointing towards) by the reproductive cycle and removed (narrow arrows pointing away) when they develop into larvae (GML). Rules govern the behaviour of the system. The 'Oxygen Intake' rule, for example, has a rate that is influenced by the surrounding oxygen saturation and temperature (hollow arrows) and its execution increases the attribute of oxygen content of the individual cod (fat arrows). Rules can have guards, meaning that a rule can only be executed if certain conditions are met. Normal growth, for example, can only take place if the cod has a sufficiently high condition (body mass relative to length). Please note that this representation is far from complete and is meant to illustrate the different types of interdependencies that can be formalised in a legible white box manner with a domain-specific modelling language.

of entities (i.e., by using attributes to transfer relevant information). For those familiar with the model, the level of verisimilitude was also satisfactory, however, this would ultimately have to be judged by a broader scope of the fisheries science community to deliver a meaningful verdict.

Experimentation with the models was also possible in the manner intended, allowing for the reproduction of wet-lab experiments for validation, general comparisons, and in particular, the comparison of competing hypotheses. The simulation of 'narratives', the mental models of fisheries scientists surveying and familiar with conditions in the field, equally supports using the model as a'thinking tool'.

To the extent this is possible when investigating CAS, the requirement of providing accessibility was also met. One example is the separation of the bio-energetic and the developmental facets of reproduction, whereby precise control of both can be maintained during further development. Another example is the implementation of alternative models for the two opposing causal directions of the relationship between parasitation and condition, thereby having the two mechanisms available for testing in different settings.

The specific DSL Ml-Rules supported the general approach very well, however, two shortcomings based on the differences between cell biology and fisheries science stood out. Regarding syntax, the number of attributes required for the model presented here was not catered to and made the more elaborate models hard to read and semantically the lack of an 'environment' repurposing of compartments for this function.

Regarding open questions about EBC ecology, the simulation study did not provide a definitive answer about the causes of its decline, although a number of suspected causes (e.g., ambient oxygen levels or parasitation) now seem less likely. The decline in growth observed in the field could only be reproduced by a severe restriction of 'appetite' which translates into a lack of energy. Therefore, the next step in this investigation will be to determine if the cod is unable to convert energy efficiently (i.e., due to a lack of micronutrients, for example) or if the cod is unable to acquire sufficient food (i.e., due to a lack of suitable prey, ability to hunt or additional reason).

Chapter IV
Establishing and Maintaining Validity

When to use iterative development? You should use iterative development only on projects that you want to succeed.

Martin Fowler

Validation of Simulation Models

A comprehensive overview of modelling and simulation terminology was published as early as 1979 and already included a definition for semantic model validation stating that validation is a:

> "Substantiation that a computerized model within its domain of applicability possesses a satisfactory range of accuracy consistent with the intended application of the model".[132]

This definition can be broken down into three aspects. Firstly 'substantiation' implies that something should be shown or rather provided proof for, meaning that validation aims to satisfy a burden of proof placed on the model. Secondly, that what shall be shown is its applicability or usefulness for a purpose. Thirdly, that this applicability and/or usefulness comes with a considerable caveat, namely that the scope of its applicability will always be limited.

This also aligns with the definition of a model given in Chapter II: "A model (M) for a system (S) and an experiment (E) is anything to which E can be applied in order to answer questions about S",[133] since this definition of a model implies that the model can be exerted in place of the real-world system to test its behaviour. This can only be true if the model is valid according to the definition given above.

Regarding the types of exertions, i.e., the experiments, ZEIGLER had pointed out that the validity of a model can not be shown independently from the experiments it is valid for. This aligns with the 'intended application' given in the definition above. He introduces the concept of the "experimental frame", which provides the set of experiments a given model is valid for, while it has not been shown to be valid for any experiment outside of this frame[134]. Bluntly put, a model of the EBC, which is valid for experiments pertaining to oxygen, having been validated for realistic asphyxiation and oxygen-dependent growth, is not automatically also valid for experiments on the impact of the lunar cycle on vertical migrations[135].

Therefore, if modelling and simulation are to be used to explore and understand the EBC, or any complex adaptive system, this requires that models be validated for the 'experimental frame' from which experiments will be applied. Where validity can not be estab-

---

[132] Stewart Schlesinger et al. "Terminology for model credibility". In: *SIMULATION* 32.3 (1979), pp. 103–104. ISSN: 0037-5497, 1741-3133. DOI: `10.1177/003754977903200304`

[133] François E. Cellier. *Continuous System Modeling*. New York, NY: Springer New York, 1991. ISBN: 978-1-4757-3924-4 978-1-4757-3922-0. DOI: `10.1007/978-1-4757-3922-0`

[134] Bernard P. Zeigler. *Theory of modelling and simulation*. New York: Wiley, 1976. 435 pp. ISBN: 978-0-471-98152-7

[135] a serious hypothesis connected to the movement of prey

lished, the lack of validity needs to be well known and part of all interpretations of results. Additionally, the general approach introduced in Chapter II[136] requires the integration of different models. To ensure continued validity, any validity for an 'experimental frame' established before integration must consequently be re-established after integration.

These considerations address the requirements of the general approach, however, additional consideration can be derived from the overall objective of informing management. In this context validation can also be approached as part of the triad of verification, validation and accreditation of simulation models[137]. By combining the concepts of validation (as introduced above), verification (i.e., is the model built correctly) and accreditation (i.e., "the official certification that a model or simulation is acceptable for use for a specific purpose.") together they attest to the accuracy of models and simulations at a standard fit for the purposes of management support. Such credibility will ultimately be required if the approach is implemented on a larger scale. In that case, ICES could take on the role of the accreditation body, as it does for stock assessment models, and verification, as argued in the sections on 'Separation of Model and Simulator – Trust and Transparency', 'Semantic Domain - Verisimilitude of Functionality' and 'Domain Specific Languages - Accessibility' in Chapter II, would be addressed by providing a formal modelling language and by separating model and simulator and thereby making the model description unambiguous and providing opportunity to thoroughly scrutinise the simulator.

Here the focus will be on providing continued validity of simulation models throughout the successive integration of processes and iterations of model building.

Maintaining Validity by Reusing Simulation Experiments

Building the first iterations of the EBC model required expertise from various branches of biology, e.g., physiology, ethology, and expertise in abiotic conditions. Each of these fields can be regarded as its own scientific domain. As argued in Chapter II, it is not the most efficient course of action to ask an individual scientist to formalise models outside their area of expertise, and developing and validating smaller submodels for the different phenomena and subsequently integrating them is a more practical approach. These models, concerned with a manageable amount of complexity, can be handled by a single or a few scientists. However, constructing a larger, more complex model from such existing models might risk "breaking" the model, i.e., it losing its semantic validity in the process [138].

The systematic reuse of simulation experiments specifications to ensure semantic validity of larger models after composition has already been explored [139]. In this approach hypotheses about the behaviour of the model over time defined in terms of temporal logic and tested by statistical model checking played a central role. Here this idea

[136] referring to the use of modelling and simulation as a means of increasing knowledge about a system, here in particular through multi-faceted modelling

[137] Osman Balci. "Verification, Validation and Accrediation of Simulation Models". In: *Proceedings of the 1997 Winter Simulation Conference*. Ed. by S. Andradöttir et al. IEEE. Piscataway, New Jersey, 1997, pp. 135–141. DOI: `10.1109/IEEESTD.1997.8685803`

[138] Claudia Szabo and Yong Meng Teo. "An Approach for Validation of Semantic Composability in Simulation Models". In: IEEE, 2009, pp. 3–10. ISBN: 978-0-7695-3713-9. DOI: `10.1109/PADS.2009.14`

[139] Danhua Peng et al. "Reusing simulation experiment specifications in developing models by successive composition - a case study of the Wnt/ß-catenin signaling pathway". In: *SIMULATION* 93.8 (2017), pp. 659–677. ISSN: 0037-5497, 1741-3133. DOI: `10.1177/0037549717704314`

has been adopted for validation experiments in general with a focus on successively expanding the experimental frame as the simulation study progresses.

Modelling the EBC has, so far, already resulted in an extensive simulation study spanning a broad range of processes, system levels and facets which has required input from several scientific domains. This has been accomplished by iteratively fusing smaller submodels into large complex ones. During this process, the semantic validity of the fused models was ensured by carefully interlinking composition and validation steps and re-using simulation experiments. In this chapter an early stage of the EBC model, exploring physiology and connected behaviour regarding vertical movement, will be used to elucidate the approach. In this example, the different aspects of respiration, energy budgets and behaviour are fused and validated as a whole after being modelled and validated individually.

Since the goal is to explore a a system that is, due to its size and expanse hard to grasp, by successively modelling and integrating processes and facets of the real-world system loosing validity of processes already 'signed off on' is not an option. The associated workflow then consists of the two tasks composition and validation.

The composition step

There are several different manners in which two or more models can be used to build a single model. Generally, although more detailed characterisations exist [140], black-box and white-box composition can be distinguished. In biology white-box composition prevails as it avoids the "inaccessible variable problem" associated with hiding the internals of black-box composition [141]. Integration, also known as fusion, a manner of white-box composition, combines models into a single unified model without redundancies and, as such, is an irreversible process. Integration appears most suitable to the general approach introduced in Chapter II since there can be no straightforward interfaces between the submodels as they are not suitably aligned (e.g., the same entity might have different attributes) to support black-box composition. At the same time, there would also be no advantage in keeping the models separate, rather, the integration process itself is another point at which ambiguities must be resolved and alignment of mental models, for example, regarding units of measurements or established simplifications, can take place.

The validation step

The validation step is based on different artefacts. These need to be accessible, i.e., the simulation model, the validation experiment, the data used for validation, and data post-processing. In the implementation given here the simulation models are defined in ML-Rules, the validation experiments are specified in the embedded domain-specific language SESSL[142] (Simulation Experiment Specification via a Scala Layer) which is an open source (Apache 2.0 license) project[143] (see

[140] Clifford Shaffer, Ranjit Randhawa, and John Tyson. "The Role of Composition and Aggregation in Modeling Macromolecular Regulatory Networks". In: IEEE, Dec. 2006, pp. 1628–1635. ISBN: 978-1-4244-0501-5 978-1-4244-0500-8. DOI: `10.1109/WSC.2006.322937`

[141] Maxwell L. Neal et al. "A Reappraisal of How to Build Modular, Reusable Models of Biological Systems". en. In: *PLoS Computational Biology* 10.10 (2014). Ed. by Ruth Nussinov, e1003849. ISSN: 1553-7358. DOI: `10.1371/journal.pcbi.1003849`

[142] Adelinde M. Uhrmacher Roland Ewald. "SESSL: A domain-specific language for simulation experiments". In: *ACM Trans Model Comput Simul* 24 (2014), p. 11. DOI: `10.1145/2567895`

[143] The SESSL source repository is available at `http://sessl.org`

Figure 29 for an example). The output is provided as csv files, and post-processing uses R[144] scripts, since this is the most widespread tool for data analysis in the fisheries science community.

Interlinking composition and validation steps

The process of successive composition is schematically depicted in Figure 28. Separately $\mathbf{m_1}$ and $\mathbf{m_2}$ are validated by executing the experiment specifications $\mathbf{Ex_1}$ and $\mathbf{Ex_2}$ respectively by comparing the simulation outputs to wet-lab results. Then $\mathbf{m_1}$ and $\mathbf{m_2}$ are fused into model $\mathbf{m_{1,2}}$ and, if need be, calibrated or otherwise refined. Next the original validation experiments are adjusted for the new model. This entails adapting formal changes such as denotation, the amount of attributes of entities or the position of the observed attributes. Now the adjusted validation experiments $\mathbf{Ex_{a1}}$ and $\mathbf{Ex_{a2}}$ can be applied to $\mathbf{m_{1,2}}$ and again the simulation outputs are compared to the respective wet-lab results. If the integration is deemed successful model $\mathbf{m_3}$, already having been validated, is fused with $\mathbf{m_{1,2}}$ and all three experiments are adapted. Finally the new model $\mathbf{m_{1,2,3}}$ is now validated using $\mathbf{Ex_{b1}}$, $\mathbf{Ex_{b2}}$ and $\mathbf{Ex_{b3}}$.

Application to the Eastern Baltic Cod

As introduced in Chapter I and reiterated in Chapter III there are numerous aspects of EBC ecology which could potentially lead to its current decline. In particular the unique conditions of the Baltic Sea, namely its stable stratification and occurrence of hypoxia and anoxia are often regarded as likely drivers for the observed development. The broad range of hypotheses including factors as diverse as fishing pressure, the decrease of sufficiently oxygenated areas and an increase in infestation levels of liver parasites have all been addressed in this thesis. One aspect which was explored in the early development of the model presented here was the interaction of physiology, stratification and the behaviour which guides vertical migration.

Submodel 1: Respiration

A SESSL experiment using ML-Rules (scala keywords are shown in blue). This simulation experiment runs the metabolism submodel (with the suitable initial solution for this specific simulation) for variations of the parameter t (temperature), o (oxygen saturation) and init (initial state) with 1 replication per combination - replications in the experimental sense are included through the amount of cod in the initial state. It reproduces the wet-lab experiment by PLANTE[145] on hypoxia tolerance.

Aquatic environments are generally oxygen limited [146] and the Baltic, as introduced in Chapter I, is particularly vulnerable to the formation of hypoxic and even anoxic areas due to strong stratification in the deeper basins, their dependence on inflow events and

---

[144] R Core Team. *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing. Vienna, Austria, 1999. URL: `https://www.R-project.org`

[145] S. Plante, D. Chabot, and J.-D. Dutil. "Hypoxia tolerance in Atlantic cod". In: *J Fish Biol* 53 (1998), pp. 1342–1356. DOI: `10.1111/j.1095-8649.1998.tb00253.x`

[146] Daniel Pauly. *Gasping Fish and Panting Squids: Oxygen, Temperature and the Growth of Water-Breathing Animals.* In collab. with Otto Kinne. Excellence in Ecology 22. Oldendorf/Luhe, 2010. ISBN: 978-3-946729-22-8

increasing eutrophication during the past decades [147]. Therefore a sound representation of respiration is needed as a baseline.

For cod the fundamental biological functions either use oxygen or result in a physiological 'oxygen debt'. The wet-lab experiments conducted by Claireaux [148] determined the influence of water temperature and oxygenation on the metabolic scope of cod. This relation was formalised into an ML-Rules model. In addition, the maximum amount of oxygen a cod can contain was based on [149]. The result is a straightforward model of the oxygen budget of cod under different oxygen saturations and temperatures. Model definition must include its limitations or, inversely, its scope of applicability. For this respiration model no limitations for environmental variables are included, meaning that unrealistic oxygen saturation and temperatures are not excluded. Although peek performance of the metabolism can not yet be validated since it has no functionality beyond respiration, mortality due to asphyxiation can be validated.

To validate realistic mortality due to asphyxiation the behaviour of the modelled cod was tested against wet-lab results by Plante[150]. In their experiments cod of different weight classes where kept at two temperatures and six different oxygen saturations for 96 hours to determine cumulative mortality (please note that this experiment has a discrete result, dead or alive, therefore no standard deviations will be given, the variance in results is covered for both wet-lab and simulation experiment by exposing 20 individuals to each respective treatment). This experiment was reproduced with the ML-Rules respiration model using SESSL. The SESSL experiment (see code given above) scans through the temperature an oxygen combinations with the different weight classes of cod. The number of cod not perished at the end of the simulation are observed and results are written in systematically organised csv files.

Since the output of SESSL is systematically organized it can be imported into any data analysis software in a systematic manner. As reasoned before the choice of data analysing software is based on the habits and customs within the community of fisheries scientist. For the asphyxiation experiment, the results from the original wet-lab experiment where extracted and the results of the simulation experiment prepared in the same manner.

The results (Figure 30) reveal a general consistency with the wet-lab results. An obvious difference between the wet-lab and simulation results is that the simulation results are comparatively regular, in particular the impact of size and temperature are not as pronounced. There are several facts which explain the resulting difference. Firstly, there is no experimental variability for the abiotic conditions in the simulation experiment, e.g., pockets of water with higher oxygen content near the surface of experimental tanks or fluctuations in temperature. Secondly, although ML-Rules includes stochasticity this is only applied to the formalised process of respiration but this is not the only factor in variability of the individuals tested in the wet-lab. Differences in efficiency of metabolism, developmental stage and even

---

[147] Daniel J. Conley et al. "Hypoxia-Related Processes in the Baltic Sea". In: *Environmental Science & Technology* 43 (2009), pp. 3412–3420. doi: `10.1021/es802762a`

[148] G. Claireaux et al. "Influence of water temperature and oxygenation on the aerobic metabolic scope of Atlantic cod (*Gadus morhua*". In: *Journal of Sea Research* 44.3 (2000), pp. 257–265. issn: 13851101. doi: `10.1016/S1385-1101(00)00053-8`

[149] Jack D. Burke. "Vertebrate Blood Oxygen Capacity and Body Weight". In: *Nature* 212 (1966), pp. 46–48. doi: `10.1038/212046a0`

[150] S. Plante, D. Chabot, and J.-D. Dutil. "Hypoxia tolerance in Atlantic cod". In: *J Fish Biol* 53 (1998), pp. 1342–1356. doi: `10.1111/j.1095-8649.1998.tb00253.x`

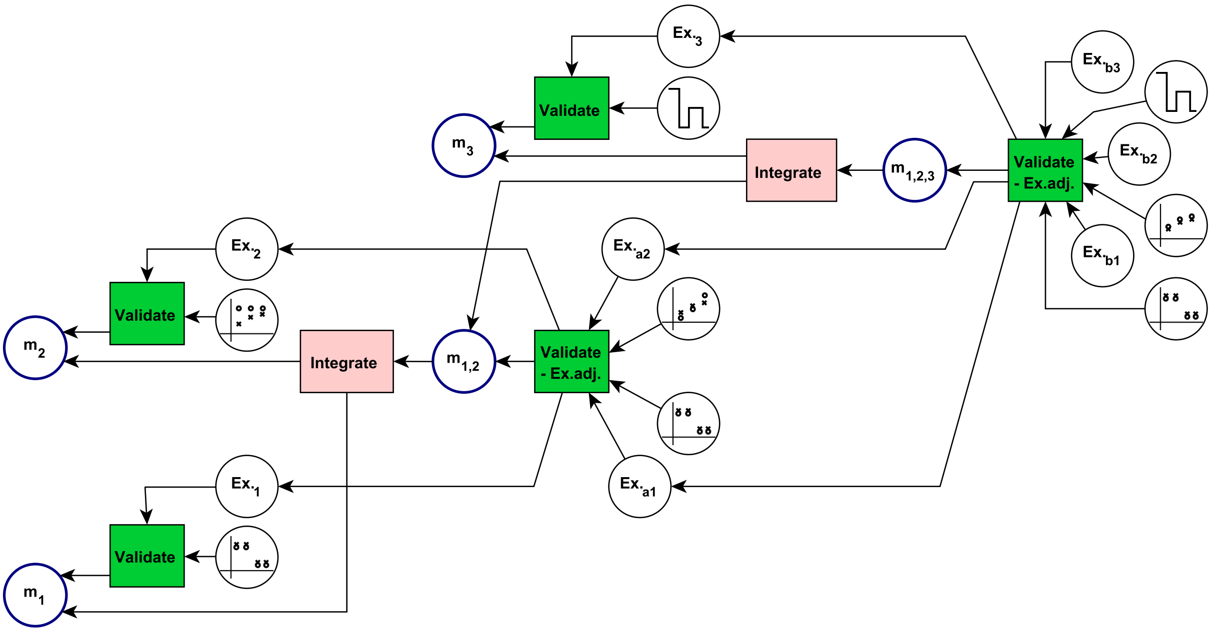


Figure 28: This figure shows the progression of successive integration and validation of simulation experiments in the form of a simplified provenance graph (full introduction to the provenance data model given in Chapter IV). Simulation models denoted by blue border, integration steps denoted by salmon colour fill and validation step by green fill (Ex.adj. stands for experiment adjustment).

Import SESSL core	`import sessl._`
Import ML-Rules binding	`import sessl.mlrules._`
Execute the following experiment	`execute {`
Include observation and parallelization	`new Experiment with Observation with ParallelExecution` `CSVOutput{`
Location and name of model file	`model = "./m1.mlrj"`
Choice of simulator	`simulator = StandardSimulator()`
Use 1 thread	`parallelThreads = -1`
1 replication of each configuration	`replications = 1`
Simulation stops at time 9600	`stopTime = 9600`
Observe at 9600	`observeAt(range(0, 9600, 9600))`
Vary between 4 different initial states	`scan("init" <~ ("20 GM(570)","20 GM(1740)","20 GM(890)","20` `GM(1790)"))`
Vary parameter t between 2 and 6	`scan("t" <~ (2,6))`
Vary parameter o for the given values	`scan("o" <~ (13.8,17.8,23.7,29.5,36.5,42.5))`
Observe attribute o of GM	`observeAttribute("GM", 0)`
Set output directory	`csvOutputDirectory(() => "result_asph_m1")`
Write observed attribute values to file	`withExperimentResult(writeCSV)`
	`})}}`

Figure 29: A SESSL experiment using ML-Rules (Scala keywords are shown in blue). This simulation experiment runs the metabolism submodel (with the suitable initial solution for this specific simulation) for variations of the parameter t (temperature), o (oxygen saturation) and init (initial state) with 1 replication per combination - replications in the experimental sense are included through the amount of (non interacting) cod in the initial state. It reproduces the wet-lab experiment by PLANTE[150] on hypoxia tolerance.

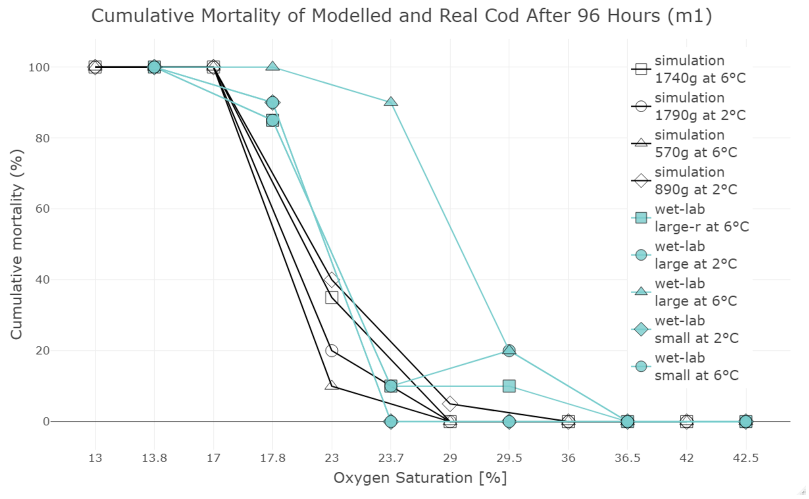


Figure 30: Comparing wet-lab and simulation results for identical asphyxiation experiment by Plante[150] using Model m1. Cod are held at different oxygen saturation for 96 hours. Wet-lab results in blue and simulation results in black.

personality and grouping of animals tested are not present in the modelled cod but are in the experimental animals.

Submodel 2: Energy budget

As introduced in Chapter I, fish stocks in ICES waters are generally managed according to the principle of MSY[151]. Therefore the goal of management is to understand and influence stock dynamics in such a way that it produces the highest possible yield per recruit (cod which have survived to become part of the population), resulting in a large number of individuals in good condition, i.e., healthy and fat. In this approach, typically, the energy budget of individual fish is treated as a balance sheet. 'Consumption' (food intake) is on one side of the equation and the other side represents the sum of 'maintenance' (basal metabolic rate), 'production' (somatic growth and maturation) and 'loss' (digestive, fecal and urinary loss) [152]. However, this is only sufficient when total biomass is of interest. Since information about condition (how plump an individual is) is also required, the model needs to be more advanced. Additionally, if more detailed mechanisms are to be explored, points to address are the speed of digestion and the different percentages of loss for particular types of food will be required.

This submodel is more complex than the respiration model. As introduced in Chapter III it formalises basal metabolic rate, both in regular and in starvation mode, and growth, both in length and in body mass. Building on this, death from starvation, the process of ingestion and stomach evacuation are formalised, which allows to model appetite and different gut transit times for particular types of food or prey. Wherever these processes are understood to be regulated by temperature this has been formalised accordingly.

For this submodel several aspects of model behaviour can be validated but here the focus will be on the validation of growth including condition. The growth of the modelled cod will be compared with the wet-lab results by Chabot[153] who tested the growth of juvenile

[151] E. S. Russell. "Some theoretical Considerations on the "Overfishing" Problem". In: *ICES Journal of Marine Science* 6.1 (1931), pp. 3–20. issn: 1054-3139, 1095-9289. doi: `10.1093/icesjms/6.1.3`

[152] Max Kleiber. *The fire of life: an introduction to animal energetics*. Wiley, 1961

[153] D. Chabot and J.-D. Dutil. "Reduced growth of Atlantic cod in non-lethal hypoxic conditions". In: *Journal of Fish Biology* 55 (1999), pp. 472–491. doi: `10.1111/j.1095-8649.1999.tb00693.x`

cod at different non-lethal oxygen saturations. In this experiment cod were held for 12 weeks being fed three times a week for one hour, this experimental design was formalised and included in the ML-Rules model.

The SESSL experiment for this model calls the ML-Rules model m_2 and lets it complete the twelve weeks of simulation time and reproduces the experiment twenty time, equivalent to the number of cod in the wet-lab experiment.

The results of the simulation experiment (Figure 31) match the wet-lab results only for the higher oxygen concentrations. Still the model can be considered valid within the limitation of not including respiration and only including the stochasticity of ML-Rules of variability.

Fused Model I: Metabolism

The integration of two physiology models is a process which in itself increases, or at least refines, scientific understanding of the system in question, since the exercises increases the amount of "...common language between modellers and experimentalists..." [154]. Meaning that the integration process, as well as the fused model itself can serve as the basis of scientific debate. This has the potential to reduce misunderstandings as the formalised models allow for more rigour in aligning understanding than verbalised mental models do. In this particular example the integration of the two submodels reaffirmed the understanding that the aquatic environment and by extension the physiology of EBC is capped by the availability of oxygen. Bringing together the balance-sheet approach of the energy budget and the wet-lab result based mechanistic approach of the respiration did not require alterations of the semantics of either model and a calibrating factor used in the original model to reduce the rate for the asphyxiation process was no longer needed. The SESSL experiment on asphyxiation was adjusted by changing the executed model to the fused one and extending the attributes of the cod to those it now holds in the model (m_1: weight; m_2: length, weight $\rightarrow$ $m_{1,2}$: length, weight and prey count).

This first check of continuing semantic validity was successful as the results of the asphyxiation experiment on the $m_{1,2}$ model are as satisfactory as the results of the m_1 model (Figure 33). But the results are still very regular which can be interpreted as showing the need to further refine the model. This situation is an example of a fused model serving as the bases for scientific debate in fisheries science. If the difference between reality (wet-lab results) and the modelled entity are lacking which phenomena is missing from the (formalised) conceptual and mental models?

The SESSL experiment to validated growth (Ex_2) was adjusted in the same manner as Ex_1 adjusting the attributes of the modelled cod to the ones it was assigned in the fused model ($m_{1,2}$).

The results of the growth experiment with the fused model show how the integration process has increased the scope of applicability

---

[154] Jonathan Cooper, Jon Olav Vik, and Dagmar Waltemath. "A call for virtual experiments: Accelerating the scientific process". In: *Progress in Biophysics and Molecular Biology* 117 (2015), pp. 99–106. DOI: `10.1016/j.pbiomolbio.2014.10.001`

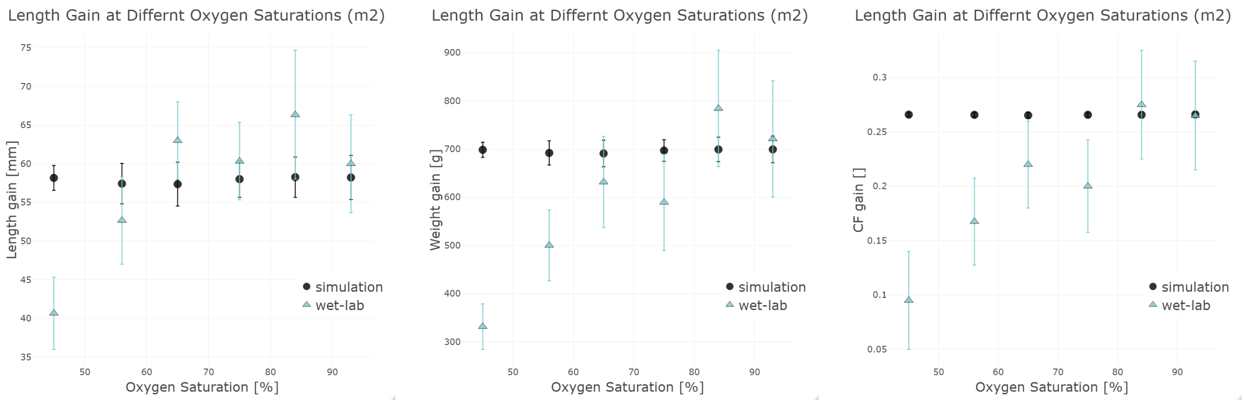


Figure 31: Comparing wet-lab by Chabot[153] and simulation results of the energy budget model for a growth experiment using Model m2. Wet-lab results in blue and simulation results in black, please note that standard deviations condition factor (CF) gain of simulated cod are to small to be visible in the graph.

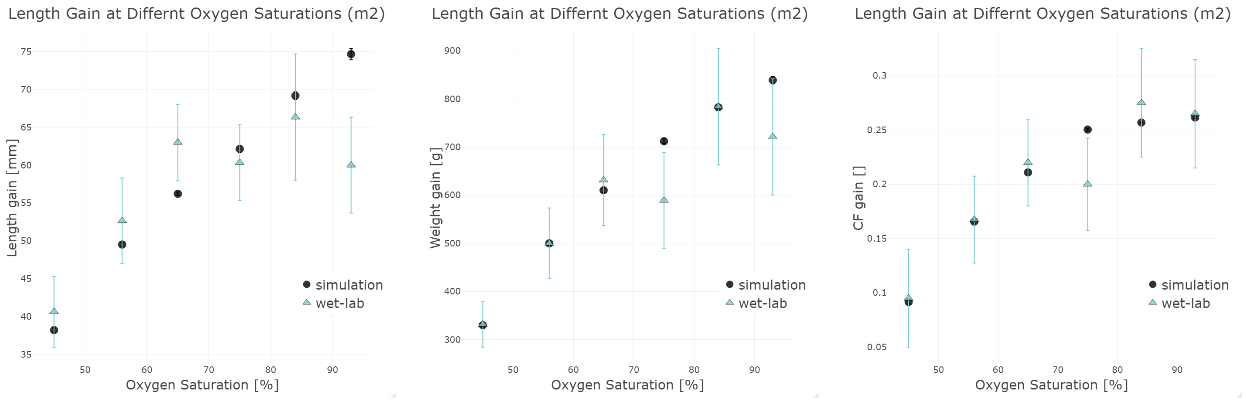
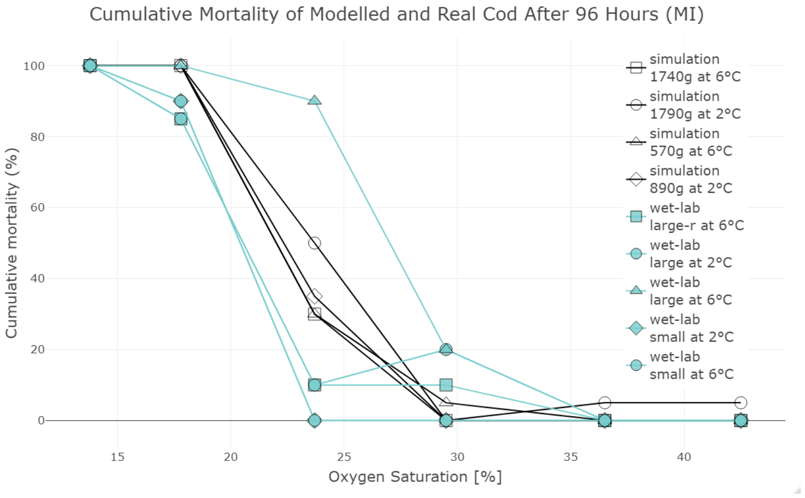


Figure 32: Comparing wet-lab and simulation results for growth experiment by Chabot[153] using Model MI. Wet-lab results in blue and simulation results in black, please note that standard deviations for simulated results are to small to be visible in the graph.

Figure 33: Comparing wet-lab and simulation results for identical asphyxiation experiment by Plante[150] using Model MI. Cod are held at different oxygen saturation for 96 hours. Wet-lab results in blue and simulation results in black.



of the modelled cod. Now the simulation results are in line with the results from the original wet-lab experiment across all oxygen saturations. But as with the asphyxiation experiment the simulated results are unrealistically regular and show the need for further refinement of the model.

Submodel 3: Behaviour

Ethology, the study of animal behaviour, to date, is rarely included into both the conceptual and the formalised models in fisheries science. But virtually all mental models do and an increasing number of current hypotheses find the need to include behaviour to fit observations. One example for the EBC is that this demersal fish used to be caught with ground fishing gear, in recent years however pelagic gear (gear used in the water column) has become the rule. To fully understand all changes therefore behaviour needs to be included in a comprehensive model.

The behaviour submodels governs the vertical movement of the EBC between volumes of water with different abiotic (oxygen saturation and temperature) and biotic (food availability) conditions. It is based on the mental models of field scientists familiar with the Baltic Sea and the EBC. It governs behaviour thus that the cod dwells as near the seabed as possible without asphyxiating or starving while additionally undertaking exploratory upward and downward dives for prey. A complementary development of abiotic conditions is included in the ML-Rules model to test for appropriate reactions. The abiotic structure, which is modelled as four stacked volumes, changes over time: primarily the abiotic conditions of the demersal volume shift from inhabitable to hypoxic back to inhabitable. This should prompt the modelled cod to dwell at the seabed while remaining above the lowest volume during the hypoxic period and undertaking random vertical dives. The SESSL experiment replicates this 23 day run for three individuals ten times.

The results (Figure 34) show the implemented behaviour both in terms of avoiding dangerous conditions as well as undertaking vertical dives for feeding. As there is no independent data to validate against these results are a verification rather than a validation, however this does not impede the workflow presented here.

Fused Model II: Juvenile Cod

Fusing the metabolism model ($m_{1,2}$) with complex behaviour (more than is now contained in model m_3) will ultimately result in the model of a juvenile cod (a model of an adult cod in contrast would require a reproductive cycle). With the example, fusing model $m_{1,2}$ with model m_3 raised a number interesting issues concerning behaviour, for example how behavioural priorities are connected to the condition of the individual. As long as behaviour is seen largely independent of the individual and based mostly on environmental conditions, known phenomena such as collective memory [155] or knowledgeable

[155] G. De Luca et al. “Fishing out collective memory of migratory schools”. In: *Journal of The Royal Society Interface* 11 (2014), pp. 20140043–20140043. DOI: 10.1098/rsif.2014.0043

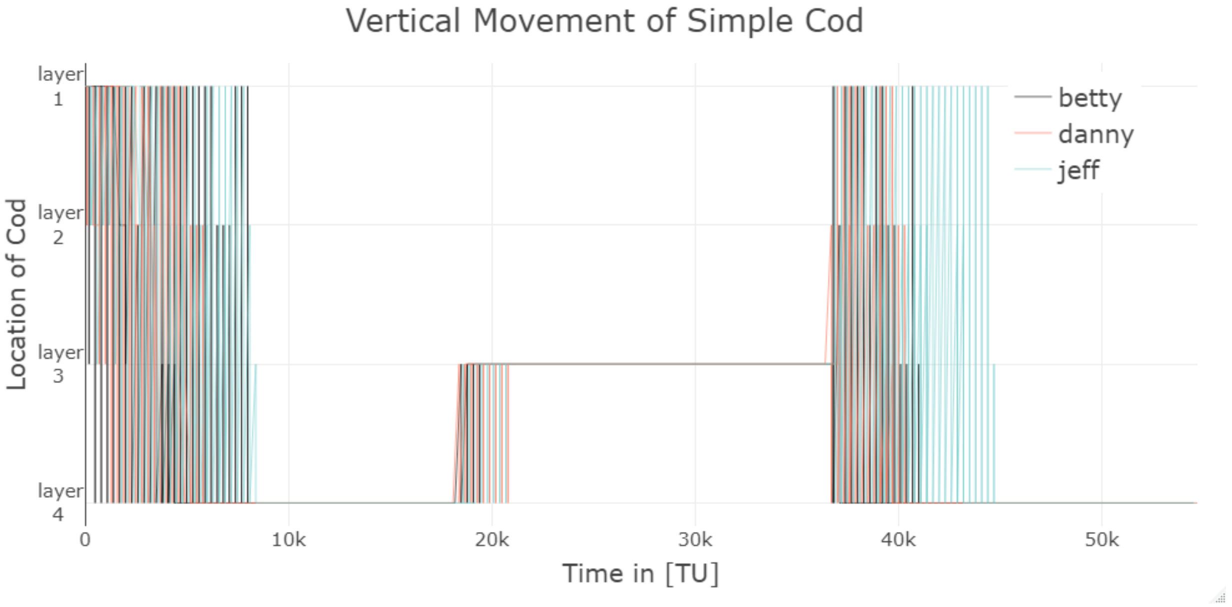


Figure 34: Behaviour in terms of vertical movement of modelled cod for three individuals with ten replications. The simulation represents 23 days of simulation time with model m_3.

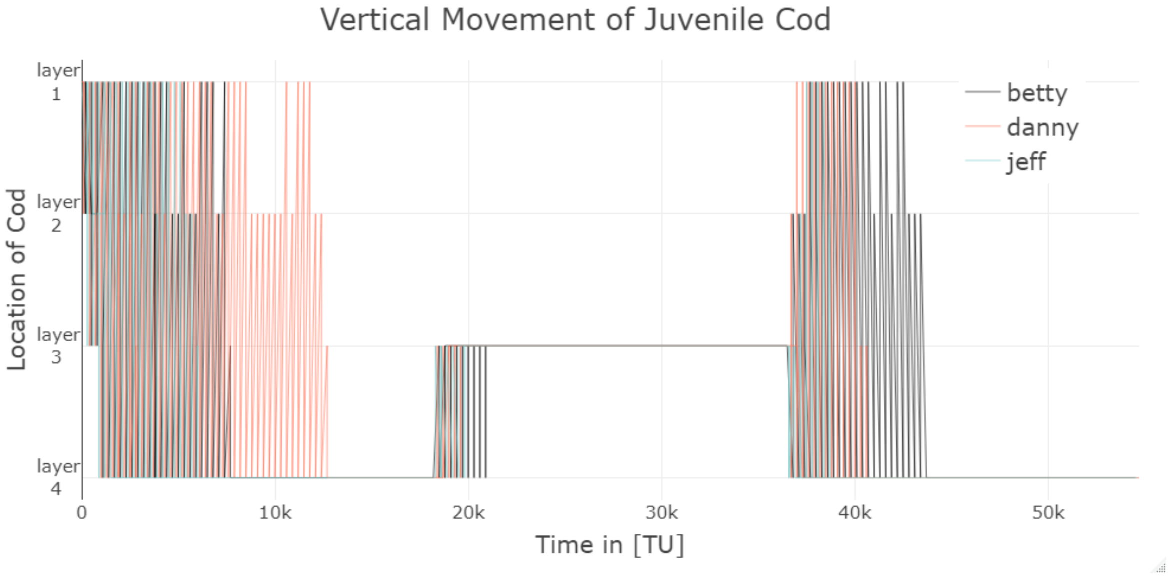


Figure 35: Behaviour in terms of vertical movement of modelled cod for three individuals with ten replications. The simulation represents 23 days of simulation time with model MII.

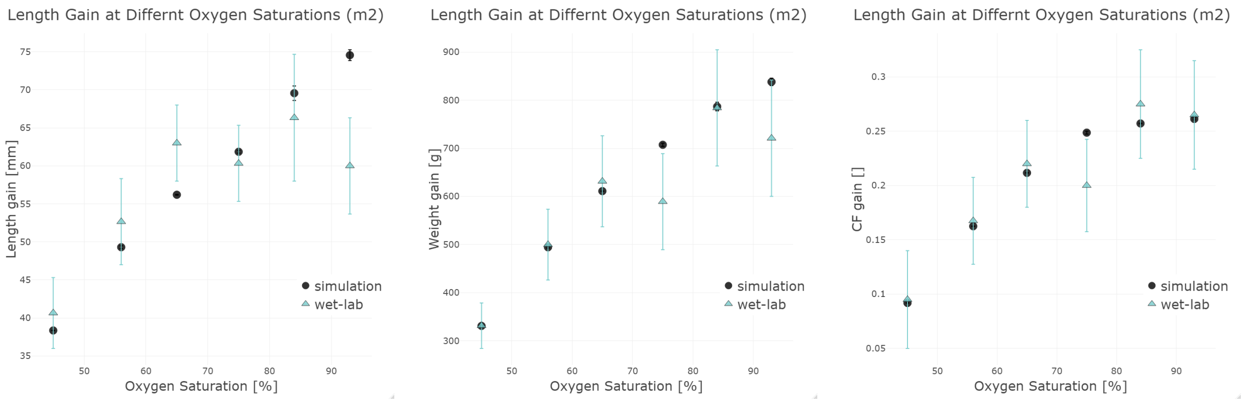
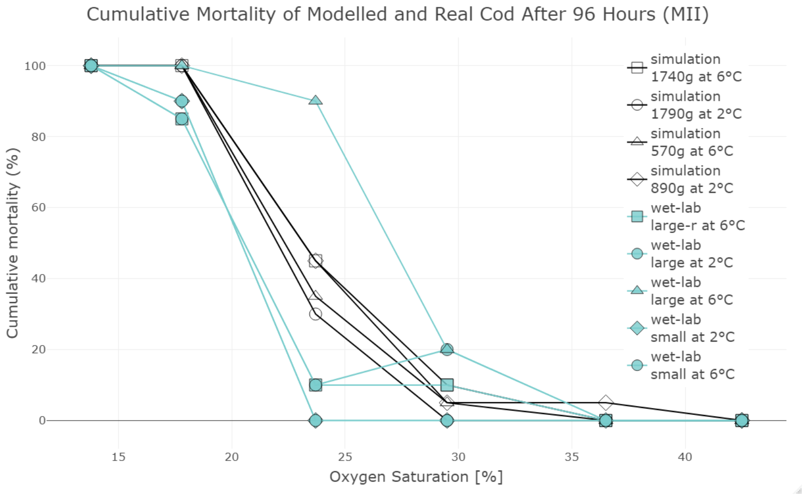


Figure 36: Comparing wet-lab and simulation results for growth experiment by Chabot[153] using Model MII. Wet-lab results in blue and simulation results in black.

Figure 37: Comparing wet-lab and simulation results for identical asphyxiation experiment by Plante[150] using Model MII. Cod are held at different oxygen saturation for 96 hours. Wet-lab results in blue and simulation results in black.



mature individuals can not be included in a comprehensive model. Another issue raised is the intricate and regulatory interdependence of energy available to the organism and energy expenditure necessary for movement i.e., behaviour and both issues raised have been a valuable input to the further development of the model.

The SESSL experiment for asphyxiation Ex_{b1} was further adjusted to account for the increase in attributes of the cod ($m_{1,2}$: length, weight and prey count → $m_{1,2,3}$: length, weight, prey count, name, cube of residence) both for the initial state and the observations. The results of the experiment (Figure 37) remain satisfactory with the same shortcomings (i.e., high regularity and decreased impact of size and temperature) as with model m_1 and $m_{1,2}$.

Adjustments of the growth experiment ($\mathrm{Ex.}_{b2}$) and the behaviour experiment ($\mathrm{Ex.}_3$) were parallel to the adjustment of the asphyxiation experiment. The results of the growth experiment (Figure 36) are not further improved from the results of the same growth experiment with model $m_{1,2}$ however, there is also no loss of semantic validity. When the model is further refined to meet the issue raised during integration of $m_{1,2}$) and m_3 and behaviour is linked with energy expenditure a loss of regularity and therefore a more realistic modelled cod can be expected. Results of the behaviour experiment (Figure 35) also show no decrease in semantic validity compared to the results of $\mathrm{Ex.}_3$.

Discussion

Carefully interlinking composition and validation steps revealed how the established experimental frame (the set of experiments a given model is valid for) and corresponding validity could be maintained by successive composition. This demonstrated the benefit of making small steps in extending and composing a model, as the first fused model can realistically depict that the growth of cod is capped by the availability of oxygen, making superfluous the calibration factors introduced in the first submodel which required a reduced rate for asphyxiation. Although the integration of metabolism and behaviour did not further refine or broaden the scope of the model, the mindfulness of the process raised important and clear issues in understanding the link between physical condition and behaviour of cod. One example is the need to integrate the oxygen status of individual cod into the rules initiating vertical dives, thereby also linking behaviour to physiological (internal) rather than only abiotic (external) factors. This might seem obvious but has implications as the development of the model goes further, new processes and connections implemented in the model will, if they impact the oxygen status because of the now established link, also impact behaviour. In this manner, the high degree of interconnectedness of processes, reflecting the properties of the real-world system is iteratively established during the simulation study.

More generally, the approach further facilitates the discourse between experts of different sub-fields by the realised 'divide and conquer' strategy of the general approach. For each integration and validation step a focus is placed on the interconnectedness of a small number of processes an whether they are modelled to the satisfaction of those most knowledgable about them. Furthermore, a basis for the discourse between modellers and experimentalists is provided by associating validation experiments with wet-lab experiments and expert knowledge, again placing a focus on a selected scope of processes.

To keep this approach organised and thereby more easy to handle, previous work on the systematic reuse of simulation experiments[139] was adapted to the purpose of maintaining validity alongside successively expanding models. In doing so, the workflow of integrating models, slightly adapting the simulation experiments which have become part of validating the simulation models and determine the experimental frame up to that point (primarily by expanding the number of attributes and sometimes entities), using these experiments to conduct validation experiments, and evaluating the results was developed. The efficiency of this workflow was supported by using the DSL SESSL for experiment specifications, as it is tailored to define experiments, adapting them is a straightforward task that only requires altering specific lines of code. Equally, the R scripts used for evaluating the results could be reused with the same type of minor adjustments. Overall this workflow then lent itself to maintaining, and implicitly documenting, a thorough overview of the experimental frame of an increasingly complex model.

Chapter V
Documentation of Simulation Studies

Ink is better than the best memory.

Chinese proverb

Documentation

Documentation of simulation models, in general, provides descriptions, explanations and/or instructions on the parts, setup, maintenance and use of the model code. Documentation of modelling and simulation studies gives additional information about experiments, iterations and reproducibility.

As introduced in Chapter II, documentation is part of the supporting infrastructure of modelling and simulation studies. Nonetheless, it is sometimes omitted because it requires additional work and is not, technically, necessary to produce modelling and simulation results. However, those modelling and simulation studies without proper documentation become less accessible and more challenging to reproduce and scrutinize, especially if they are more expansive. Since management requires trust, which requires scrutiny, and the general approach[156] introduced above requires accessibility, documentation also becomes an indispensable requirement for modelling and simulation of the EBC.

Types of model documentation

Simulation studies can be documented in several different manners, these include 'unique' natural language descriptions (i.e., the structure of the documentation is unique to the specific model), structured natural language descriptions (i.e., the documentation follows a structure which has been developed and is being reused) and standardised descriptions in a knowledge representation language. This chapter will focus on structured and standardised documentation.

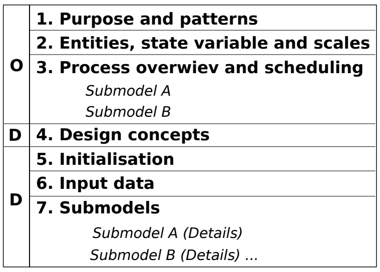
Thorough and structured documentation has several advantages: Firstly, it provides ease of use and reduces time and effort for those wishing to work with the model. Secondly, it provides independent access, introducing interested parties to the model on their own schedule and without the need for assistance by one (of possibly very few) of the developers. Thirdly, it can thereby facilitate further, independent, development. A fourth aspect is that the process of documentation itself can provide internal feedback during development. Finally, (well structured) documentation subjects the model to a high level of scrutiny, narrowing the results' application while simultaneously increasing credibility.

[156] referring to the use of modelling and simulation as a means of increasing knowledge about a system, here in particular through multi-faceted modelling

Structured Documentation of IBMs - the ODD Protocol

While continuous models, such as ODE models, generally have the same or similar structures for specifications, individual-based models exhibit a broad range of types of implementations. This makes IBMs far less accessible and poses a particular challenge to the IBM community. The original ODD protocol (Overview, Design concepts and Details)[157] was explicitly developed for the documentation of agent- and individual-based models to address this issue, as there was no standard protocol available at the time. Providing such a standard was aimed at supporting more rigorous documentation and, by organising its elements, making IBM models more accessible. It has since then been frequently used for the documentation of agent- and individual-based models and was subsequently updated for clarification, replacing the original protocol. At the same time critique about redundancies, and expressiveness when used for simple models and sub-optimal grouping was addressed[158]. A second update revised the protocol in detail to accommodate highly complex models and information such as the underlying narrative, again replacing the previous protocol[159].

The current (2020) iteration of the ODD protocol intends for seven sections (see Figure 38). The first section provides information on the purpose of the model, which then also helps define the scope of applicability. The second section gives information on entities, state variables, and scales, thereby providing the information to assess the respective counterparts in the underlying mental model. The third section does the same for the modelled processes. Together these three aspects provide an overview of the documented model. The fourth section (which is equivalent to the 'Design' part of ODD) documents the design concepts. To this end, it is meant to contain information on basic principles, emergence, adaptation, objectives, learning, prediction, sensing, interaction, stochasticity, collectives, and observation. The fifth, sixths and seventh sections make up the part on the details of the model and provide information on initialisation, input data, and submodels, respectively which are required for reproducibility and vital for continued use.



O	**1. Purpose and patterns**
	2. Entities, state variable and scales
	3. Process overwiev and scheduling *Submodel A* *Submodel B*
D	**4. Design concepts**
D	**5. Initialisation**
	6. Input data
	7. Submodels *Submodel A (Details)* *Submodel B (Details) ...*

Figure 38: Overview of the subsections of the current iteration of the ODD protocol recreated from[159].

Structured Documentation of Ecological Modelling and Simulation Studies – TRACE

Documentation of simulation studies can be classified by the intended purposes of the study and the intended audience. For example, in modelling of biological systems, the COMBINE ("COmputational Modeling in BIology NEtwork")[160] network aims to coordinate standards for modelling with the goal of scientific advancement in this particular field. It has established the MIRIAM ("Minimal Information Required In the Annotation of Models") guidelines to support consistent annotation of models [161] and the MIASE ("Minimum Information About a Simulation Experiment") [162] guidelines to support

[157] Volker Grimm et al. "A standard protocol for describing individual-based and agent-based models". In: *Ecol Model* 198 (2006), pp. 115–126. ISSN: 03043800. DOI: `10.1016/j.ecolmodel.2006.04.023`

[158] Volker Grimm et al. "The ODD protocol: a review and first update". In: *Ecol Model* 221 (2010), pp. 2760–2768. ISSN: 03043800. DOI: `10.1016/j.ecolmodel.2010.08.019`

[159] Volker Grimm et al. "The ODD Protocol for Describing Agent-Based and Other Simulation Models: A Second Update to Improve Clarity, Replication, and Structural Realism". In: *Journal of Artificial Societies and Social Simulation* 23.2 (2020), p. 7. ISSN: 1460-7425. DOI: `10.18564/jasss.4259`

[160] http://co.mbine.org/

[161] https://co.mbine.org/standards/miriam

[162] https://co.mbine.org/standards/miase

Table 4: Overview of the eight TRACE elements verbatim from Grimm et al.[167]

TRACE element	This TRACE element provides supporting information on:
1. Problem formulation	The decision-making context in which the model will be used; the types of model clients or stakeholders addressed; a precise specification of the question(s) that should be answered with the model, including a specification of necessary model outputs; and a statement of the domain of applicability of the model, including the extent of acceptable extrapolations.
2. Model description	The model, i.e. a detailed written model description. For individual/agent-based and other simulation models, the ODD protocol is recommended as standard format. For complex submodels, include concise explanations of the underlying rationale. Model users should learn what the model is, how it works, and what guided its design.
3. Data evaluation	The quality and sources of numerical and qualitative data used to parameterize the model, both directly and inversely via calibration, and of the observed patterns that were used to design the overall model structure. This critical evaluation will allow model users to assess the scope and the uncertainty of the data and knowledge on which the model is based.
4. Conceptual model evaluation	The simplifying assumptions underlying a model's design, both with regard to empirical knowledge and general, basic principles. This critical evaluation allows model users to understand that model design was not ad hoc but based on carefully scrutinized considerations.
5. Implementation verification	(1) Whether the computer code implementing the model has been thoroughly tested for programming errors, (2) whether the implemented model performs as indicated by the model description, and (3) how the software has been designed and documented to provide necessary usability tools (interfaces, automation of experiments, etc.) and to facilitate future installation, modification, and maintenance.
6. Model output verification	(1) How well model output matches observations and (2) how much calibration and effects of environmental drivers were involved in obtaining good fits of model output and data.
7. Model analysis	(1) How sensitive model output is to changes in model parameters (sensitivity analysis), and (2) how well the emergence of model output has been understood.
8. Model output corroboration	How model predictions compare to independent data and patterns that were not used, and preferably not even known, while the model was developed, parametrized, and verified. By documenting model output corroboration, model users learn about evidence which, in addition to model output verification, indicates that the model is structurally realistic so that its predictions can be trusted to some degree.

reproducibility of simulation experiments. These guidelines naturally reflect the objectives and common practices by supporting the reuse of models, requiring clarity on properties such as biological processes and information required to retrace post-processing implementation to acquire the final results. When these guidelines are adhered to, they support a thorough and efficient exchange between scientists investigating biological systems employing modelling and simulation.

Desired properties for the documentation of simulation studies that are aimed at management support are discussed as early as the 1970s, addressing the issues of providing insight, not only to for those working with the model but for those relying on the merit of its results [163]. Documentation of simulation studies to support the management of ecological systems specifically allows framing standards even more precisely. Such standards have been evaluated with regards to the evaluation of models from a policy perspective [164] and from a modelling perspective [165].

The TRACE protocol [166] ('TRansparent And Comprehensive model Evaluation') was developed as a tool (and in another attempt) to establish good modelling practice for ecological simulation studies aimed at decision support. The original developers found "current modelling practices unsatisfactory", finding that "the elements of good modeling practice have long been identified but are widely ignored." They speculate that reasons for this "might include lack of involvement from decision makers, lack of incentives for modellers to follow good practice and the use of inconsistent terminologies."[166]. Since its introduction it has been increasingly applied and accepted and based on its use cases it has been subsequently updated to more directly address model quality and credibility[167] and clarify nomenclature[168].

The TRACE protocol, applied here, contains information about: what the model is for, how it works, the type and quality of input data, the underlying concepts and hypothesis, assessment of implementation, assessment of output, analysis of sensitivity and behaviour and output corroboration, these are described in detail in Table 4. This information is structured in such a manner that, depending on the focus, different parties (e.g. project coordinators, decision-makers, or scientists both those familiar and those from different sub-domains) can access information about the model relevant to them without the need to read through the entire documentation.

Not only are different sections relevant to different parties (e.g., the section on problem formulation includes information on acceptable extrapolations, which is essential to decision-makers, but the section on model description, in most cases, is not important to decision-makers but highly important to scientists wishing to expand the model) but also to the same parties at different points in time (e.g., when a scientist wants to decide if a model might be worth their time the section on the conceptual model might be of interest if they decide they wish to expand the model they could use the subsection of the model analysis on sensitivity to gauge which processes they might

---

[163] Belva J. Cooley. "Documenting simulation studies for management". In: *ACM SIGSIM Simulation Digest* 10.3 (1979), pp. 24–27. ISSN: 0163-6103. DOI: `10.1145/1102802.1102804`

[164] Stephen P Prisley and Michael J Mortimer. "A synthesis of literature on evaluation of models for policy applications, with implications for forest carbon accounting". In: *Forest Ecology and Management* 198.1 (2004), pp. 89–103. DOI: `10.1016/j.foreco.2004.03.038`

[165] N. Crout et al. "Good Modelling Practice". In: *Developments in Integrated Environmental Assessment*. Vol. 3. Elsevier, 2008, pp. 15–31. ISBN: 978-0-08-056886-7. DOI: `10.1016/S1574-101X(08)00602-9`

[166] Amelie Schmolke et al. "Ecological models supporting environmental decision making: a strategy for the future". In: *Trends in Ecology & Evolution* 25.8 (2010), pp. 479–486. ISSN: 01695347. DOI: `10.1016/j.tree.2010.05.001`

[167] Volker Grimm et al. "Towards better modelling and decision support: Documenting model development, testing, and analysis using TRACE". en. In: *Ecological Modelling* 280 (2014), pp. 129–139. ISSN: 03043800. DOI: `10.1016/j.ecolmodel.2014.01.018`

[168] Jacqueline Augusiak, Paul J. Van den Brink, and Volker Grimm. "Merging validation and evaluation of ecological models to 'evaludation': A review of terminology and a practical approach". In: *Ecological Modelling* 280 (2014), pp. 117–128. ISSN: 03043800. DOI: `10.1016/j.ecolmodel.2013.11.009`

consider to be missing).

The TRACE protocol thereby, through the merit of its structure, provides comprehensive information about modelling and simulation studies while maintaining accessibility for the diverse audience of interested parties involved in bringing about science-based policies that affect ecosystems.

Documentation of the EBC Study

The Modelling and Simulation study conducted in the course of the work presented here is currently being documented using the TRACE and, as it is an IBM, the ODD protocol for the model description. The current iteration of the document can be found at `https://github.com/Baltic-Cod/EBC_IBM`.

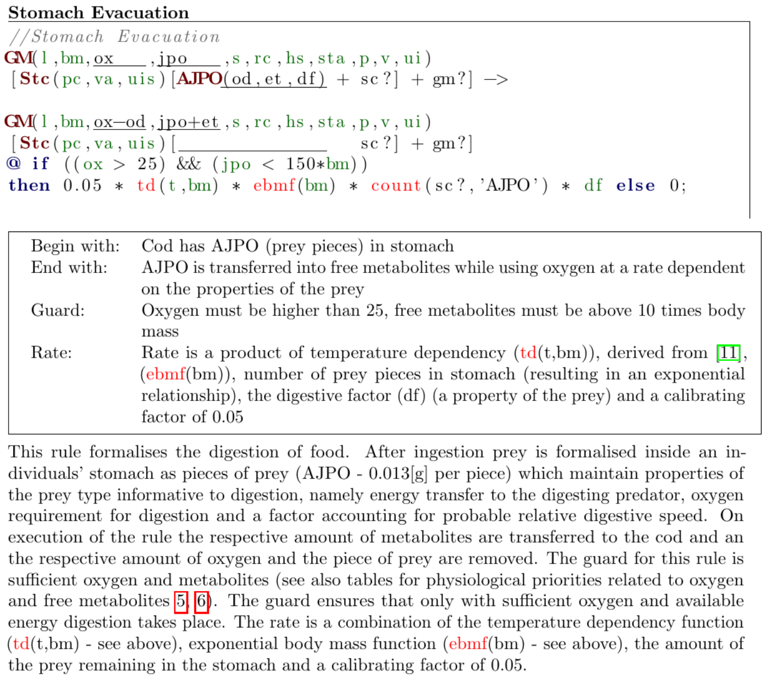
Here some examples from the sections on 2.) Model description, 6.) Model output verification and 8.) Model output corroboration will be provided.

Model Description

Since the EBC models where developed using a formal language with a defined semantic the model code is technically unambiguous in its meaning. However, it does not include additional information about aspects such decisions or input data and it can be cumbersome to read in the case of more complicated functions. Taking advantage of the rule-based syntax providing one 'unit' of code per process the model was described rule by rule. This not only provides easy access for scrutinisation but also keeps model and documentation iterations parallel.

To further support accessibility, documentation of rules is given in three parts: First, the full model code in ML-Rules syntax is specified. Next an info-box provides a concise overview over the 'before', 'after', guard and rate including, where applicable, references. At the end a long-form natural language description of the respective rules is provided including, again where applicable, already known further development requirements (e.g., currently modelled cod can only ingest prey, the process of regurgitation under stress is missing), see Figure 39 for an example.

The screenshot shows the documentation of the rule which formalises the process of stomach evacuation, this process has to utilise information about the prey pieces (AJPO) relevant to digestion. The 'oxygen demand' (od) and 'energy transfer' (et) attributes hold information on how much oxygen will be required and how much energy will be transferred when this particular type of prey is digested (a piece of fatty fish will require less oxygen and transfer more energy than a small crab consisting mostly of carapace will). Note that this rule currently ignores the *quality* of the food in terms of composition and micronutrients. The 'digestive factor' (df) attribute holds information on how quickly the specific type of prey can be digested this



Stomach Evacuation

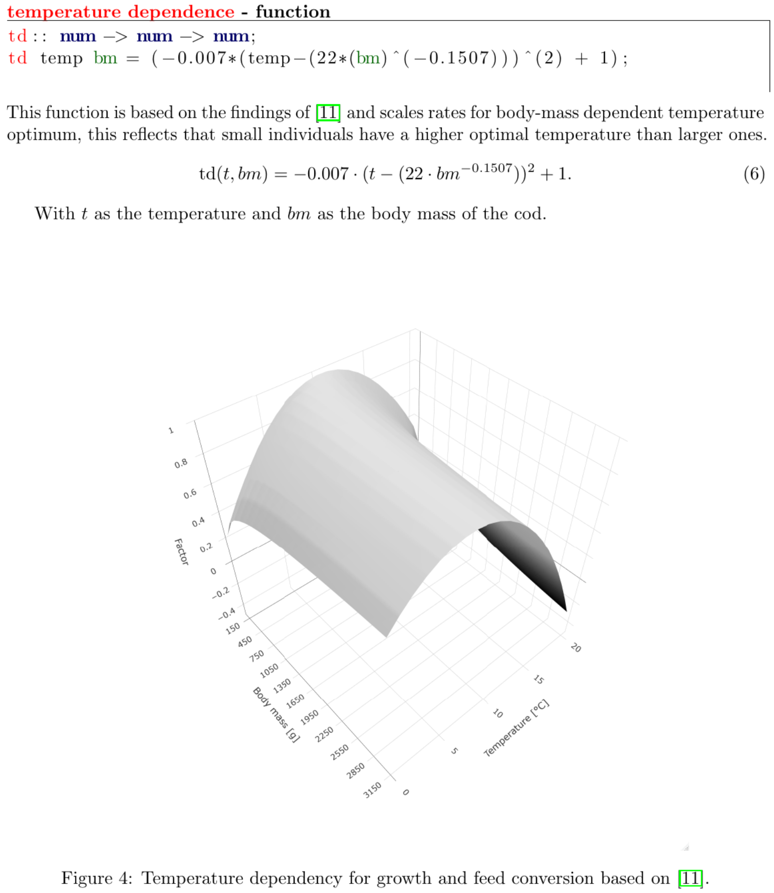
```
//Stomach  Evacuation
GM(l,bm,ox    ,jpo    ,s,rc,hs,sta,p,v,ui)
 [Stc(pc,va,uis)[AJPO(od,et,df) + sc?] + gm?] ->

GM(l,bm,ox-od,jpo+et,s,rc,hs,sta,p,v,ui)
 [Stc(pc,va,uis)[               sc?] + gm?]
@ if ((ox > 25) && (jpo < 150*bm))
then 0.05 * td(t,bm) * ebmf(bm) * count(sc?,'AJPO') * df else 0;
```

Begin with:	Cod has AJPO (prey pieces) in stomach
End with:	AJPO is transferred into free metabolites while using oxygen at a rate dependent on the properties of the prey
Guard:	Oxygen must be higher than 25, free metabolites must be above 10 times body mass
Rate:	Rate is a product of temperature dependency (td(t,bm)), derived from [11], (ebmf(bm)), number of prey pieces in stomach (resulting in an exponential relationship), the digestive factor (df) (a property of the prey) and a calibrating factor of 0.05

This rule formalises the digestion of food. After ingestion prey is formalised inside an individuals' stomach as pieces of prey (AJPO - 0.013[g] per piece) which maintain properties of the prey type informative to digestion, namely energy transfer to the digesting predator, oxygen requirement for digestion and a factor accounting for probable relative digestive speed. On execution of the rule the respective amount of metabolites are transferred to the cod and an the respective amount of oxygen and the piece of prey are removed. The guard for this rule is sufficient oxygen and metabolites (see also tables for physiological priorities related to oxygen and free metabolites 5, 6). The guard ensures that only with sufficient oxygen and available energy digestion takes place. The rate is a combination of the temperature dependency function (td(t,bm) - see above), exponential body mass function (ebmf(bm) - see above), the amount of the prey remaining in the stomach and a calibrating factor of 0.05.

Figure 39: Screenshot of the documentation of the stomach evacuation rule. Description is given in three parts:
1.) the rule in ML-Rules syntax
2.) an info-box with 'before', 'after', guard and rate
3.) a long-form natural language description



temperature dependence - function

```
td:: num -> num -> num;
td temp bm = (-0.007*(temp-(22*(bm)^(-0.1507)))^(2) + 1);
```

This function is based on the findings of [11] and scales rates for body-mass dependent temperature optimum, this reflects that small individuals have a higher optimal temperature than larger ones.

$$\mathrm{td}(t, bm) = -0.007 \cdot (t - (22 \cdot bm^{-0.1507}))^2 + 1. \tag{6}$$

With t as the temperature and bm as the body mass of the cod.

Figure 4: Temperature dependency for growth and feed conversion based on [11].

Figure 40: Screenshot of the documentation of the temperature dependence function. Description is given in three or four parts:
1.) the rule in ML-Rules syntax
2.) a natural language description of the function
3.) a mathematical description of the function
4.) where deemed useful, a figure visualising the function in the range of parameter values used in the model

is factored into the rate of the rule. The rate of the rule also takes into account the current fullness of the stomach (`count(sc?,'AJPO')`), the body mass of the cod (`ebmf(bm)`) and the relationship between body mass and optimal temperature for cod (`td(t,bm)`). The rule can only be executed if the cod has sufficient oxygen (`ox>25`) and if digestion does not 'overwhelm' its system (`jpo>150*bm`), which is a proxy for likely several regulatory mechanisms currently not included in the model.

This three-part structure, i.e., giving the equivalent information in three different forms, allows readers to choose which representation is most accessible or most useful to them at the time. This includes preferences due to training (readers not trained in coding might prefer the natural language description) or their current intentions (readers currently specifying an experiment might prefer the conciseness of the info-box). Starting at the representation most comfortable to them in that instance, they can subsequently be mapped onto each other.

The functions used in the rules are described using a similar approach. Here the info-box is replaced by a mathematical representation of the function, as the ML-Rules code of more complicated maths can be cumbersome to read. Where applicable, the information is further expanded, by a visual representation of the relationships formalised by the function, again providing the information in different forms to support accessibility, see Figure 40 for an example.

The screenshot shows the documentation of the function formalising the relationship between body mass and optimal temperature for cod (`td(t,bm)`) is documented. This relationship was based on wet-lab findings[169] and is used at several points in the model. The visual representation gives intuitive access to domain experts knowledgeable about the performance of cod in the field and exhibits the known relationship of lower optimal temperature for larger individuals.

Model Output Verification

The general approach of the study presented here aims to iteratively add and integrate facets of EBC ecology to the model to understand what is driving its current decline. As a result, early iterations lack the completeness of processes required to reproduce field observations. Nonetheless, experiments reproducing time 'at liberty', which ultimately will serve as experiments to show how well output matches observations, are conducted from an early stage to evaluate the remaining gap between the behaviour of the model and the real-world system.

Experiments are provided both as code (since the experiments specification language SESSL was used, in Scala syntax) together with an info-box relaying relevant information about the model used, the duration of the experiment, the variable which was scanned, the variables which were monitored, the interval at which the simulation is monitored and the number of replications. Finally, a natural language description is provided (this is the same structure as with validation experiments, for an example, see Figure 42)

---

[169] Björn Björnsson, Agnar Steinarsson, and Matthias Oddgeirsson. "Optimal temperature for growth and feed conversion of immature cod (*Gadus morhua* L.". In: *ICES Journal of Marine Science* 58 (2001), pp. 29–38. DOI: `10.1006/jmsc.2000.0986`

Results of simulation experiments are visualised in two manners: changes in attributes of interest (e.g., length, weight, and condition factor) over time (see Figure 41) and animations. The figure shows the results of a one-year (modelled time) at liberty growth experiment. Fifty modelled adult cod (25 female and 25 male) weighing 715g where 'set free' in the modelled Bornholm Basin environment. The cod are equipped with behaviour including reproduction and movement through the environment, and the environment has forced abiotic conditions based on survey data and is outfitted with different types of prey of varying distribution and relative abundance based on stomach sampling of cod. If the model had already contained all relevant processes, the cod should have exhibited an average weight increase of 250g. In contrast, the simulation experiment showed that their weight increase is about double, showing that factors are missing from the model.

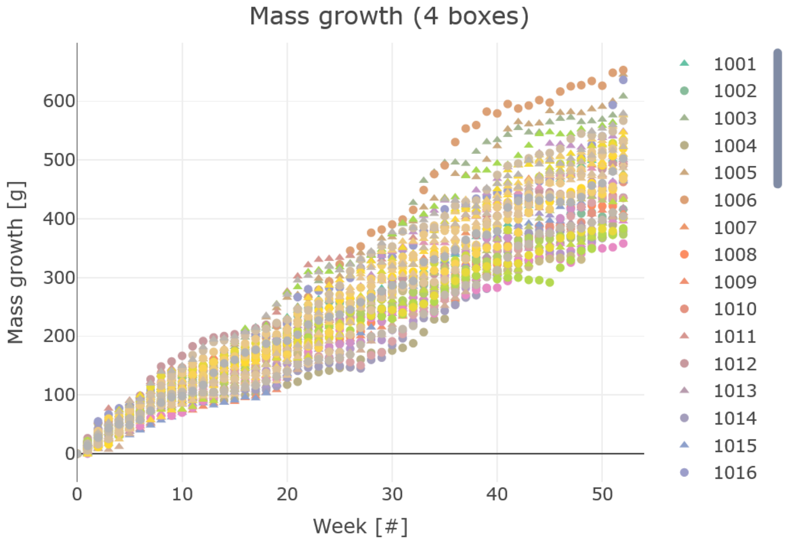


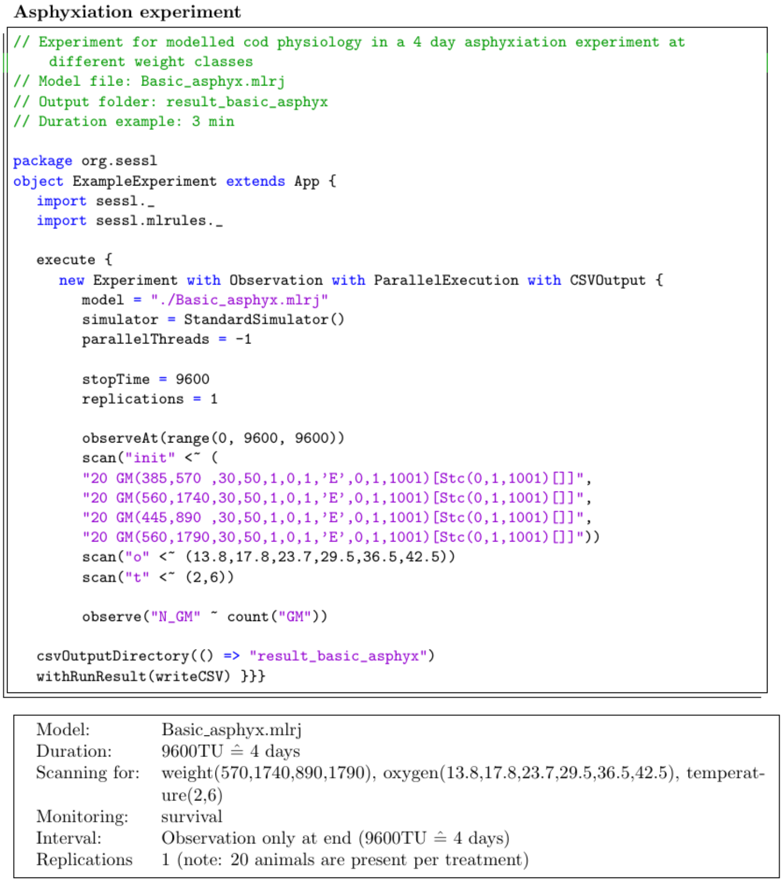
Figure 41: Visualisation of mass growth during at liberty simulation experiments of adult cod prior to the addition of parasitation to the model. 25 female (triangles) and 25 male (circles) individual cod all produced gonads and spawned. Cod weigh 715[g] at beginning of experiment, mass increase would be expected to be approximately 250[g] for real EBC.

Model Output Corroboration

Based on the approach of reuse of simulation experiments to ensure continued validity outlined in Chapter IV, regular validation experiments were conducted with different iterations of the EBC models.

Here too, experiments are provided both as code, since the experiments specification language SESSL was used, in Scala syntax, together with an info-box relaying relevant information about the model used, the duration of the experiment, the variable which was scanned, the variables which were monitored, the interval at which the simulation is monitored and the number of replications, see Figure 42. The example shows a validation experiment for asphyxiation. Reproducing a wet-lab experiment[170] cod of different weight classes are exposed to two different temperatures and six different oxygen

[170] S. Plante, D. Chabot, and J.-D. Dutil. "Hypoxia tolerance in Atlantic cod". In: *J Fish Biol* 53 (1998), pp. 1342–1356. DOI: `10.1111/j.1095-8649.1998.tb00253.x`



Asphyxiation experiment

```
// Experiment for modelled cod physiology in a 4 day asphyxiation experiment at
   different weight classes
// Model file: Basic_asphyx.mlrj
// Output folder: result_basic_asphyx
// Duration example: 3 min

package org.sessl
object ExampleExperiment extends App {
  import sessl._
  import sessl.mlrules._

  execute {
    new Experiment with Observation with ParallelExecution with CSVOutput {
      model = "./Basic_asphyx.mlrj"
      simulator = StandardSimulator()
      parallelThreads = -1

      stopTime = 9600
      replications = 1

      observeAt(range(0, 9600, 9600))
      scan("init" <~ (
      "20 GM(385,570 ,30,50,1,0,1,'E',0,1,1001)[Stc(0,1,1001)[]]",
      "20 GM(560,1740,30,50,1,0,1,'E',0,1,1001)[Stc(0,1,1001)[]]",
      "20 GM(445,890 ,30,50,1,0,1,'E',0,1,1001)[Stc(0,1,1001)[]]",
      "20 GM(560,1790,30,50,1,0,1,'E',0,1,1001)[Stc(0,1,1001)[]]"))
      scan("o" <~ (13.8,17.8,23.7,29.5,36.5,42.5))
      scan("t" <~ (2,6))

      observe("N_GM" ~ count("GM"))

  csvOutputDirectory(() => "result_basic_asphyx")
  withRunResult(writeCSV) }}}
```

Model:	Basic_asphyx.mlrj
Duration:	9600TU ≙ 4 days
Scanning for:	weight(570,1740,890,1790), oxygen(13.8,17.8,23.7,29.5,36.5,42.5), temperature(2,6)
Monitoring:	survival
Interval:	Observation only at end (9600TU ≙ 4 days)
Replications	1 (note: 20 animals are present per treatment)

Figure 42: Screenshot of the documentation for the 12 week growth experiment. Description is given in two parts:
1.) the code SESSL (Scala) syntax
2.) a info-box about the model used, the duration of the experiment, the variable which were scanned, the variables which were monitored, the interval at which the simulation is monitored and the number of replications

To validate the physiological response to persistent low levels of oxygen the wet-lab findings of [13] where reproduced in a 4 day asphyxiation experiment. Both real and simulated cod of different weight classes where kept at different temperatures and oxygen saturation in groups of twenty animals and the the resulting cumulative mortality was calculated. the simulated response of the updated model was, again, more regular than wet-lab observations but showed the same tipping point from 100% to 0% mortality (20 to 30 % oxygen saturation). When not aiming to gain insight on this exact threshold the model can therefore be deemed valid concerning death from asphyxiation.

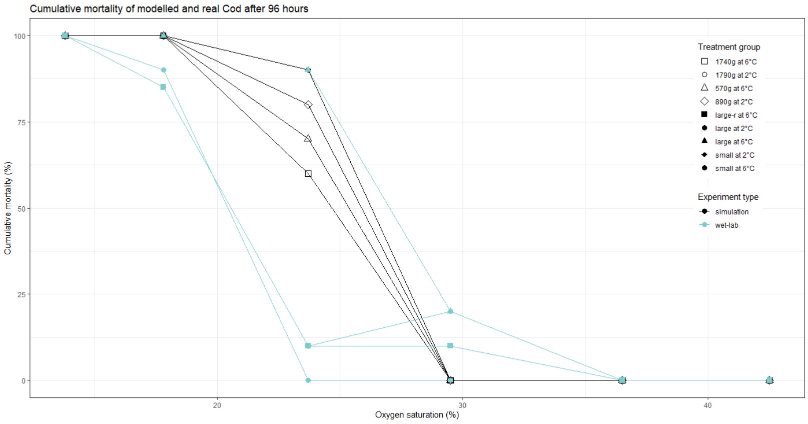


Figure 43: Screenshot of results of the validation experiment wet-lab and simulation data included in the figure. Figure show cumulative mortality after 96 hours exposed to different oxygen saturations at different temperatures for different weight classes.

saturations for 96 hours of modelled time. To provide easy access, these results are visualised alongside the respective wet-lab results and provided as a figure, see Figure 43, here comparing cumulative mortality of real and simulated cod. Although the mortality of modelled cod is slightly more regular the result is satisfactory.

Enriching TRACE with a Specialised Provenance Data Model

Documentation, such as ODD and TRACE, structures information on models and simulation studies based on their functional aspects, such as design concepts or model analysis. However, information about simulation studies can also be structured in the form of 'provenance', providing an additional perspective.

Standardised Documentation of Provenance - Prov

Provenance[171] aims to provide

"information about entities, activities, and people involved in producing a piece of data or thing, which can be used to form assessments about its quality, reliability, or trustworthiness"[172]

PROV is a set of standards recommended by the W3C (and the successor of the open provenance model OPM standard[173]). The goal of this standard is to support the exchange of provenance information on the web, of particular interest here is the data model (PROV-DM) which specifies the "core structures, forming the essence of provenance information"[174].

In practice, four concepts and their formalised representations are relevant to the provenance of modelling and simulation studies, two types of 'provenance-artefacts' and two types of relationships. 1.) Entities, such as models or data, are represented by circles or ovals 2.) Activities, such as building or merging, are represented by squares or rectangles 3.) Relationships signifying generation, and 4.) relationships signifying use, where the latter two are represented by an arrow pointing chronologically backward (see Figure 44).

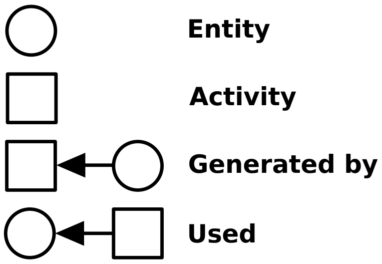


Figure 44: Overview of relevant provenance concepts. 'Provenance-artefacts': entities and activities and relationships: generations and use

Customising the Provenance Data Model for Simulation Studies

Since the provenance data model aspires to a broad range of use cases, it is naturally a general one. It has been adapted for modelling and simulation studies to be more tailored to this specific use case. To this end, the central processes and products of simulation studies were identified and mapped onto the entities and activities of the provenance data model. The main products are simulation models, simulation experiment specifications and the data resulting from executing these experiments. The development of models requires input in the form of theories and data, as do the processes of calibration, validation, and assessment of models, which are part of the iterative modelling and simulation life cycle. These processes and the "input" entities (theories and data) have a *used by* relationship and upon their

[171] Provenance as a concept was adopted from its use concerning works of art, where it tracks the chronology of their custody, location, or ownership.

[172] Paul Groth and Luc Moreau. *PROV-Overview – An Overview of the PROV Family of Documents*. Technical Report, World Wide Web Consortium. 2013

[173] Luc Moreau et al. "The Open Provenance Model core specification (v1.1)". In: *Future Generation Computer Systems* 27.6 (2011), pp. 743–756. ISSN: 0167739X. DOI: `10.1016/j.future.2010.07.005`

[174] https://www.w3.org/TR/2013/REC-prov-dm-20130430/

execution *generate* simulation experiment specifications and output data which can then be evaluated[175] [176].

In comparison to the structured natural language documentation of TRACE, this type of documentation[174] maps to directed acyclic graphs that underlie the provenance data model. This provides additional options in the form of queryability and the explicit tracking of individual steps. Thereby, for example, different versions of a model are related to each other.

Since the workflow of modelling and simulation studies often produces reoccurring patterns of entities and activities it has been suggested to provide these patterns as units to aid in the recoding and analysing of provenance information of simulation studies. Common patterns generating simulation models were identified as: Creating (requiring unspecified input and generating a simulation model), Re-Implementing (requiring a simulation model as input and generating a different one), Refining (like Re-Implementing but requiring additional input), Composing (like Refining, only requiring two or more models as input) can be supplied as one to support the recording of provenance. Common patterns generating data were identified as: Calibrating (this process also belongs to the former group as it generates a new model), Validating, Analysing, and Sensitivity Analysis, all require a simulation model, and additional input all produce a simulation experiment, simulation data and a 'success/failure' status[177].

Provenance of the EBC study

The provenance of the EBC modelling and simulation study has been structured into eight different phases, or 'studies': 1.) the development of the physiological model 2.) the development of the two aspects (physiological and behavioural tie-in) of reproduction 3.) the development of the environment which includes abiotic factors and suitable prey 4.) the combination of the juvenile physiology and reproduction into adult cod 5.) the combination of the environment and adult cod to have cod 'at liberty' 6.) the expansion of the juvenile physiology with the uptake and impact of parasitation 7.) the combination of parasitation and adult for recreated wet-lab experiments with adult cod investigating the effects of parasitation 8.) the combination of the "at liberty" set up with parasitation to recreate field conditions. A structured overview of a streamlined version of this provenance is given in Figure 45.

Recording Specialised Provenance - WebProv

The provenance standard explicitly aims to remain 'technology-agnostic' subsequently, practical implementations can take many forms. Relevant to the work presented here is an implementation that also allows for recording meta-data. The specific tool used is WebProv, a web application for inserting, accessing, and querying provenance information[178]. Users can create nodes, fill in associated metadata,

[175] Andreas Ruscheinski and Adelinde Uhrmacher. "Provenance in modeling and simulation studies — Bridging gaps". In: Winter Simulation Conference. IEEE, 2017, pp. 872–883. ISBN: 978-1-5386-3428-8. DOI: `10.1109/WSC.2017.8247839`

[176] Andreas Ruscheinski et al. *Towards a PROV Ontology for Simulation Models.* In: Provenance and Annotation of Data and Processes, 09-10 Jul 2018, London, UK. Proceedings, published by Springer International Publishing, pp. 192-195. 2018

[177] Pia Wilsdorf et al. "Automatic Reuse, Adaption, and Execution of Simulation Experiments via Provenance Patterns". In: *arXiv:2109.06776 [cs]* (2021)

[178] Kai Budde et al. "Relating simulation studies by provenance—Developing a family of Wnt signaling models". In: *PLOS Computational Biology* 17.8 (2021). Ed. by Pedro Mendes, e1009227. ISSN: 1553-7358. DOI: `10.1371/journal.pcbi.1009227`

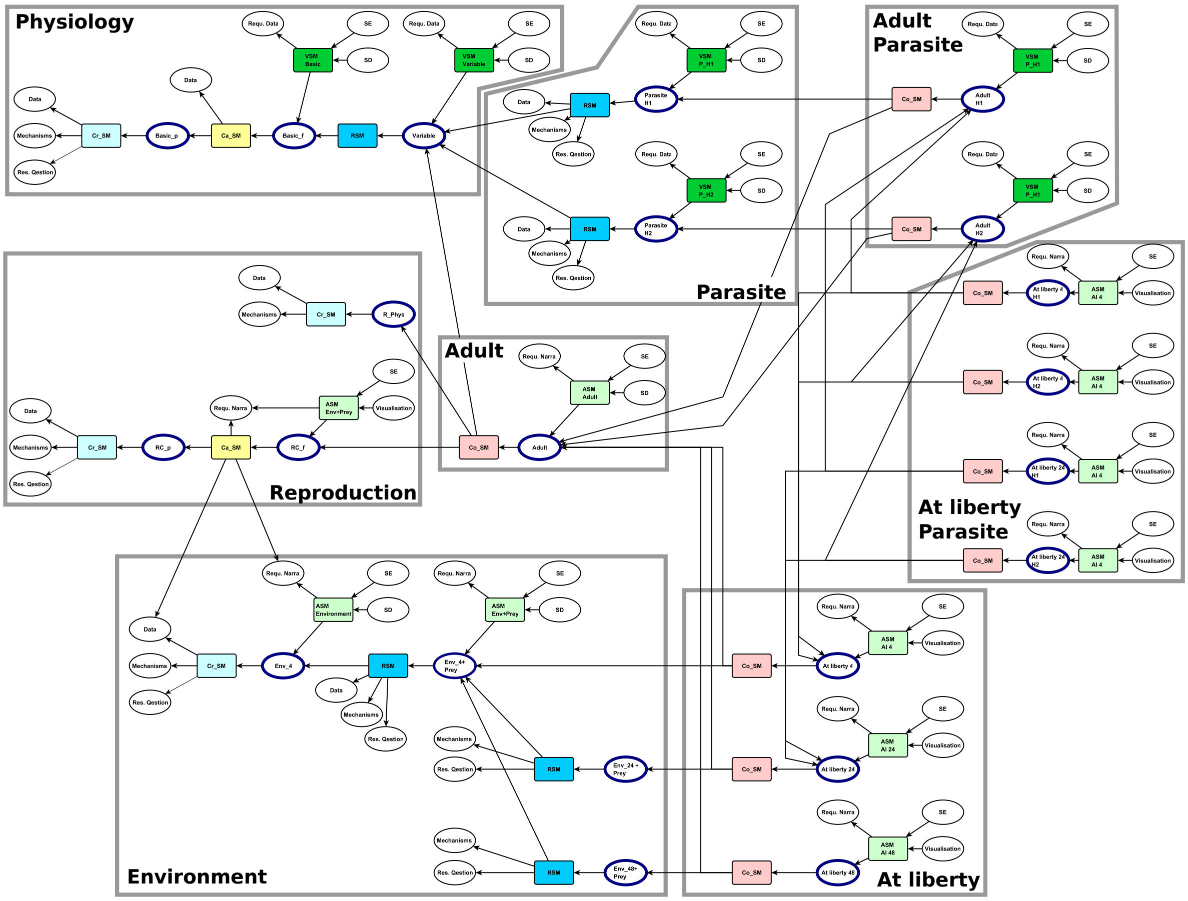


Figure 45: Overview of streamlined Provenance Graph of modelling and simulation study on Eastern Baltic Cod.

connect them via appropriate relationships, and group the resulting networks into studies. It was created to visualise the development of cellular biochemical models to make relationships between models that investigate the same or closely related topics accessible. The simplified implementation of the EBC study in WebProv is shown in Figure 47.

The WebProv frontend is based on the JavaScript framework Vue and the JavaScript library D3.js. It supports simple text queries for filtering nodes based on their metadata. All entities, activities and relationships are stored in a Neo4j database backend that handles advanced graph queries in Neo4j's Cypher language [179] and allows for JSON imports and exports.

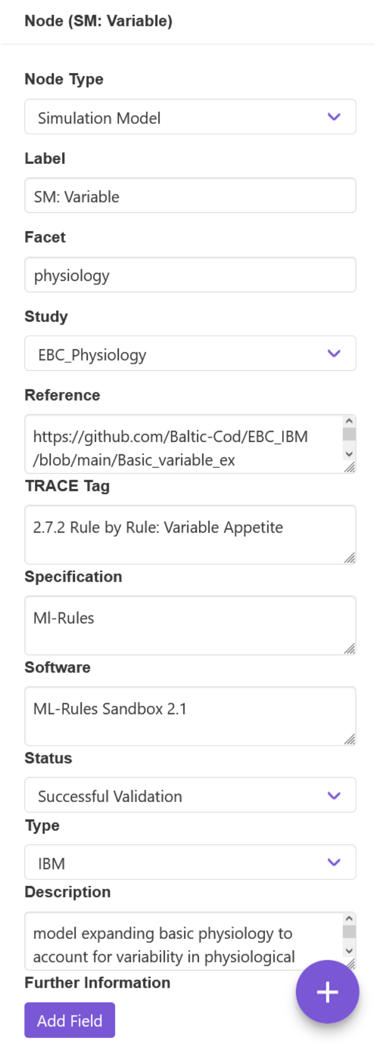
Recording Specialised Provenance for Ecology

To best accommodate the requirements of the EBC simulation study and similar ecological simulation studies, the specialisations made so far for simulation studies in general, and simulation studies of cell-biological systems specifically[178] were further built upon. Primarily the entities, activities and attached metadata were further tailored to simulation studies of aquatic ecology and the general approach. There is a considerable overlap between the requirements for cell-biology and aquatic ecology, to avoid repetition the WebProv version developed for the EBC will be presented here and the specific changes made from the previous version will be pointed out.

Entities types now are: 'Research Question' (which guides the following activities in terms of objectives), 'Assumption' (which is a weighty input and contributes to outlining the scope of applicability), 'Requirement' (which guides the following activities in terms of specific properties), 'Qualitative Model' (which guides the following activities in terms of structure), 'Simulation Model' (which is one of the two central artefacts of a modelling and simulation study), 'Simulation Experiment' (which is the other central artefact), 'Simulation Data' (which is on of the two primary artefacts produced by a modelling and simulation study), 'Simulation Visualisation' (which is the other primary artefact produced), 'Wet-Lab Data' (which is one of the two primary data sources in ecology), 'Field Data' (which is the other primary data source). Apart from the addition of 'Simulation Visualisation' and 'Field Data' the entities remained unchanged from the previous (cell biology) version.

These Entities are outfitted with metadata: All entities have the option for metadata on 'Facet' (the respective topical facet of the model), 'Description' (a text field to give a short natural language description), 'Study' (which affiliates the entity to a specific study and thereby provides additional structure, please note that while studies might explore a particular facet entities of a given study can be affiliated with more than one facet) and 'Further Information' (which allows adding additional fields when required). All entities can also have metadata in the form of a 'TRACE-Tag' this will be discussed in more detail in the next section. The entities to which a

[179] Justin J Miller. "Graph Database Applications and Concepts with Neo4j". In: *Proceedings of the Southern Association for Information Systems Conference, Atlanta, GA, USA*. vol. 2324. 2013



Node (SM: Variable)

Node Type
Simulation Model

Label
SM: Variable

Facet
physiology

Study
EBC_Physiology

Reference
https://github.com/Baltic-Cod/EBC_IBM
/blob/main/Basic_variable_ex

TRACE Tag
2.7.2 Rule by Rule: Variable Appetite

Specification
Ml-Rules

Software
ML-Rules Sandbox 2.1

Status
Successful Validation

Type
IBM

Description
model expanding basic physiology to
account for variability in physiological

Further Information

Figure 46: Screenshot of WebProv metadata for an entity of the type simulation model.

reference can assigned are: 'Qualitative Model', 'Simulation Model', 'Simulation Experiment', 'Simulation Data', 'Simulation Visualisation', 'Wet-Lab Data' and 'Field Data' to easily provide either their sources or their current availability in published form. For 'Simulation Model', 'Simulation Experiment' and 'Simulation Visualisation' both programming language and software, including the specific version, can be assigned which supports reproducibility. A 'status' denoting successful validation or calibration can be assigned to 'Simulation Model' and 'Simulation Data' quickly providing this relevant information. Here the metadata for Facet was added from the previous version to accommodate the structure of the general approach introduced in Chapter II and aid in meaningful queryability of metadata. Also, the TRACE-Tag was added to interlink provenance information with TRACE documentation. For an example of metadata see Figure 46.

The remaining category 'Type' is a different one for each entity and will therefore be given individually: the type information on 'Research Question' distinguishes between explanation and prediction, denoting the intended goal. This is an important distinction regarding the communication with policymakers as there can be a temptation to treat quantitative results of studies aimed at investigation as predictions, which is less than optimal. 'Assumptions' can be grouped into quantitative assumptions (e.g., cod will die below a condition factor of 0.6) and narratives (i.e., cod will have the urge to swim to their spawning grounds when their gonads are fully developed). Especially to support clear communication during broad investigations, this information is helpful as the two types are equally valid but have vastly different implications when discussed among experts. Correspondingly 'Requirements' are grouped into quantitative, data-based and narrative-based, again denoting the different implications. 'Qualitative Model' can be formal or informal, also supporting clear communication by supplying the level of ambiguity at the starting point. 'Simulation Models' can be associated with the type of approach such as IBMs or ODEs. This gives direct input about the classification of the model and assists in gauging which types of insights to expect and not to expect. 'Simulation Experiments' can be assigned to categories like parameter scan, sensitivity analysis, and more, relaying this procedural information. Similarly 'Simulation Visualisation' can be assigned categories such as a graph, animation, etc., in this case giving information on which types are available. Finally, 'Field Data' is grouped into surveys and experiments, implicitly providing information on suitability, specialisation and possible availability of, for example, similar data on a different species. Assigning metadata on type had been previously introduced[180], here it has been adapted for the specific requirements of ecology by adding and/or (re)defining it for all entities except 'Simulation Experiments'. This information is cumulated in Table 5.

All these entities are generated and used by activities, these are 'Build', which can be further distinguished into 'Create', 'Refine' and 'Compose' simulation models. This set of activities uses input entities

[180] Pia Wilsdorf, Fiete Haack, and Adelinde M. Uhrmacher. "Conceptual Models in Simulation Studies: Making it Explicit". In: *2020 Winter Simulation Conference (WSC)*. 2020 Winter Simulation Conference (WSC). Orlando, FL, USA: IEEE, 2020, pp. 2353–2364. ISBN: 978-1-72819-499-8. DOI: `10.1109/WSC48552.2020.9383984`

Table 5: Overview of provenance information considered useful for ecological modelling and simulation studies. Type applies differently to each entity: research question [explanation, prediction, ...], assumption [quantitative, narrative, ...], requirement [quantitative, data based, narrative based, ...], qualitative model [informal, formal, ...], simulation model [IBM, ODE, ...], simulation experiment [parameter scan, sensitivity analysis, ...], simulation visualisation [graph, animation, ...], field data [survey, experiment, ...]

Entity	Reference:	TRACE-Tag:	Specification:	Software:	Status:	Type:	Others:
	Text	TRACE (Sub)Section(s)	programming language	type, version	[successful validation, successful calibration, none]		Label Facet Description Study Information
Research Question		Yes				Yes	Yes
Assumption		Yes				Yes	Yes
Requirement		Yes				Yes	Yes
Qualitative Model	Yes	Yes				Yes	Yes
Simulation Model	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Simulation Experiment	Yes	Yes	Yes	Yes		Yes	Yes
Simulation Data	Yes	Yes			Yes		Yes
Simulation Vis.	Yes	Yes	Yes	Yes		Yes	Yes
Wet-Lab Data	Yes	Yes					Yes
Field Data	Yes	Yes				Yes	Yes

such as research questions and data and generate simulation models. The activity 'Calibrate' also generates a simulation model and requires input to calibrate against. 'Validate' uses input data and a simulation experiment to produce output data which can be used to evaluate if a model is valid for a given type of experiment. 'Analyse' also uses a simulation experiment but for a different purpose as validation and generates simulation data. 'Visualise' is, of course, a useful step to convey simulation data, however, it is of particular interest in the case of assumptions or requirements being narratives or narrative-based, respectively. For example for the EBC the formalised narratives of cod moving through the Bornholm Basin habitat based on time of year and their reproductive cycle was visualised in a 2D animation, this provided a direct feedback about model behaviour to domain scientists familiar with surveying the cod. Based on the previously identified provenance patterns[181], only 'Visualise' needed to be added while the activity 'Sensitivity Analysis' was not adopted at this stage.

Provenance and TRACE interlinked

Using the specialised provenance data model for simulation studies on aquatic ecology, the provenance of the EBC study was implemented in WebProv, based on the associated metadata, it can now be related to the TRACE documentation.

[181] Pia Wilsdorf et al. "Automatic Reuse, Adaption, and Execution of Simulation Experiments via Provenance Patterns". In: *arXiv:2109.06776 [cs]* (2021)

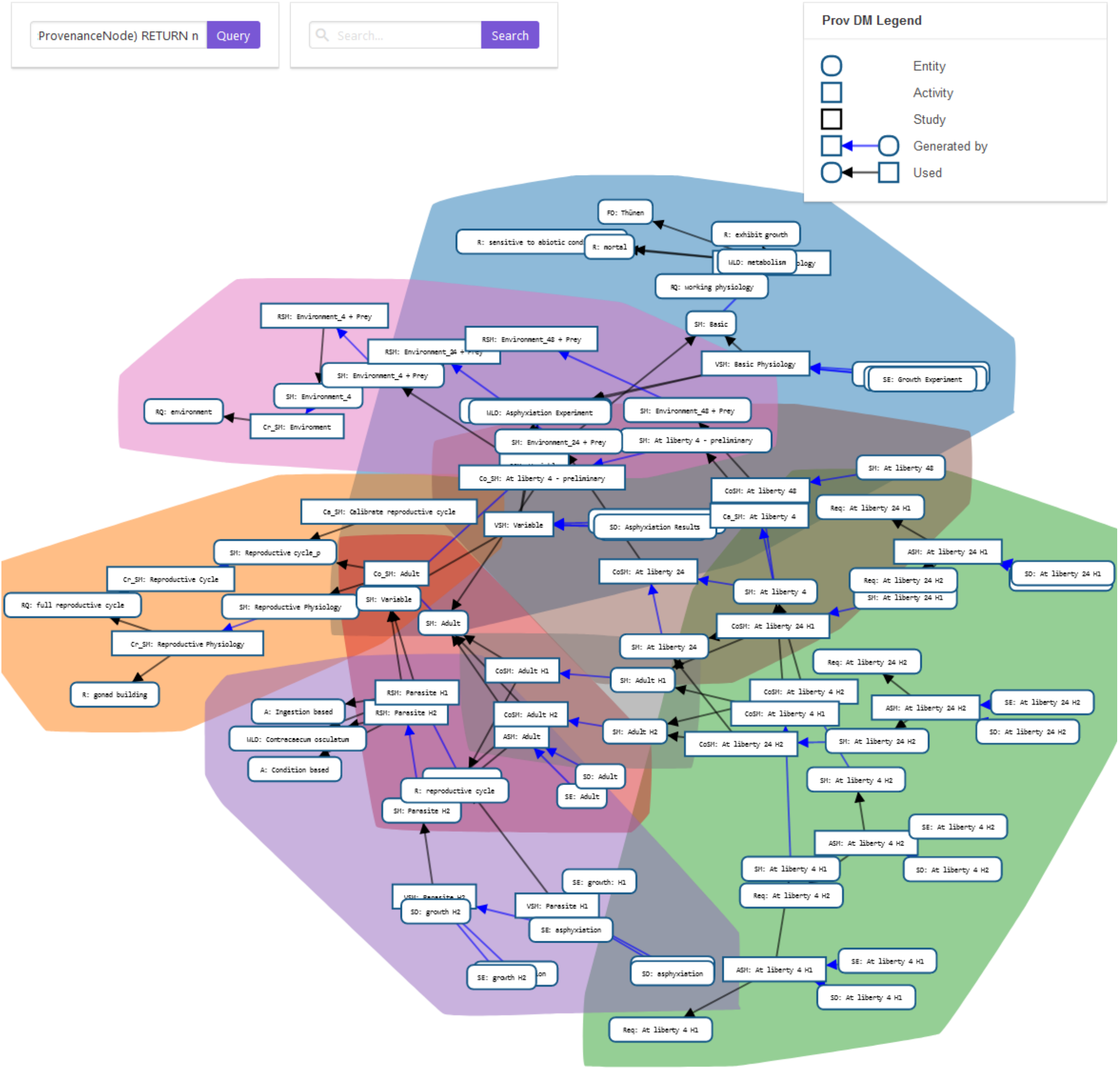


Figure 47: Screenshot of EBC Provenance Graph in WebProv

TRACE - Tags

The TRACE-Tag, introduced above, allows to link a given entity or activity of the provenance within the associated section in the TRACE document. Using metadata to connect recorded provenance with TRACE documentation produces a navigable combination of the two. Depending on the depth of information required, the provenance graph supplies a visual overview of activities that have taken place and the available artefacts. However, if the desired information goes beyond categorisations and short explanations, which is what WebProv is intended to handle, the TRACE-Tag provides a direct pointer to the in-depth information provided by the long-form TRACE documentation of the EBC study.

Queryability

As stated in the section on WebProv, the metadata of the provenance information is rendered queryable using Cypher[182]. Queries can be used to navigate and extract useful sub-graphs. As a result, metadata, in particular on the different facets each entity or activity addresses, provides the basis for "navigation" of the documentation. Here some examples shall be given of its practical application.

Naturally the entire graph can be queried, providing a quick overview of the scope of the entire study:

```
MATCH (n:ProvenanceNode) RETURN n
```

The result shown in Figure 47 shows the entirety of the graph. Please note that query results will, where applicable automatically include relationships (edges).

Another option is to query for entities or activities. For example querying for simulation models:

```
MATCH (n) WHERE n.definitionId =
"Simulation Model" RETURN n
```

The result shown in Figure 48 displays all simulation models currently included in the graph, this type of query provides a quick overview of the availability or lack of a specific entity or activity. Note that as two simulation models can not be related in the provenance graph, no relationships are shown. Depending on the perspective, such queries can be used to scrutinise (i.e., why is there only one version of the 'Adult' simulation model, was no refinement or calibration required?) or to navigate (i.e., oh, there are two competing hypotheses, H1 and H2, in the parasitation study).

But in particular the option to query the meta information by domain relevant content, for example for a specific facet, is a valuable asset. For example querying for the facet of parasitation:

```
MATCH (n) WHERE n.facet CONTAINS "parasitation"
RETURN n
```

The results are shown in Figure 49 and display all entities and

[182] https://neo4j.com/developer/cypher/

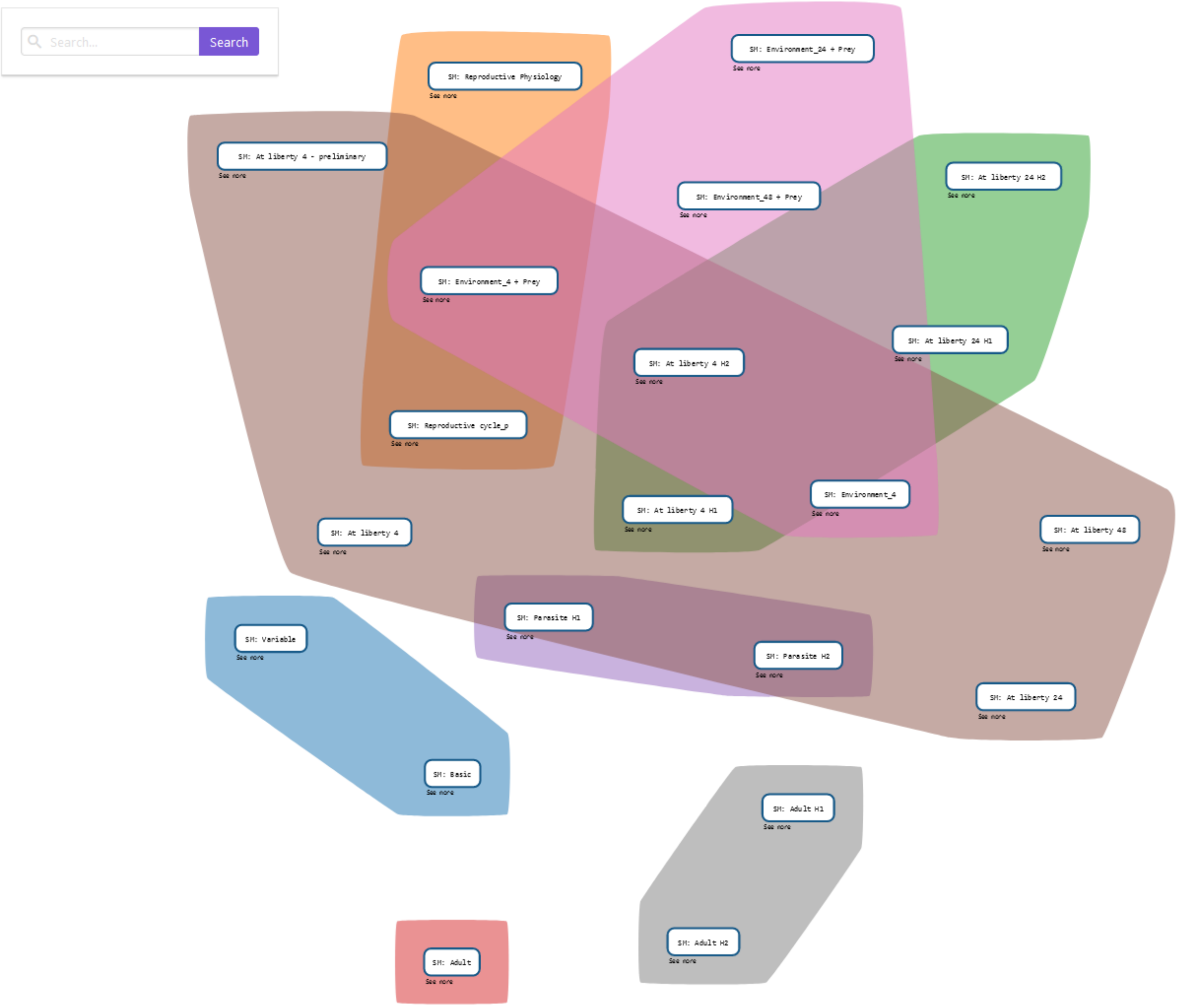


Figure 48: Screenshot of the results for the query:
`MATCH (n) WHERE n.definitionId = "Simulation Model" RETURN n`
this query returns all simulation models.

activities which pertain to the facet of parasitation. The type of simulation study suggested by the general approach introduced in Chapter II will naturally become expansive and address numerous facets in an overlapping manner, as the EBC study does. The option to query for a specific facet or combination of facets makes this intentional blending of aspects navigable. The example shows that three of the studies factor in parasitation and that the competing hypotheses, H1 and H2 (please note that working hypothesis are introduced as assumptions 'A:' in the provenance graph), have been carried forward, indicating that neither could be ruled out so far.

Subsequently the queries performed on the provenance graph can then also be used to swiftly navigate the TRACE documentation, further enhancing accessibility of expansive modelling and simulation studies.

Discussion

The work presented in this thesis is set in the general context of supporting management, which requires a basis for trust, and the general approach introduced in Chapter II aims to use modelling and simulation to integrate knowledge and align mental models, which requires a basis for exchange and scrutiny. To this end, the TRACE protocol, which addresses these exact requirements, having been developed and refined to 'establish good modelling practices for ecological simulation studies', is being followed to document the EBC simulation study.

However, this type of modelling and simulation study can benefit further from formal documentation of its provenance, placing emphasis on the succession of activities and artefacts that were used and generated. In particular, the general approach requires the development and subsequent integration of submodels and the regular realisation of sub-studies that guide the further direction of the model. Recording provenance provides a structured overview of the paths which data and assumptions have taken, allowing to gauge their impact quickly and, if need be, to identify points in the study form which an alternate route should be taken.

By adding customised metadata to the provenance graph, additional possibilities were revealed. Firstly, this allows to query the graph in a meaningful manner, in particular, querying for the facets will allow domain scientists to obtain the subset of information relevant to them. Secondly, it can be used to interlink the provenance graph with the TRACE documentation by providing a 'TRACE-Tag' which allows nodes to point towards sections of the TRACE documentation, thereby providing navigability. No one documentation tool can satisfy all demands: queryability, structure, automatisation, detail, catering to a different perspective. However, by interlinking them, each can be used to the extent useful for that particular study without having to overburden or under-use any framework beyond the applications for which it is best.

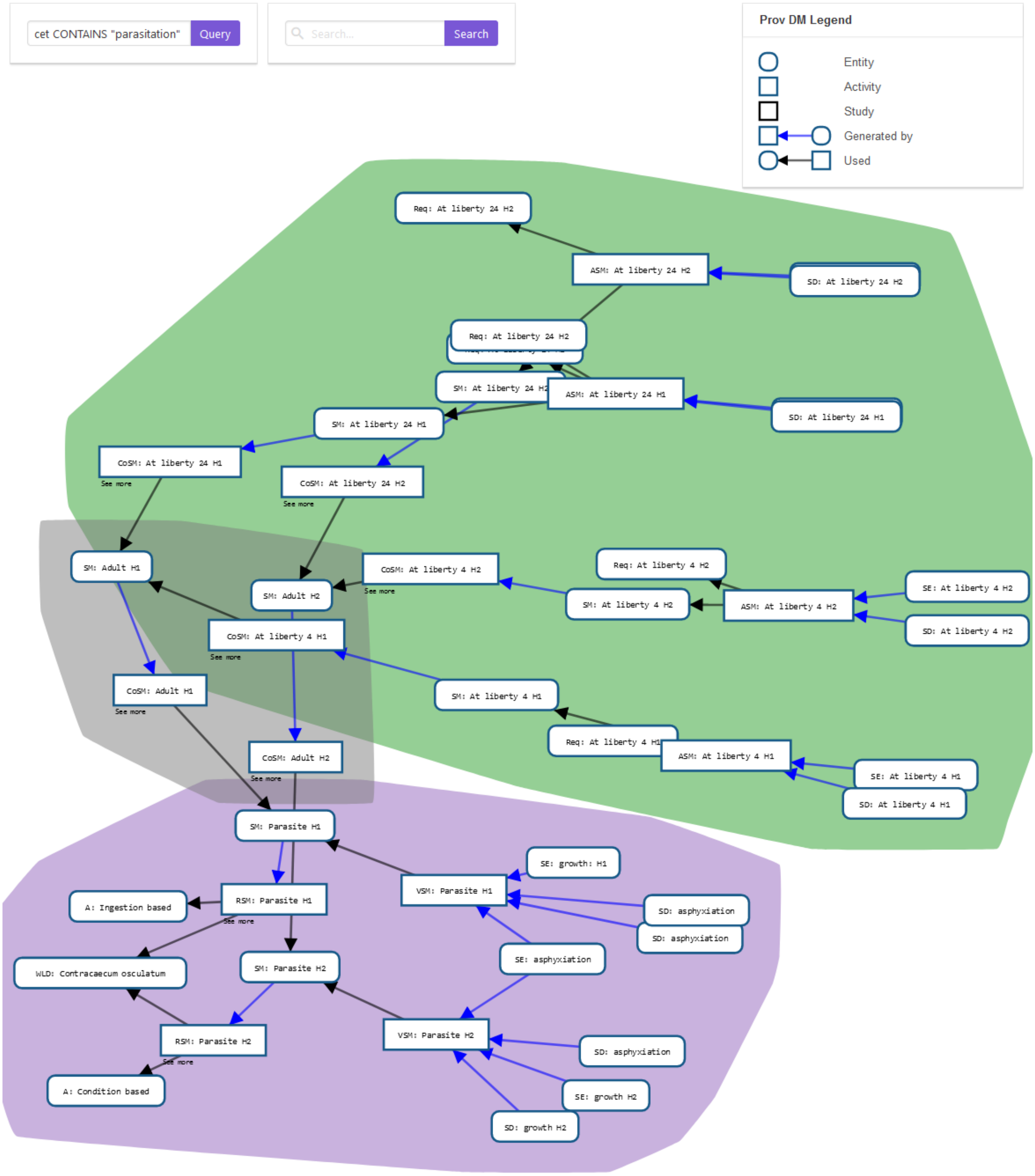


Figure 49: Screenshot of the results for the query:
`MATCH (n) WHERE n.facet CONTAINS "parasitation" RETURN n`
this query returns all entities and activities where one of the facets being considered is parasitation.

Part III

EPILOGUE

Conclusion

I may not have gone where I intended to go, but I think I have ended up where I needed to be.

Douglas Adams

Retrospective

Here, the work presented in this thesis shall be summarised and the key insights reiterated.

The Eastern Baltic cod (EBC) stock is declining, and there is currently no solid scientific basis to identify suitable measures to support its recovery. Established management methods have failed, and it is the consensus of the scientific community surrounding the stock that the drivers of this development need to be identified, and only a sound knowledge base can provide the required directionality. The dynamics of the Baltic ecosystem and the EBC are evolving under anthropogenic pressures, and the mechanisms of these changes and the new system states will have to be understood to identify appropriate points of intervention to support the recovery and subsequent consistent management of the EBC.

Knowledge of the EBC is distributed by scientific disciplines and areas of expertise across the many different mental models each individual scientist holds. Communicating and subsequently functionally and unambiguously integrating this knowledge will be required to establish the desired knowledge base. In this thesis, I showed how this could be achieved by a multi-faceted modelling approach using white-box integration and supported at different levels by appropriate modelling and simulation tools (i.e., domain-specific languages - ML-Rules as an available example), approaches (i.e., separation of model and simulator, using individual-based models, etc.) and infrastructure (i.e., a workflow to establish and maintain validity and a thorough documentation).

Formalising mental models of different facets of EBC ecology will naturally make them unambiguous, and subsequently integrating them will result in the alignment of understanding of the different scientists involved. Finally, larger models can be used to conduct experiments, and in this manner, the more extensive study can iteratively work towards honing in on a well-founded understanding of the relevant processes and interactions currently driving EBC ecology. However, this will only work in practice if these formalisations are accessible to fisheries scientists. Such accessibility can be supported by providing a modelling language with a syntax that uses familiar or at least intuitive metaphors, thereby providing a low threshold for participation and no or minimal requirement to distort mental models. At the same time, such a domain-specific language (DSL) will also have to allow for concise and succinct notation and easy composition to let the model be iteratively expanded and remain legible and accessible throughout this process.

The DSL ML-Rules used in the work presented here fulfilled most but not all of these requirements. It is suitably succinct and legible, and due to its rule-based syntax, it lends itself well to the successive composition of ever more interdependencies into even a single process. At the same time, it was evident that although some metaphors are quite similar between cell biology (for which it was developed) and aquatic ecology, e.g., a rule-based syntax using guards and rates, some metaphors are not. In particular, the lack of an 'environment', a metaphor for providing abiotic conditions as part of the semantic domain, required considerable and cumbersome workarounds. The lack of named attributes was not as detrimental, but it did reduce legibility of the later iterations for those not familiar with the model.

I also proposed how the workflow of multi-faceted modelling as a 'thinking tool' can be incorporated into the superordinate workflow of stock assessment and producing Advice[183] as conducted by ICES and its contributing institutes and scientists. Connections can be made at four key points, firstly, there is the obvious input of mental models from the scientist conducting research explicitly for ICES or in alignment with their objectives into the different models. Secondly, any gaps in data that have been uncovered by the necessity to produce working and parametrised models can advise the conduct of experiments and surveys. Thirdly, any valuable insights from alignment and updating of mental models can be included in the advisory process. Finally, any mechanisms crucial to the stock's dynamic that are uncovered can be subsequently incorporated into stock assessment models. Thus, the approach can be seamlessly embedded into the existing structures established by ICES while adding value to the data and research already surveyed and conducted by carrying out a simulation study that captures functional knowledge in an accessible manner and generates output that feeds back into supporting consistent management.

Nevertheless, such a simulation study will only produce valuable output if it can genuinely support investigating a complex adaptive system (CAS). Such systems are distinguished by consisting of large dynamic networks of interactions, being path-dependent, and exhibiting emergence and self-organisation. Investigating such systems through modelling and simulation requires sufficient verisimilitude of functionality on the part of the model. Therefore the semantic domain of the modelling language used must be well chosen. Discussed in the context of their respective alternatives, key properties were identified as: multi-levelness (mirroring this property in the ecosystem), continuous-time (avoiding artefacts and supporting the broad ranges of time-scales for processes encountered in the different system levels), stochasticity (to allow for path dependence) and micro-modelling in the form of individual-based modelling (again, supporting processes

[183] Advice, the scientific product as defined and implemented by ICES

across system levels by foregoing aggregation, accounting for heterogeneity and finally to possibly capture processes leading to emergent behaviour). [184] That being said, there is no reason why in some cases, using these approaches in combination with different ones might not be advantageous in the general approach.

Equally crucial to the value of the simulation study is the semantic validity of the produced models. Only if this can be ensured is subsequent experimentation expedient. A workflow interlinking integration and validation was developed to re-establish the validity of models for each given process after their iterative expansion. Specifically, the reuse of simulation experiments facilitated by a DSL for simulation experiment specification resulted in an efficient approach, supporting a systematic and, therefore, uncluttered workflow for this task. The process was subsequently further streamlined by also reusing the R-scripts for processing and visualising the simulation results. The DSL SESSL (Simulation Experiment Specification via a Scala Layer) applied here allowed to adapt experiment specifications with little effort, handling only minimal lines of code. In comparison to the modelling language ML-Rules, there are no reservations about its continued use for fisheries science, rather, the full scope of its feature (e.g., statistical model checking) has not been exhausted at this point.

Finally, the importance of accompanying thorough, high-quality documentation can not be overstated. Developing simulation models does not only require a considerable amount of work, but ideally, they subsequently also contain a correspondingly large amount of functional knowledge. Providing and maintaining access to this functional knowledge and exposing it to scrutiny is at the heart of the work presented here. On the level of informing management, statements on quality and applicability need to be provided to support trust and transparency, on the level of using and developing the model, information about data, decisions, and other considerations that are not made explicit by the code need to be available to provide easy and independent access. Fortunately, the TRACE protocol (TRansparent And Comprehensive model Evaluation), developed precisely for the use case presented here, already exists and is currently being used to document the EBC study.

Additionally to this documentation of functional aspects, provenance of the study using the provenance data model was also recorded. Particularly in light of iterative development and carrying along alternate hypotheses having access to an overview of the study in the form of a succession of actions and artefacts has proved to be valuable (i.e., at which point of development was a specific assumption implemented?). The provenance graph can also be equipped with metadata, here, previous work was adapted for the requirements of fisheries science. In particular, the assignment of facets and TRACE-Tags shows promise. The former allows querying the graph from the domain perspective. The latter connects provenance documentation with corresponding but often more detailed documentation within a TRACE document.

[184] The fact that ML-Rules offers all of these properties was another reason why it could be successfully applied to the EBC.

Implications for The Eastern Baltic Cod

The thorough and incremental modelling formalising of the different facets of EBC ecology has not uncovered the mechanisms responsible for its current decline. That being said, I wholeheartedly stand by the result this implies, namely that the mechanism as they have been included in the model so far can therefore be excluded from being responsible. Note that this is meant in the strict sense of both the nominal mechanisms and the exact manner in which they have been formalised. (There can be a reluctance to see such a 'negative result' as relevant[185], however investigating systems with such a multitude of open questions as the EBC might motivate a higher regard for being able to 'tick of' lines of investigation).

As was stated in Chapter III, the model can be forced to reproduce the lack of growth observed in the field by severely restricting food intake, which might, in a way, be the correct answer. Unpublished work on the Baltic food web using nitrogen isotope analysis points towards a lengthening of the food chain, reducing the amount of energy reaching the highest trophic levels and thereby the EBC. This lengthening is attributed to anthropogenic factors, albeit still requiring identification of the exact mechanisms. Two implications result from this speculation.

Firstly, how could this be shown using the approach presented in this thesis? Forcing reduced growth in the model has already pointed towards the obvious subsequent line of investigation: energy. Food availability in the model currently reproduces field data on the frequency and distribution of prey items, the next variable to quantify would be the absolute availability of biomass. This could, for example, be implemented by scaling the values deduced from survey data on prey species down to proportionally match the number of cod in a given simulation.

Secondly, what would this imply for management measures? If it is indeed the case that an overall change in the food web has been realised, unfortunately, the implications are dramatic and point toward two disparate types of approaches. On the one hand, working inside the new dynamic and, on the other hand, implementing decisive interventions. Working inside the new dynamic would require determining the new carrying capacity of the Baltic ecosystem and its associated variability and subsequent adaption of fisheries and supporting industries to the consequently reduced surplus productivity of the stock. This path would consolidate the already drastic socio-economic and cultural shift in progress along the shores of the Eastern Baltic Sea. Decisive interventions would likely need to go beyond bringing anthropogenic influences, such as eutrophication, to a halt but would rather require undoing them to allow the system to revert to its previous dynamic. This path would require some form of ecosystem engineering, and examples where such undertakings have been conducted gently and with sufficient responsiveness not to cause devastating unintended consequences are exceedingly rare.

[185] Daniele Fanelli. "Negative results are disappearing from most disciplines and countries". In: *Scientometrics* 90.3 (2012), pp. 891–904. ISSN: 0138-9130, 1588-2861. DOI: `10.1007/s11192-011-0494-7`

Prospective

There are also several avenues of pursuit for future work.

The most demanding and elaborate task will be establishing the workflow presented here in the fisheries science community surrounding the EBC. To this end, its merits will need to be demonstrated and successfully communicated. This will likely require a considerable amount of legwork to overcome the initial inertia to the point of participation becoming its own reward.

Interrelated with establishing the general approach is the development of a suitable DSL. To this end, the insights on semantic domain and the trade-off between accessibility and verisimilitude presented here can be built upon. Apart from the scientific challenge of conducting user studies and implementing a DSL tailored to the broad spectrum of requirements that need to be met, this also entails securing the necessary funding to provide a user-friendly finished product.

As has been made clear by the scientific community's repeated and not always successful efforts to establish good documentation practices for simulation studies, documentation requires additional work, which is omitted when other tasks in the study claim resources in terms of time and labour. One way to alleviate this situation is to reduce the work required for documentation by providing semi-automated tools and frameworks.

Although the visualisations produced throughout this study are useful and up to the task of relaying a considerable scope of information in a meaningful manner, they are nevertheless not state of the art. Adapting and developing more sophisticated methods of visualisation for the entire process an thereby 'taking the human into the loop' would greatly benefit the building, refinement and interpretation of models, which in turn will support efficient and thorough investigation.

Ultimately, this approach could be introduced to the management of other stocks facing similar issues and eventually, management approaches inherently geared to taking into account the interactions of sub-systems, such as ecosystem management, might adapt the approaches and tools presented here to their specific workflow.

Final Conclusion

The work presented in this thesis began deeply grounded in fisheries science, aiming to answer open questions about the EBC. It arrived at a point viewing this goal from the perspective of modelling and simulation. Working towards understanding the EBC revealed the need to develop approaches and tools that can aid in understanding the EBC not only once but unremittingly as its supporting ecosystems continues to evolve. Only if management has access to a scientific workflow that can provide uninterrupted insight and foresight will it be able to provide consistent guidance for use and stewardship.

Part IV

APPENDIX

References

[1] Madhur Anand et al. "Ecological Systems as Complex Systems: Challenges for an Emerging Science". In: *Diversity* 2.3 (2010), pp. 395–410. ISSN: 1424-2818. DOI: `10.3390/d2030395`.

[2] Gennady Andrienko et al. "Space, time and visual analytics". In: *International Journal of Geographical Information Science* 24.10 (2010), pp. 1577–1600. ISSN: 1365-8816, 1362-3087. DOI: `10.1080/13658816.2010.508043`.

[3] Jacqueline Augusiak, Paul J. Van den Brink, and Volker Grimm. "Merging validation and evaluation of ecological models to 'evaludation': A review of terminology and a practical approach". In: *Ecological Modelling* 280 (2014), pp. 117–128. ISSN: 03043800. DOI: `10.1016/j.ecolmodel.2013.11.009`.

[4] Osman Balci. "Verification, Validation and Accrediation of Simulation Models". In: *Proceedings of the 1997 Winter Simulation Conference*. Ed. by S. Andradóttir et al. IEEE. Piscataway, New Jersey, 1997, pp. 135–141. DOI: `10.1109/IEEESTD.1997.8685803`.

[5] Francois Bastardie et al. "A Review Characterizing 25 Ecosystem Challenges to Be Addressed by an Ecosystem Approach to Fisheries Management in Europe". In: *Frontiers in Marine Science* 7 (2021), p. 629186. ISSN: 2296-7745. DOI: `10.3389/fmars.2020.629186`.

[6] Paul R. Berg et al. "Adaptation to Low Salinity Promotes Genomic Divergence in Atlantic Cod (*Gadus morhua* L." In: *Genome Biology and Evolution* 7.6 (2015), pp. 1644–1663. ISSN: 1759-6653. DOI: `10.1093/gbe/evv093`.

[7] Björn Björnsson, Agnar Steinarsson, and Matthias Oddgeirsson. "Optimal temperature for growth and feed conversion of immature cod (*Gadus morhua* L." In: *ICES Journal of Marine Science* 58 (2001), pp. 29–38. DOI: `10.1006/jmsc.2000.0986`.

[8] Sieme Bossier et al. "Integrated ecosystem impacts of climate change and eutrophication on main Baltic fishery resources". In: *Ecological Modelling* 453 (2021), p. 109609. ISSN: 03043800. DOI: `10.1016/j.ecolmodel.2021.109609`.

[9] Daniel B. Botkin, James F. Janak, and James R. Wallis. "Some Ecological Consequences of a Computer Model of Forest Growth". In: *The Journal of Ecology* 60.3 (1972), p. 849. ISSN: 00220477. DOI: `10.2307/2258570`.

[10] L. W. Botsford. "The Management of Fisheries and Marine Ecosystems". In: *Science* 277.5325 (1997), pp. 509–515. ISSN: 00368075, 10959203. DOI: `10.1126/science.277.5325.509`.

[11] Kai Budde et al. "Relating simulation studies by provenance—Developing a family of Wnt signaling models". In: *PLOS Computational Biology* 17.8 (2021). Ed. by Pedro Mendes, e1009227. ISSN: 1553-7358. DOI: `10.1371/journal.pcbi.1009227`.

[12] Jack D. Burke. "Vertebrate Blood Oxygen Capacity and Body Weight". In: *Nature* 212 (1966), pp. 46–48. DOI: `10.1038/212046a0`.

[13] D.t. Campbell. "Downward Causation". In: *Studies in the philosophy of biology*. University of California Press, 1974.

[14] François E. Cellier. *Continuous System Modeling*. New York, NY: Springer New York, 1991. ISBN: 978-1-4757-3924-4 978-1-4757-3922-0. DOI: `10.1007/978-1-4757-3922-0`.

[15] D. Chabot and J.-D. Dutil. "Reduced growth of Atlantic cod in non-lethal hypoxic conditions". In: *Journal of Fish Biology* 55 (1999), pp. 472–491. DOI: `10.1111/j.1095-8649.1999.tb00693.x`.

[16] G. Claireaux et al. "Influence of water temperature and oxygenation on the aerobic metabolic scope of Atlantic cod (*Gadus morhua*". In: *Journal of Sea Research* 44.3 (2000), pp. 257–265. ISSN: 13851101. DOI: `10.1016/S1385-1101(00)00053-8`.

[17] Daniel J. Conley et al. "Hypoxia-Related Processes in the Baltic Sea". In: *Environmental Science & Technology* 43 (2009), pp. 3412–3420. DOI: `10.1021/es802762a`.

[18] Belva J. Cooley. "Documenting simulation studies for management". In: *ACM SIGSIM Simulation Digest* 10.3 (1979), pp. 24–27. ISSN: 0163-6103. DOI: `10.1145/1102802.1102804`.

[19] Jonathan Cooper, Jon Olav Vik, and Dagmar Waltemath. "A call for virtual experiments: Accelerating the scientific process". In: *Progress in Biophysics and Molecular Biology* 117 (2015), pp. 99–106. DOI: `10.1016/j.pbiomolbio.2014.10.001`.

[20] N. Crout et al. "Good Modelling Practice". In: *Developments in Integrated Environmental Assessment*. Vol. 3. Elsevier, 2008, pp. 15–31. ISBN: 978-0-08-056886-7. DOI: `10.1016/S1574-101X(08)00602-9`.

[21] G. De Luca et al. "Fishing out collective memory of migratory schools". In: *Journal of The Royal Society Interface* 11 (2014), pp. 20140043–20140043. DOI: `10.1098/rsif.2014.0043`.

[22] Tom De Wolf and Tom Holvoet. "Emergence Versus Self-Organisation: Different Concepts but Promising When Combined". In: *Engineering Self-Organising Systems*. Ed. by Sven A. Brueckner et al. Vol. 3464. Berlin, Heidelberg: Springer Berlin Heidelberg, 2005, pp. 1–15. ISBN: 978-3-540-26180-3 978-3-540-31901-6. DOI: `10.1007/11494676\_1`.

[23] Margit Eero, Massimiliano Cardinale, and Marie Storr-Paulsen. "Emerging challenges for resource management under ecosystem change: Example of cod in the Baltic Sea". In: *Ocean & Coastal Management* 198 (2020), p. 105314. ISSN: 09645691. DOI: `10.1016/j.ocecoaman.2020.105314`.

[24] Margit Eero et al. "Eastern Baltic cod in distress: biological changes and challenges for stock assessment". In: *ICES J Mar Sci* 72 (2015), pp. 2180–2186. ISSN: 1095-9289. DOI: `10.1093/icesjms/fsv109`.

[25] *Empfehlungen zur Profilierung der Einrichtungen mit Ressortforschungsaufgaben des Bundes*. Wissenschaftsrat. 2010.

[26] Josefin Engelhardt et al. "Severe thiamine deficiency in eastern Baltic cod (*Gadus morhua*". In: *PLOS ONE* 15.1 (2020). Ed. by Ilaria Corsi, e0227201. ISSN: 1932-6203. DOI: `10.1371/journal.pone.0227201`.

[27] Matthew R. Evans. "Modelling ecological systems in a changing world". In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 367.1586 (2012), pp. 181–190. ISSN: 0962-8436, 1471-2970. DOI: `10.1098/rstb.2011.0172`.

[28] Matthew R. Evans et al. "Do simple models lead to generality in ecology?" In: *Trends in Ecology & Evolution* 28.10 (2013), pp. 578–583. ISSN: 01695347. DOI: `10.1016/j.tree.2013.05.022`.

[29] Roland Ewald et al. "Flexible experimentation in the modeling and simulation framework JAMES II - implications for computational systems biology". In: *Briefings in Bioinformatics* 11 (2010), pp. 290–300. DOI: `10.1093/bib/bbp067`.

[30] James R. Faeder et al. "Rule-based modeling of biochemical networks". In: *Complexity* 10.4 (2005), pp. 22–41. ISSN: 1076-2787, 1099-0526. DOI: `10.1002/cplx.20074`.

[31] Daniele Fanelli. "Negative results are disappearing from most disciplines and countries". In: *Scientometrics* 90.3 (2012), pp. 891–904. ISSN: 0138-9130, 1588-2861. DOI: `10.1007/s11192-011-0494-7`.

[32] Matthias Felleisen. "On the expressive power of programming languages". In: *Science of Computer Programming* 17.1 (1991), pp. 35–75. ISSN: 01676423. DOI: `10.1016/0167-6423(91)90036-W`.

[33] Brian R Gaines. "General systems research: quo vadis?" In: *General Systems Yearbook* 24 (1979), p. 15.

[34] Volker Grimm et al. "A standard protocol for describing individual-based and agent-based models". In: *Ecol Model* 198 (2006), pp. 115–126. ISSN: 03043800. DOI: `10.1016/j.ecolmodel.2006.04.023`.

[35] Volker Grimm et al. "The ODD protocol: a review and first update". In: *Ecol Model* 221 (2010), pp. 2760–2768. ISSN: 03043800. DOI: `10.1016/j.ecolmodel.2010.08.019`.

[36] Volker Grimm et al. "The ODD Protocol for Describing Agent-Based and Other Simulation Models: A Second Update to Improve Clarity, Replication, and Structural Realism". In: *Journal of Artificial Societies and Social Simulation* 23.2 (2020), p. 7. ISSN: 1460-7425. DOI: `10.18564/jasss.4259`.

[37] Volker Grimm et al. "Towards better modelling and decision support: Documenting model development, testing, and analysis using TRACE". en. In: *Ecological Modelling* 280 (2014), pp. 129–139. ISSN: 03043800. DOI: `10.1016/j.ecolmodel.2014.01.018`.

[38] Paul Groth and Luc Moreau. *PROV-Overview – An Overview of the PROV Family of Documents*. Technical Report, World Wide Web Consortium. 2013.

[39] GuÐmundur J. GuÐmundsson. "THE COD AND THE COLD WAR". In: *Scandinavian Journal of History* 31.2 (2006), pp. 97–118. ISSN: 0346-8755, 1502-7716. DOI: `10.1080/03468750600604184`.

[40] Leif Gustafsson and Mikael Sternad. "Consistent micro, macro and state-based population modelling". In: *Mathematical Biosciences* 225.2 (2010), pp. 94–107. ISSN: 0025-5564. DOI: `10.1016/j.mbs.2010.02.003`.

[41] K Haase et al. "Diet of dominant demersal fish species in the Baltic Sea: Is flounder stealing benthic food from cod?" In: *Marine Ecology Progress Series* 645 (2020), pp. 159–170. ISSN: 0171-8630, 1616-1599. DOI: `10.3354/meps13360`.

[42] Stefanie Haase et al. "Validation approaches of a geolocation framework to reconstruct movements of demersal fish equipped with data storage tags in a stratified environment". In: *Fisheries Research* 237 (2021), p. 105884. ISSN: 01657836. DOI: `10.1016/j.fishres.2021.105884`.

[43] Garrett Hardin. "The Tragedy of the Commons: The population problem has no technical solution; it requires a fundamental extension in morality." In: *Science* 162.3859 (1968), pp. 1243–1248. ISSN: 0036-8075, 1095-9203. DOI: `10.1126/science.162.3859.1243`.

[44] D. Harel and B. Rumpe. "Meaningful modeling: what's the semantics of "semantics"?" In: *Computer* 37.10 (2004), pp. 64–72. ISSN: 0018-9162. DOI: `10.1109/MC.2004.172`.

[45] Jan Harff, Svante Björck, and Peer Hoth, eds. *The Baltic Sea Basin*. Central and Eastern European Development Studies (CEEDES. Heidelberg ; New York: Springer, 2011. 449 pp. ISBN: 978-3-642-17219-9.

[46] Ferdi L. Hellweger. "75 years since Monod: It is time to increase the complexity of our predictive ecosystem models (opinion". In: *Ecological Modelling* 346 (2017), pp. 77–87. ISSN: 03043800. DOI: `10.1016/j.ecolmodel.2016.12.001`.

[47] Tobias Helms et al. "Semantics and Efficient Simulation Algorithms of an Expressive Multilevel Modeling Language". In: *ACM Transactions on Modeling and Computer Simulation* 27.2 (2017), pp. 1–25. ISSN: 1049-3301, 1558-1195. DOI: `10.1145/2998499`.

[48] H.-H. Hinrichsen et al. "Spatio-temporal dynamics of cod nursery areas in the Baltic Sea". In: *Progress in Oceanography* 155 (2017), pp. 28–40. ISSN: 00796611. DOI: `10.1016/j.pocean.2017.05.007`.

[49] Peter L. Holtermann et al. "Deepwater dynamics and mixing processes during a major inflow event in the central Baltic Sea: DEEPWATER DYNAMICS DURING A MBI". In: *Journal of Geophysical Research: Oceans* 122.8 (2017), pp. 6648–6667. ISSN: 21699275. DOI: `10.1002/2017JC013050`.

[50] John E. Hopcroft and Jeffrey D. Ullman. *Introduction to automata theory, languages, and computation*. Addison-Wesley series in computer science. Reading, Mass: Addison-Wesley, 1979. 418 pp. ISBN: 978-0-201-02988-8.

[51] Peter Hornetz. "Spatio-temporal distribution and food intake of cod (*Gadus morhua*) along a depth gradient in the western Bornholm Sea". MA thesis. University of Hamburg, 2020.

[52] ICES. "Cod (*Gadus morhua*) in subdivisions 24-32, eastern Baltic stock (eastern Baltic Sea". In: (2021). Publisher: ICES. DOI: `10.17895/ICES.ADVICE.7745`.

[53] ICES. "Herring (Clupea harengus) in subdivisions 25-29 and 32, excluding the Gulf of Riga (central Baltic Sea)". In: (2021). Publisher: ICES. DOI: `10.17895/ICES.ADVICE.7767`.

[54] ICES. "Sprat (Sprattus sprattus) in subdivisions 22-32 (Baltic Sea)". In: (2021). Publisher: ICES. DOI: `10.17895/ICES.ADVICE.7867`.

[55] Arne N. Johanson and Wilhelm Hasselbring. "Effectiveness and efficiency of a domain-specific language for high-performance marine ecosystem simulation: a controlled experiment". In: *Empirical Software Engineering* 22.4 (2017), pp. 2206–2236. ISSN: 1382-3256, 1573-7616. DOI: `10.1007/s10664-016-9483-z`.

[56] Max Kleiber. *The fire of life: an introduction to animal energetics*. Wiley, 1961.

[57] Granino A. Korn and John V. Wait. *Digital continuous-system simulation*. Englewood Cliffs, N.J: Prentice-Hall, 1978. 212 pp. ISBN: 978-0-13-212274-0.

[58] Friedrich W. Köster et al. "Eastern Baltic cod recruitment revisited—dynamics and impacting factors". In: *ICES Journal of Marine Science* 74.1 (2017). Ed. by Howard Browman, pp. 3–19. ISSN: 1054-3139, 1095-9289. DOI: `10.1093/icesjms/fsw172`.

[59] Averill M Law. *Simulation modeling and analysis*. OCLC: 1200558671. Boston: McGraw-Hill, 2007. ISBN: 978-0-07-132169-3.

[60] Natalie Loick-Wilde et al. "Stratification, nitrogen fixation, and cyanobacterial bloom stage regulate the planktonic food web structure". In: *Global Change Biology* 25.3 (2019), pp. 794–810. ISSN: 1354-1013, 1365-2486. DOI: `10.1111/gcb.14546`.

[61] Alfred James Lotka. *Elements of Physical Biology*. Williams & Wilkins Company, 1925. 514 pp.

[62] Etienne Low-Décarie, Corey Chivers, and Monica Granados. "Rising complexity and falling explanatory power in ecology". In: *Frontiers in Ecology and the Environment* 12.7 (2014), pp. 412–418. ISSN: 1540-9295. DOI: `10.1890/130230`.

[63] Carsten Maus, Stefan Rybacki, and Adelinde M. Uhrmacher. "Rule-based multi-level modeling of cell biological systems". In: *BMC Syst Biol* 5 (2011), p. 166. DOI: `10.1186/1752-0509-5-166`.

[64] George A. Miller. "The magical number seven, plus or minus two: some limits on our capacity for processing information." In: *Psychological Review* 63.2 (19569), pp. 81–97. ISSN: 1939-1471, 0033-295X. DOI: `10.1037/h0043158`.

[65] Justin J Miller. "Graph Database Applications and Concepts with Neo4j". In: *Proceedings of the Southern Association for Information Systems Conference, Atlanta, GA, USA*. Vol. 2324. 2013.

[66] Marvin Minsky. *Models, Minds, Machines*. Proceedings IFIP Congress, pp 45-49. 1965.

[67] M. Mion et al. "Effect of fish length and nutritional condition on the fecundity of distressed Atlantic cod *Gadus morhua* from the Baltic Sea: POTENTIAL FECUNDITY OF BALTIC *G. MORHUA*". In: *Journal of Fish Biology* 92.4 (2018), pp. 1016–1034. ISSN: 00221112. DOI: `10.1111/jfb.13563`.

[68] Volker Mohrholz. "Major Baltic Inflow Statistics – Revised". In: *Frontiers in Marine Science* 5 (2018). ISSN: 2296-7745. DOI: `10.3389/fmars.2018.00384`.

[69] Luc Moreau et al. "The Open Provenance Model core specification (v1.1)". In: *Future Generation Computer Systems* 27.6 (2011), pp. 743–756. ISSN: 0167739X. DOI: `10.1016/j.future.2010.07.005`.

[70] Nicolas Mouquet et al. "REVIEW: Predictive ecology in a changing world". In: *Journal of Applied Ecology* 52.5 (2015). Ed. by Marc Cadotte, pp. 1293–1310. ISSN: 00218901. DOI: `10.1111/1365-2664.12482`.

[71] Maxwell L. Neal et al. "A Reappraisal of How to Build Modular, Reusable Models of Biological Systems". en. In: *PLoS Computational Biology* 10.10 (2014). Ed. by Ruth Nussinov, e1003849. ISSN: 1553-7358. DOI: `10.1371/journal.pcbi.1003849`.

[72] Stefan Neuenfeldt et al. "Feeding and growth of Atlantic cod (*Gadus morhua* L.) in the eastern Baltic Sea under environmental change". In: *ICES Journal of Marine Science* 77.2 (2020). Ed. by Henn Ojaveer, pp. 624–632. ISSN: 1054-3139, 1095-9289. DOI: `10.1093/icesjms/fsz224`.

[73] A. Nissling and L. Westin. "Egg mortality and hatching rate of Baltic cod (*Gadus morhua*) in different salinities". In: *Marine Biology* 111.1 (1991), pp. 29–32. DOI: `10.1007/BF01986341`.

[74] Denis Noble. "Claude Bernard, the first systems biologist, and the future of physiology: Systems biology and the future of physiology". In: *Experimental Physiology* 93.1 (2008), pp. 16–26. ISSN: 09580670. DOI: `10.1113/expphysiol.2007.038695`.

[75] Daniel Pauly. *Gasping Fish and Panting Squids: Oxygen, Temperature and the Growth of Water-Breathing Animals*. In collab. with Otto Kinne. Excellence in Ecology 22. Oldendorf/Luhe, 2010. ISBN: 978-3-946729-22-8.

[76] Danhua Peng et al. "Reusing simulation experiment specifications in developing models by successive composition - a case study of the Wnt/ß-catenin signaling pathway". In: *SIMULATION* 93.8 (2017), pp. 659–677. ISSN: 0037-5497, 1741-3133. DOI: `10.1177/0037549717704314`.

[77] Alan J. Perlis. "Special Feature: Epigrams on programming". In: *ACM SIGPLAN Notices* 17.9 (1982), pp. 7–13. ISSN: 0362-1340, 1558-1160. DOI: `10.1145/947955.1083808`.

[78] Maria E. Pierce et al. "Developing and validating a multi-level ecological model of eastern Baltic cod (*Gadus morhua*) in the Bornholm Basin - A case for domain-specific languages". In: *Ecological Modelling* 361 (2017), pp. 49–65. ISSN: 03043800. DOI: `10.1016/j.ecolmodel.2017.07.012`.

[79] É. E. Plagányi and D. S. Butterworth. “A critical look at the potential of Ecopath with ecosim to assist in practical fisheries management”. In: *African Journal of Marine Science* 26.1 (2004), pp. 261–287. ISSN: 1814-232X, 1814-2338. DOI: `10.2989/18142320409504061`.

[80] Benjamin Planque. “Projecting the future state of marine ecosystems, “la grande illusion”?” In: *ICES Journal of Marine Science: Journal du Conseil* 73.2 (2016), pp. 204–208. ISSN: 1054-3139, 1095-9289. DOI: `10.1093/icesjms/fsv155`.

[81] S. Plante, D. Chabot, and J.-D. Dutil. “Hypoxia tolerance in Atlantic cod”. In: *J Fish Biol* 53 (1998), pp. 1342–1356. DOI: `10.1111/j.1095-8649.1998.tb00253.x`.

[82] A. Poćwierz-Kotus et al. “Genetic differentiation of brackish water populations of cod *Gadus morhua* in the southern Baltic, inferred from genotyping using SNP-arrays”. In: *Marine Genomics* 19 (2015), pp. 17–22. ISSN: 18747787. DOI: `10.1016/j.margen.2014.05.010`.

[83] Stephen P Prisley and Michael J Mortimer. “A synthesis of literature on evaluation of models for policy applications, with implications for forest carbon accounting”. In: *Forest Ecology and Management* 198.1 (2004), pp. 89–103. DOI: `10.1016/j.foreco.2004.03.038`.

[84] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria, 1999. URL: `https://www.R-project.org`.

[85] M. C. Rojbek et al. “Forage fish quality: seasonal lipid dynamics of herring (Clupea harengus L.) and sprat (Sprattus sprattus L.) in the Baltic Sea”. In: *ICES Journal of Marine Science* 71.1 (2014), pp. 56–71. ISSN: 1054-3139, 1095-9289. DOI: `10.1093/icesjms/fst106`.

[86] Adelinde M. Uhrmacher Roland Ewald. “SESSL: A domain-specific language for simulation experiments”. In: *ACM Trans Model Comput Simul* 24 (2014), p. 11. DOI: `10.1145/2567895`.

[87] Andreas Ruscheinski and Adelinde Uhrmacher. “Provenance in modeling and simulation studies — Bridging gaps”. In: Winter Simulation Conference. IEEE, 2017, pp. 872–883. ISBN: 978-1-5386-3428-8. DOI: `10.1109/WSC.2017.8247839`.

[88] Andreas Ruscheinski et al. *Towards a PROV Ontology for Simulation Models.* In: Provenance and Annotation of Data and Processes, 09-10 Jul 2018, London, UK. Proceedings, published by Springer International Publishing, pp. 192-195. 2018.

[89] E. S. Russell. “Some theoretical Considerations on the "Overfishing" Problem”. In: *ICES Journal of Marine Science* 6.1 (1931), pp. 3–20. ISSN: 1054-3139, 1095-9289. DOI: `10.1093/icesjms/6.1.3`.

[90] Marie Plambech Ryberg et al. “Physiological condition of Eastern Baltic cod, *Gadus morhua*, infected with the parasitic nematode *Contracaecum osculatum*”. In: *Conservation Physiology* 8.1 (2020). Ed. by Steven Cooke, coaa093. ISSN: 2051-1434. DOI: `10.1093/conphys/coaa093`.

[91] R G Sargent. “Verification and validation of simulation models”. In: *Journal of Simulation* 7.1 (2013), pp. 12–24. ISSN: 1747-7778, 1747-7786. DOI: `10.1057/jos.2012.20`.

[92] Matthias Schaber, Hans-Harald Hinrichsen, and Joachim Gröger. “Seasonal changes in vertical distribution patterns of cod (*Gadus morhua*) in the Bornholm Basin, central Baltic Sea”. In: *Fish Oceanogr* 21 (2012), pp. 33–43.

[93] Fm Schade, P Weist, and U Krumme. “Evaluation of four stock discrimination methods to assign individuals from mixed-stock fisheries using genetically validated baseline samples”. In: *Marine Ecology Progress Series* 627 (2019), pp. 125–139. ISSN: 0171-8630, 1616-1599. DOI: `10.3354/meps13061`.

[94] M Schaefer. "Some aspects of the dynamics of populations important to the management of the commercial marine fisheries". In: *Bulletin of Mathematical Biology* 53.1 (1991), pp. 253–279. ISSN: 00928240. DOI: 10.1016/S0092-8240(05)80049-7.

[95] Milner B. Schaefer. "Some aspects of the dynamics of populations important to the management of the commercial marine fisheries". In: *Inter-American Tropical Tuna Commission Bulletin* 1.2 (1954), pp. 23–56. DOI: 10.1016/S0092-8240(05)80049-7.

[96] Thomas C. Schelling. "Models of Segregation". In: *The American Economic Review* 59.2 (1969), pp. 488–493. DOI: 10.1080/0022250x.1971.9989794.

[97] Stewart Schlesinger et al. "Terminology for model credibility". In: *SIMULATION* 32.3 (1979), pp. 103–104. ISSN: 0037-5497, 1741-3133. DOI: 10.1177/003754977903200304.

[98] Amelie Schmolke et al. "Ecological models supporting environmental decision making: a strategy for the future". In: *Trends in Ecology & Evolution* 25.8 (2010), pp. 479–486. ISSN: 01695347. DOI: 10.1016/j.tree.2010.05.001.

[99] A.C. Setyawan et al. "Baltic cod endohelminths reflect recent ecological changes". In: *Journal of Helminthology* 94 (2020), e155. ISSN: 0022-149X, 1475-2697. DOI: 10.1017/S0022149X20000176.

[100] Clifford Shaffer, Ranjit Randhawa, and John Tyson. "The Role of Composition and Aggregation in Modeling Macromolecular Regulatory Networks". In: IEEE, Dec. 2006, pp. 1628–1635. ISBN: 978-1-4244-0501-5 978-1-4244-0500-8. DOI: 10.1109/WSC.2006.322937.

[101] M Sokolova et al. "Spatial patterns in infection of cod *Gadus morhua* with the seal-associated liver worm Contracaecum osculatum from the Skagerrak to the central Baltic Sea". In: *Marine Ecology Progress Series* 606 (2018), pp. 105–118. ISSN: 0171-8630, 1616-1599. DOI: 10.3354/meps12773.

[102] Karl Johan Åström, Hilding Elmqvist, and Sven Erik Mattsson. "Evolution of Continuous-Time Modeling and Simulation". In: The 12th European Simulation Multiconference. ESM'98. Manchester, UK, 1998.

[103] Henrik Svedäng et al. "Compensatory Feeding in Eastern Baltic Cod (*Gadus morhua*): Recent Shifts in Otolith Growth and Nitrogen Content Suggest Unprecedented Metabolic Changes". In: *Frontiers in Marine Science* 7 (2020), p. 565. ISSN: 2296-7745. DOI: 10.3389/fmars.2020.00565.

[104] *system, n.* In: *OED Online*. Oxford University Press, 2021.

[105] Claudia Szabo and Yong Meng Teo. "An Approach for Validation of Semantic Composability in Simulation Models". In: IEEE, 2009, pp. 3–10. ISBN: 978-0-7695-3713-9. DOI: 10.1109/PADS.2009.14.

[106] ICES (2019). Benchmark Workshop on Baltic Cod Stocks (WKBALTCOD2). ICES Scientific Reports. 1:9. 310 pp. DOI: 10.17895/ices.pub.4984.

[107] Barbara Jobstmann1 Thomas A. Henzinger and Verena Wolf. "Formalisms for Specifying Markovian Population Models". In: *Reachability problems: third international workshop, RP 2009, Palaiseau, France, September 23-25, 2009: proceedings*. Ed. by Olivier Bournez and Igor Potapov. Lecture notes in computer science 5797. Springer, 2009. ISBN: 978-3-642-04419-9.

[108] Vito Volterra. "Fluctuations in the Abundance of a Species considered Mathematically". In: *Nature* 118.2972 (1926), pp. 558–560. ISSN: 0028-0836, 1476-4687. DOI: 10.1038/118558a0.

[109] Uwe Waller and Dietrich Schnack. "Development of Baltic cod eggs at different levels of temperature and oxygen content". In: *Dana* 10 (1994), pp. 163–177.

[110] Tom Warnke. “Domain-specific languages for modeling and simulation”. Dissertation. Rostock: University of Rostock, 2020.

[111] Tom Warnke and Adelinde M. Uhrmacher. “Practical expressiveness of internal and external domain-specific modeling languages”. In: *2017 Winter Simulation Conference (WSC*. 2017 Winter Simulation Conference (WSC. Las Vegas, NV: IEEE, 2017, pp. 4566–4567. ISBN: 978-1-5386-3428-8. DOI: `10.1109/WSC.2017.8248207`.

[112] Frans Willekens. “Continuous-time Microsimulation in Longitudinal Analysis”. In: *New Frontiers in Microsimulation Modelling*. Ed. by Asghar Zaidi, Ann Harding, and Paul Williamson. 1st ed. Ashgate: Routledge, 2009, pp. 413–436. ISBN: 978-1-315-24806-6. DOI: `10.4324/9781315248066`.

[113] Pia Wilsdorf, Fiete Haack, and Adelinde M. Uhrmacher. “Conceptual Models in Simulation Studies: Making it Explicit”. In: *2020 Winter Simulation Conference (WSC)*. 2020 Winter Simulation Conference (WSC). Orlando, FL, USA: IEEE, 2020, pp. 2353–2364. ISBN: 978-1-72819-499-8. DOI: `10.1109/WSC48552.2020.9383984`.

[114] Pia Wilsdorf et al. “Automatic Reuse, Adaption, and Execution of Simulation Experiments via Provenance Patterns”. In: *arXiv:2109.06776 [cs]* (2021).

[115] Eric B. Winsberg. *Science in the age of computer simulation*. Chicago: The University of Chicago Press, 2010. 152 pp. ISBN: 978-0-226-90202-9 978-0-226-90204-3.

[116] Bernard P Zeigler. *Multifacetted modelling and discrete event simulation*. Academic Press Professional, Inc., 1984. ISBN: 978-0127784502.

[117] Bernard P. Zeigler. *Theory of modelling and simulation*. New York: Wiley, 1976. 435 pp. ISBN: 978-0-471-98152-7.

[118] Bernard P. Zeigler, Alexandre Muzy, and Ernesto Kofman. *Theory of modeling and simulation: discrete event and iterative system computational foundations*. Third edition. OCLC: on1020030023. London San Diego Cambridge, MA Oxford, UK: Academic Press, an imprint of Elsevier, 2019. ISBN: 978-0-12-813370-5.

[119] S Zuo et al. “Host size-dependent anisakid infection in Baltic cod *Gadus morhua* associated with differential food preferences”. In: *Diseases of Aquatic Organisms* 120.1 (2016), pp. 69–75. ISSN: 0177-5103, 1616-1580. DOI: `10.3354/dao03002`.

List of Figures

Index

Abbreviations and Acronyms

CAS	Complex Adaptive System
COMBINE	COmputational Modeling in BIology NEtwork
EBC	Eastern Baltic Cod
EC	European Commission
EU	European Union
F	Fishing mortality
IBM	Individual Based Model
ICES	International Council for the Exploration of the Sea
MBI	Major Baltic Inflow
MIASE	Minimum Information About a Simulation Experiment
MIRIAM	Minimal Information Required In the Annotation of Models
MSY	Maximum Sustainable Yield
ODD	Overview, Design concepts and Details
ODE	Ordinary Differential Equation
OPM	Open Provenance Model
ProvDM	Provenance Data Model
SSB	Spawning stock biomass
TAC	Total Allowable Catch
TRACE	TRansparent And Comprehensive model Evaluation
WBC	Western Baltic Cod