

Konsolidering af planktondata i NOVANA

Indlæsning af mesozooplankton

Fagligt notat fra DCE - Nationalt Center for Miljø og Energi

Dato: 18. December 2023 | 68



AARHUS
UNIVERSITET

DCE - NATIONALT CENTER FOR MILJØ OG ENERGI

Datablad

Fagligt notat fra DCE – Nationalt Center for Miljø og Energi

Kategori: Rådgivningsnotat

Titel: Konsolidering af planktondata i NOVANA
Undertitel: Indlæsning af mesozooplankton

Forfattere: Hans Jakobsen og David Rytter
Institution: Aarhus Universitet, Institut for Ecoscience

Faglig kommentering: Per Andersen
Kvalitetssikring, DCE: Lars Moeslund Svendsen
Sproglig kvalitetssikring: Anne Mette Poulsen

Ekstern kommentering: Miljøstyrelsen. Kommentarerne findes her:
http://dce2.au.dk/pub/komm/N2023_68_komm.pdf

Rekvirent: Miljøstyrelsen

Bedes citeret: Jakobsen, H, Rytter, D. 2023. Konsolidering af planktondata i NOVANA. Indlæsning af mesozooplankton Aarhus Universitet, DCE – Nationalt Center for Miljø og Energi, 20 s. -- Fagligt notat nr. 2023 | 68

Gengivelse tilladt med tydelig kildeangivelse

Foto forside: Jakobsen HH: Larve af børsteormen *Polydora ciliate*

Sideantal: 20

Indhold

Sammenfatning	4
Summary	5
1 Introduktion	6
2 Metoder	8
Metodebeskrivelse	8
2.1 Aggregerede historiske data i ODA	8
2.2 Berigelse af ODA med individmålinger	10
Metodebeskrivelse	10
Beregning	10
2.3 Genberegning af individbiomasser	12
3 Afsluttende bemærkninger og konklusion	13
3.1 Aggregering af data	13
4 Referencer	14
5 Bilag	15
Bilag B1 Tabel over stationer, der er samlet og struktureret i ODA	15
Bilag B2 QA_Dublet_klassifikation_Zoo	17
Bilag B3 QA_Dublet_Classification_Zoo_19950601	18

Sammenfatning

Dette statusnotat beskriver arbejdet med at samle, overføre og påbegynde kvalitetssikring af mesozooplanktondata, som er blevet indsamlet i perioden 1973 til 2021 i forbindelse med de nationale overvågningsprogrammer. Der er i alt identificeret 304588 unikke artregistreringer, fordelt på 78 stationer. Den længste tidsserie stammer fra Århus Bugt, hvor der er data fra 33 år, stationen i Løgstør Bredning (VIB3708) har den længste tidsserie med fulde år. Ikke alle år er dækket siden 1973. Det endelige datasæt er blevet samlet i databasen ODA under DCE, hvor der yderligere er tilføjet individmålinger, som er indsamlet af laboratoriet Fishlab siden 2016 fra 776 prøver på 18 stationer. Disse målinger vil blive anvendt i beregningerne af art- og stadiespecifikke biomasser for hele det historiske datasæt af mesozooplankton på tværs af alle indsamlede prøver. Dette vil resultere i udarbejdelsen af et samlet, homogent nationalt mesozooplankton datasæt.

Summary

This status report describes the work of collecting, transferring and initiating quality assurance of mesozooplankton data collected from 1973 to 2021 as part of national monitoring programs. A total of 304,588 unique species records have been identified, distributed across 78 stations. The longest time series is from Aarhus Bay, with data spanning 33 years, and the station in Løgstør Bredning (VIB3708) holds the longest continuous yearly time series. Not all years are covered since 1973. The final dataset has been compiled in the ODA database under DCE, with additional individual measurements collected by the Fishlab laboratory since 2016 from 776 samples at 18 stations. These measurements will be used in calculations of species- and stage-specific biomasses across the entire historical dataset of mesozooplankton from all collected samples. This will result in the creation of a comprehensive, homogeneous national mesozooplankton dataset.

1 Introduktion

DCE har indgået en kontrakt med Miljøstyrelsen (MST) vedrørende indlæsning af marine mesozooplanktondata (herefter benævnt "planktondata") til ODA. Udgangspunktet var en opgavebeskrivelse fra Miljøstyrelsen fra den 2. oktober 2023 vedrørende "Konsolidering af mesozooplankton – 2023". I denne bestilling fremgår følgende:

"I Danmark udfører Miljøstyrelsen (MST) miljøovervågning i overensstemmelse med EU's havstrategidirektiv. Overvågningen har til formål at monitorere miljøtilstanden i de områder, der er omfattet af direktivet. Direktivet er opdelt i 11 såkaldte deskriptorer, der dækkes af forskellige overvågningsaktiviteter. Den danske havstrategiovervågning er beskrevet i Danmarks Havstrategi II, Overvågningsprogram. Direktivets deskriptor 1 – Biodiversitet, Pelagiske habitater – omhandler overvågning af plankton, og deskriptorens kriterium 6 beskriver, at forekomsten af plankton følger langtidsgennemsnittet. Formålet med projektet var at sikre historiske planktondata, der endnu ikke var blevet overført til fagsystemet STOQ og/eller DCE's overfladevandsdatabase ODA. Herudover var formålet at harmonisere det samlede datasæt og indlæse det i ODA. Dette omfattede: 1) indlæsning af individuelle målinger af vandloppe-dimensioner, 2) strukturering af data fra tidligere WSP-databaser, 3) identifikation af forskellige metoder i STOQ, der er blevet brugt til at beregne volumen og biomasse for de enkelte zooplanktonarter for til sidst at skabe en harmonisk sammenhæng for hele dataperioden. Det endelige produkt skal derefter være klar til MST og FDC-kvalitetssikring, som beskrevet i Jakobsen m.fl. (2023)"

Før nærværende projekt blev der i 2021 gennemført et andet projekt for MST, der primært omfattede opgaver i relation til at udvikle funktionaliteter for plankton i forhold til VanDa-databasen (fælles offentlig overfladevandsdatabase, der hostes af Danmarks Miljøportal) og at skaffe og modtage data fra historiske databaser, blandt andet fra rådgivningsfirmaet WSP.

Planktondata er blevet indsamlet under NOVANA (Det Nationale Overvågningsprogram for Vandmiljø og Natur) og fra tidligere overvågningsprogrammer gennemført af de tidligere amter, som har været gemt i flere fagsystemer. En del planktondata fandtes i fagsystemet STOQ, som lå hos Danmarks Miljøportal (DMP), men ud over dette findes der planktondata i andre databaser: tre versioner af fagsystemet Planktonsys, Fishlab ved Kirsten Engell-Sørensen.

Mesozooplankton blev rapporteret i den årlige NOVANA-rapport indtil 2011 for de marine områder (Hansen 2012). I årene derefter blev mikrozooplankton stadige rapporteret i den årlige NOVANA-rapport for de marine områder. Efterfølgende har der ikke været rapportering af mesozooplankton, da der blev konstateret fejl i kodningen i STOQ, og data er ikke blevet overført til ODA siden 2015. Dette har to konsekvenser:

1. Tidligere NOVANA-rapporteringer for de marine områder indeholder muligvis fejl.
2. Det har ikke været muligt at få adgang til data siden 2015, da overførslen mellem STOQ og ODA har været stoppet.

Der er derfor usikkerhed vedrørende omfanget af den tidligere faglige kvalitetssikring, da DCE ikke har kunnet finde en metodebeskrivelse for den udførte kvalitetssikring, der dækker perioden fra de tidligere amter. Da mesozooplankton ikke har været rapporteret i forbindelse med den årlige havrapport, er der derfor heller ikke blevet udført FDC-faglig kvalitetssikring siden da. Endelig er tidligere kvalitetssikringsmærker (som angiver, hvilken type kvalitetssikring der var udført) blevet slettet i forbindelse med rettelsen af kodningsfejlen i STOQ.

Projektet i 2023 fokuserer på at klargøre mesozooplanktondata til yderligere behandling og omfatter tre trin:

1. Overførsel af individuelle vandloppedata til ODA: Indlæsning af individuelle målinger fra vandlopper foretaget af Fishlab siden 2016, som aldrig tidligere var blevet overført til ODA. Dette skyldtes, at STOQ-systemet ikke var tilpasset til at modtage disse data.
2. Samkøring af data fra tidligere WSP-databaser (ca. 1973-2012) til ODA: Forberedelse af WSP-databaserne og import af data til den samme grundlæggende struktur i ODA, som der tidligere er anvendt for planktondata. Dette inkluderer grundig import og kvalitetssikring, herunder tilpasning af parametre og enheder. Efterfølgende skal eventuelle duplikater behandles efter samme koncept, som blev anvendt til fyto- og mikrozooplanktondata i det tidligere projekt fra 2022, som beskrevet i Jakobsen m.fl. (2023).
3. Fejlretning og genberegning af data: Gennemgang af data for fejl og implementering af nødvendige rettelser. I 2016 blev der introduceret en ny metode til opgørelse af mesozooplankton som en del af revisionen af den tekniske anvisning TA11 (Jakobsen & Møller 2015). Aarhus Universitet (AU) skal identificere forskellige metoder, der er blevet anvendt til at beregne volumen og biomasse for de enkelte arter. Der oprettes en harmonisk sammenhæng for hele dataperioden, inklusive vurdering af de anvendte "længde:vægt regressioner". Hvis nødvendigt, skal der foretages genberegninger baseret på faglige skøn.

Det endelige datasæt afstemmes og klargøres til MST- og FDC-kvalitetssikring. Projektet blev igangsat den 22. august 2023 efter MST's endelige godkendelse og blev afsluttet den 31. december 2023 med et statusnotat primo december 2023 og et afsluttende møde med Miljøstyrelsen medio december 2023.

2 Metoder

Metodebeskrivelse

Konsolidering og klargøring til en mulig fremtidig MST samt FDC-kvalitets-sikring omfatter tre trin:

1. Aggregering af datakilder til en fælles datastruktur samt identificering af mulige dubletter.
2. Berigelse af ODA med individmålinger af vandlopper og andre meso-zooplanktonarter med data fra Fishlab.
3. Genberegning af individbiomasser, hvor det skønnes nødvendigt.

2.1 Aggregerede historiske data i ODA

Planktondata fra tre versioner af Planktonsys-databaser (fra WSP, det tidligere ORBICON/Bio Consult, Ringkøbing og Aarhus amter) blev tilføjet ODA's eksisterende zooplanktondata stammende fra Miljøstyrelsens (MST) STOQ-fagsystem. Databasestrukturer på tværs af datakilderne (se tabel 1.1) blev afklaret gennem et omfattende udredningsarbejde, så data kunne indlægges i en fælles datastruktur i ODA efter samme principper som i Jakobsen m.fl. (2023).

Miljøbiologisk Laboratoriums dimensionsdata, liggende i 776 Excel-filer fra FishLab, blev tilføjet eksisterende STOQ-prøver, hvor disse manglede. I den oprindelige aftale med MST var det forudsætningen for dette udredningsarbejde, at datastrukturen var ens på tværs af datakilder. Data i STOQ foreligger i et fast struktureret format, der ikke har ændret sig siden ibrugtagningen. Derimod er de øvrige datakilder ikke løbende vedligeholdt, og datastrukturen er forskellig mellem disse. Harmoniseringen af disse data viste sig at være meget ressourcekrævende, da struktur, koder og enheder har været afvigende i de individuelle datakilder angivet i tabel 1.1, og Excel-filerne var knap så strukturerede som forventet.

Tabel 1.1. Oversigt over antallet af zooplankton prøver, som AU modtog fra de forskellige kilder.

Alias	Beskrivelse	Antal prøver
MBLZ	Miljøbiologisk Laboratorium	150
OCZ	ORBICON	5.577
RKBZ	Ringkøbing Amt	3.263
AAZ	Aarhus Amt	741
STOQZ	STOQ zooplankton	3.675

Slutproduktet efter dataaggregeringen blev en generel strukturering af alle data til en fælles ODA-databasestruktur. Det betyder, at information såsom stationer, prøver, enheder og koder er blevet harmoniseret til en fælles ODA-standard.

Der viste sig at være et stort overlap mellem planktondata på tværs af de databaser, som DCE modtog kopier af, samt en række forkerte stationsangivelser. Dette resulterede i en betydelig mængde redundante data (dubletter), som skulle udredes, før data kunne bringes i en prækvalitetssikringstilstand i ODA.

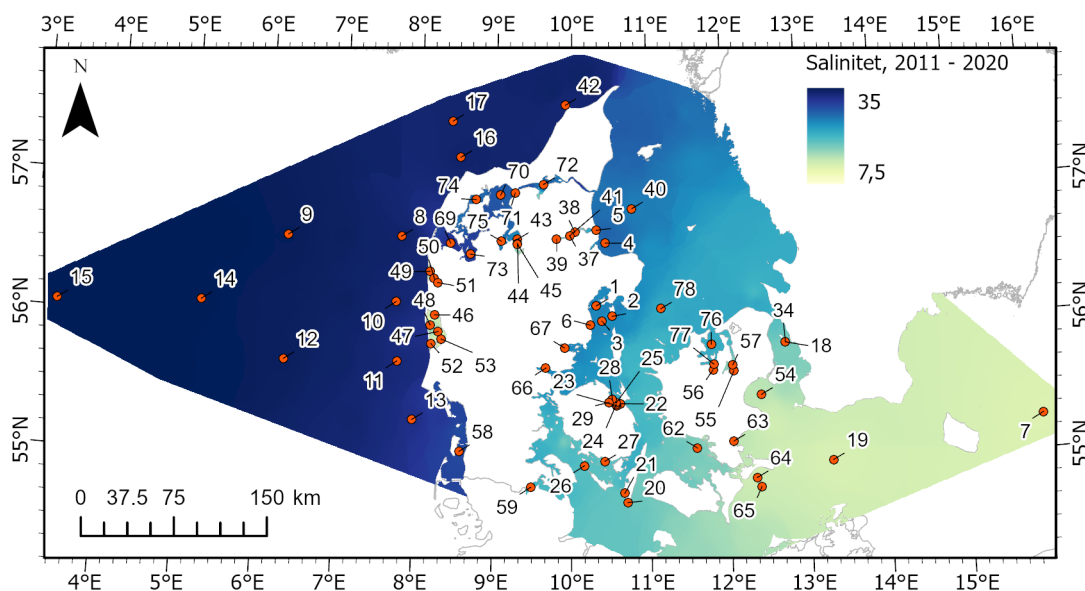
Dublethåndtering

Efter overførelse af alle data i ODA blev der genereret en liste med prøve-id og dato. Der, hvor der var flere prøver på en dato, var disse potentielle dubletter. Station og dybde, som ellers ville være naturlig at medtage i afklaringen af potentielle kombinationer af dubletter, blev ikke medtaget, da det har vist sig, at der er dubletter på tværs af stationer og dybder.

På potentielle dubletter blev nøgledata samlet såsom ID'er, antal arter, antal dimensioner, artsprofiler, dimensionsprofiler, cellekoncentration, cellevolumen og eventuelle andre parametre. For at sikre sammenlignelighed blev der foretaget en generel konvertering af arter i artsprofilen, da artskoder kan være konverteret forskelligt på tværs af datakilderne (se tabel 1.1). En liste med kombinationer af potentielle dubletter blev genereret ved at sammenligne parvist. Dermed kunne der identificeres omkring 30.000 potentielle dubletprøvepar.

Dette blev efterfulgt af en sammenligning af prøveprofilerne for de potentielle dubletter baseret på station, prøvetager, analytiker, tidspunkt, dybder, antal arter, antal dimensioner, artsprofiler, dimensionsprofiler, parametre og summeringer. Resultatet var en dublet-score (se forklaring i bilag 3). På baggrund af denne score og en manuel vurdering blev der givet et bud på dubletter station for station (76 stationskombinationer).

Bilag 2 viser eksempler på anvendelse af dubletaflklaringslisten, Bilag 3 viser et specifikt eksempel på dubletaflklaring på en given dato med underliggende data.



Figur 1.1 Kort over stationer, hvor data er tjekket for dubletter. Stationerne er angivet med nummer og refererer til tabellen i bilag 1.

Følgende prioriteringsrækkefølge anvendes:

1. Unikke prøver fik tildelt kvalitetsflaget "elektronisk godkendt" i ODA.
2. Tilsyneladende dubletter, der med sikkerhed identificeredes automatisk som unikke, fik alle kvalitetsflaget "elektronisk godkendt".

3. Dublet, der med sikkerhed kunne identificeres automatisk som kildeprøve, fik kvalitetsflaget '*elektronisk godkendt*', resten i dubletgruppen fik kvalitetsflaget "*elektronisk forkastet*" (omkring 30 % af data blev identificeret som dubletter), men opbevares fortsat i ODA.

Data blev hermed rensset for dubletter og findes i struktureret form i ODA og vil på et senere tidspunkt kunne gennemgå MST-faglig og FDC-faglig kvalitetssikring.

I alt blev der godkendt 190057 artsregistreringer fordelt på 78 stationer inden for Danmarks territorialfarvand (*figur 1.1.*)

2.2 Berigelse af ODA med individmålinger

Metodebeskrivelse

For HELCOM-området anvendes en ældre reference, Hernroth (1985), samt HELCOM COMBINE-manualen (HELCOM 2022) til beregning af mesozooplankton-biomasser. Individuelle vådvægte af forskellige taksonomiske grupper og udviklingsstadier tilføjes efter Hernroth (1985). I fravær af specifikke værdier kan en standardfaktor for et bestemt takson/udviklingsstadium anvendes, og længdemålinger af den pågældende organisme kan bruges til at beregne individuel masse i henhold til metoden Standard Size Classes (SSC) (Witek m.fl. 1966). I NOVANA-programmet er mesozooplankton blevet opmålt siden 2016 for at fastlægge en national metode til SSC.

HELCOM COMBINE-manualen (HELCOM 2022) beskriver, at dokumentationen for individuel vægtberegning bør være tilgængelig på anmodning til interkalibrering mellem laboratorier. Metoder indført i 2016 til inddeling af SSC muliggør desuden længde:vægt-regressioner til bestemmelse af individuelt kulstofindhold, men er aldrig blevet endeligt implementeret i det tidligere STOQ-system.

Længde:vægt-regression undersøger relationen mellem længden og individets vægt. For at gennemføre dette indlæses alle længdemålinger, der er indsamlet af konsulenterne, i ODA. Opmålinger af enkeltindivider er foretaget på en underprøve under oparbejdningen af mesozooplanktonprøver efter anvisningerne i de tekniske anvisninger for mesozooplankton (TA11, Jakobsen & Møller 2016).

Beregning

Når der arbejdes med mesozooplankton i NOVANA-programmet, rapporteres længden i mm, mens det følgende er baseret på længder angivet i μm . Resultatet af beregningerne kan enten være vådvægt, askefri tørvægt eller kulstofbiomasse (μg). I NOVANA rapporteres vægten som μg kulstof per individ.

Individuelle biomasser for mesozooplankton stiger ikke lineært i forhold til længden (L), hvilket betyder, at individuel biomasse (y) ikke kan beskrives lineært som:

$$y = b \times L^a \qquad \text{formel 1}$$

hvor a og b er konstanter, som bestemmes eksperimentelt og findes i litteraturen. Konstanterne a og b findes i VanDa-artsarkivet som henholdsvis ka og kb . Formlen er identisk med artsarkivets formel 19.

Når $b \neq 1$, repræsenterer b en ikke-lineær proportionalitet i datasættet, der spænder fra mindste til største længde. For at bestemme de deskriptive konstanter (a og b) eksperimentelt logaritmetransformeres længden (x) og biomassen (y) parvis (normalt ved at tage base 10-logaritmen) ud fra et datasæt med enkeltmålinger af længder med de tilhørende individbiomasser. Logaritmetransformationen ændrer forholdet mellem datapunkterne og gør det lettere at observere mønstre og sammenhænge mellem biomasse og dyrenes længder, især når dataværdierne spænder over flere størrelsesordener.

De logaritmisk transformerede data følger derfor

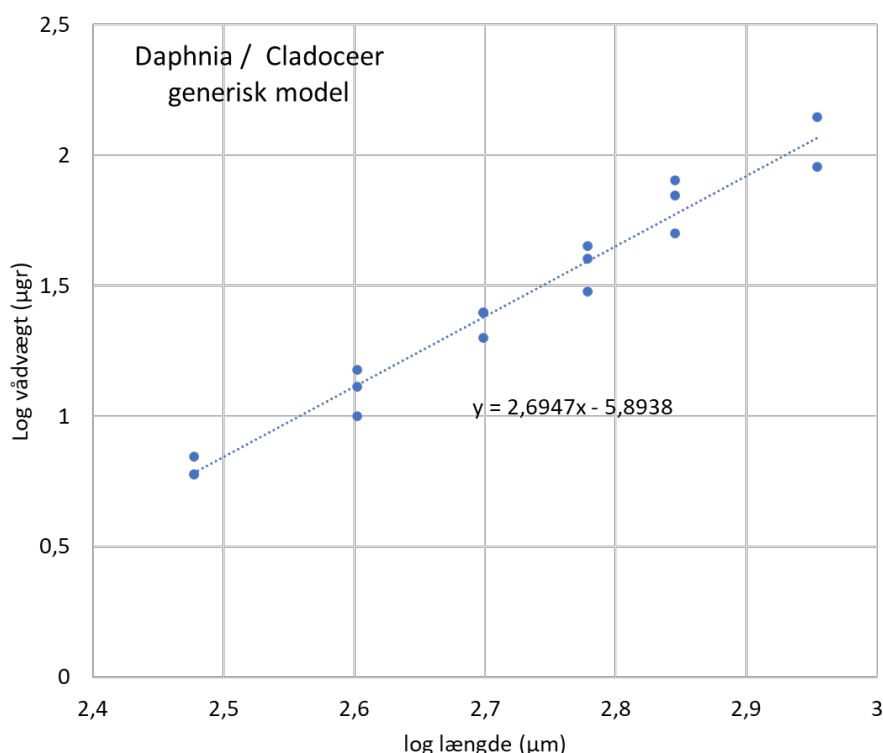
$$\log(y) = \log(b) + a \times \log(L) \quad \text{formel 2}$$

Det følger heraf, at når logaritmen fjernes på begge sider af lighedstegnet, er

$$y = 10^b \times L^a \quad \text{formel 3}$$

Hældningen er derfor uændret, og skæringen med y-aksen (biomassen) bestemmes som antilogaritmen til b . Eksemplet i figur 1.2 illustrerer forholdet mellem længden og vådvægten for alle dafnier, som blev præsenteret i Herthroths studie (1985). Det betyder, at b er afgørende for skaleringen af y-aksen, dvs. afgørende for enheden af biomassen y .

Figur 1.2 Data fra alle marine dafnier fra Herthroth (1985).



De nødvendige eksperimentelle, artsspecifikke konstanter er identificeret i litteraturen og findes i VanDa-arkivet. Værdierne i artarkivet er baseret på, at længden af dyrene (L) er givet i μm . Derfor er det kun nødvendigt at kende dyrets længde.

I TA11 (Jakobsen & Møller 2016) angives længden af rygskjoldet, men andre steder i litteraturen anvendes en mere generisk metoder, hvor hele dyrets længde benyttes i beregningen af biomassen. Konstanterne i artsarkivet findes i Mauchline (1998) og bruges til at beregne individuel tørvægt.

Eksempler

Allometrisk beregning

Baseret på figur 1.2 kan den eksperimentelle vådvægt (VV i μg) for dafnier beregnes som $VV=10^{-5,8938} \times \text{længden } (\mu\text{m})^{2,6947}$. Vær opmærksom på fortegnet af a ! Dette betyder, at vådvægten for en dafnie på $500 \mu\text{m}$ kan beregnes som $1,3 \times 10^{-6} \times 500^{2,6947} = 24 \mu\text{g}$. Ved omregning fra vådvægt til tørvægt multipliceres ifølge Kiørboe (2013) med 0,16 eller omregningsfaktoren 0,0995, hvis biomassen omregnes fra vådvægt til kulstof.

I eksemplet med en dafnie med en opmålt længde på $500 \mu\text{m}$, hvor vådvægten er $24 \mu\text{g}$, vil tørvægten derfor være $3,89 \mu\text{g}$ eller $2,39 \mu\text{g}$ kulstof.

I ovenstående eksempel regnes der i vådvægt ud fra de oprindelige data, og det er vigtigt at have fokus på enhederne. I artslisten er a og b angivet, således at omregningen altid sker fra længdemålinger (μm) til kulstof *per* individ ($\mu\text{g C}$).

Rotationsellipsoide

En polychete-larve er opmålt til $250 \mu\text{m}$ og har en diameter på $75 \mu\text{m}$. Det betyder, at volumen, dvs. vådvægten, er $= \pi \cdot (4/3) \cdot 250 \cdot 75^2 = 5,86 \times 10^6 \mu\text{m}^3$ eller $5,86 \mu\text{g VV}$. Ved omregning fra vådvægt til tørvægt multipliceres ifølge Kiørboe (2013) med 0,134 eller omregningsfaktoren 0,052, hvis biomassen omregnes fra vådvægt til kulstof.

Bemærk, at omregningen fra μm^3 til vådvægt μg^3 skal multipliceres med 10^{-6} for at omregne til μg vådvægt.

Tablet 1.2 Relevante omregningsfaktorer fra vådvægt til eksempelvis kulstof findes i reviewet af Kiørboe (2013), der præsenterer en række faste konstanter til omregning.

Grupper	%tørvægt af vådvægt	%C af vådvægt
Ctenophorer	4,0	0,26
Tunicater	5,4	0,72
Cnidarier	4,1	0,48
Chaetognather	9,3	3,6
Polycheter	13,4	5,2
Pteropoder	23	5,3
Copepoder	16,2	10
Euphasider	22,8	10,7
Amphipoder	23,9	8,4
Crusteceaer	18,3	9,6
Protozoer	-	15,4

2.3 Genberegning af individbiomasser

Data for alle prøver og individuelle opmålinger af arter er samlet i ODA, og DCE ved Ecoscience er ved at revidere artskodelistens omregningskonstanter (ka og kb) til individuel kulstofbiomasse. Dette arbejde påbegyndes efter nytår og forventes afsluttet inden udgangen af januar, hvorefter den endelige kvalitetssikring vil blive gennemført, og derefter kan overførslen til VanDa igangsættes.

3 Afsluttende bemærkninger og konklusion

3.1 Aggregering af data

Metoden til beregning af individbiomasser er beskrevet, og der er opbygget et bibliotek af artsspecifikke regressionskonstanter, der delvist er indlagt i artsregisteret. Fremadrettet mangler DCE endeligt at beregne og kvalitetssikre de individuelle arts- og stadiobiomasser og udarbejde tabeller til at bagudberegne biomasser i tid.

Overførsel af individuelle mesozooplankton-opmålinger

Fishlab-data med individmålinger fra 776 Excel-filer er blevet overført til ODA og fusioneret med eksisterende prøver fra STOQ. Udredningen var mere omfattende end forventet, men data er nu tilgængelige i ODA, og mesozooplankton-individbiomasser kan beregnes der. Der vil i begyndelsen af 2024 blive udviklet tabeller med zooplanktonbiomasse på art- og stadieniveau, som kan anvendes til at beregne biomasser bagud i tid på hele NOVANA-datasættet.

Samkøring af data fra tidligere databaser (ca. 1973-2012) med ODA (1991-2021)

Der er foretaget samkøring af data fra tidligere WSP-databaser (data fra ca. 1973-2012) med ODA (data fra 1991-2021). Et endeligt datasæt, der dækker perioden fra 1973-2021, er blevet oprettet og vil blive videreført til den endelige kvalitetssikring af zooplanktondata. Den længste tidsserie er Aarhus Bugt, hvor de ældste data er fra 1973. Det endelige datasæt indeholder i alt 304.588 datalinjer, hvor hver linje indeholder en unik artsregistrering, fordelt på i alt 78 stationer.

Fejlretning og genberegning af data

Der vil blive udført fejlretning og genberegning af data. Metoden til beregning af individbiomasser er beskrevet, og der er opbygget et bibliotek af artsspecifikke regressionskonstanter, der delvist er indlagt i artsregisteret. Fremadrettet mangler DCE stadig at beregne og kvalitetssikre de individuelle arts- og stadiobiomasser samt at udarbejde tabeller til at beregne biomassen bagud i tid.

4 Referencer

Hansen JW (2012) Marine områder 2011 NOVANA. In: Hansen JW (ed) Videnskabelig rapport fra DCE - Nationalt Center for Miljø og Energi, Book 34. Aarhus Universitet, Aarhus, <https://dce2.au.dk/pub/SR34.pdf>

HELCOM (2022) Annex C: 7 Guidelines for monitoring of mesozooplankton (2021). Manual for marine monitoring in the COMBINE program of HELCOM <https://www.helcom.fi/wp-content/uploads/2019/08/Guidelines-for-monitoring-of-mesozooplankton.pdf>

Hernroth L (1985) Recommendations on methods for marine biological studies in the Baltic Sea. In: Hernroth L (ed) mesozooplankton biomass assessment. Institute of Marine Research.

Jakobsen HH, Rytter D, Andersen P, Holbach A (2023) Konsolidering af planktondata i NOVANA: Fytoplanktondata. Aarhus Universitet, DCE - Nationalt Center for Miljø og Energi nr. 287.

Jakobsen HH, Møller EF (2016) TA11 Mesozooplankton. Tekniske anvisninger. Aarhus Universitet, DCE - Nationalt Center for Miljø og Energi.

Kjørboe T (2013) Zooplankton body composition. *Limnol Oceanogr* 58:1843-1850.

Mauchline J (1998) Biology of calanoid copepods, Vol 33. Academic Press, San Diego London Boston New York Sydney Tokyo Toronto

Witek ZG, Breul M, Wolska-Pys P, Gruszaka A, Krajewska-Soltys A, Ejsynont L, Sujak D (1996) Comparison of different methods of Baltic zooplankton biomass estimations. Proc Proceedings of the XII BMB Symposjum. Institute of Aquatic Ecology, University of Latvia

5 Bilag

Bilag B1 Tabel over stationer, der er samlet og struktureret i ODA

Station angiver stationsnavn, antal prøver er antallet af unikke prøver, der indgår i stationen, og antal år er antallet af årstal, der er registreret på stationen med første og sidste år angivet i kolonnen "Periode".

Tabel B.1 Oversigt over station, position, prøveperiodens længde (år), antal prøver samt årrække (periode). Stationer markeret med *mangler korrekte positioner.

Nummer	Station	Længdegrad	Breddegrad	År	Antal prøver	Periode
1	ARH170006	10,32	56,15	33	718	1978-2021
2	ARH170115	10,52	56,08	2	11	2002-2003
3	ARH170117	10,39	56,04	2	11	2002-2003
4	ARH190004	10,44	56,61	1	18	1998-1998
5	ARH240009	10,32	56,70	1	2	2000-2000
6	ARH250032	10,24	56,02	1	16	2006-2006
7	BRKBMPK2	15,98	55,25	18	72	1999-2021
8	DMU1023	7,77	56,64	4	5	2018-2021
9	DMU1026	6,29	56,62	4	5	2018-2021
10	DMU1035	7,72	56,17	3	4	2018-2021
11	DMU1042	7,76	55,74	3	4	2018-2021
12	DMU1046	6,30	55,72	3	4	2018-2021
13	DMU1052	7,97	55,32	3	4	2018-2021
14	DMU1075	5,20	56,12	3	5	2018-2020
15	DMU1077	3,34	56,05	3	5	2018-2020
16	DMU1130	8,53	57,22	4	5	2018-2021
17	DMU1133	8,41	57,48	4	6	2018-2021
18	DMU431	12,75	55,87	7	144	1999-2009
19	DMU444	13,30	55,00	6	111	2016-2021
20	FYN0018501	10,70	54,73	1	1	2004-2004
21	FYN0018531	10,67	54,80	1	1	2004-2004
22	FYN0018825	10,62	55,45	3	32	1990-1992
23	FYN0018841	10,57	55,43	3	58	1990-1992
24	FYN0018842	10,57	55,44	1	1	1991-1991
25	FYN0018843	10,57	55,45	2	13	1991-1992
26	FYN6300043	10,16	55,00	3	45	2004-2006
27	FYN6500051	10,42	55,03	1	19	2011-2011
28	FYN6900017	10,52	55,48	3	58	2002-2011
29	FYN6910801	10,48	55,46	1	2	2002-2002
30	*Fynsværket indtag (Odense Fjord)#1584-1	9,13	55,74	2	15	2002-2008
31	*Fynsværket udtag (Odense Fjord)#1584-2	9,13	55,74	2	15	2002-2008
32	*Glyngøre#12854	9,13	55,74	1	13	1996-1996

33	*Grarup Sø#S019001	9,13	55,74	1	13	1993-1993
34	KBH431	12,75	55,87	10	185	2011-2021
35	Ketting Nor#S057001	9,13	55,74	3	99	2001-2003
36	Mariager Fjord; ???#5142	9,13	55,74	1	1	2000-2000
37	NOR240008	9,97	56,66	1	1	1999-1999
38	NOR240016	10,04	56,69	1	1	2000-2000
39	NOR240021	9,80	56,64	1	2	2000-2000
40	NOR409	10,79	56,86	18	333	2000-2021
41	NOR5503	9,97	56,66	15	522	1991-2006
42	NOR7715	9,92	57,61	6	63	2016-2021
43	P1	9,29	56,64	5	162	2009-2013
44	P7	9,29	56,61	5	162	2009-2013
45	P8	9,29	56,60	5	156	2009-2013
46	RKB1	8,22	56,08	22	1040	1992-2014
47	RKB10	8,28	55,96	1	8	2000-2000
48	RKB12	8,17	56,00	1	8	2000-2000
49	RKB21	8,15	56,39	6	140	1990-1997
50	RKB22	8,21	56,34	5	189	1998-2004
51	RKB23	8,26	56,31	8	186	1990-2004
52	RKB4	8,19	55,87	1	26	1992-1992
53	RKB9	8,32	55,90	1	8	2000-2000
54	ROS1727	12,42	55,49	3	44	2004-2006
55	ROS2	12,08	55,67	1	2	1991-1991
56	ROS2044	11,81	55,68	1	2	1991-1991
57	ROS60	12,07	55,71	28	529	1991-2021
58	SJY3	8,58	55,09	3	43	1995-1997
59	SJYKFF2	9,49	54,84	5	116	1993-1997
60	station 3#8843	9,13	55,74	3	52	1990-1992
61	station 4 Kertinge Nor#42314	9,13	55,74	2	5	1991-1992
62	STO0101015	11,59	55,12	1	16	2011-2011
63	STO0802008	12,05	55,16	8	143	2004-2011
64	STO0901016	12,33	54,89	3	63	2001-2003
65	STO0901032	12,39	54,82	2	52	1999-2000
66	VEJ0004273	9,67	55,70	1	24	1993-1993
67	VEJ0005790	9,91	55,85	3	56	2004-2006
68	Venø Sund#12858	9,13	55,74	1	44	1996-1996
69	VIB3702-00001	8,41	56,60	17	354	1973-2003
70	VIB3708-00001	9,06	56,95	29	586	1973-2021
71	VIB3709-00001	9,25	56,97	1	21	1997-1997
72	VIB3711-00001	9,63	57,04	9	127	1988-1997
73	VIB3720-00001	8,67	56,52	11	164	1988-2008
74	VIB3723-00001	8,74	56,92	10	149	1988-1997
75	VIB3727-00001	9,08	56,62	22	521	1988-2011
76	VSJ10003	11,80	55,87	6	112	2004-2010
77	VSJ10006	11,82	55,72	7	37	1991-1997
78	VSJ20925	11,16	56,13	10	216	2000-2010

Bilag B2 QA_Dublet_klassifikation_Zoo

Filen QA_Dublet_Classification_Zoo_ALL.xlsx giver et overblik over samtlige prøver samt deres dubletstatus (Kolonne G/Active). Dubletterne er ind delt i grupper, hvor den aktive (1) er grøn, og den/de inaktive (0) er gul/gule.

Tabel B.2 Eksempel på dubletoverblik. Den fulde liste findes i filen QA_Dublet_Classification_Zoo_ALL.xlsx.

A	B	C	D	E	F	G	H	I
Date	Station	Sample_ID	Depth	System	Primary_Sample_ID	Active	Score	Group
20001012	VSJ20925	1196		14	8832	0	88	3685
20001017	ARH170006	8654		6	8654	1		3686
20001017	ARH170006	3772	770	8	8654	0	100	3686
20001017	ROS60	1174		14	1174	1		3687
20001017	VSJ20925	8833		6	8833	1		3688
20001017	VSJ20925	1197		14	8833	0	90	3688
20001018	RKB22	8015	150	6	8015	1		3689
20001018	RKB22	8014	150	7	8015	0	93	3689
20001018	RKB22	8012	150	6	8012	1		3690
20001018	RKB22	8013	150	7	8012	0	100	3690
20001019	STO0901032	827		14	827	1		3691
20001023	RKB1	7724	150	6	7724	1		3692
20001023	RKB1	7725	150	7	7724	0	99	3692
20001023	RKB1	7727	150	6	7727	1		3693
20001023	RKB1	7726	150	7	7727	0	99	3693
20001023	STO0901032	828		14	828	1		3694
20001023	VIB3727-00001	8712	250	6	8712	1		3695
20001023	VIB3727-00001	8713	250	7	8712	0	100	3695
20001024	VIB3708-00001	8762	375	6	8762	1		3696
20001024	VIB3708-00001	8763	375	7	8762	0	100	3696
20001025	VIB3702-00001	8810	325	6	8810	1		3697
20001025	VIB3702-00001	8811	325	7	8810	0	100	3697
20001026	NOR5503	8584	500	6	8584	1		3698
20001026	NOR240008	3801		8	8584	0	97	3698
20001107	RKB22	8019	150	6	8019	1		3699
20001107	RKB22	8018	150	7	8019	0	96	3699
20001107	RKB22	8016	150	6	8016	1		3700
20001107	RKB22	8017	150	7	8016	0	92	3700
20001107	ROS60	1175		14	1175	1		3701
20001108	ARH170006	8655		6	8655	1		3702
20001108	ARH170006	3772	770	8	8655	0	100	3702

Sample_ID er prøve-ID, som er unikt. Der er i alt 13.406 unikke prøve-ID'er i datasættet. Primary_Sample_ID er den prøve-ID, der er aktiv/bruges i gruppen. Hvis Sample_ID og Primary_Sample_ID er ens, vil Active være 1. Hvis der er prøver uden dubletter, er der derfor kun én linje i gruppen, f.eks. gruppe 3687 (tabel B.2). Typisk varierer dybden ofte inden for samme gruppe, hvilket er årsagen til, at dybde ikke kan anvendes i den initiale potentielle dublet-gruppering. Tilsvarende er der eksempler på varierende stationsnavne i samme gruppe, og derfor anvendes station ligeledes ikke i den initiale potentielle dubletgruppering.

Bilag B3 QA_Dublet_Classification_Zoo_19950601

Filen QA_Dublet_Classification_Zoo_19950601.xlsx er et eksempel på håndtering af dubletter på en given dato, nemlig den 01/06/1995, hvor de underliggende data er vist. Eksemplet indeholder syv prøver fra tre forskellige systemer (tabel B.3).

Dubletudredning samt manuel vurdering har identificeret, at der er én samlende prøve stammende fra STOQ (system = 14) samt tre prøver fra Orbicon (6), som samlet er indeholdt i STOQ-prøven. Derudover er der tre Århus- (8) prøver, som er direkte dubletter af de tre Orbicon- (6) prøver. De tre Orbicon- (6) prøver, som indeholder de bedste data, er derfor valgt som aktive prøver, og de fire andre prøver er følgelig deaktiveret. Som det også kan ses, er der forskellige stationsangivelser. Dette er generelt for hele datasættet og har nødvendiggjort, at alle kombinationer af prøver på en dato er blevet gennemgået for dubletter.

Tabel B.3 Eksempel på dublet overblik på datoen 01/06/1995.

Date	Station	Sample_ID	Depth	System	Primary_Sample_ID	Active	Score
19950601	NOR5503	6262		6	6262	1	
19950601	NOR240008	10065		8	6262	0	90
19950601	NOR5503	374		14	6262	0	60
19950601	NOR5503	6263		6	6263	1	
19950601	NOR240008	10066		8	6263	0	90
19950601	NOR5503	374		14	6263	0	50
19950601	NOR5503	6264		6	6264	1	
19950601	NOR240008	10067		8	6264	0	90
19950601	NOR5503	374		14	6264	0	70

Tabel B.3 viser et eksempel på datoen 01/06/1995. Eksemplet i tabel B.4 indeholder syv prøver, der kan kombineres parvis i 21 mulige kombinationer. Projektet har udviklet en score-algoritme, der implementeres i den videre analyse. Kolonne AI i scorearket i QA_Dublet_Classification_Zoo_19950601.xlsx indeholder en samlet score, der er udgangspunktet for yderligere kontrol. En score under 50 indikerer en usandsynlig dublet, en score over 70 er sandsynligvis en dublet, og en score over 90 er med stor sandsynlighed en dublet. Kombinationer i området 50-90 tjekkes manuelt, og i tvivlstilfælde udføres en faglig vurdering. Alt vurderes manuelt station for station, hvor scoren er en væsentlig hjælp i afklaringen.

Filen "QA_Dublet_Classification_Zoo_19950601.xlsx" er vedhæftet som bilag.

Tabel B.4 Eksempel på de 21 potentielle kombinationer, hvor et udvalg af sammenligningsparametre kan ses.

	A	B	C	D	E	F	G	H	I	J	K	AI
1	Date	S1	S2	S1_Station	S2_Station	S1_Depth	S2_Depth	S1_Sample_ID	S2_Sample_ID	S1_Time	S2_Time	Score
2	19950601	6	8	NOR5503	NOR240008			6263	10067	1200	1200	44
3	19950601	6	8	NOR5503	NOR240008			6264	10066	1200	1200	45
4	19950601	6	8	NOR5503	NOR240008			6262	10066	1200	1200	47
5	19950601	6	8	NOR5503	NOR240008			6263	10065	1200	1200	47
6	19950601	6	6	NOR5503	NOR5503			6263	6264	1200	1200	47
7	19950601	8	8	NOR240008	NOR240008			10066	10067	1200	1200	48
8	19950601	8	8	NOR240008	NOR240008			10065	10066	1200	1200	49
9	19950601	6	6	NOR5503	NOR5503			6262	6263	1200	1200	49
10	19950601	6	14	NOR5503	NOR5503			6263	374	1200	0	57
11	19950601	8	14	NOR240008	NOR5503			10066	374	1200	0	61
12	19950601	6	8	NOR5503	NOR240008			6262	10067	1200	1200	62
13	19950601	8	8	NOR240008	NOR240008			10065	10067	1200	1200	63
14	19950601	6	8	NOR5503	NOR240008			6264	10065	1200	1200	63
15	19950601	6	6	NOR5503	NOR5503			6262	6264	1200	1200	65
16	19950601	8	14	NOR240008	NOR5503			10065	374	1200	0	69
17	19950601	6	14	NOR5503	NOR5503			6262	374	1200	0	69
18	19950601	8	14	NOR240008	NOR5503			10067	374	1200	0	71
19	19950601	6	14	NOR5503	NOR5503			6264	374	1200	0	71
20	19950601	6	8	NOR5503	NOR240008			6263	10066	1200	1200	90
21	19950601	6	8	NOR5503	NOR240008			6264	10067	1200	1200	96
22	19950601	6	8	NOR5503	NOR240008			6262	10065	1200	1200	98
23												
24												
25	Not Doublet											
26	Doublet											
27	Doublet(Part)											
28	Diff											
29	Identical											
30												

Tabel B.5 Eksempel på artsprofil, som er en sorteret artskodeliste med tilhørende indeks, da koden kan være i brug flere gange ifm. forskellige stadier, størrelser og bestemmelsesusikkerhed. Bruges til at sammenligne to prøvers arter som en procentuel indikation.

Messages		
n	S1_Profile_Specie	S2_Profile_Specie
	901_1#1030_1#1036_1#1038_1#1043_1#1049_1#1114_1#...	901_1#1030_1#1036_1#1038_1#1043_1#1049_1#1114_1#...
	901_1#1030_1#1036_1#1038_1#1043_1#1049_1#1114_1#...	901_1#1030_1#1036_1#1038_1#1043_1#1049_1#1114_1#...
	1_1#1152_1#2347_1#2347_2	1_1#1152_1#2347_1#2347_2
	1_1#1152_1#2347_1#2347_2	1152_1#2347_1#2347_2#3865_1
	901_1#1030_1#1036_1#1038_1#1043_1#1049_1#1114_1#...	901_1#1030_1#1036_1#1038_1#1043_1#1049_1#1114_1#...
	1_1#1152_1#2347_1#2347_2	1152_1#2347_1#2347_2#3865_1

Tabel B.6 Data fra de syv prøver kan ses i de med gult markerede ark i Excel-filen QA_Doublet_Classification_Zoo_19950601.xlsx.

19950601	<u>Score</u>	374	6262	6263	6264	10065	10066	10067	Kilde_Orbicon	Kilde_Århus
----------	--------------	-----	------	------	------	-------	-------	-------	---------------	-------------

Tabel B.7 Kildedata fra de tre Orbicon- (6) prøver og de tre Århus- (8) prøver kan ses i de med gult markerede ark i Excel-filen QA_Dublet_Classification_Zoo_19950601.xlsx.

19950601	Score	374	6262	6263	6264	10065	10066	10067	Kilde_Orbicon	Kilde_Århus
----------	-------	-----	------	------	------	-------	-------	-------	---------------	-------------