

OPERATIONALISERING OG MODEL-ASSIMILERING AF SATELLITBASEREDE KLOROFYLDATA

Videnskabelig rapport fra DCE - Nationalt Center for Miljø og Energi

nr. 508

2022



OPERATIONALISERING OG MODEL-ASSIMILERING AF SATELLITBASEREDE KLOROFYLDATA

Videnskabelig rapport fra DCE – Nationalt Center for Miljø og Energi

nr. 508

2022

Andreas Holbach¹ Stiig Markager¹ Rikke Margrethe Closter² Jesper Sandvig Mariegaard² Anders Christian Erichsen² Mai-Britt Kronborg²

¹Aarhus Universitet, Institut for Ecoscience ²DHI A/S



Datablad

Serietitel og nummer:	Videnskabelig rapport fra DCE - Nationalt Center for Miljø og Energi nr. 508		
Kategori:	Rådgivningsrapporter		
Titel:	Operationalisering og modelassimilering af satellitbaserede klorofyldata		
Forfattere:	Andreas Holbach ¹ , Stiig Markager ¹ , Rikke Margrethe Closter ² , Jesper Sandvig		
Institutioner:	¹ Aarhus Universitet, Institut for Ecoscience & ² DHI A/S		
Udgiver: URL:	Aarhus Universitet, DCE - Nationalt Center for Miljø og Energi © http://dce.au.dk		
Udgivelsesår: Redaktion afsluttet:	Oktober 2022 Oktober 2022		
Faglig kommentering: Kvalitetssikring, DCE: Sproglig kvalitetssikring:	Peter Anton Upadhyay Stæhr (AU), Mads Joakim Birkeland (DHI) Anja Skjoldborg Hansen Charlotte Hviid		
Ekstern kommentering:	Miljøstyrelsen. Kommentarerne findes her: http://dce2.au.dk/pub/komm/SR508_komm.pdf		
Finansiel støtte:	Miljøstyrelsen		
Bedes citeret:	Holbach, A., Markager, S., Closter, R.M., Mariegaard J.S., Erichsen, A.C., Kronborg, MB. 2022. Operationalisering og modelassimilering af satellitbaserede klorofyldata. Aarhus Universitet, DCE – Nationalt Center for Miljø og Energi, 74 s Videnskabelig rapport nr. 508 <u>http://dce2.au.dk/pub/SR508.pdf</u>		
	Gengivelse tilladt med tydelig kildeangivelse		
Sammenfatning:	Denne rapport omhandler operationaliseringen af en ny metode for satellitbaserede klorofylestimater i komplekse kystnære områder og assimileringen af det optimerede klorofylprodukt i mekanistiske bio-geokemiske modeller.		
Emneord:	Remote Sensing, Mekanistisk Model, Dataassimilering, Fytoplankton, Klorofyl		
Layout: Foto forside:	Grafisk Værksted European Union, Copernicus Sentinel-2 imagery		
ISBN: ISSN (elektronisk): Sideantal:	978-87-7156-713-7 2244-9981 74		
Internetversion:	Rapporten er tilgængelig i elektronisk format (pdf) som <u>http://dce2.au.dk/pub/SR508.pdf</u>		

Indhold

Fo	rord		5
Sa	mmer	nfatning	6
Su	mmar	у	7
1	Proje	ektets formål	8
2	Udg	angspunkt	10
	2.1	Sentinel-3 kl5orofylprodukt for komplekse kystnære marine vandområder (AU)	10
	2.2	Match-Ups med klorofyl <i>in situ</i> data i Danmarks marine overvågningsprogram (AU) Mekanistisk modellering (DHI)	11 13
3	Moto	oder	19
U	3.1	Optimering af Sentinel-3s klorofylprodukt (AU)	19
	3.2	Operationalisering af korrektionsmetoden (AU)	25
	3.3	Optimering af DHIs model ved assimilering af satellitbaserede data (DHI)	26
4	Pog	litator	20
7	4.1	2D Satellitbaserede overfladeklorofvlprodukter på flere	57
		rumlige og tidslige opløsninger (AU)	39
	4.2	Dataassimilering af satellit klorofyldata i biogeokemiskmodel (DHI)	45
5	Disk	ussion og perspektiver	55
	5.1	Økologisk betydning af skaleringsfaktorerne brugt i optimeringen af satellitbaseret klorofyl (AU)	55
	5.2	Korrektionsmodellens afhængighed af klorofyl	59
	5.3	Dataassimilering i mekanistiske modeller (DHI)	61
6	Konl	klusion	62
7	Refe	erencer	63
Ар	pendi (CoC	x 1 Complex Coast - Sentinel-3 Chlorophyll Estimator Coa-SenCE)	67
	Man	ual for the operationalized system CoCoa-SenCe to optimize Sentinel-3 based chlorophyll estimates for the	
	0	complex Danish marine waters (AU)	67
	CoC	coa-SenCE modules	67 68

Forord

Nærværende rapport er med baggrund i en henvendelse fra Miljøstyrelsen (MST) til Aarhus Universitet (AU) og DHI om at få undersøgt muligheder ved at inddrage data fra nye teknologier (her satellitbaserede data og mekanistiske modeller) for at supplere de traditionelle vandkemimålinger i den nationale overvågning af de danske marine vandområder. Det langsigtede mål er at finde den mest optimale måde at bedømme vandområdernes økologiske tilstand i fremtiden.

Under hensyntagen af de til rådighed stående tidslige og finansielle ressourcer, blev der i dialog mellem AU og DHI udarbejdet forsknings- og udviklingsprojektet 'Operationalisering og Modelassimilering af Satellitbaserede Klorofyldata'. Projektet skal således danne grundlag og være udgangspunkt for videregående fremtidige forsknings- og udviklingsarbejde mht. inddragelse af nye teknologier i Danmarks marine overvågningsprogram. Projektet har været fulgt af en styregruppe med repræsentanter fra MST (Kristian Ege Nielsen), AU (Andreas Holbach) og DHI (Anders Chr. Erichsen). Styregruppens opgave har været at sikre fremdrift i projektet og koordinere mellem de forskellige aktiviteter. Ved et fælles møde mellem MST, AU og DHI blev de foreløbige resultater præsenteret, som udgangspunkt til en diskussion om potentielle fokusområder for fremtidigt forskningsarbejde.

I projektet har AU stået for arbejdet med satellitdata, mens DHI har været ansvarlig for assimilering af disse data i deres modeller. Flere underkapitler i rapporten indeholder faglige bidrag udelukkende fra enten AU eller DHI. Dette er kendetegnet i de tilsvarende overskrifter.

Sammenfatning

Denne rapport omhandler et fælles forsknings- og udviklingsprojekt mellem Aarhus Universitet og DHI med formålet at (1) operationalisere en ny metode til optimeringen af et satellitbaseret klorofylprodukt for komplekse kystnære områder og (2) teste og udvikle rutiner til assimilering af det optimerede klorofylprodukt i DHIs mekanistiske bio-geokemiske modeller.

Metoden til optimering af satellitbaseret klorofylovervågning gør brug af Sentinel-3s klorofylprodukt for komplekse kystnære marine vandområder, samt *in situ* klorofylmålinger i NOVANA programmet, og en ny distancevægtet rumlig regressionsmodel. Regressionsmodellen leverer rumlige lag af to skaleringsfaktorer, som bruges til at beregne klorofylkoncentration ud fra estimater af pigmentabsorption i Sentinel-3s klorofylprodukt. For de danske marine områder forbedrer det dermed optimerede klorofylprodukt sammenhængen med traditionelle *in situ* klorofylmålinger signifikant. Derudover viser skaleringsfaktorerne klare sammenhænge med lokale økologiske forhold og fytoplanktonsamfundet. Optimeringsmetoden er blevet operationaliseret således at nye satellit- og overvågningsdata fremadrettet kan integreres rutinemæssigt for at anvende optimeringsmetoden automatisk. Det operationelle system kan ligeledes levere klorofyldata egnet til assimilering i DHIs model.

Dataassimilering er en kompleks opgave, men ved eksemplet af lokalmodellen for det Nordlige Bælthav viser de udviklede rutiner, baseret på ensemble Kalman filter, potentiale til at forbedre overensstemmelsen mellem modellerede og *in situ* klorofylkoncentrationer. Forbedringen er ikke begrænset til områder, hvor der eksisterer satellitdata til assimilering, også i områder, som f.eks. Vejle Fjord, hvor dækningen af satellit data er ringere, forbedres modelresultaterne. Dataassimilering påvirker ikke alene den assimilerede variabel (her klorofyl), men også afledte variable. Mens lys blev forbedret en anelse, sammenlignet med målinger, viste næringssaltene imidlertid ikke de samme positive tendenser på alle stationer.

Projektets resultater viser at der er et stort potentiale ved anvendelse af satellitbaserede data til overvågning af klorofylkoncentrationen og udbredelse af fytoplankton i danske farvande. Desuden viser vi, at man med fordel kan forbedre mekanistiske modeller ved assimilering af satellitbaserede klorofyldatasæt. Projektet afdækker også en række potentielle fokusområder for fremtidigt forskningsarbejde, som skal adresseres for at opnå den bedst mulige merværdi af disse nye teknologier i den nationale overvågning af de danske marine vandområder.

Summary

This report results from a joint research and development project between Aarhus University and DHI with the objective to (1) operationalize a new method for optimization of a satellite-based chlorophyll product for complex and near-coastal marine water bodies, as well as (2) test and develop routines for assimilation of this optimized chlorophyll product into DHI's mechanistic biogeochemical models.

The method for optimization of satellite-based chlorophyll monitoring uses Sentinel-3's chlorophyll product for complex and near-coastal marine water bodies together with *in situ* chlorophyll data acquired in the frame of the NO-VANA program and a new distance-weighted regression model. The regression model generates spatial layers of two scaling factors, which are used to calculate chlorophyll concentrations based on estimates of pigment absorption in the Sentinel-3 chlorophyll product. The resulting optimized chlorophyll product significantly improves the relationship with chlorophyll concentrations acquired from traditional *in situ* samples for the Danish marine water bodies. Furthermore, the scaling factors show clear relationships with local ecological conditions and phytoplankton communities. The optimization method has been operationalized in a way that new satellite and *in situ* data can henceforth routinely be integrated and the optimization method applied automatically. This operational system can likewise generate chlorophyll data suitable for assimilation within DHI's model.

Data assimilation is a complex task, however, at the example of a local model for the Northern Belt Sea, we could show that the developed routines using ensemble Kalman filter have the potential to improve the agreement between modelled and *in situ* acquired chlorophyll concentrations. This improvement is not only limited to areas with available satellite data for assimilation, but also in areas without or with less coverage, such as e.g. Vejle Fjord, the agreement between model results and *in situ* data gets better. Data assimilation does not only affect the single assimilated variable, but also other related variables. In comparison with *in situ* observations, the agreement for light attenuation was improved a bit; however, for nutrients we did not observe the same positive tendencies for all reference stations.

In general, the project's results show that satellite-based data has great potential for monitoring of chlorophyll concentrations and phytoplankton distributions in Danish marine waters. Furthermore, we show that assimilation of these data into mechanistic models can considerably improve their performance. The project also identifies a couple of potential focus areas for further research with respect to optimizing the added value of integrating these new technologies into the national marine monitoring program for the Danish Seas.

1 Projektets formål

Som en del af den nationale overvågning af de danske marine vandområder (både de kystnære jf. Vandrammedirektivet og de mere åbne marine områder, jf. Havstrategidirektivet) arbejder Miljøstyrelsen på at inddrage data fra nye teknologier for at supplere de traditionelle vandkemimålinger og for at finde den mest optimale måde at bedømme vandområdernes økologiske tilstand i fremtiden.

AU har over det sidste år udviklet en ny metode til optimeringen af satellitbaserede klorofylmålinger på baggrund af en rumlig vægtet regressionsmodellering af sammenhængen mellem satellitbaserede (Sentinel-3 OLCI) og *in situ* målinger (NOVANA) for de danske marine vandområder (Holbach m.fl., 2022). Grundidéen i metoden er at bestemme to parametre ("skaleringsfaktorer"), som beskriver sammenhængen mellem pigmentabsorption og klorofylkoncentration. Dette gøres dynamisk og dækkende i rum i stedet for at bruge faste værdier over hele området. Vi forventer at disse to skaleringsfaktorer er relateret til fytoplanktonets fysiologi og artssammensætning samt koncentrationer af næringsstoffer, da disse påvirker algernes optiske egenskaber. Vores arbejdshypotese er at skaleringsfaktorerne i sig selv på sigt kan udvikles til indikatorer for et områdes eutrofieringsstatus. Eutrofieringsstatus vil ændre sig over tid, så den nye metode er afgørende for at satellitobservationer kan bruges til monitering da den signifikant forbedrer sammenhængen mellem satellitbaserede klorofylmålinger og de traditionelle *in situ* prøver.

DHI har tidligere – i samarbejde med AU – set på brugen af mekanistiske modeller og satellit data som et supplement til overvågningen i Nordsøen (Upadhyay et al. 2021). I dette projekt blev resultaterne fra satellit-analysen og de mekanistiske modelresultater sammenlignet med henblik på at vurdere anvendeligheden i den danske del af Nordsøen, hvor den traditionelle overvågning er begrænset pga. logistiske omkostninger. I nærværende projekt er formålet at udvikle et operationelt system med henblik på at anvende den nye satellitbaserede metode, udviklet af AU, systematisk og med regelmæssige opdateringer af satellitbaserede klorofyldata i de danske farvande, så de over tid rutinemæssigt kan indgå i f.eks. modelassimilering, udtræk af tidsserier på punkt og vandområdeniveau, såvel som estimater af årlige gennemsnitsværdier.

Ud over arbejdet med satellit data vil vi udvikle og teste forskellige metoder til dataassimilering af satellitbaserede klorofylmålinger i mekanistiske modeller. Som led i MSTs arbejde med de kommende vandområdeplaner har DHI udviklet en række mekanistiske biogeokemiske modeller. Disse modeller vil i dette projekt blive anvendt til at undersøge hvordan satellit data bedst muligt kan assimileres for at opnå en bedre beskrivelse af primærproduktionen i vandsøjlen og derved den økologiske tilstand i vandområderne. Projektet vil alene fokusere på at assimilere og teste satellit data for klorofyl med det formål at undersøge og optimere rumlig opløsning og frekvens af satellitdata til assimilering. Denne forskning og udvikling vil etablere et grundlag, som i fremtiden skal lede mod en fuld integrering af alle *in situ* observationer, satellitobservationer og mekanistiske modeller, og at resultaterne herfra kan forbedre den marine miljøovervågning. På længere sigt forventer vi derfor at observationer fra flere satellitbaserede parametre (f.eks. lyssvækkelse), fluorescensdybdeprofiler, droner og ferrybox målinger også vil finde anvendelse i den fremtidige operationelle miljøovervågning.

2 Udgangspunkt

2.1 Sentinel-3 klorofylprodukt for komplekse kystnære marine vandområder (AU)

Klorofyldata baseret på Sentinel-3 (S3) satellitternes Ocean and Land Color Instrument (OLCI) er offentligt tilgængelige f.eks. på ESA data portal (https://finder.creodias.eu/). De to satellitter (S3-A & S3-B) er blevet opsendt i 2016 hhv. 2018 og de leverede de første tilgængelige dataprodukter for danske områder d. 01.11.2017 og d. 13.03.2019. Den oprindelige rumlige opløsning af dataproduktet ligger på ca. 300x300 m. Da de to satellitter har forskellige orbits, samler vi alle satellitbilleder over en dag og laver re-projected composite billeder for hver dag på en 500x500 m grid-raster i UTM32N coordinate reference system (Figur 2.1**Figur 2.1**). Fra 01.03. til 31.10 fås der i gennemsnit klorofyl målinger på ca. hver 3.-4. dag per pixel (baseret på data fra 2018-2020). I den resterende tid er solvinklen for lav og der findes kun ganske få målinger, som også har større usikkerhed. Derfor vælger vi at arbejde primært med data fra 01.03.-31.10.

Der leveres et level-2 satellitprodukt på baggrund af den såkaldte Case-2 Regional CoastColour (C2RCC) processor, som baserer sig på en neural-network algoritme, der fordeler den samlede spektrale lyssvækkelse på tre grupper af optisk aktive stoffer og deres 'inherent optical properties' (IOPs) ved en bølgelængde af 443 nm: 1) absorption af fytoplanktons pigmenter (a_{pig}), 2) absorption af opløst og partikulært organisk stof (a_{dg}), og 3) spredning ved suspenderede partikler (b_{tot}) (Doerffer, 2010; Brockmann m.fl., 2016). Da algoritmen også tager højde for andre optisk aktive stoffer i vandsøjlen end fytoplankton, er C2RCC processoren velegnet til brug i komplekse kystnære marine vandområder. Pigmentabsorption og spredning ved suspenderede partikler udgives som koncentrationsværdier af både klorofyl (Chl) og total suspenderet stof (TSS) efter de følgende tilgrundliggende ligninger:

$$Chl = a_{pig}^{Chl_{exp}} * Chl_{fact}$$
 Ligning 1
 $TSS = b_{tot}^{TSS_{exp}} * TSS_{fact}$ Ligning 2

Derved bruges et set af standard skaleringsfaktorer (
$$Chl_{exp} = 1.04$$
, $Chl_{fact} = 21$, $TSS_{exp} = 0.942$ og $TSS_{fact} = 1.06$), der repræsenterer gennemsnitlige egenskaber hos fytoplankton og suspenderet materiale (Doerffer, 2010; Brockmann m.fl., 2016). De to slags skaleringsfaktorer (faktor og eksponent) beskriver en polynomfunktion mellem f.eks. Chl og a_{pig}. På log-skala svarer Chl_{exp} til hældningen, mens log(Chl_{fact}) svarer til y-afskæring i den følgende lineære funktion:

$$\log(Chl) = Chl_{exp} * \log(a_{pig}) + \log(Chl_{fact})$$
 Ligning 3

Omvendt, kan ligningerne så også bruges til at genberegne de oprindelige IOPs a_{pig} og b_{tot} ud fra level-2 satellitproduktet. Det er bekendt at skaleringsfaktorer for a_{pig} varierer i rum afhængigt af miljøforholdene og fytoplanktons økologiske tilstand (Jakobsen og Markager, 2016). I Copernicus' værktøj 'Sentinel Application Platform' (SNAP), hvori C2RCC processoren kan køres på

egen computer, kan skaleringsfaktorerne dog kun tilpasses globalt. Pga. de dynamiske og komplekse forhold i de danske kystnære marine vandområder, er et globalt set af skaleringsfaktorer ikke så velegnet til at beskrive klorofylforholdene repræsentativt.

S3 level-2 dataproduktet indeholder en række masker for potentielt fejlbehæftede værdier (wqsf datasættet). Her anvendte vi de følgende masker: 'INVA-LID', 'CLOUD', 'SNOW_ICE', 'COSMETIC', 'SUSPECT', 'HISOLZEN', 'SA-TURATED', 'HIGHGLINT', 'CLOUD_AMBIGUOUS', 'CLOUD_MARGIN', 'OCNN_FAIL'. En maske for at skelne mellem 'LAND' og 'WATER', som f.eks. anvendt af Kratzer og Plowey (2021), har vi ikke valgt at bruge for store dele af Limfjorden ville så være klassificeret som 'inland water body' og udelukket fra analysen.



Sentinel-3 Chlorophyll Composite (C2RCC) on 2019-07-01

Figur 2.1. Eksempel på et re-projected composite satellitbillede for 2019-07-01. NOVANA stationer vises som sorte punkter.

C2RCC-processoren baserer sig på to grundlæggende antagelser: 1) En homogen blandet vandsøjle og 2) uendeligt dybt vand. Derfor er C2RCCs resultater principielt ikke egnede i optisk lavvandede områder, hvor refleksionen fra havbunden kan påvirke det spektrale refleksionssignal af havets overflade signifikant. Den optiske dybde (i modsætning til den rene vanddybde) er produktet mellem lyssvækkelseskoefficienten K_d og vanddybden. Denne parameter er afgørende for, om havbundens refleksionssignal kan påvirke havoverfladens spektrale refleksion signifikant. Dvs. at en ren filtration efter vanddybde ikke er tilstrækkelig. Signifikant lagdeling i de øverste lag af vandsøjlen, som kan forårsage vertikale gradienter i klorofylkoncentration, kan have tilsvarende effekt. Alligevel kan havets lysrefleksionssignal kun give et estimat for klorofylkoncentrationen i den øverste del af overfladevande

2.2 Match-Ups med klorofyl *in situ* data i Danmarks marine overvågningsprogram (AU)

Ideelt, ville man kunne sammenstille klorofylkoncentrationer målt *in situ* med satellitobservationer fra samme sted og tidspunkt. Det er dog urealistisk at opnå i et operationelt udført nationalt overvågningsprogram, hvor monitering regelmæssigt skal dække over store marine områder, og det er derfor svær at tage hensyn til vejrforhold og hvornår satellitterne passerer. Derfor har vi her testet to toleranceniveauer i tidsforskel: 1) Begge målinger fra samme dag og 2) satellitestimat baseret på en ±5 dage glidende gennemsnit (Gifur). Dertil har vi brugt alle NOVANA klorofylmålinger i 2018-2020 fra ≤ 1 m dybde. I alt kunne vi trække 5387 dagsaggregerede klorofylkoncentrationer fra 146 stationer ud af ODA databasen mellem 2018-2020 for den øverste meter af vandsøjlen.



Sentinel-3 Chlorophyll +-5 Days Moving Avrg. (C2RCC) on 2019-07-01

Figur 2.2. Eksempel på et ±5 dage glidende gennemsnitsbillede af S3s C2RCC klorofylprodukt for 2019-07-01. Bemærk forskellen i rumlig dækning sammenlignet med Figur 2.1. NOVANA stationer vises som sorte punkter.

Ved at bruge glidende gennemsnit fås der en meget større dækning med satellitbaserede estimater i tid og rum og større sandsynlighed for match-ups mellem både *in situ* og satellitbaseret data. I tabel 2.1 kan der ses, at antallet af match-ups for hele datasættet er ca. tredoblet ved brug af ±5 dage glidende gennemsnit, mens den statistiske sammenhæng mellem begge datasæt (her Pearson korrelationskoefficient på log-transformerede data, R_{Pears}^{log}) aftager lidt i 2018 & 2019, mens den faktisk bliver bedre i 2020. Anvendelsen af glidende gennemsnit udjævner kortvarige spring op og ned i tidsserier, mens kontinuerlig vækst eller henfald af klorofyl påvirkes i mindre grad. Match-Up punkterne vises for årene 2018-2020 Figur 2.3.

Tabel 2.1. Statistik på sammenhæng mellem S3 klorofylestimater og in situ klorofylmålinger i NOVANA programmet. Tabellen viser antallet af punkter og Pearson korrelationskoefficient på log-transformerede data for årene 2018-2020, to toleranceniveauer tidsforskel, såvel som for hele datasættet og efter anvendelse af en maske for optisk lavvandede områder.

Datasæt udvalg	Parameter Tidsforskel		2018	2019	2020	Alle
	Antal Match-	Samme dag	334	334	373	1041
Hele datasættet	Ups	±5 dage glidende gennemsnit	1109	1179	1004	3292
	R _{Pears} log	Samme dag	0.58	0.58	0.38	0.52
		±5 dage glidende gennemsnit	0.55	0.55	0.40	0.50
	Antal Match-	Samme dag	237	209	250	696
Maskeret for optisk lavvandede arealer og	Ups	±5 dage glidende gennemsnit	768	719	626	2113
de første og sidste 60 dage af året	R _{Pears} log	Samme dag	0.67	0.65	0.51	0.62
		±5 dage glidende gennemsnit	0.64	0.64	0.57	0.62

Det kan ses, at der i begge tilfælde er en signifikant statistisk sammenhæng mellem S3-C2RCC baserede og *in situ* NOVANA værdier (her vist på logskala) og at sammenhængene er næsten lige gode for match-ups fra samme dag og for ±5 dage glidende gennemsnit i S3 værdierne. Alligevel er spredningen af sammenhængen meget stor og ikke tilfredsstillende mht. til brug i den marine overvågning på vandområdeniveau. Dette er også i overensstemmelse med tidligere undersøgelser ift. til S3 produkter i Østersøen (f.eks. Kyryliuk og Kratzer, 2019; Toming m.fl., 2017).



Figur 2.3. Sammenhæng mellem S3 klorofylestimater (x-akse) og *in situ* klorofylmålinger i NOVANA programmet (y-akse). Venstre: Begge målinger fra samme dag; Højre: Match-Ups med ±5 dage glidende gennemsnit af S3 værdier. Hvide punkter repræsenterer hele datasættet, mens de grønne punkter er efter anvendelsen af en maske for optisk lavvandede områder (se også afsnit 3.1.1 og Figur 3.4). Korrelationskoefficienten er beregnet efter maskering.

2.3 Mekanistisk modellering (DHI)

Som beskrevet under projektets formål, har DHI og AU samarbejdet omkring modeller og satellitdata, men i modsætning til tidligere, er formålet nu at integrere data og modeller igennem dataassimilering.

MST har fået udviklet en række mekanistiske modeller, som i dag indgår som en samlet del af MSTs modelkompleks bag vandområdeplanerne 2021-2027. Til dette projekt har vi taget udgangspunkt i en lokalmodel for det Nordlige Bælthav (NBS) (Erichsen og Birkeland, 2019 og 2020).

2.3.1 Modeludvikling

Lokalmodellen for det Nordlige Bælthav består af en 3D hydrodynamisk model, der beskriver det fysiske system; vandstand, strøm, saltholdighed og vandtemperaturer og en biogeokemiske (økosystem) model, der beskriver de styrende biogeokemiske pelagiske og bentiske parametre og processer som fytoplankton, opløst ilt, primærproduktion osv. Modelstrukturen er modulær, hvilket betyder, at den hydrodynamiske model er udviklet uafhængigt af den biogeokemiske model.

Den mekanistiske model er opsat og kalibreret for perioden 2002-2011 og valideret for perioden 2012-2016, og sidenhen udvidet til også at dække perioden 2017-2018. Modelkalibrering- og validering er rapporteret i Erichsen og Birkeland, 2019 og 2020.

2.3.2 Modelleringssystem

Lokalmodellen for det Nordlige Bælthav er baseret på modelleringssoftwaren MIKE 3 HD FM. MIKE 3 HD FM er baseret på en fleksibel mesh-tilgang, udviklet til anvendelse inden for oceanografiske, kystnære og estuarine systemer (DHI, 2017).

2.3.3 Modelområde

Modellen fra Vandplansprojektet 2021-2027 er udviklet med henblik på at beskrive de kystnære vandområder i det Nordlige Bælthav, hvorfor modellen har en forholdsvis høj horisontal opløsning i disse områder, fx med gridceller på 50-250 m i Horsens Fjord og på 300-1000m langs jyske og fynske kyster (Figur 2.4). For at optimere dataassimileringsrutinerne er der udviklet en grovere model for området med en opløsning på 500-2000 m langs kysterne inkl. Horsens Fjord (Figur 3.4.).



Figur 2.4. Modeldomæne i vandområdeplans-modellen (øverst) og dataassimilerings-modellen (nederst).

2.3.4 Den biogeokemiske model

Den biogeokemiske model er en dynamisk, mekanistisk økosystemmodel, der er koblet til den hydrodynamiske model, og det er netop assimilering af data i den biogeokemiske model, som er i fokus. Den biogeokemiske model omfatter mere end 50 tilstandsvariable hvoraf ca. halvdelen beskriver de bentiske processer (Figur 2.5). Modellen beskriver transporten med havstrømme og omsætningen af organisk stof og næringsstoffer i selve vandfasen og havbunden, herunder vækst og henfald af fytoplankton og er således delt op i to centrale moduler; den pelagiske del og den bentiske del (DHI, 2014).

Koblingen mellem de pelagiske og bentiske moduler omfatter adskillige processer (bl.a. sedimentation, resuspension, diffusion, filtration, næringsstofoptagelse i bentiske planter og mineralisering), der driver udvekslingen af opløste stoffer, partikler og organismer (Griffiths m.fl., 2017). Næringsstofferne i den pelagiske del stammer bl.a. fra "intern belastning" herunder mineralisering af organisk stof i sedimenterne. Den interne sedimentbelastning varierer efter størrelsen af de biogeokemisk tilgængelige puljer af C, N og P i sedimentet sammen med bundiltskoncentrationer, vandtemperatur og udskiftning af bundvandet. Den anvendte model integrerer således den pelagiske og den bentiske del.



State variables & processes

Figur 2.5. Skematisk præsentation af koblingen mellem pelagiske og bentiske tilstandsvariable. Det bemærkes, at diagrammet er stærkt simplificeret ift. modellens faktiske kompleksitet.

Endvidere indeholder modellen bentiske primærproducenter. Primærproducenterne er inddelt i funktionelle grupper for rodfæstede frøplanter (ålegræs), makroalger og bentiske mikroalger.

Opsætning, kalibrering og validering af den biogeokemiske model er beskrevet i flere detaljer i Erichsen og Birkeland, 2020.

Biogeokemisk model - pelagisk del

Den pelagiske del (vandsøjlen) af den mekanistiske biogeokemiske model, omfatter de to mest betydende funktionelle grupper af fytoplankton i de indre danske farvande. Derved er det muligt delvist at beskrive den sæsonmæssige variation i fytoplanktonbiomasse og i den funktionelle sammensætning. De to funktionelle grupper er:

- Kiselalger, der repræsenterer en ikke-bevægelig, silikatafhængig algegruppe med lave lysbehov, der er afhængig af turbulens for at forhindre udsynkning.
- Flagellater, der repræsenterer bevægelige neutralt flydende celler.

Fytoplankton nettovæksten er resultatet af primærproduktion minus tab. Hvor produktionen hovedsageligt styres af næringsstof- og lystilgængelighed samt temperatur, omfatter tabsprocesserne respiration, græsning og sedimentation. Fytoplanktons næringsstofafhængighed er beskrevet i to trin.

Først optages de uorganiske næringsstoffer i en intern pulje efter Michaelis-Menten kinetik for næringsstofoptagelse som en funktion af den omgivende næringsstofkoncentration. Næringsstofferne indgår herefter i algevæksten, beskrevet efter Droop quota modellen (Droop, 1968) for vækst som funktion af den intracellulære næringsstofkoncentration (Morel, 1987; Haney m.fl. 1996; Erichsen og Rasch, 2001). Detaljer om hvordan disse processer løses matematisk i modellen er tilgængelige i Lessin og Raudsepp, 2006; DHI, 2013 og DHI, 2014.

Næringsstofferne i den pelagiske model tilføres via eksterne kilder (floder/åer, punktkilder og atmosfærisk deposition), pelagisk remineralisering og fra mineralisering af organisk materiale i sedimentet.

Ved græsning og nedbrydning "omsættes" planteplankton til biomasse af zooplankton og partikulært organisk stof (detritus) samt direkte til uorganiske næringsstoffer. Detritus kan herefter sedimentere eller mineraliseres i den "mikrobielle løkke", som hovedsageligt består af bakteriedrevne processer og fører til remineralisering af opløst og partikulært organisk stof, der (gen-)forsyner fytoplankton med de uorganiske næringsstoffer N og P. Modellen anvendt i dette studie beskriver denne mineralisering ved temperaturafhængig parameterisering og inkluderer ikke specifik beskrivelse af bakteriebiomassen via en tilstandsvariabel.

De indre danske farvande tilføres endvidere betydelige mængder af opløst organisk stof (DOM) fra floder og andre ferskvandkilder. Det påvirker omsætning af organisk stof og påvirker lysforholdene. To fraktioner af opløst organisk stof er repræsenteret i modellen: labilt opløst organisk stof (LDOM) og farvet opløst organisk stof (CDOM). Hver af de tre tilstande af organisk stof (detritus, LDOM og CDOM) er repræsenteret ved tre dynamiske tilstandsvariable (OC, ON og OP). Zooplanktons (mikro- og mesozooplankton) græsning af fytoplankton kan have en regulerende effekt på fytoplanktonbiomassen og intensiteten af afgræsning kan være afgørende for om der udvikles algeopblomstring. Mesozooplankton, der repræsenterer copepoder, omfatter alle 12 aktive græsningsstadier. Væksthastighed og afhængighed af planktonkoncentrationer er baseret på energibudgettet for Acartia tonsa (Kiørboe m.fl. 1985) og den temperaturafhængige stadievarighed/væksthastighed er baseret på Hirst og Sheader, 1997, og Kiørboe og Nielsen, 1994. Modellen inkluderer ikke heterotrofer (rovdyr) på højere trofiske niveauer end zooplankton og tab ved prædation på zooplankton er integreret i zooplanktonets dødelighed. Derfor er prædation på mesozooplankton fra f.eks. fisk og gopler inkluderet ved en ekstra zooplanktondødsrate, beskrevet ved en andengradsfunktion af zooplankton biomasse.

Lysforholdene er afgørende for alle autotrofe organismer, ikke kun for fytoplankton i vandsøjlen men også for ålegræs (frøplanter), makroalger og bentiske mikroalger på havbunden. Lysdæmpningen og dermed den lodrette lysnedtrængning påvirkes af to processer i vandsøjlen, nemlig spredning af lys og absorption af lys af partikler i form af fytoplankton, dødt organisk stof, suspenderede uorganiske stoffer og farvet opløst organisk stof (Effer, 1988 og Kirk, 2000). Selvom spredning ikke "fjerner" fotoner fra vandsøjlen, har spredningen stor betydning for lysdæmpningen, fordi det øger fotonernes vejlængde og dermed sandsynligheden for, at fotoner absorberes af de absorberende komponenter i vandsøjlen.

Iltforholdene i vandsøjlen er i modellen beskrevet som funktion af udveksling af ilt med atmosfæren, den øjeblikkelige nettoproduktion af ilt ved pelagiale og bentiske primærproducenter og sedimentets iltforbrug ved mineralisering og omsætning af labilt og refraktionært organisk materiale. Derudover er ilt i de forskellige vanddybder styret af vertikal og horisontal blanding ved de hydrodynamiske processer.Biogeokemisk model – sedimentkemiske del

Den sedimentkemiske del af den biogeokemiske model indeholder en beskrivelse af sedimentet i to lag, et øvre ukonsolideret lag og et nedre konsolideret lag. I hvert af disse lag indgår fint uorganisk sediment, organisk kulstof, organisk kvælstof og organisk fosfor. Det organiske materiale i begge lag kan blive re-suspenderet ved de kræfter, som påvirker sedimentoverfladen, beregnet ud fra strømhastigheden og bølgernes påvirkning. Energipåvirkningen og dermed forskydningsspændingen er størst i strøm- og bølgeeksponerede områder. Overskrider forskydningsspændingen sedimentets kritiske forskydningsspænding, sker der en re-suspension af organisk materiale (C, N og P) samt fint uorganisk sediment. Det betyder, at modellen kan beskrive en lokal re-suspension af sediment i den eksponerede del af modelområdet. Sedimentet transporteres med strømmen, inden det sedimenterer i områder med lav fysisk energi.

Nedbrydningen af de organiske C-, N- og P-puljer i sedimentet (ved udnyttelse af oxygen eller nitrat som elektronacceptorer) frigiver N og P til sedimentets porevand. Nedbrydningshastigheden afhænger af tilgængeligheden af ilt og nitrat samt C:N-forholdet i sedimentet. Nitrat i porevandet kan denitrificeres til frit kvælstof (N₂). Uorganisk P i porevandet kan binde sig til oxideret jern (Fe³⁺), når sedimentet er oxideret. Omvendt frigives uorganisk P til porevandet, når sedimentet er "reduceret" (Fe²⁺ er den dominerende form). De uorganiske næringsstoffer i sedimentets porevand udveksler med næringsstoffer i vandfasen via diffusion og sedimentet kan dermed enten fungere som "source" eller "sink" for uorganiske næringsstoffer til vandet over sedimentet. For flere detaljer om sedimentmodulet, henvises til Rasmussen et. al., 2009.

Biogeokemisk model - bentiske primærproducenter

Bundvegetationen er i modellen repræsenteret ved fire forskellige grupper af bentiske primærproducenter:

- Flerårige makroalger karakteriseret ved fucoide arter
- Enårige opportunistiske makroalger (f.eks. filamentøse rødalger og Ulva sp.)
- Bentiske mikroalger
- Frøplanter karakteriseret ved ålegræs (Zostera marina)

Som for den pelagiske primærproduktion afhænger den bentiske primærproduktion af vandtemperatur, næringsstoftilgængelighed og tilgængelighed af fotosynteseaktivt lys, men sammenhængende mellem de forskellige faktorer og vækst er forskellig mellem grupperne. Makroalgerne har brug for hårdt underlag at fæstne sig på (sten), mens ålegræs kræver passende sediment (passende kornstørrelse og organisk kulstofindhold under 4%). En vigtig forskel mellem grupperne er, at makroalger (enårige og flerårige) kun kan udnytte uorganiske næringsstoffer fra vandfasen, hvorimod alegræs og bentiske mikroalger også kan optage næringsstoffer fra sedimentets porevand. Derfor kan ålegræs og bentiske mikroalger vokse i områder og på årstider, hvor der er lave næringsstofkoncentrationer i vandet, hvis næringsstofkoncentrationerne i porevandet er tilstrækkelig høje. Som for fytoplankton er det de interne puljer af N og P der driver væksten, beskrevet specifikt for hver af grupperne alegræs, bentiske mikroalger, og enarige -og flerarige makroalger. Modellen omfatter endvidere ophobning af næringsstoffer i frøplanter og makroalger (hvilket resulterer i lave C:N og C:P-forhold) om vinteren, hvor nærringstofferne i vandet er rigelige, men lyset begrænsende for vækst, og derved fortsat vækst i foråret når lyset er rigeligt, men næringsstofferne i vandet er opbrugt af fytoplankton (Pedersen og Borum, 1996). Endelig varierer lysafhængigheden og tab mellem de forskellige bentiske grupper. Tab omfatter respiration, forrådnelse, græsning og tab af dele af planter. Dødt organisk materiale føres dels tilbage til vandfasen og dels til sedimentets organiske puljer. Derved bidrager de bentiske primærproducenter til de organiske og derefter uorganiske næringsstofpuljer i modellen. Flere detaljer om den bentiske primærproduktionsmodel henvises til Kuusemäe m.fl. 2016.

3 Metoder

3.1 Optimering af Sentinel-3s klorofylprodukt (AU)

3.1.1 Maske for optisk lavvandede områder

Som beskrevet ovenfor indeholder C2RCCs dataprodukt ud over pigmentabsorption også IOPs for organiske stoffer og suspenderede partikler (Doerffer, 2010). Sammen med vandets absorptions- og spredningskoefficient ved 443 nm har vi så estimeret en værdi for den samlede spektrale lyssvækkelse att_{S3}^{tot} ud fra S3 data og sammenlignet dens middelværdi for hver pixel og perioden fra 2018-2020 med tilsvarende stationsmiddelværdier for K_d^{PAR} i NOVANA programmet (Figur 3.1). Der er en positiv sammenhæng (R_{Pears} = 0,77; p < 0,01) mellem begge datasæt og vi har brugt den viste lineære regressionsfunktion (y-afskæring = 0) til at beregne et arealdækkende estimat for gennemsnitlig K_d^{PAR} og så for den lokalt gældende gennemsnitlige optiske dybde (ved inddragelse af batymetrien) for hver pixel.

Figur 3.1. Sammenstilling af stationsmiddelværdier for den samlede spektrale lyssvækkelse att_{s3}^{tot} estimeret ud fra S3 C2RCC produktet (x-akse) og K_d^{PAR} beregnet ud fra *in situ* målinger i NOVANA programmet (yakse) for perioden 2018-2020. En lineær regressionsfunktion (grøn linje) er tilpasset. Stationer med vanddybder >=3,5 m vises som blå punkter.

Average Kd Estimation



I lavvandede arealer er det S3-baserede K_d estimat dog mere usikkert, selvom den tilgrundliggende samlede IOP-værdi skulle være mere robust end a_{pig} alene (Doerffer, 2010). Derfor har vi kigget på både 1) vanddybdeafhængige afvigelser mellem S3-C2RCC baserede klorofylkoncentrationer og *in situ* NO-VANA prøver (Figur 3.2) og 2) tilsvarende afvigelser ift. den optiske dybde (Figur 3.3). Der vises absolutte afvigelser normeret til den tilsvarende in situ koncentration. Dvs. at en værdi af 0,5 kan betyde både 50 % over og undervurdering af in situ koncentration med en S3 baseret værdi.

Figur 3.2. Absolutte og med *in* situ værdierne normerede afvigelser (1 \triangleq 100% afvigelse) mellem Chl_{S3} og *in situ* Chl_{NOV} værdier for alle Match-Ups, vist som boxplots over vanddybde rundet til 1 m intervaller. Op til den øverste grænse af boksene indeholdes 75 % af værdierne. Ved vanddybder større end 3,5 m er afvigelsen mindre end 1 for >75 % af alle Match-Ups.

Figur.3 3. Absolutte og med *in* situ værdierne normerede afvigelser ($1 \triangleq 100\%$ afvigelse) mellem Chl_{S3} og *in situ* Chl_{NOV} værdier for alle Match-Ups, vist som boxplots over den optiske dybde rundet til heltal. Op til den øverste grænse af boksene indeholdes 75 % af værdierne. Ved optiske dybder større end 2,5 er den absolutte relative afvigelse mindre end 1 for >75 % af alle Match-Ups.





Vi fandt at >75 % af match-up værdierne ligger under 100 % relativ afvigelse ved vanddybder større end 3,5 m. Ligeledes, aftager relativ afvigelse med stigende optiske dybder og er <100 % for >75 % af match-ups ved optiske dybder større end 2,5. Disse to tærskelværdier valgte vi så til at beregne en rumlig maske for at filtrere arealer med pålagt høj usikkerhed i S3-C2RCC baserede klorofylkoncentrationer fra (Figur 3.4).



Optically Shallow Mask

Figur 3.4. Kort over den afledte maske for optisk lavvandede områder i de danske marine farvande. De røde arealer indeholder pixels hvor vanddybden er <3,5 m og/eller den optiske dybde er <2,5. Ifølge vores analyser er S3-baserede klorofylestimater i disse arealer pålagt stor usikkerhed og derfor maskeret fra den videregående databehandling.

Ved anvendelse af denne maske mistes der ca. en tredjedel af match-up punkterne, men til gengæld bliver den statistiske sammenhæng mellem de to datasæt bedre (Tabel 2.1). For hele datasættet fra 2018-2020 og de to toleranceniveauer i tidsforskel er der 696 match-up punkter tilbage på samme dag, mens der forbliver 2113 match-up punkter ved brug af ±5 dage glidende gennemsnit på S3 data. I alt er der match-up data fra 103 NOVANA stationer tilbage efter anvendelse af den beregnede maske. Begge match-up datasæts viser en korrelationskoefficient på $R_{Pears}^{log} = 0,62$. Pga. den meget større dækning med match-ups og den kun lille forskel i statistisk sammenhæng valgte vi at arbejde videre med ±5 dage glidende gennemsnitsværdier efter at have maskeret for alle punkter der lå enten ved vanddybde <3,5 m eller ved optisk dybde <2,5.

3.1.2 Distance-vægtet regressionsmodellering

Vi kunne se i Figur 2, at der er en signifikant sammenhæng mellem klorofylkoncentrationer i *in situ* prøver og S3-baserede estimater med anvendt parametrisering af standard skaleringsfaktorer, men med en betydelig spredning. Derfor er der ikke generelt mulig at anvende S3s level-2 klorofylprodukt direkte til det marine overvågning. Men vores hypotese er, at der findes valide regressionsfunktioner med specifikke sæt af skaleringsfaktorerne Chl_{exp} og Chl_{fact} for hvert punkt i rum. Disse specifikke sæt af skaleringsfaktorer vil vi estimere med distance-vægtet regressionsmodellering (Gholizadeh og Robeson, 2016; Charlton og Fotheringham, 2009). I metoden vægtes værdierne i regressionsfunktionen skal gælde. I nærværende projekt operationaliserer vi så denne metode på baggrund af *'through-water cost-distance'*.

Through-water cost-distance mellem to punkter estimeres med den korteste distance fra referencepunkt til målpunkt hvor hver pixel, klassificeret som vand, kan passeres i 8 retninger, mens hver pixel får tildelt en 'conductance' værdi på følgende måde:

$$Cond = \begin{cases} 0; if Land\\ 1 - \frac{d_{50}}{d_{50} + d_{coast}}; if Water \end{cases}$$
 Ligning 4

Her svarer d_{50} [m] til den distance fra kysten d_{coast} [m] hvor Cond bliver 0,5. Vi har sat d_{50} = 1000 m.

Ved Cond = 1 svarer den virkelige distance til cost-distance, mens 1 km ved f.eks. Cond = 0,5 ville svare til en cost-distance på 2. At passere pixels med Cond = 0 resulterer i uendelig cost-distance og er derfor ikke muligt. På den måde bliver afstandsmål estimeret udelukkende igennem vandet mens fysiske barrierer, som f.eks. smalle indgange i fjorde, betyder en højere cost-distance.

På baggrund af de beregnede cost-distancer mellem målpunkt og match-up referencepunkterne tildeles hvert punkt en specifik vægt som afledes af en Gaussisk funktion (Figur 8). Vi valgte at basere de lokale regressioner på de ca. 250 nærmeste match-up måleværdier. Det nøjagtige antal af brugte match-up punkter er mindre i tilfældet af, at der ikke var 250 punkter med valide cost-distancer (f.eks. i en isoleret fjord). Men antallet kunne også være >250 i tilfældet af flere punkter med samme cost-distance (f.eks. fra samme NOVANA station). Bredden af den Gaussiske vægtfunktion er dynamisk og blev sat til middelværdien af cost-distancer for alle indgående NOVANA stationer. På den måde tilpasses vægtningen til den lokale rumlige fordeling af tilgængelige match-ups. Vi har illustreret vægtningsprocessen for NOVANA station 'VEJ0006870' i Figur 3.5 og Figur 3.6. Der findes 269 match-ups fra de tætteste ni stationer. Gennemsnit af cost-distancer for disse ni stationer ligger på 49,08, som så bliver bredden af den Gaussiske vægtfunktion. Dette eksempel viser også effekten af den valgte cost-distance som rumligt afstandsmål: cost-distancen igennem Lillebælt bliver forholdsvis stor pga. dens smalle passage. Derfor ligger de udvalgte nærmeste punkter alle nord for Lillebælt. På den måde kan kraftige gradienter, som ofte findes hen over smalle forbindelser mellem vandmasser, altså godt afspejles i regressionsmodellen.



Weighting Example: VEJ0006870

Figur 3.5. Illustration af den rumlige udvalg af match-up punkter ift. cost-distance fra målpunktet station 'VEJ0006870' (grøn). De ni inkluderede stationer (grøn & blå) repræsenterer i alt 269 (>250) match-up datapunkter.

Figur 3.6. Illustration af tildeling af regressionsvægter (w_{Reg}) ift. cost-distance fra målpunktet station 'VEJ0006870' (grøn). De ni inkluderede stationer (grøn & blå) repræsenterer 269 match-up punkter. Bredden af den Gaussiske vægtfunktion (Kernel width: w_{Kernel}) er sat til middelværdien af cost-distancer til målpunktet for de ni inkluderede stationer.

Distance-Weight Function



I Figur 37 vises der through-water cost-distance vægtede regressionsmodeller for fire udvalgte NOVANA stationer hen over studieområdet: DMU1044 (Nordsøen), VIB3708-00001 (Limfjorden), DMU403 (Kattegat), VEJ0006870 (Lillebælt). Der ses tydelige forskelle i punkternes placering ift. skyen af alle match-up punkterne. De tildelte vægter vises med varierende punktstørrelser. De grønne linjer repræsenterer de lokalt gældende regressionsfunktioner. På log-skala svarer deres hældning til Chl_{exp}, mens y-afskæring repræsenterer log(Chl_{fact}).

0.20

0.50

0.50

Distance-Weighted Regression Distance-Weighted Regression DMU1044 VIB3708-00001 200.0 200.0 $ChI = (a_{S3}^{pig})^{1.72*52}$ Chl = (a_{S3}^{pig})^{2.28}*202 50.0 50.0 20.0 20.0 Chl_{NOV} [µg/L] Chl_{NOV} [µg/L] 5.0 5.0 2.0 2.0 1.0 1.0 0.5 0.5 0.2 0.2 0.50 0.05 0.10 0.20 0.05 0.10 a₅₃^{pig} [1/m] a^{pig}₅₃ [1/m] Distance-Weighted Regression DMU403 Distance-Weighted Regression VEJ0006870 200.0 200.0 $ChI = (a_{S3}^{pig})^{2.08} \cdot 118$ $ChI = (a_{S3}^{pig})^{1.53} \pm 55$ 50.0 50.0 20.0 20.0 Chl_{NOV} [µg/L] Chl_{NOV} [µg/L] 5.0 5.0 2.0 2.0 1.0 1.0 0.5 0.5 0.2 0.2

0.05

0.10

0.20 a_{S3}^{pig} [1/m]

Regressionsmodellerne kan også bruges til at beregne modelrelateret usikkerhed igennem prediction standard error. Områder ind imellem de grønne linjer viser denne usikkerhed, som er afhængig af både regressionsfunktionens fitkvalitet og apig. Usikkerheden er mindst i centeret af den tildelte sky af referencepunkter og større ved lavere og højere værdier. Derfor får hvert eneste klorofylestimat tildelt sin egen specifikke usikkerhed at skaleringsfaktorerne er tidslig stationære. For at teste validiteten af denne antagelse har vi først anvendt modellen på alle match-up data i 2018-2019 (Figur 3.8). Til venstre vises der skyen af alle beregnede regressionsfunktioner mellem apig og Chl på baggrund af match-up punkterne, mens der til højre vises billedet efter anvendelsen af disse regressionsfunktioner. Bemærk: ved anvendelsen af skaleringsfaktorerne i ligning 1 bliver apig regnet om i et Chl koncentrationsestimat (x-aksen). Korrektionsmodellen fører til betragtelig forbedring ift. sammenhængen mellem S3-baserede klorofylestimater og in situ referenceværdier, som kan ses i stigningen af korrelationskoefficient R_{Pears} fra 0,63 til 0,69.

0.05

0.10

0.20

a^{pig} [1/m]

0.50

Figur 3.7. Visualisering af through-water cost-distance vægtet regressionsmodellering for 4 udvalgte NOVANA stationer henover studieområdet: DMU1044 (Nordsøen), VIB3708-00001 (Limfjorden), DMU403 (Kattegat), VEJ0006870 (Lillebælt). De grønne punkter er de ~250 nærmeste match-up punkter og deres størrelse illustrerer deres tilsvarende vægt i beregningen. De grønne linjer viser de lokalt tilpassede regressionsfunktioner sammen med usikkerhedsområdet på ±2*prediction standard errors. I baggrunden vises alle match-up punkterne i grå og den globale regressionsfunktion i orange.

Figur 3.9. Anvendelse af through-water cost-distance vægtet regressionsmodellering med alle match-ups fra 2018-2019, men for at estimere skaleringsfaktorer for match-up punkterne i 2020. Bemærk: ved anvendelsen af skaleringsfaktorerne i ligning 1 bliver a_{pig} (venstre x-akse) regnet om i et Chl koncentrationsestimat (højre x-akse).



Bagefter brugte vi så alle match-up fra 2018-2019 for at estimere skaleringsfaktorer for alle match-ups i 2020 (Figur 3.9). Vi kan se, at der fås en lignende forbedring af sammenhængen mellem S3-baserede klorofylestimater og *in situ* referenceværdier (R_{Pears} stiger fra 0,57 til 0,64), men her brugte vi helt uafhængige data fra de to tidligere år. Dette resultat viser, at modellen er ret stabilt hen over årene og kan anvendes operationelt for at opnå en signifikant forbedring i klorofylestimater fra S3s C2RCC klorofylprodukter.



For at få det mest optimale resultat brugte vi så til sidst alle match-up punkterne fra hele perioden 2018-2020 for at beregne den endelige regressions model (Figur 3.9). Ved brug af dette endnu større datasæt kunne vi opnå en forbedring af R_{Pears} fra 0,62 til 0,69. Fremadrettet brugte vi derfor alle disse matchup punkter for at anvende modellen på hele studieområdet.



Figur 3.10. Anvendelse af through-water cost-distance vægtet regressionsmodellering med alle match-ups fra 2018-2020. Bemærk: ved anvendelsen af skaleringsfaktorerne i ligning 1 bliver a_{pig} (venstre x-akse) regnet om i et Chl koncentrationsestimat (højre x-akse)

Figur 3.8. Anvendelse af through-water cost-distance vægtet regressionsmodellering på alle match-ups fra 2018-2019. Bemærk: ved anvendelsen af skaleringsfaktorerne i ligning 1 bliver a_{pig} (venstre x-akse) regnet om i et Chl koncentrationsestimat (højre x-akse)

3.2 Operationalisering af korrektionsmetoden (AU)

Operationaliseringen af korrektionsmetoden er sat op som en række trin og koder, der skal følges og køres. Pakken med alle nødvendige koder og filer kan fås fra A. Holbach efter anmodning. Vi har døbt hele processen for 'Complex Coast Sentinel-3 Chlorophyll Estimator' (CoCoa-SenCE). Proceduren skal kun køres helt igennem, når der skal suppleres med nye satellitdata, som endnu ikke er indeholdt i systemet. Lige nu indeholder pakken alle relevante Sentinel-3 data fra 2018-2021.

3.2.1 Download af \$3s OLCI level 2 produkt

Til download af S3 OLCI level 2 produkter bruger vi Copernicus CREODIAS portalen (https://finder.creodias.eu/), hvor der skal oprettes egen profil. På denne side kan der søges satellitprodukter for et afgrænset område ved hjælp af en *.kml fil. Her bruger vi en firkantet polygon, som indeholder Danmarks *exclusive economic zone* (EEZ), som findes i mappen 'CoCoa-SenCE/Download_and_Data-Handling/DK_AOI.kml'. Selve filen åbnes ved drag&drop på CREODIS finder. Så vælges det ønskede tidsrum i felterne 'observed' og tilpasses felterne nævnt i Tabel 3.1.

Tabel 3.1. Indstillinger på CREODIAS portalen (https://finder.creodias.eu/) til download af
S3 OLCI level 2 produkter.

Felt	Valg
Collection	Sentinel-3
Instrument	OLCI
Processing level	LEVEL2
Product type	WFR
Timeliness	Non Time Critical

Ved at trykke på 'Search' og bagefter 'copy all as urls' kan man gemme filnavne for alle tilsvarende satellitbilleder som *.txt fil. Som eksempel ligger der en fil for juni 2021 i mappen 'CoCoa-SenCE/Download_and_Data-Handling/ S3_OLCI_L2_DK_2021-06.txt'.

Nu abnes Python script 'Download_Sentinel3_CoCoa-SenCE.ipynb' i f.eks. Anaconda & Jupyter Notebook. I skripten tilpasses adgangsoplysninger ('password' og 'username') og fil forbindelser til ens eget system (felt: '#Adapt by user') og skripten køres for at downloade alle udvalgte satellitbilleder automatisk. For juni 2021 ligger tilsvarende *.zip-filer i mappen 'CoCoa-SenCE/Sen3_OLCI_L2_DK'. *.zip filerne unzippes og lægges ind i samme mappe.

3.2.2 Daglige composite billeder og ±5 dage glidende gennemsnit

For at beregne daglige gennemsnitsbilleder åbnes Python script 'MaskingAndDailyAverage_Sentinel3_500m.ipynb' i Anaconda & Jupyter Notebook. Der tilpasses filstier og datoerne for databehandling (felt: '#Adapt by user') og koden køres for at samle alle billeder fra samme dato ind i et 500x500 m grid henover de danske marine områder. Daglige composite billeder gemmes som GeoTiff filer i mappen 'CoCoa-SenCE/Sen3_DailyComposites'.

For at beregne et nyt sæt af ±5 dage glidende gennemsnit af de daglige composite billeder abnes R-script 'CoCoa-SenCE/Download_and_Data-Handling/MovingAvrgs.R'. I koden tilpasses så kun årrækken hvor der skal beregnes glidende gennemsnitsbilleder under '#Adapt by user', hvorefter koden kan køres automatisk. Vær dog opmærksom på, at eksisterende glidende gennemsnitsbilleder bliver overskrevet, hvis ikke tidsrummet tilpasses udelukkende til de nye Sentinel-3 data.

3.2.3 Korrektion og dataudtræk

Korrektionsmetoden og flere muligheder for dataudtræk er operationaliseret i form af tre R-shiny apps, som er moduler til

- 1. **kalibrering og validering** af metoden ('CoCoa-SenCE/App_Cali-Vali/app.R'),
- 2. systematisk **anvendelse** af korrektionsmetoden ('CoCoa-SenCE/App_Appli/app.R') og
- 3. **rumlige og tidslige dataudtræk** ('CoCoa-SenCE/App_Ext-ract/app.R').

For at kunne køre disse R-Shiny apps succesfuldt, skal der opsættes R og R-Studio på eget system med de følgende pakker: shiny, shinyjs, raster, terra, sp, gdistance, tmap, rgeos, rgdal, bfsl. Så åbner man det ønskede skript og klikker på 'Run App' for at starte den grafiske brugergrænseflade i web-browseren. En vejledning til at bruge modulerne findes i Appendix 1.

3.3 Optimering af DHIs model ved assimilering af satellitbaserede data (DHI)

Et af formålene med dette projekt er at integrere satellitdata og mekanistiske modeller via dataassimilering. Der tages udgangspunkt i lokalmodellen for det Nordlige Bælthav (NBS), som er en af modellerne i MSTs modelkompleks bag vandområdeplanerne 2021-2027 (se afsnit 0). Dataassimileringen sker i den biogeokemiske del af modellen, hvor Sentinel-3 satellitdata af klorofyl assimileres ind i modellen. Sentinel-3 satellit data af klorofyl er behandlet og optimeret af Aarhus Universitet som beskrevet ovenfor (se afsnit 3.1). Dataassimileringen i modellen sker via metoden Dataassimilering med Ensemble Kalman Filter. Herunder følger først en teoretisk beskrivelse af metoden, dernæst et afsnit med beskrivelse af den praktiske arbejdsgang omkring dataassimilering.

3.3.1 Dataassimilering med ensemble Kalman filteret

En model af et dynamisk system, eksempelvis et havområde, kan forbedres ved brug af måledata. Målinger kan forbedre modellen på forskellige måder; eksempelvis ved modelkalibrering hvor modellens parametre estimeres "offline" ved at gentage simuleringen og sammenligne modellens resultater med målinger. Integreres målinger i stedet løbende i modellen som en del af simuleringsprocessen, hvorigennem tilstanden estimeres, kaldes det dataassimilering. Bemærk skelnen mellem modellens parametre θ (≈indstillinger) og modellens tilstand, som for en havmodel er værdien af modellens prognostiske variable (f.eks. salt, temperatur, strøm og klorofyl) i alle beregningsceller samlet i en vektor x_k , hvor k angiver tidsskridtet.

Modellen

Modellen kan karakteriseres ved dens fremskrivningsoperator Φ_M som fremskriver tilstanden x_{k-1} ét tidskridt til tilstanden x_k :

$$x_k = \Phi_{\mathsf{M}}(x_{k-1}, u_k, \theta)$$

Her er θ modellens parametre (f.eks. vindfriktionskoefficient) og u_k er eksterne påvirkninger ofte kaldet forcering i form af randbetingelser og påvirkning fra atmosfæren (vind, lufttryk, nedbør etc). For havmodeller er tilstandsvektoren x typisk stor ($n \approx 1e^7$) og Φ_M ikke-lineær. Det er vanskeligt at assimilere data i en meget stor ikke-lineær model, så for bedre at forstå metoden, behandler vi først et simplere lineært system.

Kalman filter

Lad os betragte et simpelt lineært dynamisk system med en lille tilstandsvektor (tænk eksempelvis på et tog på et spor, hvor x indeholder 2 værdier: position og hastighed):

$$x_k = F x_{k-1} + B u_k + \varepsilon_k$$

Her er *F* fremskrivningsmatricen, u er forceringen, *B* relaterer forceringen til tilstanden (tænk eksempelvis på Bu_k som effekten af at trykke på speederen eller bremse). Endeligt repræsenterer ε_k processtøj, som vi antager er normalfordelt med middelværdi 0 og kovarians *Q*.

Systemet kan observeres ved

$$y_k = Hx_k + \delta_k,$$

hvor *H* er en observationsoperator (matrix i det lineære tilfælde), y_k er observationen og δ_k observationsstøjen, som antages normalfordelt med middelværdi 0 og kovarians *R*. Ofte gør *H* ikke andet end at "vælge" det rigtige element i tilstandsvektoren.

Den sande tilstand x_k kan desværre ikke bestemmes eksakt, men vi kan estimere den og vi betegner vores bedste estimat \hat{x}_k og den tilsvarende kovarians P_k . Vi kan fremskrive vores bedste estimat og dets kovarians given det foregående skridts \hat{x}_{k-1} og P_{k-1} , ved hjælp af modellen:

$$\hat{x}_k = F \ \hat{x}_{k-1} + B \ u_k$$
$$P_k = F P_{k-1} F^T + Q$$

Kalman Filteret (KF) (Kalman, 1960), som hjalp Apollo missionen til månen og i dag er integreret i alle navigationssystemer, kan hjælpe os med at finde den "bedste" løsning for dette system ved at kombinere den støjfyldte procesligning og den støjfyldte observation. Algoritmen består af to stadier: fremskrivning og analyse angivet med henholdsvis "f" og "a" herunder (?er udeladt herfra):

Fremskrivning:

$$x_k^f = F x_{k-1}^a + B u_k$$
$$P_k^f = F P_{k-1}^a F^T + Q$$

Analyse (når observationer er tilgængelige):

$$x_k^a = x_k^f + K(y_k - Hx_k^f)$$
$$P_k^a = P_k^f - KHP_k^f$$

Hvor Kalman gain matricen er givet ved

$$K = P^f H^T (HP^f H^T + R)^{-1}$$

Ovenstående ligninger udgør Kalman filteret for det tidligere beskrevet lineære system. Metoden benytter produktet af multivariate Gaussiske fordelinger til at give den optimale løsning. Løsningen er "optimal" i den forstand at den minimerer variansen, hvis følgende er opfyldt 1) modellen er lineær, 2) al støj er normalfordelt med middelværdi nul og 3) modellens og målingernes støj er uafhængige. Derfor kaldes den Best Linear Unbiased Estimator (BLUE).

Differensen (y - Hx) mellem observationen y og modellen i samme position Hx betegnes innovationen. Opdateringen er en lineær kombination af modellens fremskrivning (prior) og observationerne (likelihood).

Ensemble Kalman Filter

Uheldigvis er det ikke muligt at bruge Kalman Filteret for store ikke-lineære modeller som DHIs havmodeller, men man kan bruge metoder der tilnærmer sig Kalman Filter og bruger de samme ligninger til opdatering af tilstanden. Den centrale forskel er approksimationen af modelfejlkovariansen P, som beskriver modellens støj eller usikkerhed. Den kan ikke bestemmes eksakt for store ikke-lineære modeller. Ensemble Kalman Filteret (EnKF) (Evensen, 1994) tilnærmer P ved hjælp af en række forskellige modelrealisationer der tilsammen benævnes et ensemble. En enkelt modelrealisation i et ensemble kaldes et medlem.

Tilstandsvektorerne for alle medlemmer kan samles i en matrix *E* med størrelsen $n \times m$, hvor *n* er tilstandsvektorens størrelse (antallet af celler i havmodellen gange antallet af prognostiske variable) og *m* er antallet af medlemmer (eksempelvis 10 eller 20). Middeltilstandsvektoren betegnes \overline{x} og forskellen fra dette middel betegnes anomalimatricen:

$$A = E - \overline{x}$$

A er altså et udtryk for modellens "fejl" og derfor kan vi estimere modelfejlkovariansen P med

$$P = \frac{1}{m-1}AA^T$$

Det er en helt central antagelse i EnKF at P kan estimeres via anomalimatricen ved ovenstående formel. I praksis er det derfor også afgørende, hvordan modellens ensemble genereres. Sammenlignet med andre dataassimileringsmetoder, hvor der benyttes en simplere approksimation af P (eksempelvis "optimal interpolation" OI), giver EnKF den fordel at P her approksimeres ved hjælp af modellens egen dynamik og følgelig beskriver, hvordan modellens værdier i forskellige positioner og forskellige prognostiske variable "hænger sammen". Opdatering bliver derfor mere "fysisk" konsistent og kan bedre "accepteres" af modellen.

Tabel 3.2.		
Symbol	Beskrivelse	
Ε	Model ensemble (m medlemmer med tilstandsstørrelse n)	
x	Ensemble middelværdi	
Α	Ensemble anomalier $A = E - x$	
Р	Modelfejlskovariansen $P = \frac{1}{m-1}AA^T$	
у	Observationer (med p observationspunkter)	
H	Observationsmatricen (afbildning af observationer på tilstandsvektor)	
R	Kovariansen af observationsusikkerheden (diagonal hvis observationerne er	
	ukorrelerede)	
K	Kalman gain $K = PH^T(HPH^T + R)^{-1}$	

I Tabel 3.2 ses en opsummering af de vigtigste sym-boler der bliver brugt i de følgende afsnit.

EnKF algoritmen

Fremskrivningsstadiet i EnKF består af fremskrivning af de enkelte ensemblemedlemmer hver for sig:

Fremskrivning:

$$x_{-}(j,k)^{\wedge}f = \Phi_{M}(x_{j,k-1}^{a}, u_{j,k}, \theta_{j,k}), \quad for \ j = 1, ..., m$$
$$P_{k}^{f} = \frac{1}{m-1}A_{k}^{f}(A_{k}^{f})^{T}$$

Analyse: når observationer y_k er tilgængelige, kan modellen opdateres, hvilket sker for ensemblemiddel \overline{x} og anomalier *A* hver for sig:

$$\overline{x}_{k}^{a} = \overline{x}_{k}^{f} + K\left(y_{k} - H\overline{x}_{k}^{f}\right)$$
$$A_{k}^{a} = A_{k}^{f}T_{R}$$

Hvor K stadig er den velkendte Kalman gain matrix:

$$K = P^f H^T (HP^f H^T + R)^{-1}$$

og T_R er ensembletransformationsmatricen som skal konstrueres således at følgende er opfyldt:

$$T_R: \qquad P^a = P^f - KHP^f, \qquad A^a \mathbf{1} = \mathbf{0}$$

Hverken *P* eller *K* konstrueres fuldt i praksis, i stedet benyttes en række algebraiske omskrivninger som gør at problemet bliver mere effektivt at løse på en computer. I MIKE 3 FM benyttes udelukkende såkaldte square-root filtre (ESRF), som er deterministiske i modsætning til de oprindelige EnKF formuleringer (Evensen, 1994), som brugte stokastisk perturberede målinger. Dataassimileringsmodulet i MIKE 3 FM tillader forskellige EnKF algoritmer, men i dette projekt anvendes Ensemble Transform Kalman Filter (ETKF) (Bishop, 2001), hvor transformationsmatricen er symmetrisk:

$$T_R = [I + (HA^f)^T R^{-1} (HA^f)]^{-\frac{1}{2}}$$

Ensemblemodellering

Som tidligere nævnt er det afgørende vigtigt for EnKF at modellens ensemble repræsenterer modelusikkerheden. I praksis genereres en model ensemble ved at variere forskellige modelinput, men først er det nyttigt at spørge sig selv hvad modellens fejlkilder er:

- a. Begyndelsesbetingelsesfejl (initialbetingelser)
- b. Forceringsfejl (udefra kommende data) fra randbetingelser, meteorologiske forceringer (herunder lys), tilførsler, mm.
- c. Modelformuleringsfejl så som ligninger, diskretiseringsfejl, parametriseringer (dvs. modellen selv)
- d. Parameterfejl (indstillinger sat af modelløren) (eksempelvis bundfriktion, temperaturafhængigheder, væksthastigheder, græsningsrater mm.)

DA modulet i MIKE 3 FM understøtter beskrivelse af fejl i forceringer (b) og i udvalgte parametre (d). Begyndelsesbetingelsesfejl kan håndteres via præprocessering af modelløren. DA modulet understøtter ikke interne modelfejl (c) stammende fra de styrende ligninger, unøjagtige parametriseringer eller for grov modelopløsning.

Modelfejl

Modelfejl har ligesom modellen selv en tilstandsvektor. Modelfejlens tilstandsvektor kaldes ξ og indeholder en række tal svarende til den rummelige diskretisering af den pågældende modelfejl. Det kunne eksempelvis være en rummeligt varierende ændring af vindhastigheden over et modeldomæne – hvis vindhastighedsfejlen er diskretiseret på et net med 6x5 værdier, så har modelfejlens tilstandsvektor en størrelse på 30.



Modelfejlen er forskellig for hvert ensemblemedlem og kan interpoleres ned på modellens net, når den skal benyttes. Hvis en modelfejl i stedet skal variere en konstant modelparameter, vil dens tilstandsvektor have størrelsen 1.

Figur 3.11. Eksempel på modelfejlsnet med 6x5 værdier ovenpå et MIKE FM modelområde ud fra Norges kyst.

Modelfejlens fremskrivningsmodel

Typisk antages det at modelfejlen kan ændre sig over tid og at den ændring kan beskrives med en første ordens autoregressiv proces AR(1):

$$\xi_k = \Phi_{AR(1)}(\xi_{k-1}, \eta_k) = D\xi_{k-1} + \eta_k$$

Diagonalmatricen, *D*, indeholder koefficienterne for den autoregressive proces (karakteriseret ved den valgte tidslige halveringstid), og den tilførte støj η_k antages at være Gaussisk fordelt med middelværdi 0 og fejlkovarians Q_k^{η} , som er bestemt ud fra den valgte rummelige korrelationslængdeskala.



Den samlede EnKF algoritme i MIKE 3 FM

Opsummerende kan vi nu beskrive den endelige MIKE 3 FM dataassimileringsproces:

Fremskrivning: Både fejlmodellen og MIKE modellen fremskrives ét tidsskridt hvert medlem for sig ved:

$$\begin{split} \xi_{j,k}^{f} &= \Phi_{AR(1)} \big(\xi_{j,k-1}^{a}, \eta_{j,k} \big), \qquad for \, j = 1, \dots, m \\ x_{j,k}^{f} &= \Phi_{FM} \big(x_{j,k-1}^{a}, u_{j,k}, \theta_{j,k}, \xi_{j,k}^{f} \big), \qquad for \, j = 1, \dots, m \end{split}$$

Hvor de eksterne forceringer u og/eller modellens parametre θ modificeres af modelfejlene ξ .

Analyse: Når observationer y_k er tilgængelige, kan den samlede model (hvor tilstandsvektoren betegnes $z = [x, \xi]$) opdateres, hvilket sker for ensemble-middel \overline{z} og anomalier *A* hver for sig:

$$\overline{z}_{k}^{a} = \overline{z}_{k}^{f} + K(y_{k} - H\overline{z}_{k}^{f})$$
$$A_{k}^{a} = A_{k}^{f}T_{R}$$

Hvor Kalman gain matricen, K, modelfejlskovariansen, P_k^f , og transformationsmatricen, T_R , er defineret ved

$$K = P^{f}H^{T}(HP^{f}H^{T} + R)^{-1}$$
$$P_{k}^{f} = \frac{1}{m-1}A_{k}^{f}(A_{k}^{f})^{T}$$

Figur 3.12. Eksempel på en AR(1) proces beskrivende variabiliteten af vindhastighedsfejlen i et punkt som funktion af tiden. Værdierne, som varierer omkring nul, bruges til at perturbere vindhastigheden i modellen.

$$T_R = [I + (HA^f)^T R^{-1} (HA^f)]^{-\frac{1}{2}}$$

Metoden benyttes således til løbende, når nye observationer bliver til rådighed, at forbedre modellens beskrivelse af systemets tilstand. Men da antallet af observationspunkter p typisk er lille i forhold til størrelsen n af modellens tilstandsvektor x, er dette ikke en nem opgave. Og da vi ydermere ofte kun har "råd" til en lille ensemblestørrelse m, er der typisk brug for nogle numeriske "tricks" for at lykkes i praksis. En lille ensemblestørrelse fører til en prøvetagningsfejl (en: sampling error) som blandt andet ses i modelfejlskovariansen P, som beskriver sammenhænge på tværs af rum og mellem forskellige variable, ved at områder langt væk fra hinanden kan se korrelerede ud uden de er det i den virkelige verden. Her kan man benytte lokalisering, som begrænser størrelsen på modelfejlskovariansen P gradvist i takt med at afstanden bliver større.

Opsætning af dataassimilering med ensemble Kalman Filter i Nordlige Bælthav modellen

Til dataassimileringen anvendes Sentinel-3 satellit-klorofyldata behandlet og optimeret af Aarhus Universitet (se afsnit 3.1). Satellitdataene assimileres ind i den biogeokemiske del af den mekanistiske model for Nordlige Bælthav (Figur 3.13 og afsnit 0).



Data er konverteret fra NetCDF til grid-data med fast tidsskridt på 24 timer. De daglige data er placeret kl. 12:00. I tilfælde hvor det oprindelige datasæt havde flere tidsskridt i samme døgn, er der taget et middel. Til dataassimileringen blev der først afprøvet et satellit-datasæt med ±5 dages glidende gennemsnit, hvilket gav en god datadækning, men til gengæld begrænsede modellens dynamik, fordi der i perioder blev assimileret mod samme værdi i op

Figur 3.13. Placering af modeldomænet Nordlige Bælthav i de indre danske farvande (område med tyk linje) samt det anvendte udsnit af satellitdata. til 10 dage. Satellit-datasættet er derfor efterfølgende udskiftet til daglige data uden ±5 dages glidende gennemsnit. Det resulterer i mindre datadækning på de enkelte dage, men til gengæld får modellen nu lov til tage sig af udviklingen mellem observationerne, hvilket den med sin interne dynamik forventes at kunne gøre mere realistisk.

Inden den egentlige dataassimilering kan påbegyndes, skal der skabes en stokastisk variation i modellens ensemble-medlemmer. Herefter starter dataassimileringen og dermed arbejdet i at finjustere hvor kraftigt dataassimileringen tager fat i samspil med bevarelse af modellens stabilitet. Disse to trin er beskrevet nedenfor.

Stokastisk variation og ensemble medlemmer

En væsentlig del af den dataassimilering, som ligger bag rutinerne implementeret i MIKE 3 FM, ligger i at skabe en variation i ensemble-medlemmerne omkring den centrale kalibrering. Variationen introduceres med en stokastisk variation, en såkaldt modelfejl, som beskriver usikkerheden på den enkelte modelparameter, der undersøges. I MIKE 3 FM kan der introduceres modelfejl på modelparametre, randbetingelser mm. (se afsnit 3.3.1 "Ensemblemodellering").

Usikkerheden på en modelparameter beskrives med en standardafvigelse, og for at sikre at den ikke bliver negativ eller urealistisk høj, kan øvre og nedre grænser defineres. Udbredelsesparameteren beskriver hvor stor variationen skal være over tid - sættes den lavt (sekunder, minutter og timer) varierer konstanten meget mellem hvert tidsskridt og sættes den højere (dage og uger) er den tidslige variation lavere.

Variationen tillægges kalibreringskonstanten additivt eller multiplikativt:

- Additiv: Modelkonstant + variationen
- Multiplikativ: Modelkonstant * (1+variationen)

I nærværende projekt har vi identificeret 8 modelkonstanter i den biogeokemiske model, som gav en passende variation i modellens klorofylkoncentration med 10 ensemble medlemmer (Tabel 3.3 og Figur 14).

En række andre kalibreringskonstanter som typisk anvendes for at justere den biogeokemiske models klorofylniveau er også undersøgt, men blev vurderet til at have mindre/ingen effekt på variationen af klorofylkoncentrationen herunder f.eks.:

Tabel 3.3. Introduktion af stokastisk variation	n på udvalgte modelkonstanter.
---	--------------------------------

Modelkonstant	Værdi	Standard afvigelse	Øvre og nedre grænse for variationen	Udbredelses parameter	Additiv eller multiplikativ
Fytoplankton vækstrate, kiselalger (1/d)	3	0,5	0,75 til 9.0	7 dage	Multiplikativ
Fytoplankton vækstrate, grønalger (1/d)	3	0,5	0,75 til 9.0	7 dage	Multiplikativ
Fytoplankton sedimentationshastighed (m/d)	0,3	0,5	0,03 til 0,57	7 dage	Multiplikativ
Zooplankton maks. græsningsrate (1/d)	1,5	0,5	0,15 til 2,85	7 dage	Multiplikativ
Zooplankton dødsrate, 1. orden (m3/g C/d)	6	0,5	0,6 til 11,4	7 dage	Multiplikativ
Zooplankton dødsrate, 0. orden (1/d)	0.0075	0,5	0,00075 til 0,014	7 dage	Multiplikativ
Detritus mineraliseringsrate (1/d)	0,2	0,5	0,02 til 0,38	7 dage	Multiplikativ
Detritus sedimentationshastighed (m/d)	1,5	0,5	0,15 til 2,85	7 dage	Multiplikativ

- Temperaturafhængighed for fytoplanktonets vækstrate (StDev = 0,1; 7 dages udbredelse; Additiv)
- Silikat halvmætningskoncentration for optag i kiselalger (StDev = 0,5; 7 dages udbredelse; Multiplikativ).

Ud over kalibreringskonstanter har solindstråling og randværdier stor betydning for fytoplanktonets produktion og dermed klorofylkoncentrationen, og disse data er forbundet med en vis grad af usikkerhed. Det er derfor også forsøgt at inkludere stokastisk variation ift. solindstrålingen samt koncentrationen af ammonium, nitrit+nitrat og fosfat på modellens 3 rande:

- Solindstråling 2D kort med tidslig variation på 1 time (StDev = 0,5; 7 dages udbredelse; Multiplikativ). En stokastisk variation på solindstrålingen gav en stor variation i klorofylkoncentrationen. Men selv efter en del forsøg på at justere på indstillingerne forblev modellens resultat meget ustabil, og derfor ikke anvendt mere i dette projekt.
- De 3 randes (Lillebælt, Storebælt og Kattegat) koncentrationer af ammonium, nitrit+nitrat og fosfat (StDev = 2; 7 dages udbredelse; Multiplikativ). Undersøgt hver for sig og samlet. Med en høj standardafvigelse på 2 var der synlig effekt på variationen i klorofylkoncentrationen, men det var på bekostning af en meget høj og ustabil variation af de tre næringsstoffer, og derfor ikke anvendt mere i dette projekt.

Den stokastiske variation af ensemble medlemmerne er først undersøgt ved såkaldte Free-run kørsler, som er kørsler uden dataassimilering. Ved en Freerun kørsel uden modelfejl og med 1 ensemble-medlem fås samme resultat som i en normal kørsel (baseline). Ved at tilføje modelfejl og øge antallet af ensemble-medlemmer til f.eks. 10, skabes, afhængigt af de anvendte modelfejl, en stokastisk variation blandt ensemble-medlemmerne. Free-run kørsler anvendes til at vurdere om der er skabt en tilpas variation omkring den gennemsnitlige tilstand af de parametre der er i fokus, f.eks. klorofylkoncentrationen, som i dette projekt.

Det ses, at der ved de valgte modelfejl (Tabel 3.3) er skabt en variation i de tilsvarende modelkonstanter (Figur 3.14), som tilsammen skaber variationen i ensemble-medlemmernes klorofylkoncentration (Figur 3.15). Variationen er generelt størst i perioder hvor fytoplanktonets produktion ikke er begrænset af lys (vinter) eller næringsstoffer (sommer).


Figur 3.14. De stokastiske variationer hos 10 ensemble medlemmer (grå linjer) samt deres gennemsnit (sort linje) for de 8 anvendte modelkonstanter (Tabel 3.3).



Figur 3.15. 10 ensemble-medlemmers (State 1-10, lysegrå linjer) variation i overflade-klorofylkoncentrationen ved en af de centrale kalibreringsstationer, samt deres gennemsnit (mørkegrøn linje) og koncentrationen i baseline kørslen uden stokastisk variation (lysegrøn linje). Desuden vises måledata (røde firkanter).

Specifikke indstillinger til dataassimilering med Ensemble Kalman Filter Hvor kraftigt dataassimileringen virker, kan justeres med et antal forskellige indstillinger. En af de vigtigste er målingernes usikkerhed (standard afvigelse). Denne parameter er afhængig af selve målingens usikkerhed (inklusive repræsentationsusikkerhed ift. modellens gridceller), men også i høj grad en kalibreringsparameter, idet den bl.a. også afhænger af observationshyppighed og -dækning, og endvidere skal afstemmes i forhold til modelfejlene.

Satellittens usikkerhed kan analyseres med regressionsmodellen. For Nordlige Bælthav-modelområde er der ikke stor forskel på usikkerheden i de enkelte pixels, og satellitdatas usikkerhed er analyseret samlet for området (Figur 3.16). For at afspejle den større usikkerhed af satellitdata ved høje klorofylkoncentrationer, er der i modellen anvendt en standardafvigelse på 0,5 op til 12 µg/l, hvorefter standardafvigelsen stiger lineært som funktion af klorofylkoncentrationen (Figur 3.16). Det betyder i praksis at lave værdier i satellitdata assimileres kraftigere end de højere værdier.

Hvis målingen er meget forskellig fra modellen (f.eks. fejlagtigt datapunkt), kan dataassimileringsopdateringen blive urealistisk stor og potentielt føre til at modellen bliver ustabil eller går ned (stopper). Det er derfor muligt at se bort fra målepunkter med stor afvigelse fra modellen med en specificeret tærskel. I dette projekt er anvendt en tærskelværdi på 15 µg/l.

Figur 3.16. Satellitdatas 95 %konfidensinterval (y-akse: µg/L) beregnet for estimerede klorofylkoncentrationer (x-akse) i det Nordlige Bælthav modeldomæne (venstre). Den i modellen anvendte måleusikkerhed (standardafvigelse) som funktion af satellitdatas klorofylkoncentration(højre).



Hvor kraftigt dataassimileringen virker, afhænger også af hvor længe vi tillader at dataassimileringen varer omkring hver observation. Satellitdata er meget begrænset og unøjagtig om vinteren. Derfor assimileres der kun i perioden fra 1. marts til 31. oktober. Hver observation anvendes i modellen i et "vindue" omkring målingens tidspunkt. I dette studie har vi anvendt et vindue på 4 timer. Øges vinduet bliver assimileringen kraftigere. Herudover kan der indstilles på tidsskridt mellem hver assimilering, og i dette studie har vi anvendt et tidsskridt på 6 gange modellens overordnede tidsskridt på 5 min. dvs. 30 min. Assimileres der i hvert tidsskridt over et længevarende vindue, bliver modellen meget tung at køre - sættes tidsskridtet for højt, bliver assimileringen mindre kraftig.

Ved anvendelse af Ensemble Kalman Filter vil assimilering ofte reducere ensemblespredningen for meget, dvs. at den stokastiske variation bliver meget lille. Denne effekt kan minimeres ved anvendelse af R-faktor-anomalier. En R-faktor-anomali større end 1 virker ved at øge måleusikkerheden ved opdatering af anomalierne i forhold til ved opdatering af ensemblemiddelværdien. Dette giver en større spredning imellem de enkelte ensemble-medlemmer. I dette projekt er anvendt en værdi på 2,0.

Modellens tilstandsvariabel for klorofyl er sammen med måledata log-transformeret før og efter assimilering, for at undgå at modellens klorofyl bliver negativ.

Figur 3.17. Eksempel på effekten af lokalisering. Satellitdata fra en dag (16-03-2018) med lav dækningsgrad (venstre). Modellens opdatering af klorofyl-tilstandsvariablen, hvor effekten af lokalisering med en horisontal korrelationslængdeskala på 5 km tydelig ses (højre). Den rumlige støj i fejl-kovariansen, der ofte omtales som falske langdistancekorrelationer, kan reduceres ved såkaldt lokalisering, hvor effekten af opdateringen gradvis mindskes væk fra observationspunktet ved at gange med en konisk funktion. I dette projekt er der anvendt en Gaussisk funktion og en horisontal korrelationslængdeskala på 5 km (Figur 3.17). Effekten af at køre uden lokalisering eller med f.eks. en øget korrelationslængdeskala på 20 km er også undersøgt. Dette gav dog et meget mere ustabilt resultat. En helt ny udvikling inden for dataassimileringsmodulet er anvendelsen af vertikal lokalisering, hvor effekten af assimileringen gradvis mindskes i vertikal retning væk fra observationspunktet. Dette er dog så nyt, at det ikke har kunnet inddrages i dette projekt, men er noget, der vil give mening at undersøge nærmere ved evt. efterfølgende projekter.

4 Resultater

4.1 2D Satellitbaserede overfladeklorofylprodukter på flere rumlige og tidslige opløsninger (AU)

I metodeafsnittet viste vi anvendelse af through-water distance-vægtet regressionsmodelleringen på alle match-up punkterne mellem det S3-baserede klorofylprodukt og *in situ* referenceværdier fra NOVANA programmet i tidsrummet 2018-2020. Men modellen kan lige så godt anvendes alle andre steder. Formålet med denne del af Projektet er at få et arealdækkende modelprodukt for at korrigere for S3-baserede klorofylprodukter hen over hele studieområdet i de danske marine farvande. Derfor beregnede vi en through-water costdistance matrix mellem alle reference match-up punkter og alle pixels på det definerede 500x500 m raster hen over studieområdet. På baggrund af denne distance-matrix kunne vi beregne estimater for begge skaleringsfaktorer og hver pixel.

Hovedproduktet af den nye korrektionsmetode er to rumlige modeller af skaleringsfaktorerne Chl_{exp} og Chl_{fact} (Figur 4.1). Vi valgte at anvende en maske baseret på en konveks polygon rundt om alle match-up punkterne med en buffer på 20 km. Ude for dette område ville estimaterne være usikre pga. ekstrapolering i områder med manglende dækning af *in situ* målinger. I kortene kan der ses tydelige rumlige mønstre i udbredelsen af skaleringsfaktorerne. Sammenhængene med miljømæssige forhold og skaleringsfaktorerne bliver diskuteret i kapitel 5.1.

De to rumlige modeller for skaleringsfaktorerne kan herefter anvendes til at beregne korrigerede S3-baserede klorofylestimater på forskellige skalaer i rum og tid ved brug af ligning 1. Flere muligheder er blevet implementeret i det operationelle system, som vises i det følgende. Ift. til økologisk tilstandsvurdering af vandområder er vækstsæsonsgennemsnit i klorofylkoncentration et afgørende måltal (MST, 2021). Øverst i Figur 4.2 vises der som eksempel vækstsæsonsgennemsnit (maj-september) af klorofylkoncentrationen i 2020.

Bemærk at nogle kystnære områder forbliver hvide pga. anvendelse af masken for optisk lavvandede områder (se kapitel 3.1.1). Den relative forskel mellem det ukorrigerede og korrigerede S3 produkt, $Rel_{dev} = (Chl_{S3}^{Raw} - Chl_{S3}^{Corr})/Chl_{S3}^{Corr}$, vises nederst i Figur 4.2. Her kan der ses, at det ukorrigerede produkt har en systematisk tendens til at overestimere klorofylkoncentrationer i Nordsøen og Kattegat/Skagerrak, mens en systematisk underestimering sker inde i Limfjorden, Bælthavet, og Østersøen. Stort set afspejler den viste relative forskel det omvendte rumlige mønster af gennemsnitsklorofylkoncentrationer øverst i Figur 4.2. Høje klorofylkoncentrationer betyder både kraftig pigment-packaging og en forskellig pigment-sammensætning i algesamfundet og kan så forklare det observerede mønster.



Scaling Factor log(Chl_fact)



Figur 4.1. Rumlige modeller af de to skaleringsfaktorer Chl_{exp} og Chl_{fact} som resultat af through-water cost-distance vægtet lineær regressionsmodellering. Disse værdier bruges efterfølgende til at korrigere de S3 baserede klorofylestimater. Cirkler er NOVANA stationerne, der indgik i beregningen.

Chlorophyll S3 (Corr.) Average for Year: 2020; Months: 5-6-7-8-9



Chlorophyll S3 (Rel. Deviation) Average for Year: 2020; Months: 5-6-7-8-9



Figur 4.2. Korrigeret S3 baseret klorofylprodukt for vækstsæsonsgennemsnit (marts-september) [µg/L] henover de danske marine farvande i 2020 (øverst). Relativ afvigelse mellem det rå og det korrigerede produkt (nederst). NOVANA stationerne, der indgår i beregningerne er vist som cirkler. I Nordsøen og Kattegat/Skagerrak forekommer der systematisk overestimering af klorofyl med det oprindelige S3 C2RCC produkt, mens der i områderne Limfjorden, Bælthavet og inde i Østersøen er systematisk underestimering.

Disse sammenhænge og deres effekter på den klorofylspecifikke absorptionskoefficient diskuteres nærmere i afsnit 5.1. I Figur 4.3 vises der en sammenstilling med tilsvarende vækstsæsonsgennemsnit værdier for alle NOVANA stationer og årene 2018-2020 uden og med anvendelse af den distance-vægtede regressionsmodel. De korrigerede værdier giver sammenhænge meget tættere på 1:1 linjen og med signifikant højere korrelationskoefficienter (R). Disse resultater er standardudtræk af det operationelle system og kan beregnes for brugerdefinerede tidslige sammenstillinger af værdier.



Figur 4.3. Sammenstilling af vækstsæsonsgennemsnit (månederne 5-9) i klorofylkoncentrationer [μ g/L] ud fra ukorrigerede (Raw: orange/rød) og korrigerede (Corr: blå) S3-baserede estimater (Chl_{S3}) med *in situ* baserede værdier (Chl_{NOV}) i årene 2018-2020. De korrigerede værdier giver en sammenhæng tættere på 1:1 linjen (hældning: S) og med højere Pearson korrelationskoefficienter (R). Denne sammenstilling er et standardudtræk af det operationelle system og kan beregnes for en række tidslige sammenstillinger af værdier.

En anden mulighed med det operationelle system er at lave dataudtræk på baggrund af rumligt definerede vandområder, såsom WFD-, HELCOM- og OSPAR -områder (assessment units). I Figur 4.4 vises den korrigerede vækstsæsongennemsnit (maj-september) for WFD-områderne i årene 2018-2020. Baggrundsdata kan dog også trækkes ud som *.csv fil med estimater for standardafvigelsen og rumlig dækning inden for hvert område.



Figur 4.4. Vækstsæsongennemsnit (5-9) [µg/L] af korrigerede S3 klorofylværdier for WFD-områderne i årene 2018-2020.

Korrektionsmodellen kan også anvendes på de ±5 dage glidende gennemsnitsbilleder. På den måde kan der ekstraheres specifikke tidsserier for alle punkter i studieområdet. I Figur 4.5 vises der f.eks. tidsserierne for de før udvalgte fire NOVANA stationer (Figur 3.7) fra 2018-2021, sammen med de tilsvarende *in situ* referenceværdier, den modelbaserede usikkerhed, og ukorrigerede S3 klorofyldata. Denne tidsserieudtræk er en funktion implementeret i det endelige operationelle system og kan laves for alle 103 NOVANA stationer, der indgår i regressionsberegningerne. Tidsseriedata kan også trækkes ud som *.csv fil.

Foruden at udvikle et operationelt system til korrektion og anvendelse af klorofyl værdier ud fra Sentinel-3, handlede nærværende projekt om at vise anvendeligheden af satellitdata i DHIs model igennem dataassimilering. Til dette formål har vi beregnet daglige estimater hen over studieområdet for 2018 med en rumlig opløsning på 500x500 m opløsning. DHI fik to datasæt at arbejde med; (1) et baseret på korrigerede daglige composite billeder og (2) et baseret på ±5 dages glidende gennemsnit. Disse datasæt er så blevet assimileret i DHIs model (se afsnit 3.3) og resultaterne er beskrevet i følgende kapitel.



Figur 4.5. Ekstraherede tidsserier for fire udvalgte NOVANA stationer hen over studieområdet (se de tilsvarende kalibreringer i Figur 3.7): DMU1044 (Nordsøen), VIB3708-00001 (Limfjorden), DMU403 (Kattegat), VEJ0006870 (Lillebælt). De grønne linjer viser ±5 dage glidende gennemsnit og den modelbaserede usikkerhed (Confidence Interval: CI) af estimaterne for korrigerede klorofylkoncentrationer. Orange linjer viser ukorrigerede S3 data og punkterne repræsenterer NOVANA *in situ* målinger. Tidsrum på grå baggrund indeholder de første og sidste 60 dage af året og indgik ikke i kalibrering af modellen.

4.2 Dataassimilering af satellit klorofyldata i biogeokemiskmodel (DHI)

Dataassimileringen baseret på Ensemble Kalman filteret (se afsnit 3.3), medfører at assimilerede observationer af en enkelt parameter eller tilstandsvariabel potentielt kan påvirke alle andre tilstandsvariable i hele modeldomænet. I det følgende ser vi på den direkte effekt af dataassimilering af satellitbaseret klorofyl samt de afledte effekter på lyssvækkelseskoefficienten og næringsstoffer. Stationen VEJ0006870 er anvendt som en central station i arbejdet med dataassimilering, derudover vises resultater fra 8 andre stationer i modeldomænet (Figur 27). Bemærk at stationerne ARH170117, FYN6940622 VEJ0004273 og VEJ0006489 ikke er dækket af det korrigerede S3-satellitprodukt pga. masken for optisk lavvandede områder.

Satellitdata dækker perioden 01-03-2018 til 31-10-2018. For at skabe spredning blandt ensemble-medlemmerne startes modellen 01-01-2018 og kører året ud, på trods af der ikke er data til assimilering de sidste to måneder. De statistiske analyser i følgende afsnit er baseret på perioden med satellitdata.





4.2.1 Effekt på klorofyl a

Ved dataassimilering af satellitklorofyl justeres modellens klorofylkoncentration mod de observerede værdier, således at modellens resultat nærmer sig det observerede mønster (Figur 4.7). Det ses af figuren at de høje modelværdier den 20. marts 2018 i de to vestlige fjorde (Vejle og Horsens Fjord) samt den nordlige bugt (Århus bugt) reduceres med dataassimileringen hvorimod området rundt om Samsø har en øget klorofylkoncentration.





Modellens resultat af dataassimilering reflekterer det observerede mønster fra satellitten, dog uden at de meget høje værdier helt kystnært (de røde farver) observeret fra satellitten er assimileret direkte ind. Ved at anvende Ensemble Kalman filter assimileres der også i områder med huller i satellitdata, såfremt området befinder sig inden for den definerede lokalisering med horisontal korrelationslængdeskala på 5 km (se afsnit 3.3.1 "Specifikke indstillinger til dataassimilering med Ensemble Kalman Filter").

Med de valgte indstillinger for dataassimileringen, som beskrevet i afsnit 3.3.1 "Opsætning af Dataassimilering med Ensemble Kalman Filter i Nordlige Bælthav modellen" giver dataassimileringen en forholdsvis kraftig assimilering mod satellitdata, men dog lidt mindre kraftigt mod de højere værdier af satellitdata (Figur 4.8 til Figur 4.10). En risiko ved dataassimilering er at modellens resultater bliver ustabile eller at modellen går ned (modellen stopper). Der ses da også kortere perioder med ustabilitet, hvor modellens klorofylkoncentration pludselig stiger voldsomt, uden at det er skabt af høje observerede værdier (f.eks. station FYN6940622 i starten af juli). Modellen går dog ikke ned og finder hurtigt tilbage i stabilt leje, men der er plads til yderligere justering af dataassimileringens indstillingsmuligheder.

Det ses desuden, at den stokastiske variation i ensemble-medlemmerne reduceres betydeligt i perioder med god datadækning, hvorimod der skabes større spredning blandt ensemble-medlemmerne i perioder med lavere dækningsgrad, eftersom modellen får 'lov' til at udvikle sig mere frit i disse perioder.



Figur 4.8. Effekten af dataassimilering (Mean state, mørkegrøn linje) på klorofylkoncentrationen sammenlignet med baseline uden dataassimilering (lysegrøn linje). Desuden vises de 10 ensemble-medlemmers variation (State 1-10, lysegrå linjer) i klorofylkoncentrationen, satellitdata anvendt til dataassimilering (blå trekanter) samt NOVANA-måledata anvendt til validering (røde firkanter).

Med ophør af dækningen af satellitdata, og dermed også assimilering, i modellen den 31-10-2018 finder modellen tilbage til baseline-tilstanden. I november og december, hvor der ikke assimileres, ses en generel ringe overensstemmelse mellem model og valideringsmålinger fra NOVANA.

Som en ligefrem effekt af dataassimilering reduceres modellens RMSE og BIAS sammenholdt med satellittens observationer ved otte af ni stationer (Figur 4.12). Ved stationen i den smalle Vejle Fjord (VEJ0004273) er der ikke dækning fra satellitten, og den indgår derfor ikke i statistikken her, selvom modelresultaterne viser effekt af dataassimileringen også i Vejle Fjord (Figur 4.10, figur anden øverst).

Analyseres effekten af dataassimilering mod NOVANAS målte klorofyl a, som her anvendes som valideringsdata, falder RMSE på syv stationer og er uændret på to stationer. AT RMSE ikke ændrer sig på FYN6700053 (Figur 4.12), som ligger tæt på randen i Storebælt, skyldes at målingerne fra stationen også anvendes som randdata på den rand. BIAS mindskes på fire stationer, øges på fire stationer og er også her uændret på stationen nær randen til Storebælt. Det er interessant at bemærke at stationen i Vejle Fjord (VEJ0004273), som ikke er dækket af satellitdata, får en forbedret beskrivelse af klorofylkoncen-trationen med dataassimilering idet både RMSE og BIAS reduceres.



Figur 4.9. Effekten af dataassimilering (Mean state, mørkegrøn linje) på klorofylkoncentrationen sammenlignet med baseline uden dataassimilering (lysegrøn linje). Desuden vises de 10 ensemble-medlemmers variation (State 1-10, lysegrå linjer) i klorofylkoncentrationen, satellitdata anvendt til dataassimilering (blå trekanter) samt NOVANA-måledata anvendt til validering (røde firkanter).



Figur 4.10. Effekten af dataassimilering (Mean state, mørkegrøn linje) på klorofylkoncentrationen sammenlignet med baseline uden dataassimilering (lysegrøn linje). Desuden vises de 10 ensemble-medlemmers variation (State 1-10, lysegrå linjer) i klorofylkoncentrationen, satellitdata anvendt til dataassimilering (blå trekanter) samt NOVANA-måledata anvendt til validering (røde firkanter). Station VEJ0004273 ligger i den smalle Vejle Fjord og er derfor ikke dækket af satellitdata.



Figur 4.11. RMSE (øverst) og BIAS (nederst) beregnet for satellit-klorofyldata anvendt til assimilering sammenlignet med modellens resultat for baseline (sort bjælke) og kørsel med dataassimilering (EnKF14c, blå bjælke). Bemærk enheden er mg/l x 10⁻³.



Figur 4.12. RMSE (øverst) og BIAS (nederst) beregnet for klorofyl a valideringsdata fra NOVANA sammenlignet med modellens resultat for baseline (sort bjælke) og kørsel med dataassimilering (EnKF14c, blå bjælke). Bemærk enheden er mg/l x 10⁻³.

4.2.2 Effekt på lyssvækkelseskoefficienten

Klorofylkoncentrationen påvirker direkte lysets svækkelse i vandet sammen med en række andre lyssvækkende parametre. I modellen er lyssvækkelseskoefficienten beskrevet som summen af effekten fra klorofyl a, partikulært organisk kulstof, farvet opløst organisk kulstof (CDOC), partikulært uorganisk suspenderet stof samt en resterende baggrundslyssvækkelse fra bl.a. vandet selv. En forbedret beskrivelse af modellens klorofylkoncentration må altså forventes også at give en forbedret beskrivelse af lyssvækkelseskoefficienten. Effekten vil dog være afhængig af i hvor høj grad klorofylkoncentrationen styrer lyssvækkelsen i forhold til de andre lyssvækkende parametre.

Som ventet påvirker den ændrede klorofylkoncentration lyssvækkelsen (Figur 4.13 og Figur 4.14). Generelt giver dataassimileringen en lidt bedre beskrivelse af lyssvækkelsen i forhold til valideringsdata fra NOVANA, og RMSE mindskes på syv ud af ni stationer (Figur 4.15). Dog øges BIAS på syv ud af ni stationer.



Figur 4.13. Effekten af dataassimilering (Mean state, mørkegrøn linje) på lyssvækkelseskoefficienten sammenlignet med baseline uden dataassimilering (lysegrøn linje). Desuden vises NOVANA-måledata anvendt til validering (sorte firkanter).



Figur 4.14. Effekten af dataassimilering (Mean state, mørkegrøn linje) på lyssvækkelseskoefficienten sammenlignet med baseline uden dataassimilering (lysegrøn linje). Desuden vises NOVANA-måledata anvendt til validering (sorte firkanter).





Figur 4.15. RMSE (øverst) og BIAS (nederst) beregnet for valideringsdata fra NOVANA af lyssvækkelseskoefficienten sammenlignet med modellens resultat for baseline (sort bjælke) og kørsel med dataassimilering (EnKF14c, blå bjælke).

4.2.3 Effekt på næringsstoffer

Ved dataassimilering med Ensemble Kalman filter opdateres ikke alene modellens tilstandsvariabel for klorofyl men også andre af modellens afledte variable. Som tidligere nævnt opdateres der heller ikke kun i punktet med data, men i hele modelområdet både horisontalt og vertikalt, hvor dataassimileringsmetoden finder en korrelation til observationspunktet. Indtil videre har vi fokuseret på effekten af assimilering af klorofyl i overfladen. I dette afsnit kigger vi kort på effekten af afledte variable i form af nitrat, total kvælstof (TN), fosfat og total fosfor (TP) såvel ved overfladen som ved bunden.

Der ses tydelige effekter på nitrat, total kvælstof, fosfat og total fosfor, her eksemplificeret med station VEJ0006870 (Figur 4.16 og Figur 4.17). Resultatet antyder dog ikke umiddelbart en forbedret beskrivelse af disse parametre, nærmere tværtimod. Målet i nærværende projekt har været at forbedre klorofylresultaterne, og resultaterne for næringssaltene har ikke været i fokus. Der skal fremadrettet arbejdes videre med at få en bedre sammenhæng mellem dataassimilering af klorofyl og samspillet med næringsstofferne, som for nuværende påvirkes af dataassimileringen, men ikke i alle situationer på en hensigtsmæssig måde.



Figur 4.16. Effekten af dataassimilering (Mean state overflade: Mørkegrøn linje og Mean state bund: Mørkeblå linje) på nitrat, total kvælstof, fosfat og total fosfor sammenlignet med baseline uden dataassimilering (overflade: Lysegrøn linje og bund: lyseblå line). Desuden vises NOVANA-måledata anvendt til validering (overflade: Sorte firkanter og bund: røde trekanter).



Figur 4.17. RMSE (øverst) og BIAS (nederst) beregnet for målinger sammenlignet med modellens resultat for baseline (sort bjælke) og kørsel med dataassimilering (EnKF14c, blå bjælke). I nævnte rækkefølge er der vist resultater for nitrat, total kvælstof, fosfat og total fosfor.

5 Diskussion og perspektiver

5.1 Økologisk betydning af skaleringsfaktorerne brugt i optimeringen af satellitbaseret klorofyl (AU)

Vores resultater viser tydelige rumlige mønstre for begge skaleringsfaktorer Chl_{exp} og Chl_{fact} anvendt til at korrigere de rå S3-klorofylprodukter (Figur 4.1). Disse mønstre antages at hænge sammen med de lokale økologiske forhold, især koncentrationen af næringsstoffer og lys, som påvirker fytoplanktonets sammensætning og fysiologi (Jakobsen og Markager, 2016). Ratioen mellem S3 a_{pig} og klorofylkoncentrationen afspejler fytoplanktonsamfundets *in vivo* klorofyl-specifikke absorptionskoefficient a* ved 443 nm:

$$a^*(443 nm) = \frac{a_{pig}^{443}}{Chl} \left[\frac{m^2}{mg}\right]$$
 Ligning 5

Ligning 1 (S. 10) beskriver sammenhængen mellem a_{pig} og Chl igennem begge skaleringsfaktorer som hældning (Chl_{exp}) og y-afskæring (log(Chl_{fact})) af de tilsvarende lineære regressionsmodeller på log-skala (Figur 3.10). Tilpasning af skaleringsfaktorerne Chl_{exp} og Chl_{fact} påvirker endvidere de estimerede værdier for a* (443 nm) på følgende måde:

$$a^*(443 nm) = \frac{a_{pig}^{443}}{Chl} = \frac{Chl^{\left(\frac{1}{Chl_{exp}}-1\right)}}{\frac{1}{Chl_{fact}}} \left[\frac{m^2}{mg}\right]$$
 Ligning 6

Figur 5.1 viser den beregnede fordeling af gennemsnitlige a* (443 nm) over gennemsnitsklorofylkoncentrationer (marts-oktober) i årene 2018-2020. Den blå linje i figuren viser sammenhængen ved brug af default skaleringsfaktorerne ($Chl_{exp} = 1.04 \& Chl_{fact} = 21$), mens den orange linje repræsenterer en empirisk modelfunktion for selve sammenhængen ved 444 nm fundet af Stæhr og Markager (2004). Det kan ses, at vores korrektionsmodel med rumligt tilpassede skaleringsfaktorer fører til en fordeling af a* (443 nm) meget tæt på resultatet i Stæhr og Markager (2004), mens defaultsættet af skaleringsfaktorerne kun repræsenterer minimale ændringer i a* (443 nm) og ligger langt væk fra en realistisk fordeling.

Figur 5.1. Fordeling af klorofylspecifikt absorptionskoefficient ved 443 nm over S3-baserede estimater for klorofylgennemsnitskoncentrationer (marts-oktober) i 2018-2020 (grønne punkter). Den orange linje viser en tilsvarende model funktion (Stæhr og Markager, 2004) for 444 nm baseret på *in vivo* målinger i prøver fra åbne, kystnære og fjord vandområder. Den blå linje viser sammenhængen når der bruges default sættet af skaleringsfaktorer (Chl_{exp} = 1.04 & Chl_{fact} = 21). Specific Pigment Absorption Coefficient @ 443 nm



Pérez et al. (2021) beskriver pigment-packaging og pigmentsammensætning, som de to afgørende variabler i fytoplanktonsamfundet mht. til sammenhængen mellem a* og klorofylkoncentrationen. Pigment-packaging er igen afhængig af den intracellulære klorofylkoncentration og cellens størrelse. Produktet af den gennemsnitlige cellediameter og den intracellulære klorofylkoncentration (indikator for pigment-packaging) har tidligere vist en signifikant negativ korrelation med a* i danske farvande (Stæhr et al., 2004). Flere forskellige miljøparametre, men især lys og tilgængelighed af næringsstoffer påvirker fytoplanktoncellernes klorofylindhold (Stæhr et al., 2004). Næringsstoftilgængelighed i kystnære områder og især inde i fjordene er større og leder til dannelse af store celler med en høj intracellulær klorofylkoncentration. Det fører til selvskygning af de enkelte klorofylpigmenter og så til en mindre klorofyl-specifik absorptionskoefficient.

Vi har sammenlignet det rumlige modelresultat for gennemsnitlig a* (443 nm) med NOVANA-baserede gennemsnitsestimater for TN og TP for alle inkluderede NOVANA stationer i perioden 2018-2020. Figur 5.2 viser disse sammenhænge. For begge næringsstoffer kan vi se signifikante, men ikke-lineære, negative sammenhænge. Dette resultat er i overensstemmelse med den ovennævnte hypotese og viser at høj tilgængelighed af næringsstoffer fører til mindre a* (443 nm). **Figur 5.2.** Sammenhæng mellem a* (443 nm) og NOVANA-baserede gennemsnitsestimater for TN, TP for alle inkluderede NO-VANA stationer i perioden 2018-2020.



For årene 2018-2020 fik vi adgang til NOVANA programmets fytoplankton data (pers. komm. Hans Jacbosen). Disse data brugte vi så til at beregne estimater for gennemsnitlig biovolumen, biomasse, cellediameter (Equivalent Spherical Diameter: ESD) og produktet af ESD og den interne klorofylkoncentration (c_i) for hver station og over hele tidsrummet. Dette gjorde vi, da regressionsmodellen for nuværende er tidslig stationær. Der var 10 fytoplankton NOVANA stationer der overlapper med det korrigerede S3 klorofylprodukt efter anvendelse af masken for optisk lavvandede områder. I Figur 5.3 vises der en sammenstilling mellem gennemsnitlig a* (443 nm) og disse parametre. Både biovolumen og biomasse viser en tydelig negativ sammenhæng, lige som der kunne ses for f.eks. TN-koncentration i Figur 5.2. Cellediameter og dens produkt med ESD (pigment-packaging) viser dog ingen klare sammenhænge.



Figur 5.3. Sammenhæng mellem a* (443 nm) og NOVANA-baserede gennemsnitsestimater for fytoplankton biovolumen, biomasse, cellediameter, og dens produkt med celleintern klorofylkoncentration (proxy for pigmentpackaging) for alle stationer i perioden 2018-2020.

Figur 5.4. Sammenhæng mellem a* (443 nm) og NOVANA-baserede gennemsnitsestimater for fytoplanktonklassernes relative andele af hele samfundet for alle stationer i perioden 2018-2020.



De to stationer med mindste pigment-packaging værdier ligger i Skive og Mariager Fjord, som også viser de to største biomassekoncentrationer. Som nævnt ovenfor ville vi have forventet at se en negativ korrelation mellem pigment-packaging og a* (443 nm).

Pigmentsammensætningen af fytoplanktonsamfundet er også relateret til dets artsammensætning (f.eks. Henriksen m.fl., 2002). Derfor har vi undersøgt sammenhængen mellem gennemsnitlige estimater for a* og de gennemsnitlige relative andele af dominerende fytoplanktongrupper (her diatomeer (kiselalger) og dinophyter (flagellater)) i samfundet (Figur 5.4). Kiselalger og flagellater er også de fytoplanktongrupper, der tages højde for i DHIs biogeokemiske model (se kapitel 2.3.4). Vi ser to modsatte, men statistisk ikke signifikante (p>0.1), sammenhænge med en tendens til mindre andele af kiselalger og højre andele af flagellater ved stigende værdier for a* (443 nm). Disse to trends ville passe sammen med dem som er vist i Figur 5.1. Diatomeer er kendt for at foretrække høje næringsstofkoncentrationer, hvor de så kan vokse hurtigt og danne store biomasser, mens auto- og heterotrophe dinophyter ofte trives ved lavere eksterne koncentrationer af næringsstoffer (Griffiths m.fl., 2016). Sammenhængene med arlige gennemsnitsværdier fra fytoplanktonprøverne fremstår dog ikke helt så klare som for vandkemien. Grundene kan være, at der (1) kun er 10 overlappende stationer til rådighed, et statistisk ikke repræsentativ sæt og, at (2) fytoplanktonprøverne repræsenterer en integreret vandprøve for de øverste 10 m af vandsøjlen, mens satellitten kun ser signalet fra det øverste lag (ca. 1-5 m, afhængig af lyssvækkelsen). Især under forhold med udprægede vertikale gradienter forventes der altså signifikant afvigelse mellem fytoplanktonprøver og satellitobservationer. Hyppig lagdeling er meget sandsynlig i de indre fjorde som Mariager og Skive Fjord.

Samlet set viser resultaterne dog tydelige sammenhænge mellem resultaterne af den rumlige korrektionsmodel og de lokale økologiske forhold (næringsstof- og klorofylkoncentrationer). Den rumlige fordeling af skaleringsfaktorerne og de dermed estimerede klorofylspecifikke absorptionskoefficienter vurderer vi derfor til potentielt at kunne fungere som indikatorer for vandets økologiske tilstand, i stedet for at de udelukkende bruges til estimering af klorofylkoncentrationer.

Fremover bør det også undersøges, hvor vidt korrektionsmodellen kan udvides til også at repræsentere sæsonmæssige variationer af de økologiske forhold. Dette vil kræve en repræsentativ rumlig og tidslig dækning med matchups mellem satellit og *in situ* data. Herved kan de rumlige cost-distancer udvides med en komponent som repræsenterer den tidslige distance, hvorved skaleringsfaktorerne kan beskrives dynamisk i både rum og tid. Vi kan for nuværende ikke vurdere, om dækningsgraden med match-up data er tilstrækkeligt i hele studieområdet til en sådan analyse.

En anden mulighed kunne være at estimere korrektionsmodellens skaleringsfaktorer på baggrund af et andet distancemål, f.eks. baseret på økologisk similaritet. Dvs. at man vægter punkterne i regressionsmodellen mht. deres similaritet (f.eks. ift. salinitet, TN, TP, batymetri, isolationsgrad fra det åbne hav, osv.) i stedet for deres rumlige afstand fra hinanden. På den måde kunne korrektionsmodellen muligvis forbedres især under forhold med udprægede rumlige gradienter. Derudover vil antallet af match-ups med høje vægter ved rumlig isolerede steder forøges, idet der ville findes andre lignende steder, som for nuværende ligger langt væk fra hinanden i den cost-distance baserede model.

5.2 Korrektionsmodellens afhængighed af klorofyl referencedata (AU)

Den udviklede model for korrektionen af satellitbaserede klorofyldata afhænger af input af in situ klorofyl referencedata for at kunne kalibrere begge skaleringsfaktorer Chlexp og Chlfact med cost-distance vægtet regressionsmodellering. Med den nuværende rumlige og tidslige densitet af referencedata fra NO-VANA programmet var det muligt at opnå en signifikant forbedring af de satellitbaserede klorofylestimater. Vi kunne vise, at kalibrering af Chlexp og Chlfact med data fra 2018&2019 har en lige så god forbedringseffekt på data fra 2020. Derfor antager vi at skaleringsfaktorerne ikke ændrer sig nævneværdigt over tid. Spørgsmålet er dog, på hvilken tidsskala der sker signifikante ændringer i f.eks. lokal eutrofieringsstatus, som så vil kunne påvirke modellen. Pga. den korte periode af tilgængelige S3 satellitdata (fra 2017) er det endnu ikke muligt at vurdere det optimale antal foregående år til kalibrering af modellen. Men vi kan undersøge effekten af antallet af indgående referencedata på modelresultatet. Derfor har vi tilfældigt undladt en stigende andel af *in situ* referencedata fra 2018-2020 fra regressionsberegningen. Et mindre antal af in situ referencedata vil ændre på de lokale estimater af skaleringsfaktorerne og dermed også påvirke de satellitbaserede klorofylestimater ifølge ligning 1.

Figur 5.5 viser udviklingen af Pearson-korrelationskoefficienten mellem Chlexp, log(Chlfact) og Chl estimeret på baggrund af 1) det fulde in situ referencedatasæt og 2) et reduceret datasæt i reduktionstrin af 10% af alle referencedata. For hvert reduktionstrin er der blevet beregnet 10 scenarier med tilfældigt udvalgte reduktioner. Som forventet kan der for alle tre parametre ses en tydelig aftagende effekt på korrelationskoefficienten når man anvender færre referenceprøver. For alle tre parametre gør effekten sig gældende allerede ved en reduktion på 20 % af referenceprøverne, den er dog mest udpræget for Chl_{fact}. Dvs. at en reduktion af *in situ* reference prøver fremadrettet ville have konsekvenser for at kunne beregne gode nye rumlige modeller for skaleringsfaktorerne. Især med hensyn til mulige yderligere forbedringsmuligheder af korrektionsmodellen, til f.eks. også at repræsentere den tidslige dynamik, vil vi endnu ikke anbefale at bruge dette resultat til at mindske overvågningsindsatsen for in situ klorofylprøver. Det gælder om først at undersøge og udnytte mulighederne med den nuværende in situ data dækningsgrad for at udvikle de bedst mulige modeller til satellitbaseret klorofylovervågning. I princippet vil der på sigt være et potentiale for at reducere in situ prøvetagningen og i stedet anvende satellit estimater.

I praksis er det dog tvivlsomt, om det giver en relevant fordel, da sejlads og prøvetagning alligevel skal udføres af hensyn til andre parametre i overvågningen. Det er således kun selve udgiften til laboratorieanalysen man vil spare. Derfor skal man nok se satellitbaserede estimater af klorofyl som en metode der har et stort potentiale for at forbedre datagrundlaget. Dog gælder det, at hvis kombinationen af satellitbaserede estimater for klorofyl og assimilering af data i modeller, i kombination kan gøre, at modellerne kan give en interpolation i tid og rum af alle vandkemiske parametre, så vil man kunne reducere prøvetagningen.

Chl exp 1.0 0.8 0.6 R_{Pears} with Full Datasel 0.4 0.2 0.0 -0.2 100 % 90 % 70 % 60 % 50 % 30 % 20 % 80 % 40 % 10 % Random Fraction of Data Used log(Chl_fact) 1.0 --0.8 R_{Pears} with Full Dataset 0.6 0.4 0.2 100 % 90 % 80 % 70 % 60 % 50 % 40 % 30 % 20 % 10 % Random Fraction of Data Used Chl 1.00 0.98 0.96 RPears with Full Dataset 0.94 0.92 06.0 0.88 0.86 100 % 90 % 80 % 70 % 60 % 50 % 40 % 30 % 20 % 10 % Random Fraction of Data Used

Figur 5.5. Effekten af trinvis tilfældig reduktion af match-up datasættet for den cost-distance baserede regressionsmodel for de to skaleringsfaktorer Chl_{exp} (øverst), Chl_{fact} og korrigerede Chl estimater (nederst). Der vises boxplots af Pearson-korrelationskoefficienter mellem værdierne på baggrund af det fulde datasæt og et reduceret datasæt i 10 % reduktionstrin. For hvert reduktionstrin er der beregnet 10 tilfældige reduktionsscenarier.

5.3 Dataassimilering i mekanistiske modeller (DHI)

I dette projekt har vi arbejdet med at udvikle og teste forskellige metoder til dataassimilering af satellitbaserede klorofylmålinger i mekanistiske biogeokemiske modeller. Dataassimilering er ikke trivielt og kræver en række parameterindstillinger, som både adresserer hvordan variabiliteten blandt ensemble-medlemmerne skabes, hvilke parametre der assimileres, hvor kraftigt der assimileres mod data, i hvor lang en periode der assimileres mod data og i hvor stort et område der assimileres med fuld styrke.

Vi har med projektet vist, at de udviklede rutiner har potentiale til at assimilere satellit data, og at den parameter (her klorofyl), der assimileres, kan tilpasses i modellen, således at både RMSE og BIAS overordnet set reduceres. Det er naturligvis især gældende på de stationer, hvor der eksisterer satellitdata til assimilering, men selv i mere lukkede områder, som f.eks. Vejle Fjord, reduceres fejlen i modellen ved sammenligning med NOVANA målinger for assimileringsperioden 01-03-2018 til 31-10-2018.

I nærværende projekt har der primært været fokus på klorofyl. Der har kun kort været kigget på afledte effekter på andre variable, som lys (Kd) eller næringsstoffer (uorganisk og organiske). Projektet har dog vist, at dataassimileringen ikke alene påvirker den assimilerede variabel (her klorofyl), men også påvirker afledte variable. Mens lys imidlertid blev forbedret en anelse sammenlignet med målinger, viste næringssaltene imidlertid ikke de samme positive tendenser på alle stationer.

Metoden til dataassimilering i MIKE 3 FM er endnu ny, og er i første omgang udviklet til brug i hydrodynamiske modeller. Den biogeokemiske model bygger imidlertid på en række mere komplicerede sammenhænge, og der er derfor behov for en fortsat tilpasning af de udviklede rutiner til assimilering af biologiske parametre som klorofyl og biogeokemien.

I en evt. forsættelse af projektet er det derfor en klar anbefaling at fortsætte med assimilering af klorofyl, men samtidig fokusere på afledte effekter på andre variable og gradvist at supplere med assimilering af flere parametre, som f.eks.:

- Assimilering af iltmålinger i bundvand: Især bundkoncentrationer af næringssalte påvirkes af varierende iltforhold, og derfor vil den samlede beskrivelse af næringsstofferne muligvis kunne forbedres igennem assimilering af ilt.
- Assimilering af lys: Direkte målinger af lys kunne få betydning for både den samlede økologiske tilstand samtidigt med at fordeling af lys i vandsøjlen potentielt vil kunne påvirke klorofyl i flere dybder.
- Assimilering af uorganiske næringssalte: Da især DIN er begrænsende i sommerperioden, vil en forbedret beskrivelse af DIN koncentrationer i både overflade- og bundvandet potentielt påvirke modellens evne til at modellere klorofyl i vandsøjlen.

6 Konklusion

I nærværende projekt har vi (1) operationaliseret en ny metode til optimering af satellitbaserede klorofylmålinger for de komplekse og kystnære danske marine områder og (2) ved eksemplet af lokalmodellen for det Nordlige Bælthav udviklet og testet rutiner til assimilering af disse nye klorofyldata i DHIs mekanistiske bio-geokemiske model.

Projektets resultater viser, at der er et stort potentiale ved anvendelse af satellitbaserede data til overvågning af klorofylkoncentrationen og udbredelse af fytoplankton i danske farvande. Desuden viser vi, at man med fordel kan forbedre mekanistiske modeller ved assimilering af satellitbaserede klorofyldatasæt. Projektet afdækker også en række potentielle fokusområder for fremtidigt forskningsarbejde, som bør adresseres for at opnå den bedst mulige merværdi af disse nye teknologier i den nationale overvågning af de danske marine vandområder:

- 1) Korrektionsmodellen for satellitbaserede klorofyldata viser tydelige sammenhænge med de lokale økologiske forhold (næringsstof- og klorofylkoncentrationer). Derfor vurderer vi, at korrektionsmodellen også kunne fungere som indikator for vandets økologiske tilstand i stedet for, at den udelukkende bruges til estimering af klorofylkoncentrationer. Fremover bør det også undersøges, om korrektionsmodellen kan udvides til at repræsentere sæsonmæssige variationer af de økologiske forhold, eller om andre distancemål, som f.eks. økologisk similaritet, i regressionsproceduren kunne optimere modellens resultater i endnu højere grad.
- 2) Ved siden af at arbejde videre på at optimere assimileringsrutiner for klorofyl i DHIs model, bør der samtidig fokuseres på afledte effekter på andre variable. Processernes beskrivelse i modellen kunne forbedres på flere fronter ved gradvist at supplere med assimilering af; (a) iltmålinger i bundvand, for at forbedre den samlede beskrivelse af næringsstofferne; (b) lysmålinger, der har betydning for den samlede økologiske tilstand og klorofylkoncentrationer i flere dybder; (c) målinger af uorganiske næringssalte i både overflade- og bundvandet, der potentielt vil påvirke modellens evne til at modellere klorofyl i vandsøjlen.

7 Referencer

Bishop CH, Etherton B, Majumdar SJ, 2001. Adaptive sampling with the ensemble transform Kalman filter. part I: theoretical aspects. Mon. Wea. Rev., 129, 420-436.

Brockmann C, Doerffer R, Peters M, Stelzer K, Embacher S, Ruescas A, 2016. Evolution of the C2RCC neural network for Sentinel 2 and 3 for the retrieval of ocean colour products in normal and extreme optically complex waters. Living Planet Symposium, ESA Special Publication 740, 54.

Charlton M, Fotheringham S, 2009. Geographically Weighted Regression. White Paper. National Centre for Geocomputation, National University of Ireland Maynooth.

DHI, 2013. MIKE 21/3 Ecological Modelling. MIKE 21/3 ECO Lab FM module. Short description. DHI Water Environment Health, Hørsholm, Denmark, 14 pp. Retrieved January 6, 2013, from: http://www.dhisoftware.com/Download/DocumentsAndTools/~/media/Microsite_MIKEbyDHI/Publications/PDF/Short%20descriptions/MIKE21 3_FM_EL_Short_Description.ashx.

DHI, 2014. DHI 3 Algae and Sediment Model. ECO Lab Template. ScientificDescription. DHI Water Environment Health, Hørsholm, Denmark, 134 pp.

DHI, 2017. MIKE 21 & MIKE 3 Flow Model FM. Hydrodynamic and Transport Module. Scientific Documentation http://manuals.mikepoweredbydhi.help/2017/Coast_and_Sea/MIKE_321_FM_Scientific_Doc.pdf

Doerffer R, 2010. OLCI Level 2 Algorithm Theoretical Basis Document Ocean Colour Turbid Water. Document Ref: S3-L2-SD-03-C11-GKSS-ATBD. url: https://step.esa.int/docs/extra/OLCI_L2_ATBD_Ocean_Colour_Turbid_Water.pdf.

Droop MR, 1968. Vitamin B12 and marine ecology IV: the kinetics of uptake, growth and inhibition in Monochrysis lutheri. J Mar Biol Ass UK 48, 689-733.

Effer SW, 1988. Secchi disk transparency and turbidity. Journal of environmental engineering 114(6), 1436-1447.

Erichsen AC, Rasch PS, 2001. Two- and Three-dimensional Model System Predicting the Water Quality of Tomorrow. Proc. of the Seventh International Conference on Estuarine and Coastal Modelling. Spaulding ML (Ed), American Society of Civil Engineers 2001.

Erichsen AC, Birkeland M, 2019. Development of Mechanistic Models. Mechanistic Model for Northern Belt Sea. Hydrodynamic model documentation. DHI rapport (projekt nr. 11822245).

Erichsen AC, Birkeland M, 2020. Development of Mechanistic Models. Mechanistic Model for Northern Belt Sea. Technical documentation on biogeochemical model. DHI rapport (projekt nr. 11822245). Evensen G, 1994. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte-Carlo methods to forecast error statistics, J. Geophys. Res. 99, 10143–10162.

Gholizadeh H, Robeson SM, 2016. Revisiting empirical ocean colour algorithms for remote estimation of chlorophyll-a content on a global scale. International Journal of Remote Sensing, 37(11), 2682-2705, doi: 10.1080/01431161.2016.1183834.

Griffiths JR, Hajdu S, Downing AS, Hjerne O, Larsson U, Winder M, 2016. Phytoplankton community interactions and environmental sensitivity in coastal and offshore habitats. Oikos 125, 1134-1143, DOI: 10.1111/oik.02405.

Griffiths JR, Kadin M, Nascimento FJA, Tamelander T, Törnroos A, Bonaglia S, Bonsdorff E, Brüchert V, Gårdmark A, Järnström M, Kotta J, Lindegren M, Nordström MC, Norkko A, Olsson J, Weigel B, Žydelis R, Blenckner T, Niiranen S & M Winder, 2017. The importance of benthic-pelagic coupling for marine ecosystem functioning in a changing world. Global Change Biology 23:2179-2196. doi:10.1111/gcb.13642.

Haney JD, Jackson GA, 1996. Modeling phytoplankton growth rates. Journalof Plankton Research 18, 63-85.

Henriksen P, Riemann B, Kaas H, Munk Sørensen H, Lang Sørensen H, 2002. Effects of nutrient-limitation and irradiance on marine phytoplankton pigments. Journal of Plankton Research 24(9), 835–858, DOI: 10.1093/plankt/24.9.835.

Hirst, Sheader, 1997. Are in situ weight-specific growth rates body-size independent in marine planktonic copepods? A re-analysis of the global syntheses and a new empirical model. MEPS 154, 155–165.

Holbach A, Upadhyay Stæhr S, Upadhyay Stæhr P, Markager S, 2022. Optimization of chlorophyll concentration retrieved from Sentinel-3 OLCI through spatially resolved estimation of empirical scaling factors. *In preparation*.

Jacobsen HH, Markager S, 2016. Annual dynamics of carbon to chlorophyll ratio in temperate coastal waters. Limnology and Oceanography 61, 1853-1866, DOI: 10.1002/lno.10338.

Kalman RE, 1960: A new approach to linear filtering and prediction problems. J. Basic. Eng. 82, 35-45.

Kratzer S, Plowey M, 2021. Integrating mooring and ship-based data for improved validation of OLCI chlorophyll-a products in the Baltic Sea. International Journal of Applied Earth Observation and Geoinformation 94, 102212, DOI: 10.1016/j.jag.2020.102212.

Kirk JTO., 2000. Light and photosynthesis in aquatic ecosystems. Cambridge University Press 2. Edition 2000, pp. 401.

Kiørboe T, Møhlenberg F, Hamburger K, 1985. Bioenergetics of the planktonic copepod Acartia tonsa: relation between feeding, egg production and respiration, and composition of specific dynamic action. Mar. Ecol. Prog. Ser. 26, 85-97.

Kiørboe T, Nielsen TG, 1994. Regulation of zooplankton biomass and production in a temperate, coastal ecosystem. I. Copepods. Limnol Oceanogr 39, 493-507.

Kuusemäe K, Rasmussen EK, Vergés PC, Flindt MR, 2016. Modelling stressors on the eelgrass recovery process in two Danish estuaries. Ecological Modelling 333, 11-42.

Kyryliuk D, Kratzer S, 2019. Evaluation of Sentinel-3A OLCI Products Derived Using the Case-2 Regional CoastColour Processor over the Baltic Sea. Sensors 19, 16, 3609, doi: 10.3390/s19163609.

Lessin G, Raudsepp U, 2006. Water quality assessment using integrated modeling and monitoring in Narva Bay, Gulf of Finland. Environmental Modelling and Assessment 11, 315-33.

Miljøstyrelsen (MST), 2021. Retningslinjer for vurdering af overvågningsresultater og klassificering af tilstand i kystvande, Vandområdeplan 2021 – 2027. URL:

https://mst.dk/media/222319/vurderingtilstandkystvandevp3.pdf.

Morel FMM, 1987. Kinetics of nutrient uptake and growth in phytoplankton. J. Phycol. 23, 137-150.

Pedersen MF, Borum J, 1996. Nutrient control of algal growth in estuarine waters: nutrient limitation and the importance of nitrogen requirements and nitrogen storage among phytoplankton and species of macroalgae. Mar. Ecol. Prog. Ser. 142, 261–272.

Pérez GL, Galí M, Royer S-J, Gerea M, Ortega-Retuerta E, Gasol JM, Marrasé C, Simó R, 2021. Variability of phytoplankton light absorption in stratified waters of the NW Mediterranean Sea: The interplay between pigment composition and the packaging effect. Deep Sea Research Part I: Oceanographic Research Papers 169, 103460, doi: 10.1016/j.dsr.2020.103460.

Rasmussen EK, Petersen OS, Thompson JR, Flower RJ, Ayache F, Kraiem M, Chouba L, 2009. Model Analyses of the Future Water Quality of the Eutrophicated Ghar El Melh Lagoon (Northern Tunisia). Hydrobiologia 622, 173-193, DOI: 10.1007/s10750-008-9681-9.

Stæhr PA, Henriksen P, Markager S, 2002. Photoacclimation of four marine phytoplankton species to irradiance and nutrient availability. Marine Ecology Progress Series 238, 47-59, DOI: 10.3354/meps238047.

Stæhr PA, Markager S, 2004. Parameterization of the chlorophyll a-specific in vivo light absorption coefficient covering estuarine, coastal and oceanic waters. International Journal of Remote Sensing 25(22), 5117-5130, DOI: 10.1080/01431160410001716932.

Stæhr PA, Markager S, Sand-Jensen K, 2004. Pigment specific in vivo light absorption of phytoplankton from estuarine, coastal and oceanic waters. Marine Ecology Progress Series 275, 115–128, DOI: 10.3354/meps275115.

Toming K, Kutser T, Uiboupin R, Arikas A, Vahter K, Paavel B, 2017. Mapping Water Quality Parameters with Sentinel-3 Ocean and Land Colour Instrument imagery in the Baltic Sea. Remote Sensing, 9(10), 1070, doi: 10.3390/rs9101070.

Upadhyay S, Markager S, Erichsen AC & Closter RM, 2021. North Sea Monitoring of chlorophyll. Environmental monitoring and assessment based on satellite data, mechanistic modelling and in situ sampling – performance and perspectives. Monitoring related to the Marine Strategy Framework Directive. DHI report (projektno. 11824515).

Appendix 1 Complex Coast - Sentinel-3 Chlorophyll Estimator (CoCoa-SenCE)

Manual for the operationalized system CoCoa-SenCe to optimize Sentinel-3 based chlorophyll estimates for the complex Danish marine waters (AU)

This manual describes an operationalized system to apply a distanceweighted linear regression method on the Sentinel-3 OLCI level-2 product for ocean colour in turbid water, publicly available on the Creodias platform of the Copernicus programme. This new method enables the spatially continuous estimation of two scaling factors, required to derive chlorophyll (Chl) concentrations from the algorithm's output on the inherent optical property (IOP) of phytoplankton pigment absorption at 443 nm (a_{pig}):

$$Chl = a_{pig}^{Chl_{exp}} * Chl_{fact}$$

The two scaling factors are called Chl-Exponent (Chl_{exp}) and Chl-Factor (Chl_{fact}) and are $Chl_{exp} = 1.04$ and $Chl_{fact} = 21$ by default (Doerffer et al., 2010). These two scaling factors are, however, known to be variable, in particular in complex marine environments with considerable gradients across several water quality parameters. Therefore, we developed a method to estimate the spatial distribution of these scaling factors across the Danish marine waters in order to improve the retrieval of Chl concentrations from the Sentinel-3 satellite Chl products. The theoretical background of this method, as well as required data pre-processing is described in the main report.

Overview of the operational process

The operational process consists of 2 main steps:

- 1. Sentinel-3 OLCI level-2 data download, daily composite resampling, and ±5 days moving average composite calculation
- 2. Complex Coast Sentinel-3 Chlorophyll Estimator (CoCoa-SenCE). The package of files and codes are available from A. Holbach (anho@ecos.au.dk) on request.
 - a. Calibration & Validation Module
 - b. Application Module
 - c. Extraction Module ('CoCoa-SenCE/App_Extract/app.R')

The first main step includes manual identification of relevant satellite images, as well as automatized download and daily composite resampling through Python scripts. This is followed by an automatized calculation of temporarily smoothed ± 5 days moving average composite S3 images through an R-script. The process is described in section 3.2 of the main report.

The second main step is facilitated through three modules of graphical user interfaces, programmed as R-Shiny applications. These can be run in an R-environment and we recommend use of R-Studio. The following R-packages are required to successfully run the R-Shiny apps: shiny, shinyjs, raster, terra, sp, gdistance, tmap, rgeos, rgdal, bfsl.

CoCoa-SenCE modules

Calibration & validation module

To run the app, open the file 'CoCoa-SenCE/App_CaliVali/app.R' in R-Studio and press 'Run App' to start the user interface in a web browser.

User interface and options

In this module, the user can inspect the default calibration and validation procedure for the cost-distance weighted regression modelling for optimizing Sentinel-3 based chlorophyll estimates, as well as calculate customized calibration and validation scenarios. The latter, however, is only recommended for advanced users!



Figure M.1. Screenshot of Graphical User Interface of the CoCoa-SenCE Calibration & Validation Module.

Chose Calibration

 Default: Year ranges, included months, and number of nearest neighbors used in the regression modelling are preset and cannot be changed. The system will load the pre-processed set of calibration/validation points, which speeds up the processing.

- **Custom (Advanced Users Only!):** Year ranges, included months, and number of nearest neighbors used in the regressing modelling can be adapted by the user.
- Select Processing Type
 - **Calibration:** This function is used to calibrate the distanceweighted regression modelling procedure for 1) visualization of different scenarios and/or 2) later usage in the **Application Module**.
 - Validation: This function is used to validate the distanceweighted regression modelling procedure, where the dataset is split into a calibration and a validation dataset. The outputs are only used for visualization/exploration. To make a new calibration effective for later processing, processing type must be set to Calibration.
- Selected Calibration Year-Range: For the Default calibration, the years 2018-2020 are used. For the default validation, the years 2018-2019 are used for calibration. Under Custom Calibration, the year ranges can be adapted by the user.
- Selected Validation Year-Range: This option is only available for the Validation Processing Type. For the **Default** validation, 2020 is used for validation.
- Selected Months: For the Default calibration and validation, months 3-10 are used. Under Custom calibration and validation, the included months can be adapted by the user.
- No. of Nearest Neighbors: For the Default calibration and validation, 250 nearest neighbors are used for the regression modelling. Under Custom calibration and validation, this number can be adapted by the user.
- **Refresh NOVANA Data Selection:** When the calibration and/or processing type are changed, it is necessary to refresh the data selection to make the changes effective.
- **Refresh Sentinel-3 Data:** This option is only available for the 'Custom' calibration. It is necessary to extract respective Sentinel-3 data for the custom selection of NOVANA data. This process involves a lot of data and takes some time!
- **Calculate Distance-Weighted Regression:** Distance-weighted regression is calculated for the specific data selection. This process involves a cost-distance calculation between all pairs of points, and a subsequent linear regression modelling for each point. In particular, the cost-distance calculation takes some time.

- Save Calibration Settings for Application (Advanced Users Only!): This option is only available for the 'Custom' calibration. The calibration settings are saved for subsequent usage in the Application Module. This will only have an effect, if the **Processing Type** has been set to 'Calibration', before the regression calculation.
- Plot Type
 - Linear Regression: This plot type generates two scatterplots. On the left side, the relationship of pigment absorption (x-axis) vs. in situ Chl measurements (y-axis) is shown. The plot includes one or all calculated linear regression functions, as it responds to the specific station selection (see next option). The right plot shows the relationship between estimated Chl concentrations after application of the regression models (x-axis) vs. in situ Chl measurements (y-axis). Respective Pearson correlation coefficients are indicated in the plot (Figur 3.10).
 - **Chl Exponent & Chl Factor:** This plot type generates a map, where one of the two scaling factors is represented for each included NOVANA monitoring station on a color-scale.
 - Select Specific Station: This option is only available for the Linear Regression plot type. Users can select either all stations (see example in Figur 3.10) or a specific station from the selected calibration or validation dataset. Specific stations can be selected both by station name in the drop-down menu and by clicking in the vicinity of the desired point on the map. A specific station selection will cause the left part of the linear regression plot to only show the regression line for the selected station together with the involved pairs of points. The size of the points indicates respective regression weights (see example in Figur 3.7).
 - **Data as '*.csv':** Users may download the resulting dataset as *.csv table (Table 1). The table includes for each selected NO-VANA match-up the station name, coordinates in UTM32N coordinate reference system, year, day of the year (DOY), as well as the data for regression analysis: Chl Exponent (Chl_exp), log of Chl Factor (log_Chl_fact), in situ Chl concentration (Chl_NOVANA) in μ g/L, uncorrected Sentinel-3 Chl concentration (Chl_S3_corr), as well as the lower and upper bounds of the 95%-confidence interval of corrected Chl estimates (Chl_S3_corr_lwr, Chl_S3_corr_upr).

Application module

To run the app, open the file 'CoCoa-SenCE/App_Appli/app.R' in R-Studio and press 'Run App' to start the user interface in a web browser.

User interface and options

In this module, the user can apply the default or a customized calibration on pre-processed satellite data. Pre-processing of the raw Sentinel-3 OLCI level-2 product for ocean colour in turbid water is described in the main report.


Figure M.2. Screenshot of Graphical User Interface of the CoCoa-SenCE Application Module.

Chose Calibration

- Default: The Default calibration is the recommended one described in the main report. All regression parameters are already calculated and are ready for application on a customized selection of satellite images.
- Custom (Advanced Users Only!): In case that you have calculated a customized calibration in the Calibration & Validation Module, you can apply this calibration here. First, Calculate Cost-Distance Matrix between all monitoring stations and each pixel of the study area and then Calculate Distance-Weighted Regression. These operations require considerable computation time.
- **Select Application Year:** Select a specific year, for which you intend to apply the regression model.
- Select Temporal Aggregation Level: The regression model can be applied on both Daily images and on aggregated images By Months. Under the first option, a specific day of the year can be selected. The latter option will first calculate mean values for each of the selected months, and then the mean from all selected months.
- **Refresh Sentinel-3 Data:** After having set a new data selection, the new data must be loaded by clicking **Refresh Sentinel-3 Data**.

- Calculate Daily Corrected Images: This option is only available for the Custom calibration. The custom calibration will be applied an all daily ±5 days moving average composite S3 images for the selected year. The corrected images are saved as individual GeoTIFF files in the folder 'CoCoa-SenCE/Custom/Daily_Corr/XXXX', where XXXX is the selected year. These can then be accessed in the Extraction Module.
- Save Custom Regression Parameters (Advanced Users Only!): This option is only available for the Custom calibration. New regression parameters derived from the custom calibration will be saved in the folder 'CoCoa-SenCE/Custom'
- Plot Type
 - **Sentinel-3 Raw:** This option generates a map with aggregated Chl concentrations based on the default setting of scaling factors in the original Sentinel-3 OLCI level-2 data product (Figur 4.2).
 - **Sentinel-3 Pigment Absorption:** This option generates a map with aggregated pigment absorption values from the original Sentinel-3 OLCI level-2 data product.
 - **Sentinel-3 Corrected:** This option generates a map with aggregated corrected Chl concentrations based on pigment absorption and the two spatial models of scaling factors, derived from cost-distance weighted regression modelling (see example in Figur 4.2).
 - **Relative Deviation:** This option generates a map showing the relative deviation Chl^{dev} between the raw and the corrected Chl products: $Chl^{dev} = \frac{(Chl^{raw} Chl^{corr})}{Chl_{raw}}$. (see example in Figur 4.2)
 - **Chl-Exponent:** This option generates a map of the spatial model for the scaling factor Chl_{exp} (see example in Figur 4.1).
 - **Log(Chl-Factor):** This option generates a map of the spatial model for the scaling factor Chl_{fact} on the log-scale (see example in Figur 4.1).
 - Validation: This option generates a scatter-plot between satellite- and *in situ*-based Chl estimates of the selected temporal aggregation (see example in Figur 4.3). This plot is only available for monthly aggregated temporal selections.

Extraction module

To run the app, open the file 'CoCoa-SenCE/App_Extract /app.R' in R-Studio and press 'Run App' to start the user interface in a web browser.

User Interface and Options

In this module, the user can use the default or custom calibration settings to make specific time-series and spatial data extractions. Time-series can be extracted for individual NOVANA stations, and spatial extractions can be performed for the spatial assessment units of (1) the water Framework Directive, (2) HELCOM, and (3) OSPAR.



Figure M.3. Screenshot of Graphical User Interface of the CoCoa-SenCE Extraction Module.

Chose Calibration

- **Default:** The Default calibration is the recommended one described in the main report.
- **Custom (Advanced Users Only!):** In case that a customized calibration was applied and saved and daily corrected images were calculated in the **Application Module** you can use this calibration also here in the extraction module.

• Chose Extraction Mode

- **Time-Series:** For each NOVANA station involved in the model it is possible to extract a corresponding time-series (see example in Figur 4.5).
- Assessment Units: Three sets of spatial assessment units (Water Framework Directive: WFD, OSPAR, and HEL-COM) are implemented for data extraction (see example in Figur 4.4).
- Select Extraction Years: A year-range must be specified for either time-series extraction or temporarily aggregated mean value extraction across spatial assessment units.
- Select Temporal Aggregation Level: This option is only available for Time-Series extraction mode. The time-series can consist of daily values or monthly aggregated means.

- Select Months: This option is only available for Assessment Units extraction mode. A specific range of months, e.g. growing season from March-September, can be selected for the temporal aggregation.
- **Refresh Data Selection:** This button must be clicked so a new data selection is getting effective in the plot. For **Time-Series** extraction mode, data is extracted for all stations in parallel. This requires computation time but enables to visualize several time-series without refreshing data.
- Plot Element(s)
 - **Sentinel-3 Raw:** This option adds Chl concentrations based on the default setting of scaling factors in the original Sentinel-3 OLCI level-2 data product to the plot.
 - **Sentinel-3 Corrected:** This option adds Chl concentrations based on pigment absorption and the two spatial models of scaling factors, derived from cost-distance weighted regression modelling to the plot.
 - **S-3 Corr. Uncertainty:** This option is only available for **Time-Series** extraction mode. It adds the 95% confidence-interval of the local regression model to the time-series.
 - **NOVANA:** This option is only available for **Time-Series** extraction mode. It adds Chl concentrations based on *in situ* NOVANA data to the time-series.
- Select Specific Station: This option is only available for Time-Series extraction mode. Users must select a specific station either by station name in the drop-down menu or by clicking in the vicinity of the desired point on the map.

OPERATIONALISERING OG MODEL-ASSIMILERING AF SATELLITBASEREDE KLOROFYLDATA

Denne rapport omhandler operationaliseringen af en ny metode for satellitbaserede klorofylestimater i komplekse kystnære områder og assimileringen af det optimerede klorofylprodukt i mekanistiske bio-geokemiske modeller.

ISBN: 978-87-7156-713-7 ISSN: 2244-9981